Interaction Proxemics: Combining Physical Spaces for Seamless Gesture Interaction

Tilman Dingler[1], **Markus Funk**[1], **Florian Alt**[2]
[1]University of Stuttgart – VIS (Pfaffenwaldring 5a, 70569 Stuttgart, Germany)
[2]University of Munich – Group for Media Informatics (Amalienstrasse 17, 80333 München, Germany)

Figure 1: Providing seamless proxemic interaction for document reading: users perform text selection through touch (left), flick through pages using fine-grained finger-gestures (middle), and zoom from afar using coarse hand gestures (right).

ABSTRACT

Touch and gesture input have become popular for display interaction. While applications usually focus on one particular input technology, we set out to adjust the interaction modality based on the proximity of users to the screen. Therefore, we built a system which combines technology-transparent interaction spaces across 4 interaction zones: touch, fine-grained, general, and coarse gestures. In a user study, participants performed a pointing task within and across these zones. Results show that zone transitions are most feasible up to 2 m from the screen. Hence, applications can map functionality across different interaction zones, thereby providing additional interaction dimensions and decreasing the complexity of the gesture set. We collected subjective feedback and present a user-defined gesture set for performing a series of standard tasks across different interaction zones. Seamless transition between these spaces is essential to create a consistent interaction experience; finally, we discuss characteristics of systems that take into account user proxemics as input modality.

Author Keywords

Interaction; distance; proxemics; gestures.

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation (e.g. HCI): User Interfaces–Interaction styles.

PerDis '15, June 10 - 12, 2015, Saarbruecken, Germany
2015 ACM. ISBN 978-1-4503-3608-6/15/06$15.00
DOI: http://dx.doi.org/10.1145/2757710.2757722

INTRODUCTION

Over the past years we have seen a major change from conventional keyboard and mouse interaction to touch interaction and on-screen gestures. Touch input, for example, applies not only to mobile phones or tablets, but also to Desktop systems (cf. Windows 8). Today, taps, pan, and pinch gestures are established touch gestures that users expect from touch UIs. Meanwhile, gesture input has become popular in games and entertainment systems as more game consoles and televisions include means for sensing gestures (Kinect, Wii).

Researchers and practitioners increasingly deploy such technologies to create novel experiences in different contexts, including at work, at home, and in public [2, 7]. However, they are usually deployed with a limited set of interaction modalities, not accounting for the variable proximity between user and screen. We envision future displays to be aware of distant spaces and be flexible enough to assign different meanings and functionalities to user proximities – independent of technology and modality. A system that is aware of the room layout, yet does not focus on interaction modalities is Illumiroom [12]. Space-aware system create new challenges like a seamless user experience. Future applications should utilize technologies and modalities transparently, i.e., users should at best not even be aware of technology or modality changes.

In our work we address this by focusing on seamless touch and gesture-based interaction. In particular, we believe our research to be meaningful for environments with large screens that users can interact with both from near the screen as well as from a distance. An example is an open workspace, where a wall display could be used as a personal workplace, a shared space for discussions with colleagues, or as a meeting space where people are located up to several meters away.

At the focus of this research, we explore the combination of different interaction spaces and their perceived effectiveness. We first *define four interaction spaces* (Figure 2): (1) a touch area which allows for direct screen interaction, (2) a fine-grained gesture area in close screen proximity, (3) a general gesture area and (4) a coarse gesture area covering the space afar. We *implemented a prototype* that allows for seamless desktop interaction between big gestures, supporting both coarse gestures from afar as well as more fine-grained gestures and touch interaction. We then *evaluated our system* with regard to possible applications in a large screen environment, focussing on performance and the creation of an appropriate gesture set.

One scenario, where different proximities may be appropriate, is comfortably reading a document (Figure 1). At a medium distance for general reading, a zooming gesture may be performed with two hands where parts of the document are magnified. Closer to the screen, pages can be scrolled or turned by a one-finger swipe gesture, whereas highlighting text may be best done by direct touch of that passage. Another scenario applies to public displays or demo displays at exhibits, where there is much space available in a multi-user environment. People standing in the background may be assigned different interaction functions than people in the front. Deliberate application design can make users move between spaces and trigger user-to-user interaction [5].

The contribution of this work is threefold:

1. We introduce a framework for structuring the space in front of a screen that enables seamless interaction across 4 zones.
2. We present a set of user-defined gestures across this interaction space.
3. We conclude with a discussion about using user proxemics as an additional input modality for applications.

BACKGROUND AND RELATED WORK

Our work draws from several strands of prior research, most notably, interaction with large screens as well as gestures.

Interaction with Large Screens

Numerous techniques for interaction with large displays have been investigated in the past, mostly focussing on touch, smartphone-based interaction, and mid-air gestures [7]. For example, Rekimoto et al. [22] presented *PreSense*, a finger sensing input where the keypad is augmented with conductive tops to allow touch detection before a key is actually pressed and show previews of the action's effects. The approach has later been applied to mobile phones by Holleis et al. [11]. Further techniques have been proposed to use smartphones as pointing devices [3, 24] and to manipulate content on public displays [1, 4]. Mid-air gestures are getting more popular, in particular for playful applicaitons [19, 27, 29].

Despite a multitude of techniques being available, there are only few examples, where multiple techniques are used in parallel, for example *MirrorTouch* [18]. In general, display providers decide for one particular technique and this choice strongly depends on the task for which a display was designed [16, 17]. With novel technologies emerging we envision to see more multimodal user interfaces in the future. As a result, there is a need to focus more on the techniques itself rather than on the underlying technology.

Attempts to do so include modeling interaction with large screens. Streitz et al. [20, 26] provide a model of interaction with large displays in the context of "Hello.Wall". They identify three zones: an ambient zone, a notification zone, and a cell interaction zone. This model is mainly used to define the interactions offered, and the kind of information to be shown, on the display. Vogel et al. [28] refined the original zone model by further dividing the cell interaction zone into subtle and personal interaction zones. Many of these models follow the notion of proxemics, as introduced by Hall [10]. Greenberg et al. [9] identified a set of proxemic dimensions – distance, orientation, identity, movement, and location. Based on this, Marquardt et al. [15] presented a proximity toolkit that allows for simplifying the exploration of interaction techniques by providing proxemic information between people and devices. Further concrete examples that apply proxemics in interaction design include the work of Gellersen et al. [8], who looked into how knowledge about the spatial relationship of devices could be used to support spontaneous interaction. Rekimoto [21] proposed Pick-and-Drop, a manipulation technique that could be used to transfer information between different close physical objects. Shell et al. [23] introduced the notion of attentive user interfaces that use proximity to recognize human attention and react accordingly.

This non-conclusive list of examples shows the potential of such models and the proxemics notion when it comes to designing novel user interfaces. However, we argue that interaction in the display vicinity needs to be described in a more fine-grained way and the focus needs to be shifted towards the technique. In our work we draw upon this prior work by applying this notion to seamless gesture interaction and enhance the number of interaction dimensions.

Gestures and Gesture Sets

Gestures – "a motion of the body that contains information" [14] – have been extensively researched in the context of pen- and touch-based interfaces but research on using mid-air gestures in front of displays is relatively scarce. Gestures can in general be categorized into epistemic (tactile or haptic exploration of the environment), ergotic (manipulating objects), and semiotic gestures (communicating meaningful information) [6]. The latter two are particularly interesting for gesture-based interaction with displays, as they can be used for direct manipulation of virtual objects on the screen (ergotic) as well as for the execution of commands (semiotic). There are no standards, or even commonly accepted mid-air gesture sets for interacting with displays, so researchers and developers usually need to create their own gesture sets.

When considering the development of gesture sets, the following should be kept in mind: firstly new gesture sets should ideally draw upon existing operations that users are familiar with. Popular examples include the pan or pinch gesture users known from smartphones. Secondly, gestures belonging to a set should be coherent [30], thus making it easier for viewers to understand which gestures are supported by a system. Gestures also need to be easy to recognize for the system and at the same time be easy to teach.

To come up with an appropriate gesture set we follow methods of prior work. Wobbrock et al. let users design gestures for the use on interactive tabletops [31]. Kray et al. investigated gestures that allow for connecting mobile devices with large interactive surfaces, such as public displays and tablets [13]. Our methodology is different from the aforementioned in that we let users design gestures from scratch rather than using reference gestures. In this way we aimed at not constraining the users' creativity. Our result is a collection of user-defined gestures that applies to multiple, physical interaction zones in front of a screen.

INTERACTION SPACES AND DIMENSIONS

In a preliminary investigation we examined 10 work places in an office environment to assess different distances, in which screen interaction makes sense. Only workplaces with external screens measuring more than 24 inches in diameter were considered. We took measures of the space between screen and user in a front sitting position (i.e. engaged in a writing task), a rear sitting posture (i.e. flicking through images) and the total distance from screen to where the work place ended. Subjects kept a mean distance of 54 cm (SD=9.03), thus being able to operate mouse and keyboard in front of them. The rear sitting posture created a mean distance of 81.8 cm (SD=11.16) between user and screen, and the actively utilized space behind the screen was measured as 192.3 cm (SD=40.58). From these observations we derived a proximity concept distinguishing 4 interaction spaces (see Figure 2):

1. Touch zone: Direct screen interaction.
2. Fine-grained gesture zone: up to 0.5 m in front of screen.
3. General gesture zone: between 0.5 m and 2 m.
4. Coarse gesture zone: more than 2 m afar

Each zone offers a different granularity of interaction finesse. The idea is to move seamlessly between these spaces to transition from rather coarse to high-precision gestures. Hence, these 4 zones are not competitive, but complementary.

Touch Area

Touch – performed on the screen itself – presents the most direct interaction: it allows for direct manipulation of screen content and comprises additional dimensions, such as multiple hands, fingers and the amount of pressure applied. Common touch gestures include tap, drag, swipe, pinch and rotation. A well-known limitation is the *fat-finger problem* where fingers occlude parts of the screen during touch [25]. By lifting their fingers from the screen, users automatically enter the fine-grained gesture area. Thus, interaction can seamlessly continue while occluded spaces are revealed. Direct output modalities include visual, auditory, as well as haptic feedback.

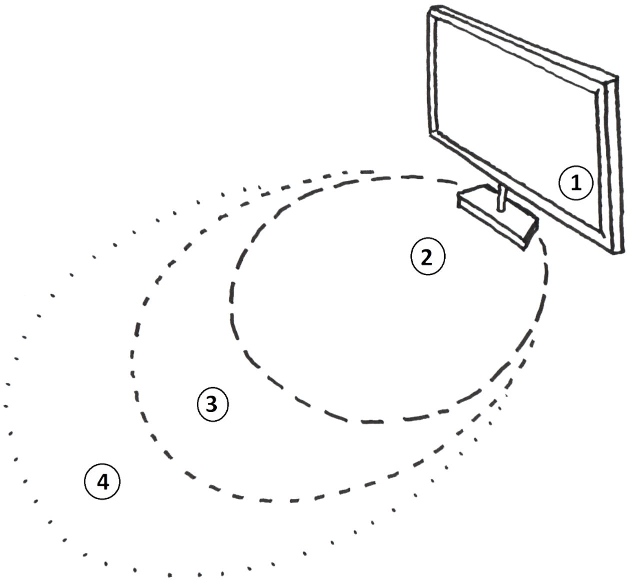

Figure 2: Four spatial zones to provide seamless gesture interaction: 1) Touch area, 2) fine-grained gesture area, 3) general gesture area and 4) coarse gesture area.

Fine-grained Gesture Area

This zone describes the space in immediate proximity to the screen. It basically begins where the touch is lifted and covers the general area where mouse, keyboard and other input devices are usually placed. Due to its close proximity, visual feedback can be given in high resolution and the input interaction itself can be performed by multiple hands and fingers. Since interaction happens in immediate proximity to the screen, a hovering gesture before the tab is introduced. In applications, this can be used to display tool tips, preview effects, or pre-select objects on screen. Furthermore, 10-finger interaction allows for fine-grained control and high-precision gestures. Common gestures include tap, click, swipe, pinch, rotation and grabbing. Feedback can be given visually on screen or via audio. In regards to a work setting user scenarios, this area would cover the space between user and screen where the user sits directly in front of the screen.

General Gesture Area

This zone describes the space between directly in front of the screen and the medium-far back of the space. This is where users usually comfortably sit (leaned-back) when they are reading a longer document, watching a video, or sitting and discussing an issue together with a colleague. Visual feedback needs to take this further proximity into account and display information in a more adjusted way. Users may interact with gestures using both hands. Feedback can be provided visually using medium-sized text and icon sizes as well as through audio. In a work and home environment this zone covers the area where users are engaged with tasks rather concerned with content consumption that does not require fine-grained interaction.

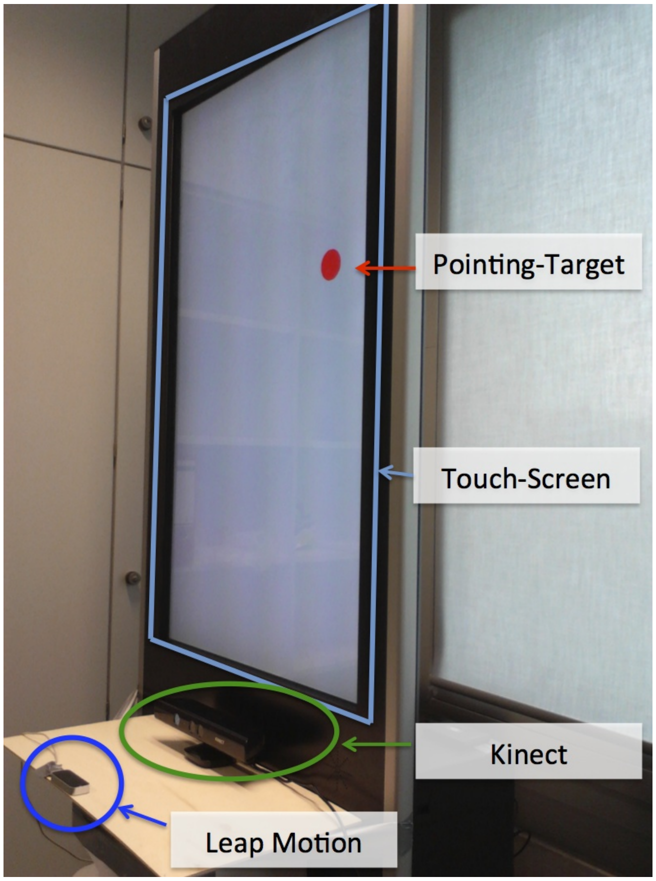

Figure 3: Setup of the prototype used during the user study. To cover the four interaction zones we employed a touch screen, a LEAP controller, and a Microsoft Kinect.

Coarse Gesture Area

This zone covers the physical space in front of the screen beyond the directly presumed interaction area. It includes the back of the room comprising scenarios where multiple people may be hovering over some content shown on the screen. Whole body tracking can be applied throughout the room and may sense people approaching or withdrawing from the screen. Scenarios include the living room where the TV can be controlled comfortably from the couch as well as public displays where people may be pouring over each others' shoulders in the back. Visual feedback should take into account the distance and display content in sufficiently enlarged manner. By detecting approaching people, the screen's content can actively attract people and invite them to interact, but also turn off automatically when people leave to save energy. Audio feedback may be used as well, although volume levels will need to be adjusted in case of different proximities in multi-user scenarios.

ZONE EVALUATION AND USER STUDY

To evaluate the feasibility of these four interaction zones we built a working prototype and conducted a user study to collect both quantitative and qualitative data. The quantitative part of the study included a pointing task whereas the qualitative part consisted of semi-structured interviews in which we asked participants to explore user-defined gestures involving the four interaction zones. We recruited 14 participants (11 male, 3 female) with a mean age of 26 (SD=3.33). We reached out to potential participants through University mailing lists. Most participants had a background in communication engineering or computer science. No one reported suffering from any type of color blindness. The initial briefing, conducting the experiment and defining the gestures took about 40 minutes per participant.

Apparatus

We created a standing working place as a demo scenario (Figure 3). To represent zone 1, we used a touch screen, placed on top of a small desk. The second zone was covered by a LeapMotion sensor, which used the forefinger as mouse input. For zone 3, we used the Kinect to track the participants' arms. When a participant pointed at the screen using the arm, the mouse cursor was moved accordingly. The LeapMotion and the Kinect were adjusted to create a seamless transition between zone 2 and zone 3. When the user was more than 2 meters away from the Kinect, zone 4 was activated. The measuring of the distance was done using the Kinect SDK. The software shows circle-shaped targets on the screen at random positions. Targets are colored according to the zones in which they can be reached: zone 1 – red, zone 2 – blue , zone 3 – black, zone 4 – green. Additionally, the cursor adjusts to the color of the zone in which the user is currently interacting.

Method

We split the user study into two parts: 1) a pointing task for collecting quantitative data and 2) an explorative part to define a gesture set for possible applications across different interaction zones. The first part was conducted using a repeated measures design with the interaction zones as independent variable. The order of the interaction zones was counterbalanced for each participant via Balanced Latin Square. As objective measures we collected the time a participant needed to reach the target depending on distance between subsequent targets and target size. For the second part, we collected subjective gesture ideas and user feedback through a semi-structured interview and a questionnaire.

Procedure

After signing the consent form, participants were asked to fill in a short questionnaire about demographics, their workplace, and their experiences in gesture-based input technology. Subsequently, we introduced them to the prototype and gave them sufficient time to make themselves familiar with the setup depicted in Figure 3. When participants felt comfortable with the prototype, we asked them to perform a series of pointing tasks: targets started popping up at different positions on the screen with different color codings which represented the corresponding zones in which they could be reached. To avoid confusion and the need for a visual legend, the experimenter additionally announced the target zone verbally. Targets were created within zones as well as across, 50 times each, in a way such that target sizes and distances were balanced to ensure equal difficulties across tasks. The algorithm evenly varied the diameter of the target between 10 and 100 pixels and avoided target overlap as well as edge contact.

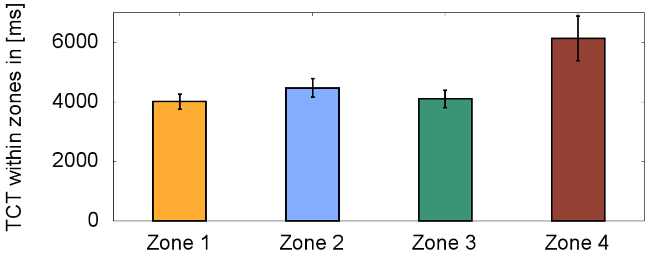

Figure 4: Task completion time for the pointing task within zones. Error bars depict standard error.

The second part of the study dealt with the creation of a gesture se for three application scenarios:

- **Navigation in documents**
  Actions: zoom, scroll, forwards, backwards, selection
- **Controlling applications**
  Actions: selection, copy, paste, cut, drag, drop, exit
- **Map-navigation**
  Actions: rotate, move, zoom out, zoom in

In a semi-structured interview we asked participants to invent and perform a 3D gesture for each action in each zone, which we recorded via video and which we transcribed. After defining gestures for each action in each zone, participants filled in a final questionnaire assessing the general gesture appropriateness with regard to the four zones. Participants rated the gesture appropriateness for each of the three application scenarios on a Likert scale from 1 to 5, where 1 was considered as 'not suitable at all' and 5 as 'very suitable'.

Results

In the initial questionnaire, most participants stated to be sitting in front of a monitor in their usual work environment. One participant sometimes used a standing-workplace. As far as their experience with gesture-based input technology went, one participant stated playing Nintendo Wii on a daily basis, while another participant was using a LeapMotion about once a week. Two of the participants were left-handed.

Task Completion Time

We statistically compared the TCT within the zones and across the zones using a one-way ANOVA. We used the Fisher's LSD correction to prevent type I errors in every analysis. Concerning TCT within-zones pointing tasks, the repeated measures ANOVA showed a statistically significant difference $F(3,652)$=4.831, p=.002. The post-hoc tests revealed that the differences between all zones compared to zone 4 were significant (all p<.05). Figure 4 shows the average TCT of all zones. The TCT was fastest in zone 1 (M=4007.7 ms, SD=3284.5 ms) followed by zone 3 (M=4105.8 ms, SD=3737.4 ms) and zone 2 (M=4467.8 ms, SD=3924.1 ms). The highest TCT was achieved in zone 4 (M=6133.7 ms, SD=9611.2 ms). However, we found no significant differences between zone 1, 2 and 3.

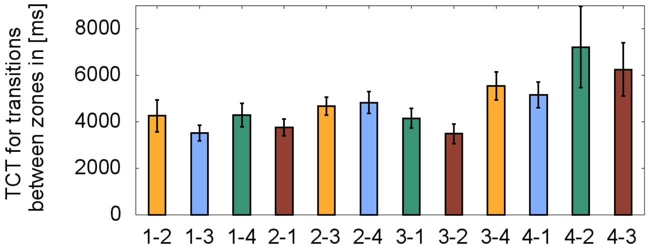

Figure 5: Task completion time for the pointing task across zones. Captions describe the zone transition. Error bars depict standard error.

Representation	Navigating in documents	Controlling applications	Navigating in maps
Zone 1	1.571	1.500	1.714
Zone 2	3.357	3.142	2.500
Zone 3	2.285	1.785	1.928
Zone 4	3.142	2.785	2.500

Table 1: Results of interaction categories in different zones (Mean Likert scale values, 1=not suitable, 5=very suitable.)

Regarding the TCT across zones, we compared the difference for each transition towards the screen and back. A repeated measures ANOVA revealed a significant difference $F(15,1299)$=2.536, p<.001. The post-hoc tests showed that the differences for the transition between zone 2 and zone 4 were significantly different (p=.031). All other differences in transitions between the same two zones were not significant. Figure 5 shows the average TCT of all transitions. The task completion time was between 3492.2 ms (SD=2947.1 ms) and 7223.7 ms (SD=13708.1 ms).

Gesture Set

We transcribed and summarized the user-defined gestures of all 14 participants to determine the most frequent gesture for each action in each zone. In the following, we present the gestures according to the three presented scenarios. Sketches of the defined gestures are attached in the appendix.

Application Control

Actions for generally controlling applications were: *select*, *copy&paste*, *cut*, and *exit*. Selection was preferred to be performed via direct touch (zone 1), finger tipping (zone 2) and hand tipping (zones 3, 4). Copy&paste was done via a lingering touch (zone 1) or by forming a fist (zones 2, 3, 4) over the selected object, dragging it to its target position and releasing it by opening the fist. For cutting an object participants defined a scissor-gesture (zones 2, 3, 4) and used a lingering two-finger selection in zone 1. The exit gesture was mainly agreed on by moving the cursor with a pointing finger (zones 1, 2) or hand (zones 3, 4) to the upper left corner of the screen.

Document Interaction

For navigating in documents, participants were asked to define gestures for the actions: *zoom*, *scroll*, *forward&backward*, and *select*. The gestures for zone 1 were fine-grained finger gestures as they were known from conventional applications: pinch, pan, swipe and a lingering touch for text selection. Interestingly, a fist gesture moving back or forth was chosen to perform zoom in zone 2, whereas for zone 3 and 4, participants chose to use both hands. Scrolling was done by finger (zone 2) and hand moving (zone 3, 4). Going forward or backward was represented by a directional hand motion in zone 2, 3, and 4. Text selection in zone 2 was done using two fingers, whereas a lingering open hand in zone 3 and 4 would start the selection gesture.

Map Navigation

Besides panning left and right, navigation in maps requires the actions *rotate* and *zoom*. For rotating, users defined a gesture rotating 2 fingers onscreen (zone 1) and an open hand rotation in zones 2, 3, and 4. Participants traditionally performed zoom by pinching in zone 1, whereas they preferred a single fist (zone 2) and two open hands (zone 3, 4), same as in document interactions.

Interestingly, gestures do not differ across zone 3 and 4, which clusters the gesture set into touch, fine-grained and general gestures from a distance. Table 1 shows the results of the Likert scale questionnaire regarding the different interactions in the different zones for each type of application. Participants liked the interaction in Zone 2 and Zone 4 more than the interaction in zone 1 and zone 3.

DISCUSSION

Results of the pointing task experiment indicate that selecting a target within zones takes a similar amount of time as selecting a target across zones, thus rendering actions across zones feasible. We argue, that mapping different actions to different zones will reduce the available actions per zone while adding additional interaction dimensions and therefore decrease the complexity of user interfaces. To make the mapping of actions and zones easy to learn and memorable, the mapping should be chosen with human intuition in mind. In the following we will discuss the implications and considerations of creating systems and applications supporting interaction proxemics.

To establish the boundaries of the 4 interaction zones, we were motivated by observations made in an office environment. The second part of our study showed, however, that gestures performed in zone 3 and 4 were quite similar, hence leaving us with a functional trisection of space in front of a screen: touch, fine-grained and coarse gestures. When defining gestures for close zones, fine-grained hand and finger gestures turned out to be feasible. Our study showed that participants defined mostly finger gestures when being close to the screen. Particularly when performing high precision tasks, finger gestures provide a higher granularity. At the same time, gestures using the entire hand or other full body parts are more feasible in distant zones. Study participants chose the same gestures for zone 3 and 4 using the entire hand or both. Hence, there seems to be a saturation of gesture variety when a certain distance is reached. Further studies will need to be conducted to assess exactly where this maximum distance starts and how that distance may relate to the screen size. However, in distant spaces we have the increasing challenge of accidental interaction: input errors are more likely by performing accidental finger gesture, which is why less fine-granular gestures may be preferred. Fine-grained actions, such as selecting text in documents or cutting and pasting objects should be mapped to zones that are close to the display. Such tasks require a higher input precision as well as more accurate feedback when it comes to legibility of potentially small content changes. For example, when selecting text with high precision in a distant zone, a user will have difficulties viewing the selected text when displayed in small font. In contrast, we found that participants defined low-precision hand gestures for actions in zone 3 and 4. Hence, coarse gestures performed in these zones should focus on triggering rather simple functions, such as exiting an application or page turning during reading. Further, the mere presence of a person could be considered as an application input. Despite being constrained by the technology (Kinect and LEAP) that we used to track user gestures, the development of more sophisticated tracking technologies will enable seamless zone transitions, to a point where users might not even be aware anymore of particular zones, but rather perform intuitive gestures depending on screen proximity.

In our gesture study we noticed that users were generally biased by established metaphors. In recent years, interacting with touch technology has clearly established some conventions and mental models which people are able to transfer to other scenarios. Our study participants had a fairly good understanding and expertise with the proposed technologies, hence resulting gestures might have been influenced by such bias. However, we argue that developers building applications for pervasive displays, should consider space and users' proxemics as input modality and allow for both fine-grained and coarse gestures. As our results show, cross-zone interaction is both feasible and adds additional dimensions of interaction to an application, thereby spreading functionality across space and reducing potentially ambiguous interactions. In what way this might support learnability and memorability of application functions, should be studied in more detail.

CONCLUSION

In this paper we present an approach to partition gesture interaction into 4 spatial interaction zones based on the proximity between user and screen. Results from a two-part user study show that there is potential for mapping distinct system functions to different spatial areas. We identify a set of user-defined gestures for 3 different application types in each zone and argue that the granularity of gesture accuracy decreases with growing distance between user and screen. Seamless transitions are necessary in between these zones in order to avoid user flow disruption and to fluently integrate application functionalities across zones. Insights from our study are pointers for application developers who want to take into account user proxemics as input modality. In future work we will explore such applications using proximity to spread application functionality across space.

ACKNOWLEDGEMENTS

We thank our participants and Mina Metias for their support during the study. Special thanks go to Carla Harris for preparing the drawings. The authors acknowledge the financial support of the Future and Emerging Technologies (FET) programme within the 7th Framework Programme for Research of the European Commission, under FET grant number 612933 (RECALL).

REFERENCES

1. Alt, F., Kubitza, T., Bial, D., Zaidan, F., Ortel, M., Zurmaar, B., Lewen, T., Shirazi, A. S., and Schmidt, A. Digifieds: Insights into deploying digital public notice areas in the wild. In *Proc. MUM'11*, ACM (2011), 165–174.
2. Alt, F., Shirazi, A. S., Kubitza, T., and Schmidt, A. Interaction techniques for creating and exchanging content with public displays. In *Proc. CHI'13*, ACM (2013), 1709–1718.
3. Ballagas, R., Rohs, M., and Sheridan, J. G. Sweep and point and shoot: Phonecam-based interactions for large public displays. In *CHI EA '05*, ACM (2005), 1200–1203.
4. Boring, S., Baur, D., Butz, A., Gustafson, S., and Baudisch, P. Touch projector: Mobile interaction through video. In *Proc. CHI'10*, ACM (2010), 2287–2296.
5. Brignull, H., and Rogers, Y. Enticing people to interact with large public displays in public spaces. In *Proc. INTERACT'03* (2003), 17–24.
6. Cadoz, C. Le geste canal de communication homme/machine: la communication instrumentale. *TSI. Technique et science informatiques 13*, 1 (1994), 31–61.
7. Davies, N., Clinch, S., and Alt, F. *Pervasive Displays - Understanding the Future of Digital Signage*. Synthesis Lectures. Morgan and Claypool Publishers, 2014.
8. Gellersen, H., Fischer, C., Guinard, D., Gostner, R., Kortuem, G., Kray, C., Rukzio, E., and Streng, S. Supporting device discovery and spontaneous interaction with spatial references. *PUC 13*, 4 (May 2009), 255–264.
9. Greenberg, S., Marquardt, N., Ballendat, T., Diaz-Marino, R., and Wang, M. Proxemic interactions: the new ubicomp? *interactions 18*, 1 (Jan. 2011), 42–50.
10. Hall, E. T., and Hall, E. T. *The hidden dimension*. Anchor Books New York, 1969.
11. Holleis, P., Huhtala, J., and Häkkilä, J. Studying applications for touch-enabled mobile phone keypads. In *Proc. TEI '08*, ACM (2008), 15–18.
12. Jones, B. R., Benko, H., Ofek, E., and Wilson, A. D. Illumiroom: Peripheral projected illusions for interactive experiences. In *Proc. CHI'13*, ACM (2013), 869–878.
13. Kray, C., Nesbitt, D., Dawson, J., and Rohs, M. User-defined gestures for connecting mobile phones, public displays, and tabletops. In *Proc. MobileHCI '10*, ACM (2010), 239–248.
14. Kurtenbach, G., and Hulteen, E. A. Gestures in human-computer communication. *The art of human-computer interface design* (1990), 309–317.
15. Marquardt, N., Diaz-Marino, R., Boring, S., and Greenberg, S. The proximity toolkit: prototyping proxemic interactions in ubiquitous computing ecologies. In *Proc. UIST '11*, ACM (2011), 315–326.
16. Memarovic, N., Langheinrich, M., Cheverst, K., Taylor, N., and Alt, F. P-layers – a layered framework addressing the multifaceted issues facing community-supporting public display deployments. *ToCHI 20*, 3 (July 2013).
17. Müller, J., Alt, F., Michelis, D., and Schmidt, A. Requirements and design space for interactive public displays. In *Proc. MM'10*, ACM (2010), 1285–1294.
18. Müller, J., Bailly, G., Bossuyt, T., and Hillgren, N. Mirrortouch: Combining touch and mid-air gestures for public displays. In *Proc. MobileHCI '14*, ACM (2014), 319–328.
19. Müller, J., Walter, R., Bailly, G., Nischt, M., and Alt, F. Looking glass: a field study on noticing interactivity of a shop window. In *Proc. CHI'12*, ACM (2012), 297–306.
20. Prante, T., Röcker, C., Streitz, N., Stenzel, R., Magerkurth, C., van Alphen, D., and Plewe, D. Hello.Wall – Beyond Ambient Displays. In *Adjunct Proc. Ubicomp'03* (2003), 277–278.
21. Rekimoto, J. Pick-and-drop: a direct manipulation technique for multiple computer environments. In *Proc. UIST'97*, ACM (1997), 31–39.
22. Rekimoto, J., Ishizawa, T., Schwesig, C., and Oba, H. Presense: interaction techniques for finger sensing input devices. In *Proc. UIST'03*, ACM (2003), 203–212.
23. Shell, J. S., Selker, T., and Vertegaal, R. Interacting with groups of computers. *Comm. ACM 46*, 3 (2003), 40–46.
24. Shirazi, A. S., Winkler, C., and Schmidt, A. Flashlight interaction: A study on mobile phone interaction techniques with large displays. In *Proc. MobileHCI'09*, ACM (2009), 93:1–93:2.
25. Siek, K. A., Rogers, Y., and Connelly, K. H. Fat finger worries: how older and younger users physically interact with pdas. In *INTERACT '05*. Springer, 2005, 267–280.
26. Streitz, N., Röcker, C., Prante, T., Stenzel, R., and van Alphen, D. Situated Interaction With Ambient Information: Facilitating Awareness and Communication in Ubiquitous Work Environments. In *Proc. HCI International'03* (2003).
27. Ten Koppel, M., Bailly, G., Müller, J., and Walter, R. Chained displays: Configurations of public displays can be used to influence actor-, audience-, and passer-by behavior. In *Proc. CHI'12*, ACM (2012), 317–326.
28. Vogel, D., and Balakrishnan, R. Interactive public ambient displays: Transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proc. UIST'04*, ACM (2004), 137–146.
29. Walter, R., Bailly, G., and Müller, J. Strikeapose: Revealing mid-air gestures on public displays. In *Proc. CHI'13*, ACM (2013), 841–850.
30. Wigdor, D., and Wixon, D. *Brave NUI world: designing natural user interfaces for touch and gesture*. Morgan Kaufmann, 2011.
31. Wobbrock, J. O., Morris, M. R., and Wilson, A. D. User-defined gestures for surface computing. In *Proc. CHI'09*, ACM (2009), 1083–1092.

APPENDIX

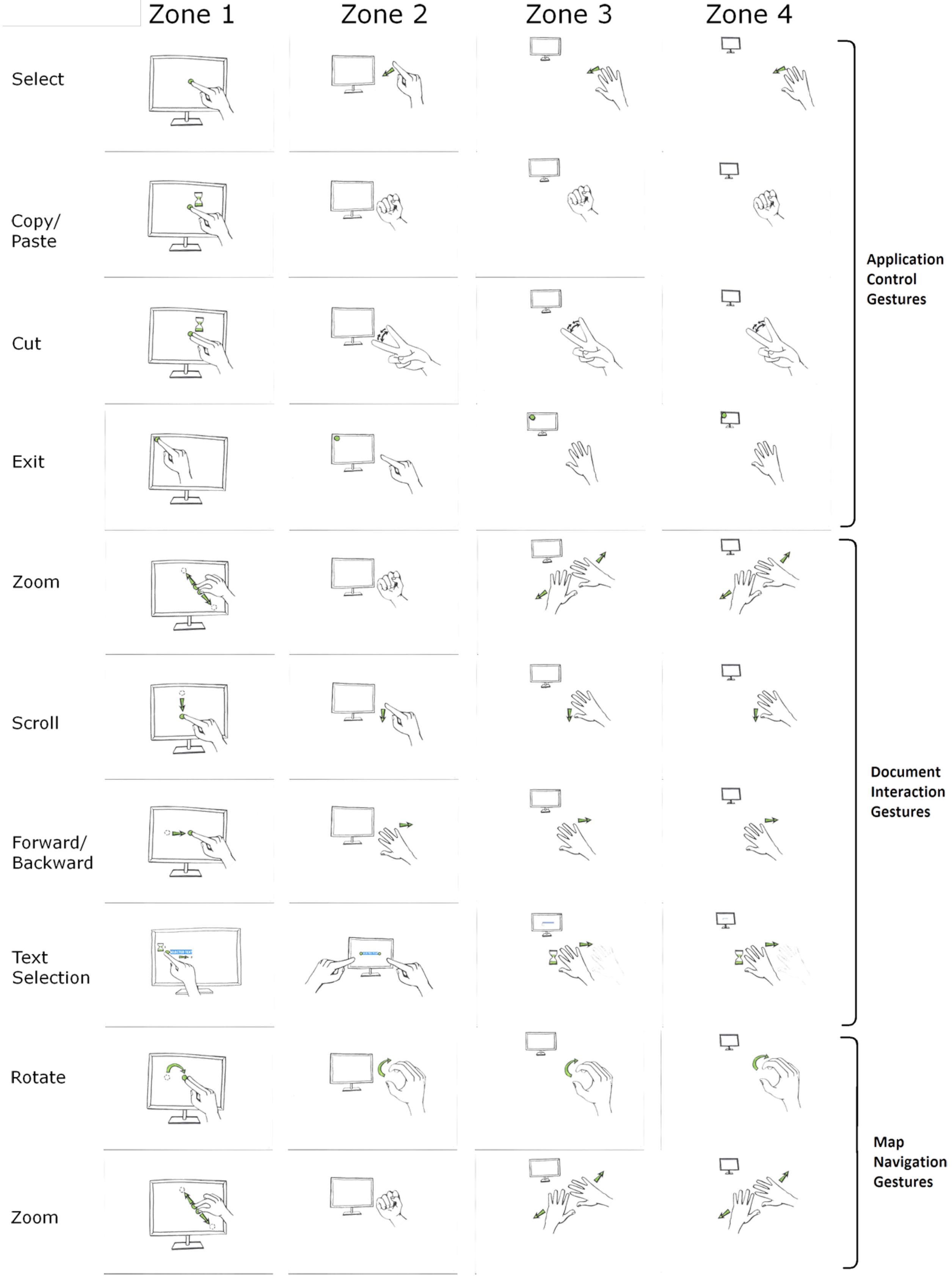

Figure 6: User-defined gesture set for interaction in documents.