Proxemic Interaction: Designing for a Proximity and Orientation-Aware Environment

Till Ballendat, Nicolai Marquardt, Saul Greenberg

Department of Computer Science
University of Calgary, 2500 University Drive NW
Calgary, AB, T2N 1N4, Canada
[tballend, nicolai.marquardt, saul.greenberg]@ucalgary.ca

ABSTRACT

In the everyday world, much of what we do is dictated by how we interpret spatial relationships, or *proxemics*. What is surprising is how little proxemics are used to mediate people's interactions with surrounding digital devices. We imagine *proxemic interaction* as devices with fine-grained knowledge of nearby people and other devices – their *position, identity, movement*, and *orientation* – and how such knowledge can be exploited to design interaction techniques. In particular, we show how proxemics can: regulate *implicit* and *explicit* interaction; trigger such interactions by continuous movement or by movement of people and devices in and out of discrete proxemic regions; mediate simultaneous interaction of multiple people; and interpret and exploit people's directed attention to other people and objects. We illustrate these concepts through an interactive media player running on a vertical surface that reacts to the approach, identity, movement and orientation of people and their personal devices.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces – Input devices and strategies.

General terms: Design, Human Factors

Keywords: Proximity, proxemics, location and orientation aware, implicit interaction, explicit interaction

INTRODUCTION

Spatial relationships play an important role in how we physically interact, communicate, and engage with other people and with objects in our everyday environment. *Proxemics* is Edward Hall's theory of these interpersonal spatial relationships [8]. It describes how people perceive, interpret and use distance, posture and orientation to mediate relations to other people, and to the *fixed* (immobile) and *semi-fixed* (movable) features in their environment [8]. Proxemic theory correlates physical distance with social distance (albeit in a culturally dependent manner): *intimate*
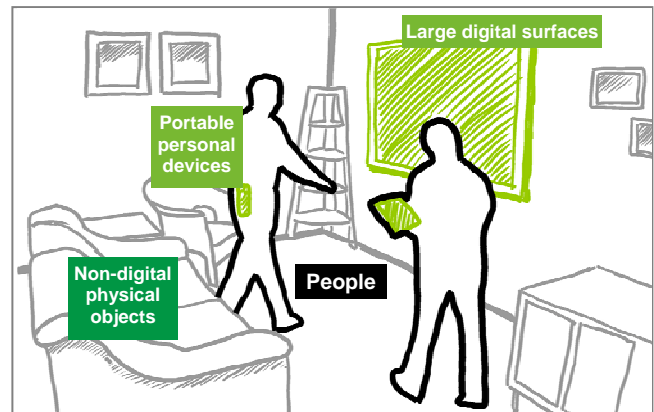

Figure 1. Proxemic interactions relate people to devices, devices to devices, and non-digital physical objects to people and devices.

6-18", personal 1.5-4', social 4-12', and *public 12->25'* distances. As the terms suggest, the distances lend themselves to a progression of interactions ranging from highly intimate to personal, to social and then to public. Each distance also defines a *close* and *far phase* that affects that interaction [8].

Hall emphasizes the role of proxemic relationships as a form of people's implicit communication – a form of communication that interactive computing systems have yet to understand. In spite of the opportunities presented by people's natural understanding of proxemics, only a relatively small number of research installations – usually within Ubiquitous Computing (Ubicomp) explorations – incorporate spatial relationships within interaction design.

Yet these installations are somewhat limited. For example, a variety of systems trigger activity by detecting the presence or absence of people within a space, e.g., reactive environments have devices in a room react to presence [2], or digital surfaces that detect and react to a device within a given range [14] [15]. While useful, this is a crude measure of proxemics, as it only considers distance as a binary value, i.e., within or outside a given distance. True proxemics demand fine-grained knowledge of people's and device's continuous movement in relationship with each other, and how this would affect interaction. Two projects stand out here [11] [21]; both have a vertical digital surface reacting to people's distance from it to control the information displayed. We take their work even further, where we extend previous notions of *proxemic interactions*.

Our contributions consider the *complete ecology* present in a small space Ubicomp environment (illustrated in Figure 1): the relationships of people to devices, of devices to devices, and of non-digital objects to people and devices. For this, we exploit continuous knowledge of *distance*, *orientation*, *movements*, and *identity* as part of an extended notion of proxemics to drive the possible interactions. Building upon Vogel's [21] and Ju's [11] work, we demonstrate how proxemic information can regulate both *implicit* and *explicit* interaction techniques within a realistic application, either based on continuous movement, or by movement in and out of discrete proxemic zones. By implicit, we mean actions the computer takes based on its interpretation of implied user actions *vs.* explicit control actions stated by the end user. We explain how proxemic interactions consider aspects of the *fixed* and *semi-fixed feature environment*, and how they extend attentive interfaces. Proxemic interactions also extend beyond pairwise interaction and consider one person or multiple people in relation to an *ecology* of multiple devices and objects in their nearby environment.

We illustrate these concepts with the design of an interactive vertical display surface that recognizes the proximity of surrounding people, digital devices, and non-digital objects. Our example application is an interactive home video media player centered around a vertical surface in a living room. It *implicitly* reacts to the approach and orientation of people, and their personal devices and objects. Depending on the distance of people to the display and their movements, the application implicitly changes information displayed on the screen, and reacts by implicitly triggering application functions. Furthermore, we explain how *explicit* interaction is supported from these varying distances to the interactive display surface.

The remainder of the paper is structured as follows. After summarizing related work, we provide a scenario of people using our proxemic media player. Next, we introduce four dimensions describing the possible proxemic relationships involving people and their things. We then introduce concepts for designing proxemic interactions in Ubicomp, which we illustrate via our proxemic media player. We close with a brief description of our implementation.

RELATED WORK
We sample related work out of two research areas: interactive wall surfaces that sense the presence of nearby devices and of people to mediate implicit and explicit interaction, and devices that sense the presence of other devices to mediate connectivity and information exchange.

Proximity-Aware Surfaces and Displays
The majority of HCI research involving digital wall displays explores direct touch or gestural interaction, but otherwise ignores proximity. Some techniques do expect people to be at a certain distance from the display to work (e.g., ray casting, or pick and drop [14]), but this is just a function of where people have to stand for the technique to work.

Several early works considered how a spatially-aware mobile device would interact with a large digital surface. Notably, Chameleon [6] was a palmtop computer aware of its position and orientation. When used relative to a vertical display, Chameleon's contents would vary depending on its spatial orientation to that surface. Similarly, Rekimoto's spatially-aware M-Pad mobile device behaved like a click-through toolglass whose attributes affect the nearby items on the surface [14].

Somewhat later, several researchers considered vertical surfaces that react to the spatial presence of people. For example, Shoemaker [18] introduced techniques for a person to directly interact with digital content on a vertical wall surface through real or virtual shadows. The person's movement in the space and resulting changes of the shadow projections become part of the interaction. Hello.Wall [13] introduced the notion of 'distance-dependent semantics', where the distance of an individual from the wall defined the interactions offered and the kind of information shown. Technically, Hello.Wall could discriminate people's rough positions as three spatial zones. Vogel et al. [21] took this concept even further, where they directly applied Hall's theory to define four proxemic zones of interaction. From far to close, these ranged from ambient display of information, then to implicit, then subtle, and finally personal interaction. A major idea in their work – developed even further by Ju [11] – is that interaction from afar is public and implicit, and becomes more private and explicit as people move towards the surface.

Researchers have also considered a person's proximity to a small display. Lean and Zoom, for example, used the distance between the user's head and a notebook display to control a zoom effect [9]: the smaller the distance, the larger the displayed content.

As mentioned earlier, we extend this prior work by exploiting continuous distance, orientation, movement and identity to tune surface interaction, where we incorporate multiple people and features of the fixed and semi-fixed environment as a complete ecology.

Device to Device Connectivity Via Proximity Sensing
A major problem in Ubicomp is how to control the connectivity of devices. Consequently, various researchers have considered how spatial distance can be used to connect devices. Most approaches define a single discrete spatial region – which often depends on the sensing technology used – where a connection (or user interaction leading to a connection) is triggered when the spatial regions between devices overlap. With Smart-its friends [10], such a connection can be established once two devices sense similar values through attached sensors (such as accelerometers). By shaking a pair of devices simultaneously, an inter-device connection can be established. Want [22] introduced the technique of detecting nearby objects and devices through attached RFID tags, while Rekimoto [16] combined RFID and infrared for establishing device connectivity. These techniques are powerful for connecting de-

vices that are in very close proximity or – like in many cases – are even directly touching one another. Swindells [19] introduced a similar technique that worked from a larger distance, where he applied it to the gesturePen for initiating remote pointing for device selection. We extend this prior work, where we contribute techniques that go beyond a binary device connection state: we introduce techniques that move from *awareness* at a larger distance, to gradually revealing of higher level of detail, to *direct interaction* for transferring digital information between devices.

Spatial relations have also been used to mediate the information exchanged between devices. For example, Kray's *group coordination negotiation* [12] introduced spatial regions around mobile phones. Their scenario used these regions to negotiate exchange of information with others and to visualize the regions on a tabletop. Depending on how devices were moved in and out of three discrete regions, the transfer of media data between the devices is initiated. We extend their approach to interaction around large surfaces, where the degree of shared information between devices depends not only on their relative distance, but also orientation.

Gellersen's RELATE Gateways [7] provided a spatial-aware visualization of nearby devices. A graphical map showed the spatial room layout, and icons indicated the position of other nearby devices. Alternatively, icons at the border of a mobile device screen represented the type and location of surrounding devices (see also [16]). We extend this notion with: visualizations that include proximity-dependent level of detail, and with techniques that move from awareness to direct interaction depending on a person's distance and orientation to the display.

THE PROXEMIC MEDIA PLAYER APPLICATION
We use the example of people interacting with a home media player application located in a living room. Later sections, which present concepts for designing proxemic interactions, will use episodes from this scenario to anchor the discussion.

Our scenario follows Fred who is approaching the display from a distance. We explain how the system supports Fred's implicit and explicit interaction with the digital surface as a function of his distance and orientation. The primary interface of the interactive media player application supports browsing, selection, and playback of videos on a large wall-mounted digital surface: a 52 inch touch-sensitive SmartBoard from Smart Technologies, Inc. (Figure 2, top). A Vicon motion capture system tracks, via reflective infrared markers, the location and orientation of nearby people, objects, and other digital devices. All equipment is situated in a room that resembles a domestic living room.
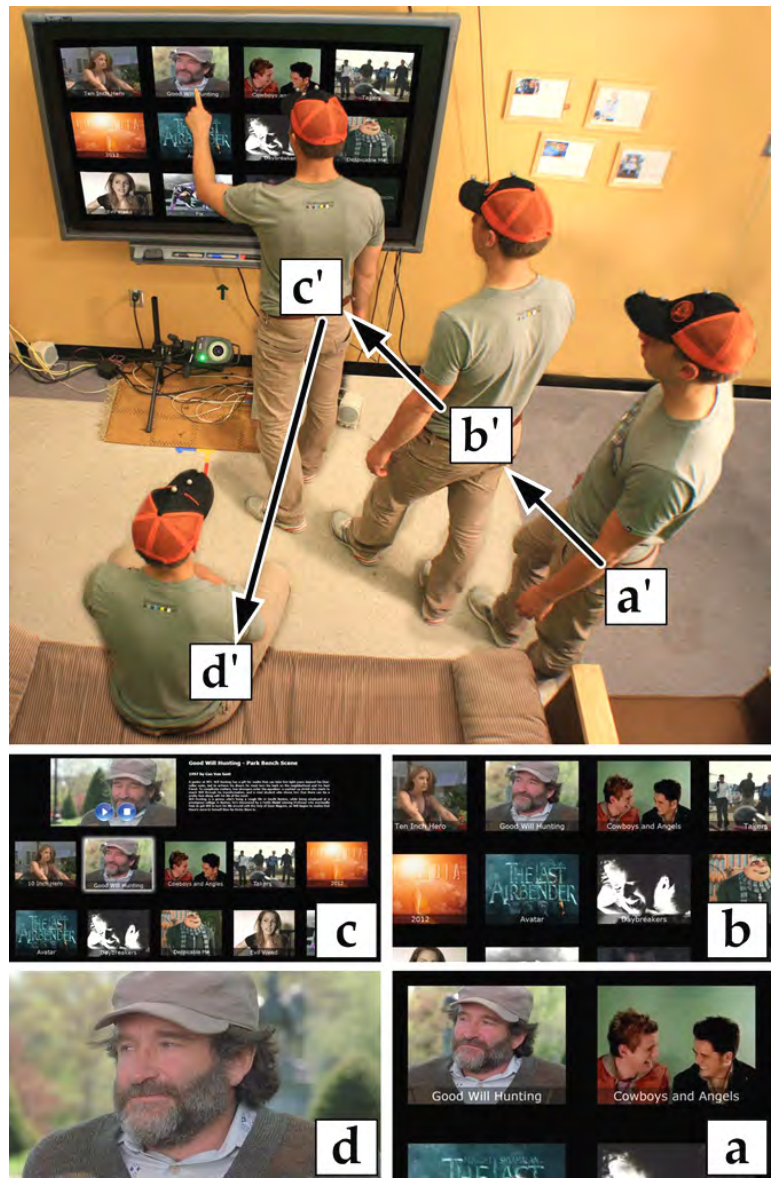

Figure 2: Proxemic Interaction: a) activating the system when a person enters the room, b) continuously revealing of more content with decreasing distance of the person to the display, c) allowing explicit interaction through direct touch when person is in close distance, and d) implicitly switching to full screen view when person is taking a seat.

Figure 2 (top) shows Fred approaching the display at four distances (a' – d'), while the four scenes at the bottom shows what Fred would see at those distances. Initially, the proxemic media player is 'asleep' as the room is empty. When Fred enters the room at position (a'), the media player recognizes Fred and where he is standing. It activates the display, shows a short animation to indicate it is activated, and then displays four large video preview thumbnails held in Fred's media collection (Figure 2a). As Fred moves closer to the display (b'), the video preview thumbnails and titles shrink continuously to a smaller size, thus showing an increasing number of available videos (2b). When Fred is very close to the surface (c'), he can select a video directly by touching its thumbnail on the

screen. More detailed information about the selected video is then shown on the display (2c), which includes a preview playback that can be played and paused (2c, top), as well as its title, authors, description and release date (2c, right). When Fred moves away from the screen to sit on the couch (d'), his currently selected video track starts playing in fullscreen view (2d). If Fred had previously seen part of this video, the playback is resumed at Fred's last viewing position, otherwise it starts from the beginning.

Fred tires of this video, and decides to select a second video from the collection. He pulls out his mobile phone and points it towards the screen (Figure 4b). From its position and orientation, the system recognizes the phone as a pointer, and a row of preview videos appears at the bottom of the screen (as in Figure 4b). A visual pointer on the screen provides feedback of the exact pointing position of Fred's phone relative to the screen. Fred then selects the desired videos by flicking the hand downwards, and the video starts playing. Alternately, Fred could have used a non-digital pen to do the same interaction (Figure 4a).

Somewhat later, Fred receives a phone call. The video playback automatically pauses when he answers the phone (Figure 3b), but resumes playback after he finishes the call. Similarly, if Fred turns away from the screen to (say) read a magazine (Figure 3a), the video pauses, but then continues when Fred looks back at the screen.

As Fred watches the video while seated on the couch, George enters the room. The title of the currently playing video shows up to at the top of the screen to tell George what video is being played (Figure 6a). When George approaches the display, more detailed information about the current video becomes visible at the side of the screen where he is standing (Figure 6b). When George moves directly in front of the screen (thus blocking Fred's view), the video playback pauses and the browsing screen is shown (Figure 6c). George can now select other videos by touching the screen. The view changes back into full screen view once both sit down to watch the video. If Fred and George start talking to each other, the video pauses until one of them looks back at the screen (Figure 3c).

Fred takes out his personal portable media player from his pocket. A small graphic representing the mobile device appears on the border of the large display, which indicates that media content can be shared between the surface and portable device (Figure 5a). Fred moves closer to the surface while pointing his device towards it; the graphic on the surface responds by progressively and continuously revealing more information about the content held on the media device (Figure 5b). When Fred moves directly in front of the surface while holding the device, he sees large preview images of the device's video content, and can then transfer videos to and from the surface and portable device by dragging and dropping their preview images (Figure 5c). The video playback on the large screen resumes as Fred puts his portable device back in his pocket and sits down on the couch. When all people leave the room, the application stops the video playback and turns off the display.

While this media player is a simple application domain, it provided a fertile setting to develop and explore concepts of proxemic interaction. In the next section we introduce the dimensions of input that are essential for designing proximity aware interfaces. Then we will discuss the details of proxemic interaction concepts associated with a single person or multiple people interacting with a large digital surface.

DIMENSIONS OF PROXEMIC RELATIONSHIPS

While many dimensions are used by people to mediate their interpersonal proxemic interactions, we identify four dimensions as essential if a system is to determine the basic proxemic relationships between entities (people, digital devices, and non-digital objects): *position*, *orientation*, *movement*, and *identity*. These four dimensions are part of our *extended notion of proxemics* that differs from Hall's understanding of discrete proxemic zones that are based primarily on the actual spatial distance between individuals.

Position of an entity can be described in *absolute* or *relative* terms. For the absolute position we have to know the distance of the entity from a defined fix point in the space. Once such a fixed point in space is defined, the absolute position of every entity can be described as the three dimensional position relative to this fixed point. Relative position, on the other hand, can be determined from knowing the spatial relationship between two entities (e.g., between a person and object), and does not require a common fixed point of reference. Through the knowledge of absolute or relative position, we can calculate information about *distance* (e.g., imperial or metric units) between objects and people.

Orientation provides the information about which direction an entity is facing. This makes sense only if an entity has a well-defined 'front' (e.g., a person's eyes, the point of a pencil). Similar to location, we can differentiate between the absolute orientation of an entity (e.g., described through yaw, pitch, and roll) or relative orientation (e.g., a quantitative description such as "this person is facing that object"). From orientation, determine where a ray cast from one entity would intersect with another entity (*ray casting)*.

Movement lets us understand the changes of position and orientation of an entity over time. This also means we can calculate the *velocity* of these changes. These movements, for example, reveal how a person is approaching a particular device or object.

Identity uniquely describes the entities in the space. The most detailed information provides the exact identity of a person or object (e.g., "Fred", "Person A", "Fred's Cell phone"). Other less detailed forms of identity are possible, such as identifying a category precisely (e.g., "book", "person"), or roughly ("non-digital object"), or even affiliation to a group (e.g., "family member", "visitor").
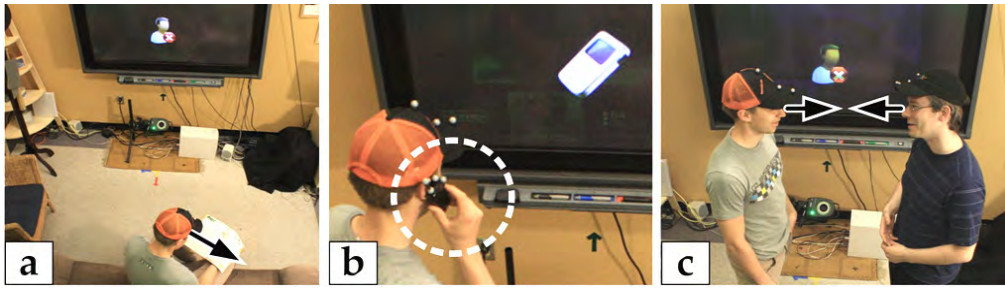
Figure 3: Integrating attentive interface behaviour: pausing the video playback when the person is (a) reading a magazine, (b) answering a call, or (c) talking to another person

DESIGNING FOR PROXEMIC INTERACTION

We now describe concepts of applying these four input dimensions in meaningful ways to people's *proxemic interactions* with Ubicomp systems. To ground our explanation, we highlight particular examples from the scenario that illustrate how each concept can be applied.

Incorporating the Fixed- and Semi-fixed Feature Space

One promise of Ubicomp is to situate technology in people's everyday environments, in a way that lets people interact with information technology in their familiar places and environment. Dourish framed this concept as *embodied interactions* [5]; technology that is seamlessly integrated into people's everyday practices, rather than separated from them. Context-aware computing is one outcome of this, where some kind of context-aware sensing [17] provided devices with knowledge about the situation around them. This sensing usually involved measuring a coarser subset of our dimensions, e.g., very rough positions, and other factors such as noise, light, or tilting. We contribute to this by introducing the notion of having context-aware systems mediate embodied interaction by understanding the proxemic relationships (as defined by our dimensions) of people to the fixed- and semi-fixed feature space [8] surrounding them.

For an interactive system (such as the interactive wall display in our media player application), knowledge about the *fixed feature space* includes the layout of the fixed aspects of the room, such as existing walls, doors and windows. It also includes knowledge about fixed displays – such as a digital surface – located in this environment. For instance, the knowledge about the position of the fixed entrance doors allows our system to recognize a person entering the room from the doorway, and then take implicit action by awaking from standby mode. Similarly, knowing the position of the fixed display means that the interface on that display can react as a person approaches it.

Semifixed features in the environment include all furniture, such as bookshelves, chairs, and tables whose position may change over time. While it is somewhat object-dependant, semi-fixed features often remain at specific locations, but are per se movable objects that people rearrange to adapt to changed situations (such as moving a group of chairs around a table). Unlike fixed features whose position needs

to be configured only once, knowledge about the positions of semi-fixed features will have to be updated over time as changes are noticed.

Knowledge of semi-fixed features can also mediate interaction. To illustrate this point, we compare two stages of a person relative to the media player's interactive surface: approaching from a distance (see Figure 2, position a') and watching the video when seated at the semi-fixed couch (Figure 2 position d'). The actual distance of the person relative to the surface is similar in both situations, yet they suggest very different forms of interaction. The fact that the person is seated on a couch or chair facing the display becomes an indicator for watching the video. Yet standing at the same distance and then moving closer to the screen is used to infer that the person is increasingly interested in getting more information about the available videos in the media collection. (Of course, inferences may not always be correct. This will be discussed later).

Thus, information about distance and orientation of a person relative to the fixed and semi-fixed feature space provides cues that can mediate implicit interactions with the system.

Interpreting Directed Attention to People and Objects

Proxemic interactions can be used to extend the concept of *attentive user interfaces* (AUIs) that are designed to "support users' attentional capacities." [20]. In AUIs, the system reaction depends on whether a person is directing his or her attention to the device that holds the system (usually through detection of eye gaze) [20]. We take this AUI concept one step further, where we also incorporate information about: what entity a person is attending, and the importance of distance and orientation in that context.

Attending to the system itself occurs if the device reacts to how it is being looked at. This is how most traditional AUIs work. We include an example of this behaviour [20] in our media player application: the system plays the video as long as at least one person faces the large display, but pauses when that person looks away for a length of time.

Attention to other surrounding objects and devices. We enrich the concept of AUIs by including how a person's directed attention to *other* surrounding objects of the semi-fixed feature space can trigger implicit system reactions. In our system, the fact that a person is holding and facing towards a newspaper (shown in Figure 3a) provides cues about the focus of this person's attention, i.e., the system infers that Fred is reading, and pauses video playback until Fred stops reading and looks back at the screen. If Fred had a similar gaze to (say) a bowl of popcorn, the video would not have paused.

A shift of attention can also be suggested by the relative distance of an object to the person. For example, our system detects when Fred is holding his mobile phone close to his ear (as shown in Figure 3b). It infers that Fred is having a phone conversation, and pauses the video until Fred moves his phone away from his head. The measurement of relative distance of phone to the person's head, as well as their orientation towards each other, provided the necessary information for the system to implicitly react to this situation.

Attention to other people. We can discriminate how one person attends other people as a means to trigger implicit system reactions. For example, consider Fred and George when they turned towards each other to converse (see Figure 3c). Our scenario illustrated how the system implicitly reacts to this situation by pausing the video. However, by knowing that they are in conversation (rather than just knowing that they are looking away from the display), the system could have just turn down its volume.

Supporting Fine Grained Explicit Interaction

Instead of *implicitly* reacting to a person's proxemic relation to other semi-fixed environment objects, these relationships can also facilitate a person's *explicit* forms of interaction with the system. We introduce the concept of using physical objects as *mobile tokens* that people can use to mediate their explicit interaction with an interactive surface. The meaning of these tokens is adjusted based upon the token's distance and orientation to other entities in the space.

To illustrate this concept, consider the explicit interaction in our scenario where Fred pointed his cell phone or a pencil at the surface to view and select content. The way this works is that all mobile tracked objects are interpreted as *mobile tokens*. Three units of information caused our system to interpret that token as a pointing device: it is held in *front* of a person, it is roughly *oriented* towards the display, and it is within a particular *distance* from the display. Indeed, we showed how two quite different devices can serve as similar tokens: the pen in Figure 4a, and the mobile phone in Figure 4b. We emphasize that we are *not* using any of the digital capabilities of the mobile digital phone to make this inference. Rather (and as with the physical pen) we are using only the knowledge of its position and orientation to switch to a certain interaction mode. Thus, the particular proxemic relationship between a person and a mobile token is interpreted as a *method of signaling* [3], as discussed in Clark's theory of pointing and placing as forms of communication. Further, the specific orientation
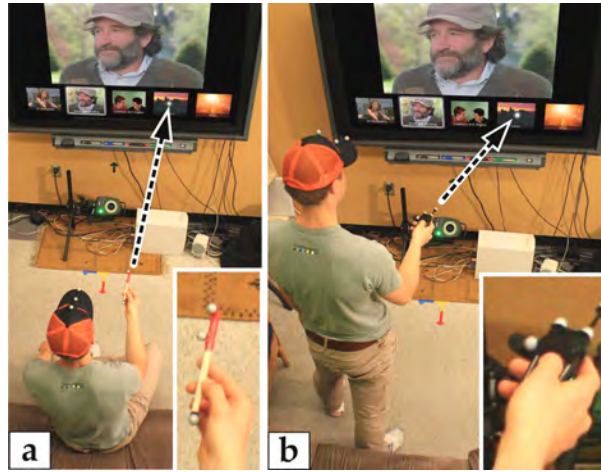

Figure 4. Explicit interaction triggered through distance and orientation between a person and digital / non-digital physical artefacts: a) pen, b) cell phone.

and distance of the token to other devices (e.g., the large display) are interpreted to establish an *intrinsic connection* [3] to control that particular device.

A key advantage is that the use of these mobile tokens as identifiers can disambiguate similar looking gestures. For example, a gesture recognition system cannot tell if the intent of a person pointing their hand towards the screen is to interact with the screen, or that it is just a gesture produced as part of a conversation. Mobile tokens, on the other hand, create a specific context to disambiguate and interpret gestures, where it uses the distance and location of the objects relative to the person and other objects to infer a certain explicit interaction mode.

Many of these behaviours can be triggered by approximate knowledge of proxemic relationships. Yet having exact knowledge is helpful for minimizing errors that can occur where the system misinterprets a person's manipulation of a mobile token as an explicit action. For example, consider a person playing with a pen in their hand vs. pointing the pen at the screen to select an item. If proxemic measures are reasonably precise, the triggering event could rely solely on the pen being a specific distance from the person's body and a specific orientation towards the screen for a particular length of time.

Another example includes the multiple meanings held by a mobile token. Consider how the meaning of the mobile phone depended on its proxemic relation to its holder and to the display. The distance of the phone to a person's head indicated an ongoing phone conversation, while holding the same device in front and towards the display shifts its meaning to an interaction pointer.

For the actual explicit interaction with the digital video content displayed on the large surface, the person can move the position of the mobile token. Changes of the orientation angle allow fine grained positioning of a pointer icon on the screen, while fast acceleration downwards can be used for selection.

Interpreting Continuous Movements or Discrete Proxemic Zones

Another concept is that the behaviours of proxemic interfaces can react to the position and distance of its entities as either *continuous movements,* or as movements in and out of *discrete proxemic zones*.

For *continuous movement,* the calculated distances between people and devices function as input variables that continually affect the interactive system's behaviour. For example, as a person approaches a screen of the media player appli-
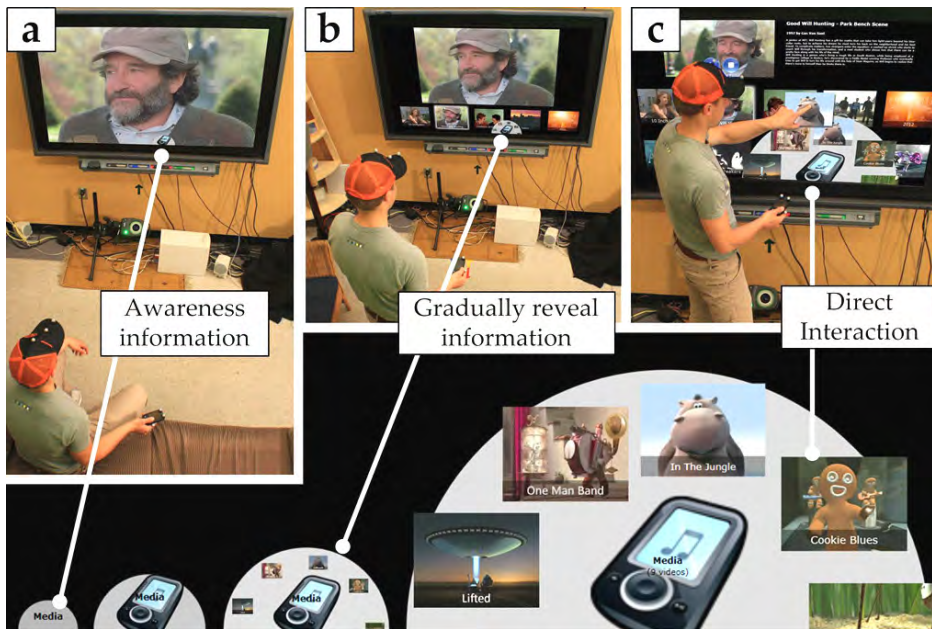
Figure 5. Proximity mediates device to device interaction: from awareness to direct interaction.

rate distances. For example, we use a 15-20% hysteresis tolerance for proxemic regions around the interactive wall display (percentage of the region dimension) to avoid this rapid switching.

Moving From Awareness to Direct Interaction

Next, we can combine both continuous movements and discrete proxemic zones to design system interfaces that move fluently from awareness to direct explicit interaction. Two examples illustrate this combination.

Our media player begins by providing peripheral awareness information about its capabilities and content when a person enters the room. The system detects the presence of the person at a distance (around 4m), activates the display, displays a welcome animation, and plays a subtle acoustic signal. This indicates to the person that the system is active. Here, we used a discrete proxemic zone around the digital display that triggers this activation behaviour. At this point, if the person just walks pass the display, or does not face the display, the media player application would revert to sleep mode. If, however, the person does move closer, the system shows preview images of video content, where it gradually reveals more preview items on the screen as the person approaches the screen. Here, we use the continuous mapping of distance to the size and quantity of preview items shown. When the person stands within reach of the screen, we enter another discrete zone: direct touch interaction. At that distance, the person can use their hands for direct touch interaction with the screen content; thus the continuous resizing of the displayed preview thumbnails stops as it would otherwise make selection difficult.

So far, we have focused on implicit and explicit interactions mediated through changes of a person's distance and orientation relative to the large digital surface. Interactions, however, increasingly take place in an environment comprising an ecology of devices – from shared large displays to portable personal devices. Using our four proximity dimensions, we can recognize nearby devices and thus facilitate using them in conjunction with one another. This opens new possibilities for interaction, communication, and information exchange. However, to make sense of device interaction, people require awareness of device interconnections and a means to move into direct interaction over them.

To explain, we illustrate *device-to-device proxemic awareness and interactions* with the interactive vertical display, where the surface reacts to nearby portable devices carried by a person. This time the system reacts to distance, con-

cation, the number of visible video preview thumbnails shown continually increase with distance (see Figures 2a,b). To do this, the system gradually resizes the preview images to a smaller size (zoom out effect); thus more content is visible as the person approaches the screen. Depending on the situation, an inverse behaviour might be applied, where the system actually *zooms into* the content to make it larger when the person is approaching the screen (similar to *Lean and Zoom* [9]). Another example of continuous mapping of distance as an input regulator are the awareness icons of nearby digital devices (visible in Figure 5). These icons grow continuously, from a small circular icon indicating its presence, to a large area on the screen that displays rich content and allows direct touch interaction with it (as in the progression from Figure 5a-c).

With *discrete proxemic zones* we can divide the space into discrete regions. When a person enters or leaves the thresholds of these zones, certain actions are triggered in the system. Indeed, the use of zones is inspired by the interpersonal proxemic distance zones defined by Hall [8], and others have applied zones as a way to mediate interaction with public ambient displays [21] and digital whiteboards [11].

Our media player uses discrete zones in several ways. We use it to trigger an associated implicit action (e.g., we activate a display screen when entering the room). We also use zones to allow certain forms of explicit interaction (e.g., switching to an interface that allows direct touch interaction when the person is standing in close distance to the screen).

A problem associated with discrete zones occurs when the interface rapidly switches back and forth between two states; this occurs when the person stands exactly at a border of one of the discrete zones. This is solved via the concept of a *hysteresis tolerance*: the entry and exit point of each region are not at the same distance, but are two sepa-
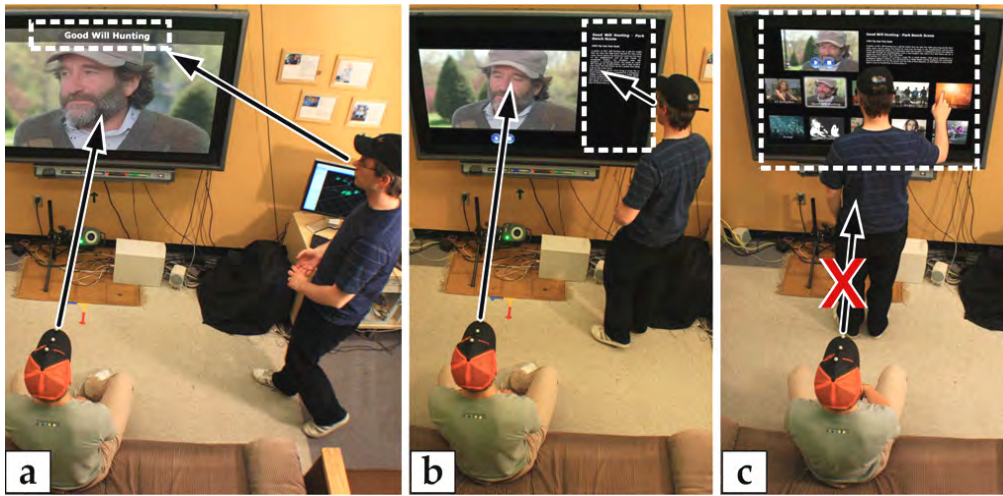
Figure 6: Mediating between multiple people: a) incoming person sees basic information such as video title; b) as one moves closer, the split view provides a more detailed video description; c) when within reach of the display, the person gets full control.

tinuous movement, and orientation of a person's *portable digital device* when approaching the media player displayed on the surface. Again, we illustrate how we use discrete zones and continuous movements to move from awareness to direct interaction.

When a person takes a portable media player out of their pocket while sitting at a distance, the system recognizes the device and indicates a possible interaction through a visual icon at the border of the display (visible in Figure 5a). This icon represents the portable device, where it indicates to the person that there is now an opportunity to share content between the large surface and the portable device. While this icon visualization is inspired by earlier approaches (e.g., [7] [16]) for visualizing spatial relationship between devices, it differs in how it incorporates proxemic distance and orientation information leading from awareness to direct interaction. If the person then orients this portable device towards the large screen, more detailed information about that device and its contents becomes visible. Depending on the orientation between the device and the large surface, the icons *continuously* and *instantly* update their position at the border of the interactive wall screen, so that they always face the direction of the portable device. As the person moves the personal device closer to the large display, even more details about the content (e.g., titles) become visible and the preview thumbnails are shown at a larger size (Figure 5b). When the person holding the device is within reach of the interactive screen (i.e., a discrete zone is entered), the size of the icon grows to a large area of the screen (visible in Figure 5c). The icon not only provides detailed information about the content of the device, but also allows full direct touch interaction. The person can now drag and drop video items from the portable device to the large surface and vice versa. When putting the device back in the pocket the visualization immediately disappears and the media player continues its playback.

Leveraging People's Identity
The concepts introduced so far only require knowledge about "a person" approaching the display, but they do not require the actual identity of a person. We now discuss examples that leverage the knowledge about the actual identity of individuals.

History. Knowing which person is interacting with the system is used to continue activities that this person began in the past. For instance, when a person enters the room and immediately sits down, the media application will resume playback of a last video that a person previously watched but did not finish.

Personalization. The media player could save one's settings as a personal profile. This can include personal configurations, idiosyncrasies of how the system should respond to that particular person, and that person's media content. For example, when a particular person approaches the display, our media player would then display content out of that person's media library.

Safeguards. Identifying the person interacting with the system can also function as a safeguard to restrict access. For instance, children may only be allowed to access the media player application during pre-defined time slots, or access to available media content could be restricted to movies rated for their age.

Mediating People's Simultaneous Interaction
Proxemic interactions should also mediate the interaction of multiple people in the same space. In the simplest case, as long as all people are in the same proxemic state relative to the display's surface, the system's behaviour could be similar to the proxemic interactions introduced for a single person interacting with the surface. In reality, however, we expect people to be in different proxemic stages, where the system would need to reason about how it should mediate its behaviour to reflect people's simultaneous interaction possibilities.

Merging multiple proxemic distances. In situations where people have different proxemic distances to the interactive display of our application, the system can be designed to individually address people's diverse proxemic needs, albeit as a compromise.

For example, we saw George enter the room while Fred was watching a video. George wants to know what was being played, while Fred wants to keep watching. To compromise between these needs, the system displayed the title of the currently playing video at the top of the screen, thus

subtly informing George while still letting Fred watch without too much distraction (Figure 6a). If George sits at the couch or on a chair, the title disappears.

If George approaches the screen instead of sitting down, the display animates and splits off a small region of the screen. This region provides further information of the video being played: its description, author information, and the release date (Figure 6b). The positioning of this region also depends on George's spatial relation to the display – if he moves between the left to right side, the information panel smoothly animates to that side of the display.

When both people are in the same proxemic state, the views merge. For instance, both people can watch the video in full screen when seated, or both can explore and choose from the videos available when standing in front of the display.

Handling conflicts. When multiple people are present within a proximity-aware application, situations will arise where the system has to handle two *conflicting* individual possibilities. For example, consider the scenario situation of Figure 6c: Fred is sitting in front of the large display watching a movie, while George moves directly in front of the display to browse a media collection.

Several strategies are possible to handle these situations. The system could favour the person in closer proximity; e.g., George standing directly in front of the display would have priority over Fred sitting at a larger distance. This is the solution shown in Figure 6c, where George gets full access to the media library to select videos; a strategy that makes sense as Fred's view is already blocked. Alternately, the system could have given the video player priority, disallowing George's interaction, where they would have to resolve this through social means (e.g., both standing up to make a selection). Or the system could create some kind of composite view, i.e., by moving the video so that Fred could still see some of it, while still giving George interactive controls in the blocked part of the screen.

Differences in Perceiving and Interpreting Proxemics

People's perception of proxemic relationships is influenced by gender, cultures, age, work hierarchies, and other factors [8]. These differences also affect the design of proxemic interactions. Imagine a system that requires people to stand in very close proximity to each other to collaboratively interact with an interactive surface, e.g., to exchange digital documents. This close proximity might be perceived as adequate by some, but as too intimate by others. Therefore, the design of proxemic interactions has to consider these variations in proxemic perception.

In this regard, our implementation – while fully functional – serves just as an example that illustrates design possibilities. We do not suggest that our media player is the ideal, nor that it achieves the perfect balance between adjudicating proxemic information and implicit or explicit interaction.

IMPLEMENTATION

We briefly describe the technical setup and software implementation behind our system.

Tracking position and orientation of people and objects. Similar to [21], we use a Vicon infrared camera tracking system [www.vicon.com] to acquire fine-grained sensing information about people, objects, and digital devices moving around the interactive wall display. Six cameras emitting infrared light are placed around the SmartBoard to capture the position of passive infrared reflective markers. These markers are attached to tracked objects, digital devices, and people (peoples' head positions are tracked via hats with reflective markers attached). With this setup, we can detect markers up to a distance of four meters around the wall display. The Vicon software returns the triangulated position of all detected markers.

However, our implementation goes beyond this low level data acquisition. Our Proximity Toolkit [4] wraps the Vicon real time raw data, where it transforms that data into a much more usable form available to the programmer via an object-oriented API. Internally, it maintains a 3D model of all fixed and semi-fixed features, and of moving tagged entities in the environment. Entity positions are updated at 50 Hz. Programmers use the Proximity Toolkit's API to receive detailed information about the relative and absolute proximity and orientation between identified people, objects, and devices (including ray-casting information) [4]. Our toolkit also includes extensive but easy to use configuration options for specifying the fixed and semi-fixed features of the environment.

Distributed access to proxemic information. The Ubicomp ecology includes multiple digital devices – such as a personal portable media player – that also need to be notified about their movements in the environment and the position and orientation of nearby objects and people. Our system maintains a distributed data structure (provided by the *.NetworkingGT Toolkit* [1]) shared over a wireless network connection. This data structure contains precise information about the position and orientation of the tracked objects. We designed a hierarchical data structure that stores information like distance, angle, proxemic areas, and file transfers for every device. Both the client and server application subscribe to the relevant values and are instantly notified about changes. The applications running on the mobile devices can then trigger actions in response to sensed spatial movements and proxemic relations.

Alternative tracking technologies. We recognize that the Vicon tracking system is expensive and thus not a realistic platform for commercialization. However, we believe that technology for sensing proximity and orientation will soon be cheaply available to the public at a lower and more reasonable price. In particular, 3D depth cameras measuring the *time of flight* can provide marker-less position- and orientation detection of arbitrary objects in a 3D space; such cameras will soon be cheaply available as part of game consoles. What is more important is that our design

concepts for proxemic interactions apply are independent of the tracking technology, as long as the technology returns the four dimensions of proxemic relationships: position, orientation, movement, and identity.

CONCLUSION

We contribute extended notions of proxemic interaction, which is based on fine grained sensing of nearby people, objects, and digital devices. Through a scenario, we showed how proxemic interactions enable a multitude of implicit and explicit interactions with an interactive vertical display. In particular, we explained how knowing the continuous movement of an approaching identified person along with the position, orientation, and usage of identified digital devices and objects can be exploited in interface design, e.g., how the system should implicitly respond to proxemic entities and how the system can afford opportunities for explicit interactions. We generalize these as concepts for the design of proxemic interactions, all illustrated with episodes extracted from our fully-functional proximity-aware interactive media player.

We will continue our work in several ways. This includes: how proxemic interactions can mediate people's interaction with particular devices (e.g., digital cameras, picture frames); how proxemic interactions can facilitate inter-device connectivity and information transfer; and how security / access between devices can be done by knowing who, what, and where people are within an environment. We are also interested in investigating the scalability of these device ecologies, where many people and devices of different types may enter and leave the environment.

The largest unsolved issue in proxemic interaction is how one can configure the 'rules of behaviour', i.e., how the system should react to the proxemic information it gathers. While computers can take action based on its inference of the proxemic dynamics, it will sometimes get it wrong. Creating meaningful behaviours and repairing mistakes [11] will, we believe, become a central issue in the design of such systems. Even with this caveat, we believe that proxemic interactions will become a powerful way to realize embodied interaction, where – ideally – the system naturally responds to people's social expectations and practices in their everyday environments, and where mistakes are easily repaired [11] or of little consequence.

REFERENCES

1. Alwis, B.D., Gutwin, C., and Greenberg, S. GT/SD: performance and simplicity in a groupware toolkit. *Proc. of EICS '09*, ACM (2009), 265-274.

2. Buxton, W.A.S. Living in Augmented Reality: Ubiquitous Media and Reactive Environments. In K. Finn, A. Sellen and S. Wilber, eds., *Video Mediated Communication*. Lawrence Erlbaum Associates, 1997, 363–384.

3. Clark, H.H. Pointing and Placing. In S. Kita, ed., *Pointing. Where language, culture, and cognition meet*. Erlbaum, Hillsdale, N.J., USA, 2003, 243-268.

4. Diaz-Marino, R. and Greenberg, S. The proximity toolkit and ViconFace: the video. *Ext. Abstracts of CHI '10*, ACM (2010), 4793-4798.

5. Dourish, P. *Where the Action Is: The Foundations of Embodied Interaction*. The MIT Press, 2001.

6. Fitzmaurice, G.W. Situated information spaces and spatially aware palmtop computers. *Comm. ACM 36*, 7 (1993), 39-49.

7. Gellersen, H., Fischer, C., Guinard, D., et al. Supporting device discovery and spontaneous interaction with spatial references. *Pers. Ubiq. Comput. 13*, 4 (2009), 255-264.

8. Hall, E.T. *The Hidden Dimension*. Doubleday, 1966.

9. Harrison, C. and Dey, A.K. Lean and zoom: proximity-aware user interface and content magnification. *Proc. of CHI '08*, ACM (2008), 507-510.

10. Holmquist, L., et al. Smart-Its Friends: A Technique for Users to Easily Establish Connections between Smart Artefacts. *Proc. of Ubicomp '01*, Springer (2001), 116.

11. Ju, W., Lee, B.A., and Klemmer, S.R. Range: exploring implicit interaction through electronic whiteboard design. *Proc. of CSCW '08*, ACM (2008), 17-26.

12. Kray, C., Rohs, M., Hook, J., and Kratz, S. Group Coordination and Negotiation through Spatial Proximity Regions around Mobile Devices on Augmented Tabletops. *Proc. of TABLETOP '08*, IEEE (2008), 1–8.

13. Prante, T., Röcker, C., Streitz, N., et al. Hello. Wall–Beyond Ambient Displays. *Adj. Proc. of Ubicomp '03*, (2003).

14. Rekimoto, J. Pick-and-drop: A Direct Manipulation Technique For Multiple Computer Environments. *Proc. of UIST '97*, ACM (1997), 31-39.

15. Rekimoto, J. A multiple device approach for supporting whiteboard-based interactions. *Proc. of CHI '98*, ACM (1998), 344-351.

16. Rekimoto, J., Ayatsuka, Y., Kohno, M., and Oba, H. Proximal Interactions: A Direct Manipulation Technique for Wireless Networking. *Proc. of INTERACT '03*, (2003).

17. Schilit, B., Adams, N., and Want, R. Context-Aware Computing Applications. *Proc. of HotMobile '94*, IEEE (1994).

18. Shoemaker, G., Tang, A., and Booth, K.S. Shadow reaching: a new perspective on interaction for large displays. *Proc. of UIST '07*, ACM (2007), 53-56.

19. Swindells, C., Inkpen, K.M., Dill, J.C., and Tory, M. That one there! Pointing to establish device identity. *Proc. of UIST '02*, ACM (2002), 151-160.

20. Vertegaal, R. and Shell, J.S. Attentive user interfaces: the surveillance and sousveillance of gaze-aware objects. *Social Science Information 47*, 3 (2008), 275-298.

21. Vogel, D. and Balakrishnan, R. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. *Proc. of UIST '04*, ACM New York, NY, USA (2004), 137-146.

22. Want, R., Fishkin, K.P., Gujar, A., and Harrison, B.L. Bridging Physical and Virtual Worlds with Electronic Tags. *Proc. of CHI '99*, ACM (1999), 370-377.