Cuenesics: Using Mid-Air Gestures to Select Items on Interactive Public Displays

Robert Walter[1] robert.walter@tu-berlin.de
Gilles Bailly[2] gbailly@mpi-inf.mpg.de
Nina Valkanova[1,3] nina.valkanova@upf.edu
Jörg Müller[1,4] joerg.mueller@acm.org

[1]Quality and Usability Lab, Telekom Innovation Laboratories, TU Berlin, Germany
[2]Graphics, Max-Planck-Institut fuer Informatik, Saarbruecken, Saarland, Germany
[3]Music Technology Group, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain
[4]Alexander von Humboldt Institute for Internet and Society, Berlin, Germany

Figure 1. People performing mid-air gestures to select items on interactive public displays. The system was installed at three locations: (L1) on a large wall-projection in a coworking space, (L2) on a rear-projection screen in a students' cafeteria, (L3) on an LCD at a venue's opening event

ABSTRACT

Most of todays public displays only show predefined content and do not allow users to change it. We argue that interactive public displays would benefit from immediately usable mid-air techniques for choosing options, expressing opinions or more generally selecting one among several items. We propose a design space for hand-gesture based mid-air selection techniques on interactive public displays, along with four specific techniques that we evaluated at three different locations in the the field. Our findings include: 1) if no hint is provided, people successfully use *Point+Dwell* for selecting items, 2) the user representation could be switched from *Mirror* to *Cursor* after registration without causing confusion, 3) people tend to explore items before confirming one, 4) in a public context, people frequently interact inadvertently (without looking at the screen). We conclude by providing recommendations for designers of interactive public displays to support immediate usability for mid-air selection.

Author Keywords

Interactive public displays; Selection Techniques; Mid-air gestures; Immediate Usability

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

MobileHCI'14, September 23 - 26 2014, Toronto, Canada.

ACM 978-1-4503-3004-6/14/09 $15.00.
http://dx.doi.org/10.1145/2628363.2628368

INTRODUCTION

The focus of public displays is transitioning from static broadcast displays to interactive displays. In particular, this enables displays to become multi-purpose [12] by enabling users to choose between different content and applications, thereby increasing their utility. However, multi-purpose public displays also raise new challenges for the interaction [12]: How to enable users to switch between applications and games, browse content or express their opinion (e.g. voting [27])?

Interactive public displays raise challenges for both the choice of the *input modality* and the design of the *interaction technique*. Traditional input devices like keyboards and mice are typically not available in public spaces. Touch screens can also be inappropriate due to hygienic reasons or because the display is installed at locations that are not within the physical reach of passers-by [29]. Finally, remotely using an input device for input (e.g. smart phone) does not allow users to interact instantly, as they have to fetch their phone and install/launch the corresponding app before.

Selection techniques for public displays have to fulfill certain requirements that are very different from those on personal computers or mobile phones. On personal devices, users usually perform many thousands of selections, causing selection time and error rate to be the most important factors. In contrast, most users of public displays are expected to be first time users [29], only performing relatively few consecutive selections. Thus, immediate usability becomes much more important than efficiency. Furthermore, users may give up if they do not immediately succeed with the interaction [17].

In this paper we investigate how to support users to select items on public displays using mid-air gestures. Mid-air gestures have several advantages for interactive public displays

[29]. They allow users to interact from a distance, may foster discoverability and collaboration and are fun to perform [19, 26, 29]. However, it is not clear how to design immediately usable mid-air selection techniques for public displays.

To guide this process, we present a design space with five dimensions: *Selection Gesture*, *Confirmation Gesture*, *Input Space*, *Layout*, and *User Representation*. The design space aims at highlighting important design choices and helps designers to categorized existing techniques. We illustrate the applicability of our design space by comparing mid-air selection techniques used in the context of distant displays.

As an application for mid-air selection on public displays we implemented an interactive polling tool, that allows users to make a vote on a given topic. We conducted a field study to evaluate four selection techniques and investigate the impact of user representation and input space on immediate usability. We deployed the system at three different locations for a total period of three weeks. We report on our findings of three studies, which include:

- After a guided registration, no further instruction is required to allow people to successfully select and confirm items using *Point+Dwell*. Some people tend to perform other confirmation gestures than dwelling (push, grasp, wave, etc).
- The user representation can be switched from *Mirror* to *Cursor* after registration without confusing users.
- People explore the options before confirming one item. In our studies, they selected 2.5 (out of 6) items in average before confirmation.
- In a public setting, people may interact inadvertently with the display, without even looking at it.

We believe that these findings, our design space, and the proposed selection techniques are useful for researchers and practitioners of public displays and digital signage. For researchers, our techniques provide a baseline for developing novel selection techniques. The availability of immediately usable mid-air selection techniques may support the investigation of more complex interactions with public displays. Both users and practitioners can benefit from multi-purpose public display applications.

RELATED WORK

We first discuss the challenges of interactive public displays, highlighting different phases of interaction and then discuss related work on techniques for choosing options with a focus on mid-air techniques.

Interaction with Public Displays

Before users can interact with the system, they need to (1) notice the display, (2) understand that the system is interactive, (3) understand how to interact, and finally (4) be motivated to interact. The challenging aspect is that the system only has a couple of seconds or minutes for communicating how it works [29], otherwise users may just leave. Additionally, a public display generally has only "one shot": Users may give up if they do not immediately succeed with the interaction [17].

Noticing Displays

Sometimes people do not even notice public displays. Huang [9] has shown that most public displays receive very few glances. Mëller [20] finds that people often expect nothing useful on public displays and speculates that therefore, they may intentionally ignore them. These works, as well as [10, 11, 18] investigate some factors that attract attention for public displays, like installation height, motion, context, etc.

Communicating Interactivity

Once passers-by have noticed the display, they need to understand that it is interactive. For communicating interactivity, techniques like *call-to-action* [14] have been proposed. The *Honeypot* effect [5] is effective in demonstrating interactivity when other users already interact with the system.

Understanding Initiation of Interaction

When users have understood that a display is interactive, they need to understand how to initiate interaction, or more specifically how to *register* [30] a gesture. While this appears apparent for touch-screens [22] (touching the screen), it is less obvious for mid-air gestures. Walter et al. [29] have investigated how to provide hints for mid-air gesture registration. They found that a descriptive textual hint, combined with an icon representation of the gestures presented at the bottom of the screen (space division) is most effective.

Understanding Interaction Techniques

After these steps, users need to understand *how* to interact with the display, e.g., how to select an item. This question has not yet been addressed for mid-air gestures in the context of public displays. Therefore, we take inspiration from other device classes, such as desktop computers and large displays.

Selecting Items

Items are often selected through menu bars, context menus, toolbars, palettes, etc. In particular, menus received a lot of attention in the literature including empirical studies (e.g. [21]), predictive models or various interaction techniques (e.g. [15]). However, most of these works focus on dedicated input devices (e.g. keyboard or mice) rather than mid-air gestures.

Several menus have been designed to exploit various input or output modalities such as multi-touch [1, 16], pen interaction [25], mobile devices [31, 7], remote control [2] or virtual reality [4]. In [8] a set of five pre-defined mid-air selection gestures for large displays is evaluated in a laboratory study regarding their *intuitiveness* and *effectiveness*, showing that *Dwelling* is the most *intuitive* gesture for selection. In our work we build on that finding and evaluate different designs of selection techniques in a field study, focusing on *one* specific selection gesture (*Point+Dwell*).

Mid-air Gestures and Large Displays

While we are not aware of existing work on selecting items using mid-air gestures on public displays, a few systems focusing on large displays have been proposed [8, 13, 24, 28, 3].

For instance, Bailly et al. [3] proposed a mid-air menu technique for large displays. Users extend their fingers to select a command in linear menus in front of a distant display and the system counts the number of fingers. However, these different menus are not well suited for public displays because they are not compatible with immediate usability since they require explicit teaching.

DESIGN SPACE

We propose a design space of hand-gesture based mid-air selection techniques for public displays. It aims at helping designers and practitioners to design selection techniques by describing and comparing their main features. This design space has been derived from the analysis of differences and similarities of existing mid-air selection techniques in NUI frameworks, commercial products and related work.

We distinguish three phases for selecting items on a public display using mid-air gestures (see Figure 9):

1. *Registration:* Users express their intention to make a selection by performing a specific action to display a list of available items. This phase has been particularly investigated in *StrikeAPose* [29]. In this paper we distinguish different modalities of registration.
   - *Modeless:* Selection is possible at any time. As soon as a users steps in front of the screen, they are instantly able to select an item. This registration technique is useful when the selection itself is an essential element of the system, as in [27].
   - *Quasimode:* For the *quasimode*, or *user-maintained* registration, a dedicated action (e.g. a posture) has to be actively maintained by the user to allow for selection (similar to modifier keys on keyboards). This technique has been proposed in [29].
   - *Modal:* After registration, the application switches to a special mode, that is only dedicated to the act of selection. The list of options is displayed once a dedicated registration gesture has been performed until the selection has been confirmed or explicitly canceled.
2. *Selection:* The display shows all the different selectable items, and users select and can explore them.
3. *Confirmation:* Once users have decided on an item, they confirm that selection in this phase. Only after this phase a command is triggered to the system.

While several design spaces of selection techniques and menus have been proposed in the literature (e.g. [23]), they mainly focus on mouse or keyboard-based interaction. As mentioned in the related work section, *public displays* is a specific context requiring to revisit existing techniques to accommodate immediate usability [19, 29]. In this design space we focus on mid-air hand gestures. We do not intend to cover other body-based selection techniques (e.g. finger-based [2] or position-based techniques [27]). Furthermore we only focus on selection and confirmation (phase 2. and 3. above). In the following, we discuss the different dimension of our design space.

D1: *Selection Gesture* (Figure 2). During selection, users may explore different options, before confirming one. While several hand gestures have been proposed, we focus our analysis on easily guessable gestures in favor of immediate usability.

Figure 2. Selection Gesture: 1) Point, 2) Swipe

1) *Point*: Cursor-based pointing: users move the representation of their hands over the desired item in order to change the selection (similar to hovering with mouse on items before clicking). Among the systems we analyzed (Table 1), pointing is the most commonly used selection technique.
2) *Swipe*: Users swipe with their hand in mid-air over a certain region on the screen in order to change the current selection. This can be compared to a *scrolling* mechanism.

D2: *Confirmation Gesture* (Figure 3). After users selected the item, they finally confirm their selection, which marks their final decision. The following hand gestures may be used to confirm selections.

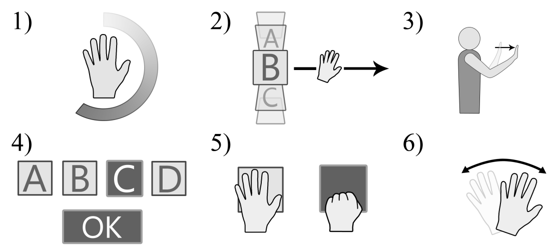

Figure 3. Confirmation Gesture: 1) Dwell, 2) Swipe, 3) Push, 4) Pointing, 5) Grip, 6) Wave

1) *Dwell*: Users rest their hand over an item. This is one of the most commonly used confirmation gestures. To define an appropriate dwell time is challenging as it may be perceived too short by novices and too long by expert users [6].
2) *Swipe*: Users swipe their hand in mid-air over an item in order to confirm their selection. This technique may be combined with both *Swiping* (different direction) and *Pointing* for selection (D1).
3) *Push*: Similar to *Swipe*, a quick movement along a dedicated axis is performed to confirm a selection. An additional DoF (depth) is used for that confirmation gesture, which enables the technique to be combined with both *One-Dimensional* and *Two-Dimensional* layouts (see D4).
4) *Point*: Users point in a second stage to a dedicated item (similar to touching *OK* button on touch screens). This approach reduces the necessity for dwell times but could hamper immediate usability, as a second pointing stage is required.
5) *Grip*: Users perform a grip or grasp motion hovering over the selected item in mid-air.
6) *Wave*: Users wave with their hand over the selected item.

D3: *Input Space* - As we focus on hand gestures, there are mainly two ways to extend the arm in order to invoke a selection (Figure 4).

Figure 4. Input Space: 1) In Front of User, 2) Around User

1) *In Front of User*: Users reach out their arm to the front in order to select items, as they would typically do for pointing gestures (D1).
2) *Around User*: Users may also reach out to their sides (left, right and top) to reach for items that are available for selection. If a *Silhouette* or *Avatar* user representation (D5) is applied, using this input space reduces occlusion with the users' representation and the selectable items.

D4: *Layout* - Different ways to layout items in the selection are possible that mainly differ in dimensionality (Figure 5).

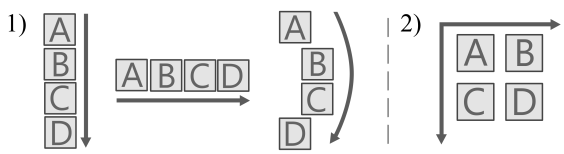

Figure 5. Layout: 1) One-dimensional, 2) Two-dimensional

1) *One-dimensional*: Selections can only be changed in one direction. Users move their hand along a straight or curved axis to select items. While one-dimensional layouts typically require less constrained hand movements than two-dimensional ones, the number of selectable items at a time is smaller.
2) *Two-dimensional*: Users exploit two degrees of freedom in hand movement to change selections, typically x, and y-direction. Such a layout is used in the *Kinect Hub* menu system on the *Microsoft Xbox 360*.

D5: *User Representation*. Mid-air interaction techniques mostly require some sort of visual feedback representing the user (Figure 6). This dimension has already been investigated for communicating the interactivity of a public display [19].

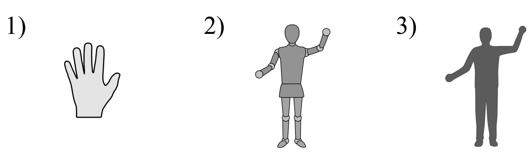

Figure 6. User Representation: 1) Cursor, 2) Avatar, 3) Mirror Image

1) *Cursor*: The user is reduced to a small pointer or cursor on the screen that moves according to the users' hands. This kind of representation only requires a small amount of screen space. It may not be immediately obvious to which of the both hands the cursor is attached. Users have to trial-and-error for both hands and observe the feedback in order to determine the correct hand. Moreover, in a multi-user scenario, which is typical for public displays [19], it may be challenging for users to identify the cursor that belongs to them.
2) *Avatar*: The user is represented by a virtual avatar that mimics their motions. Not only the hand movements are mapped to the representation of the user, but also the movements of other joints. As a downside, significantly more screen space is required for that representation. While this technique does not help to notice interactivity [19] of the display, it may work well after the registration phase, when the user is already aware that the display is interactive.
3) *Mirror Image*: The user is represented by a mirror image, or sometimes only by a simplified version of it: a silhouette on the screen. Unlike an *Avatar*, not just the movements but also the shape or outline and the colors of the representation correspond to the user. This representation provides the highest fidelity. It also works best for communicating the interactivity of a public display [19].

	D1		D2						D3		D4		D5		
System	Point	Swipe	Dwell	Swipe	Push	Point	Grip	Wave	Front	Around	1D	2D	Cursor	Avatar	Mirror
Kinect Hub [1]	✓		✓						✓			✓	✓		
Kinect Adventures [1]	✓		✓						✓		✓		✓	(✓)	
Dance Central [1]	✓			✓						✓	✓		✓		(✓)
Sonic Free Riders [1]		✓		✓					✓		✓		✓		
Your Shape [1]	✓		(✓)			✓				✓	✓				✓
Video-place [13]	✓		✓							✓	✓				✓
Samsung Smart TV	✓						✓		✓			✓	✓		
K4W [2] SDK [2]	✓				✓		(✓)		✓			✓	✓		(✓)
OpenNI NiTE	✓	(✓)	✓	(✓)	(✓)			(✓)	✓			✓	✓		

Table 1. Categorization of techniques according to the design space (✓: technique is applied; (✓): technique is applied secondarily)

Table 1 categorizes a set of mainly commercial products and systems according to the dimensions of this design space. The features that are used primarily by a system are checked (✓) in the table. Secondary techniques are checked with a lighter shade (✓). For example *Kinect Adventures* primarily uses a *Cursor* user representation, but also shows an *Avatar* representation in the background (secondary).

[1] Available for Microsoft XBox 360
[2] Microsoft Kinect for Windows

As shown in Table 1, there is already a large variety of techniques among this exemplary set. A commonly used standard or gestural language for mid-air selecting has *not* yet been established. Also it is not yet clear, which of the techniques support immediate usability best. For this reason we investigated the dimensions of this design space in a series of three studies. In an initial pilot study, we focus on gestural input, represented by the first two dimensions (*Selection* and *Confirmation* gesture).

PILOT STUDY

To better understand the mental model of users, we conducted a preliminary pilot study with 23 participants. The goal of this study was to identify immediately usable actions for *Selection Gesture* (D1) and *Confirmation Gesture* (D2), based on a *Mirror Image* representation (D5) and *Around User* as input space (D3). We presented participants a large real-time silhouette representation of themselves, along with a minimalistic body-centric menu-like overview of selectable items (Figure 7). We conducted the study as a Wizard of Oz experiment: after observing the registration gesture, the wizard manually activated the menu showing the items.

Participants were asked to select the highlighted item and confirm their selection, using what they think would be the most "intuitive" actions while no further hint was provided. They were free to try multiple actions consecutively. The experimenter highlighted that there are no wrong or right actions, and asked the participants to *think aloud* and freely express their thoughts and considerations during the task. After participants stated to have finished the task, the experimenter proceeded with a semi-structured interview asking to describe the actions they performed in detail.

To see if selecting items at different positions in the selection would afford different gestures for selection and confirmation, we randomly changed the highlighted item after each participant. We counted the frequencies of all observed gestures. We analyzed both *first* gestures and alternatives of actions that people investigate *after* performing the first gesture.

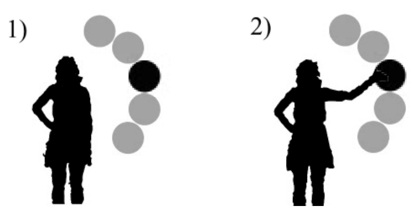

Figure 7. *Mirror Image* user representation with five items attached. Participants were asked to demonstrate how they would select and confirm the highlighted item.

Results

The following gestures and actions were used by the participants in order to select items and confirm their selections.

Selection Gesture: For a large majority of participants (22 of 23) we observed *Point* to be the most frequently performed gesture for selection: Users moved their hand over the highlighted item to signal their choice (Figure 7.2). Even after trying different actions consecutively, people sticked to pointing as a general behavior, only altering in hand postures. Considerable variations include pointing with 1) open hand, palm facing towards screen (18 of 23), 2) only index finger extended (3 of 23), 3) palm turned away from the screen (4 of 23). Six unique hand postures while pointing were observed. No effect of the item position on the performed selection gesture was identified.

Confirmation Gesture: Confirming [8], we observed *Dwell* to be the most frequently performed first action (15 of 23 participants) for confirming a selection. Still, there was a large variety of observed gestures including *Grip* (4 of 23) in different variations: grip once and dwell, grip repetitively, *Plucking* (grip and pull towards themselves). Plucking was also among the most popular alternative actions that people performed after their first action (3 of 23). Furthermore, we observed people repetitively tilting their palm towards and away from the screen (3 of 23). In total we noted 15 unique gestures for confirmation. We observed no effect of the item position on the performed confirmation gesture.

We have learned, that most people would point and dwell to select items if no hint is provided. For our design (Figure 7), different item position did not afford different gestures. Yet, it is not clear if other designs would trigger different user behavior. We decided to implement *Point+Dwell* as selection and confirmation gesture and investigate further dimensions of the design space in a iterative design study.

ITERATIVE DESIGN STUDY

The goal of this user study was to investigate two dimensions of the design space: input space (D3) and user representation (D5). We are particularly interested in *Mirror Image* and *Cursor* user representations in the selection phase. We iteratively improved two techniques that use two different user representations (D5) to improve immediate usability of both. For the registration phase we use a technique proposed in *StrikeA-Pose* [29]: Users are represented by a virtual *Mirror Image* on the screen, and an animated *Text+Icon* label describing the registration gesture is shown (compare Figure 9 left). As proposed [29] we use the *Teapot* gesture for registration: User may use their left (right respectively) hand to touch their left (right) hip. In the initial state of the study, the registration was implemented as *quasimode*: As users release their hip, the selectable items would disappear again.

Selection Techniques

We used two selection techniques that were inspired by commercial systems as a starting point, and iteratively adjusted them during the study to improve immediate usability.

Cursor-Front: This technique uses a *Cursor* user representation and an *In Front of User* input space. It was inspired by the *Kinect Hub* menu system for *Xbox 360* (Figure 8 left). While the *Kinect Hub* version uses a *Two-Dimensional* layout, we decided to use a *One-Dimensional* one to remain consistent with the other technique. Additionally to the cursor, there is a real-time PiP (picture-in-picture) *Mirror Image* representation of the user shown in the bottom right corner, with the hands highlighted. For the registration phase, a *Mirror Image* user representation is used, that is switched to *Cursor* for the selection phase.

Mirror-Around: This technique uses a *Mirror Image* representation and an *Around User* input space. It was inspired by the *Your Shape* menu system for *Xbox 360* (Figure 8 right). Since we also use a *Mirror Image* user representation in the registration phase, the visualization does not change after registration.

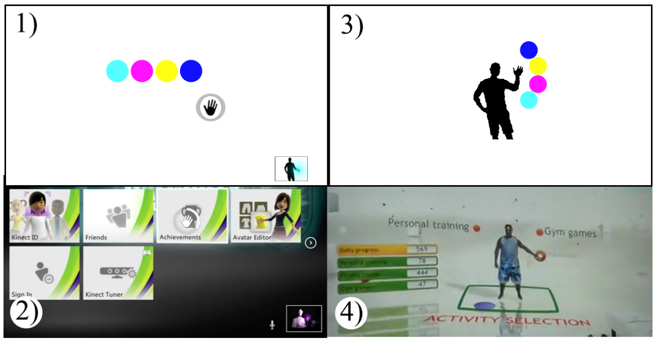

Figure 8. The two iterative design study conditions in their initial states: 1) Cursor-Front, 3) Mirror-Around, and in-game screenshots of the two corresponding systems that they were inspired by: 2) Kinect Hub, 4) Your Shape.

Apparatus and Participants

The system was installed for four days in a passage in close proximity to the main entrance of a university building. We randomly invited passers-by in the entrance to participate in a five-minute experiment. In total 51 passers-by participated in the test. For the entire time of the interaction the system logged the raw depth video, a screen capture, a camera recording and various user events to a text file. We noted user comments during the test and conducted a semi-structured interview at the end of each test. The questions included 1) how they selected the item, 2) what problems they encountered, 3) what improvements they suggested, and 4) which technique they preferred and why. Based on observations and user feedback, we adjusted the initial techniques on a daily basis.

Instructions and Task

We informed each participant, that the study is about interactive public displays and selecting items. As we focus on immediate usability, no further specific information was provided to resemble a field-situation. Each participant was asked to initially select an arbitrary item (color) using one of the two techniques (switched after each participant). After this first trial, we introduced the participants to the other technique and asked them to select ten consecutive items with each of the two techniques.

General Learnings

Modal Registration is less confusing: We observed problems with the *quasimode* registration, because it requires both hands. Especially in the first trial participants unintentionally released their hand from their hip while trying to select an item with the other hand. Some users even tried to use the same hand that they used for registration (e.g. touched their hip with the right hand and tried to make the selection with the right hand as well, consequently releasing their hip and causing the menu to disappear). It appears to be too complex for novice users to use both hands for two different sub-tasks (one for registration and one for selection). For this reason we introduced *modal* registration: after performing the registration gesture, the system switched to selection mode. The registration posture may be released for selecting. Using modal registration also allows to use one hand only for both registration and selection. Switching from quasimode to modal registration appeared to introduce less confusion to participants, especially in the first trial. As a downside, this step introduces another dwell time for entering the selection mode.

Registration Gesture should be meaningful: During the interviews some participants mentioned that while they understood and performed the registration gesture (touching the hip with one hand), they kept asking themselves for the reason: "I was confused. Do I have to touch my hip for calibration?", "I didn't know why it wanted me to touch my hip.", "Why touching the hip?", "Why do I have to touch my hip before the colors appear?". It appears that people do not wonder about the registration itself, but about the unusual gesture which is generally not used for human communication. For this reason we decided to change the gesture to a more meaningful one: raising the hand [27]. Though this gesture is less subtle, it's more clear for users of the system and onlookers, that they raise their hand in order to communicate with the system. This gesture is socially used to gain attention, so people may wonder less why they have to do it to initialize the selection.

Cursor-Front - Learnings

Dual-Cursor makes selection easier: Participants reported that they were confused about which hand they should use for selection: "You have to try which hand you can use!". We observed both behaviors, using the same hand that was used for registration and using the opposite one. For this reason we implemented two cursors to allow users to decide which hand to use. They could either use their left hand to control the left cursor (displaying a left hand), or their right hand to control the right cursor (displaying a right hand). This change appeared to reduce the confusion about what hand ought to be used for selection.

Representation switch does not cause confusion: As mentioned earlier for this technique, the representation of the user switches from *Mirror Image* to *Cursor* after the registration (compare Figure 9.1 to 9.2). During the switch, the mapping between hand position in space and pixel coordinates on the screen was preserved: the cursor appeared at the exact same position on the screen as the hand of the users' mirror representation. Initial concerns that this could confuse users were not confirmed. However, it is not yet clear if this could become a problem in a multi-user scenario.

PiP hint is ignored: After the first ten participants, we observed that seven of them did not notice the Picture-in-Picture (PiP) hint showing a *Mirror-Image* representation of the user at all (see Figure 8 left). Two participants noticed but ignored it, and only one actually looked at it. One even reported that he was distracted by it. We did not observe a changed user behavior after removing the picture.

Mirror-Around - Learnings
Dual-Selection causes confusion: We observed the same confusion about what hand to be used as in the *Cursor-Front* technique. Initially, the items appeared on the opposite side than the one that was used for registration (Figure 8.3), suggesting to use the opposite hand. Still we observed that participants used the same hand that they used for registration also for the selection (for example trying to use the right hand to reach for items that are displayed at the left side of their body). For this reason, as an analogy to the dual-cursor, we decided to put two identical, mirrored sets of items to both sides of the user. However this lead to even more confusion, as people were not sure if it makes a difference if they select items from the left or right side. To avoid constraining the free choice of the user what hand to be used to select items, we moved the items from the side to the top of the users' representation (compare Figure 8.3 to 10.4). At this position the selection can be reached with both hands equally well, allowing users to freely choose which hand they use for selection.

We have iteratively developed two selection techniques using different user representations (D5) and input spaces (D3) that support immediate usability on public displays. We have learned that 1) modeless registration causes confusion, 2) the registration gesture should be meaningful to users in the given context, 3) users should be able to freely choose which hand to use, 4) a switch of the user representation after registration does not necessarily cause confusion, 5) PiP hints may be ignore by users. To investigate if these findings also apply in an ecologically valid in-the-wild setting and to explore additional combinations of user representation (D5) and input space (D3), we decided to conduct a field study.

FIELD STUDY

We conducted a field study to evaluate the techniques and analyze user behavior in an ecologically valid setting. We deployed the system at three different locations (Figure 1):

Coworking Space (L1): We installed the system for two weeks in the cafeteria of Betahaus Berlin, a well visited coworking space. Visitors mainly include entrepreneurs, freelance software developers, as well as artists and designers.

Student's Cafeteria (L2): The system was also installed for four consecutive days in a large student's cafeteria. The deployment location was mainly visited around noon by students or staff of the Technical University of Berlin and Berlin University of the Arts, as well as locals living nearby.

Venue Opening (L3): On the occasion of an opening event of a large campus for startups and technology-driven companies in the city of Berlin, we deployed the system for one evening in a passage close to the entrance.

System: As an application for our selection techniques, we implemented an interactive public survey tool, that allows users to place a vote for one out of six different options to a given topic. We adapted these topics to the particular location in order to increase the participation rate. For example at the coworking space, we invited passers-by to "Tell us your passion!" (written in large letters in the upper area of the display, see Figure 1). The selectable options in the coworking space were 1) *Dev & Code*, 2) *DIY*, 3) *Media*, 4) *Electronics*, 5) *Art & Design*, and 6) *Business*. While for the registration phase a *Text+Icon* hint was provided, no further instructions for selection and confirmation were presented (Figure 9). The order of the items was randomly altered on a daily basis for all locations. As users choose one of the options, the system takes a photo of the their face and adds it to a tile-based poll visualization [27], that shows the distribution of users across the different options (Figure 1).

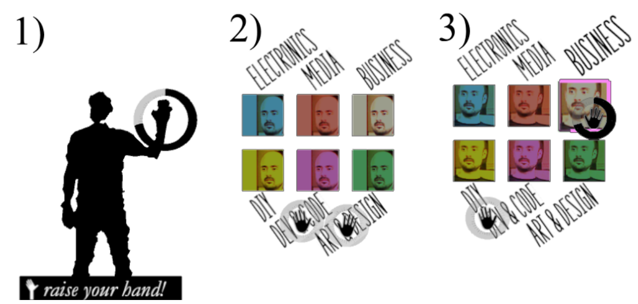

Figure 9. The three phases of interaction - 1) registration: a user represented by a *Mirror Image*, a registration-gesture hint is shown at the bottom. 2) selection: user moves their hand over an item. 3) confirmation: user dwells over desired item; displayed Technique is *Cursor-Front*

Implementation: The system is using the *Kinect* camera to capture passers-by. For visualization, the *Processing*[3] library is used. The software implementation is based on the prototypes of the previews studies. It uses the *OpenNI*[4] framework and the skeletal tracking features of the *NiTE* middleware.

Techniques: We tested four techniques (Figure 10) in total. *Cursor-Front* and *Mirror-Around* are the results of the iterative design study. The aim was to evaluate two dimensions of the design space (User Representation and Input Space). As User Representation (D5) we use *Cursor* and *Mirror Image*. As Input Space (D3) we use *In Front of User* and *Around User*. This dimension also affects the item layout (D4): if the input space is *In Front of User*, items can be arranged in a *Two-dimensional* layout, while for the other case the items are arranged in a *One-Dimensional* layout. The condition was changed randomly every ten selections.

Data Analysis: We recorded a screen capture, the raw depth video from the sensor, as well as various user events in a log file for the entire time of the deployment in all the locations. The most important automatically captured events in the log file include when users 1) step in front of the display, 2) perform the registration gesture, 3) make a selection, and 4) leave the display. During that time two researchers gathered data from on-site observations and semi-structured interviews with users. In particular we were interested in 1) if users are able to select the item they want, 2) what they think how to select items, and 3) how they knew how to select items. The recorded videos and material were later reviewed. Each instance of a user performing a selection was annotated by a researcher. Finally data for each selection include 1) which item / option was selected, 2) which condition was used, 3)

[3]http://processing.org/
[4]http://www.openni.org/

System	D1		D2						D3		D4		D5		
	Point	Swipe	Dwell	Swipe	Push	Point	Grip	Wave	Front	Around	1D	2D	Cursor	Avatar	Mirror
Cursor-Around (C1)	✓		✓							✓	✓		✓		
Cursor-Front (C2)	✓		✓						✓			✓	✓		
Mirror-Around (C3)	✓		✓							✓	✓				✓
Mirror-Front (C4)	✓		✓						✓			✓			✓

Table 2. Techniques that were investigated in the field study mapped to the design space

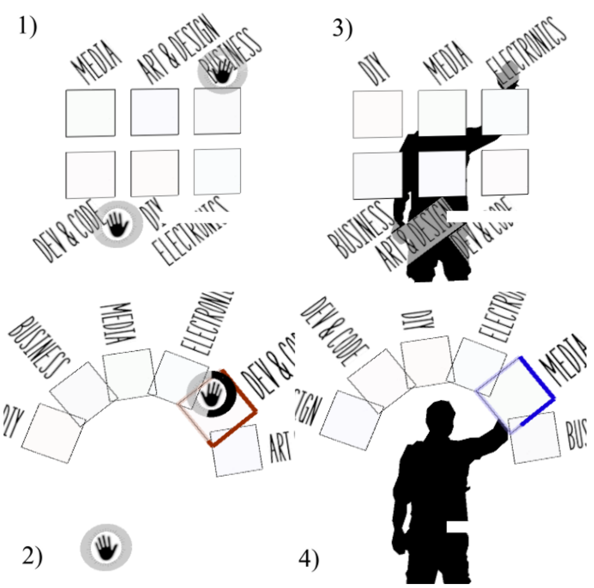

Figure 10. The four fieldstudy conditions: 1) *Cursor-Front*, 2) *Cursor-Around*, 3) *Mirror-Front*, 4) *Mirror-Around*

which hand was used for registration / selection, 4) how long it took for users to register / select / confirm, 5) which hand postures / confirmation gestures they use.

We report on 405 unique observed selections, excluding 1) repetitive selections of one user, 2) cases of wrongly recognized registration gestures, 3) researchers of the team and other people that are familiar with the system, and 4) unintended user interaction from the data.

Results

People explore the system before selecting: The average time from stepping in front of the screen until registering for selection was 13 seconds, using similar hints as in [29]. After registration, most users confirmed a selection within 10 seconds. This time includes 1) understanding how to select items, 2) exploration of the options and 3) temporal demand to execute selection gestures, as well as 4) dwell time (2 seconds). In average people selected 2.5 items before they finally confirmed one. They spent about 5 to 7 seconds for that exploration process. We did not observe an effect of the condition on the selection time, or on the amount of exploration.

No hint for selection required: One question in the interview was "Did you successfully select the item you wanted?" 47 (of 49) participants reported that they did, which indicates that throughout all techniques, most people were able to use the system and select the items they want. In the data we could not identify one significantly least, or most frequently selected position in the selection. Randomly altering the order of the options did not affect the survey results. Both indicates that selections were not random. We also asked people how they knew how to select an item. The majority of interviewed people (33 of 49) reported that they were able to make the selection based on their "intuition" or "gut feeling" and by just trying. Five participants reported that someone else explained them how to use the system, four participants reported to have experience with the *Kinect* and one stated he has learned from observing other people how to interact with the system. However, the majority (39 of 49) of people report that did not observe other people before trying to interact with the system.

Representation switch does not cause confusion: As in the lab study, for the cursor conditions (*Cursor-Front*, *Cursor-Around*) the user representation was switched from *Mirror* to *Cursor* after registration. We did not observe a difference in selection time (required time from registration to confirmation) and conversion (percentage of people that actually select an item after registration) between the techniques. Both indicates that this switch of the user representation does not cause confusion also in a field scenario. This confirms findings from the laboratory study and shows that they also apply in a field scenario.

People perform different confirmation gestures: We asked interviewees to recap the interaction and explain us what steps they went through until they finally confirmed a selection. While the last step (confirmation) was not explicitly mentioned by a majority or participants, most of the others (17) reported to have "rested" their hand over the corresponding item. 12 of them explicitly refer to the dwell-time visualization (Figure. 9, 10) and mentioned that they have rested their hand until "the circle was fully charged". However, as observed in the pilot study, some people tend to perform other confirmation gestures than dwelling. Two participants reported to have performed a *Push*, and one mentioned he performed a *Grip* gesture to select the item. In the annotated videos we observed that 90% of users *Point+Dwell* to select items, as suggested by the visualization. 4% perform a *Push* gesture, 3.5% decide to *Grip* and another 1% *Wave* over the item.

People stick to one hand: From those users that could potentially use both of their hands (none occupied), 80% decide to use the same hand for registration as for selection. Even if they could use the left hand to better reach an item at the left side of their body, they would still use the right hand, if they already have registered with that one. Surprisingly, we did not observe an effect of the position of the item on the used hand. It appears that mainly the choice of the registration hand determines what hand would be used for selection.

Reasons of errors: We observed 25% of all registrations to be *unintended*: people inadvertently performed the registration gestures without even looking at the screen. In most of these cases they were raising their hand while pointing to the menu of the cafeteria in order to place an order next to our screen. About 28% of all users were carrying objects (mugs, bottles, phones, tablet computers, bags etc.) in one hand. This constraint leads to inadvertent selections where people were pointing with one hand at an item, but unintentionally selected another item with the other hand (lifted as well). This scenario was observed 11 times, while in five of these cases, one hand of the user was occupied. We only observed these accidental selections when the *In Front of User* input space was active. Unlike to *Around User*, people use this space as a resting position for their hands (e.g. arms crossed in front of body).

RECOMMENDATIONS

From the insights and findings we collected during our three studies we provide the following recommendations for designers of interactive public displays, that want to allow users to select options using mid-air gestures.

No hint for point+dwell required: In all of the three studies we observed that people would point and dwell at items that they want to select, if no further hint or instruction is provided. This confirms the finding of [8] and shows that this also applies in the field. Given one of the four presented selection techniques, designers are free omit hints as "move your hand to the item" for the selection and confirmation phase.

Design for one-handed interaction: Many people in public spaces carry objects in their hands. To allow people with one hand occupied to interact with the display, we recommend to only require one hand for a successful interaction. From the the field study we learned that users tend to use one hand only, so this behavior should be supported. However, we also observed that some people do switch hands, so the application should also not *require* to stick to one hand. The best solution could be to use one hand only, that can be switched at any time, i.e. all items are within reach of both hands.

Handle unintentional interactions: Especially in public space, we can not expect that users who gesture in front of the display are actually addressing it. Many people may unintentionally trigger actions (especially registration gestures) on the screen. We are not aware of an appropriate registration gesture that would avoid this problem. The application should either be designed to cope with unintentional interactions, or use techniques to detect if peoples are addressing the screen with their gestures (e.g. face detection or gaze tracking).

Support exploration: Most users of interactive public displays are first-time users, who are not familiar with the interface and the options to choose from. In the field study we observed that users in average explore 2.5 (out of 6) items before they select one. This behavior should be supported for example by using higher dwell times in the beginning.

User representation may be switched: From the laboratory study and the field study, we have learned that it is not necessary to stick to one user representation for the entire interaction. In our studies we did not observe that a switch from *Mirror Image* to *Cursor* after registration would introduce any confusion to users. If the mapping between hand position and screen coordinates is preserved during the switch, the user representation may be changed after registration. This way, designers could take advantage of both representations: high fidelity and strong communication of interactivity of the *Mirror Image* representation, as well as reduced screen space consumption of the *Cursor* representation.

CONCLUSION

In this paper we presented a design space for hand-gesture-based mid-air selection techniques on interactive public displays. From two laboratory studies, we derived four different selection techniques supporting immediate usability. These technique vary in dimensions of *Input Space* (D3) and *User Representation* (D5). In a field study we showed that the four techniques perform equally well in an in-the-wild scenario, so designers may choose freely among them.

Our findings allow designers of interactive public displays to create applications that go beyond playful interaction. Allowing users to easily make selections or change contents, allows public display applications to evolve towards multi-purpose applications. We believe that this will make the platform of interactive public display more attractive to users, practitioners and advertisers.

Yet, the design space of hand gesture-based mid-air selection techniques for interactive public displays is not yet fully explored experimentally. As the main scope of the field studies in this paper is on the dimensions of *Input Space* and *User Representation*, further investigations on the other dimensions is yet to be done. Besides, future work is necessary to show if the techniques also apply in a wider demographical and cultural context, as the public installations of this work were addressed to a defined audience.

ACKNOWLEDGMENTS

This research was funded by the European Institute of Innovation and Technology, the German Federal Ministry of Education and Research (grant number 01IS12056), the Max Planck Center for Visual Computing and Communication and Cluster of Excellence Multimodal Computing and Interaction of Saarland University. Special thanks go to Betahaus Berlin, Factory Berlin and TU Berlin Mensa for facilitating our field study locations and to Ines Ben Said, Céline Coutrix, Tiare Feuchtner and Moritz Behrens for their valuable support.

REFERENCES

1. Bailly, G., Lecolinet, E., and Guiard, Y. Finger-count; radial-stroke shortcuts: 2 techniques for augmenting linear menus on multi-touch surfaces. In *CHI'10*, ACM (2010), 591–594.

2. Bailly, G., Vo, D.-B., Lecolinet, E., and Guiard, Y. Gesture-aware remote controls: guidelines and

interaction technique. In *ICMI'11*, ACM (2011), 263–270.

3. Bailly, G., Walter, R., Müller, J., Ning, T., and Lecolinet, E. Comparing free hand menu techniques for distant displays using linear, marking and finger-count menus. In *INTERACT'11*, Springer (2011), 248–262.
4. Bowman, D. A., and Wingrave, C. A. Design and evaluation of menu systems for immersive virtual environments. In *VR'01*, IEEE (2001), 149–.
5. Brignull, H., and Rogers, Y. Enticing people to interact with large public displays in public spaces. In *In Proceedings of the IFIP International Conference on Human-Computer Interaction (INTERACT 2003* (2003), 17–24.
6. Cockburn, A., and Gin, A. Faster cascading menu selections with enlarged activation areas. In *GI'06*, Canadian Information Processing Society (2006), 65–71.
7. Francone, J., Bailly, G., Lecolinet, E., Mandran, N., and Nigay, L. Wavelet menus on handheld devices: stacking metaphor for novice mode and eyes-free selection for expert mode. In *AVI'10*, ACM (2010), 173–180.
8. Hespanhol, L., Tomitsch, M., Grace, K., Collins, A., and Kay, J. Investigating intuitiveness and effectiveness of gestures for free spatial interaction with large displays. In *PerDis'12*, ACM (2012), 6:1–6:6.
9. Huang, E. M., Koster, A., and Borchers, J. Overcoming assumptions and uncovering practices: When does the public really look at public displays? In *Springer Pervasive'08* (Berlin, Heidelberg, 2008), 228–243.
10. Jacucci, G., Morrison, A., Richard, G. T., Kleimola, J., Peltonen, P., Parisi, L., and Laitinen, T. Worlds of information: designing for engagement at a public multi-touch display. In *CHI'10*, ACM (2010), 2267–2276.
11. Ju, W., and Sirkin, D. Animate objects: how physical motion encourages public interaction. In *PERSUASIVE'10*, Springer (2010), 40–51.
12. Kostakos, V., Kukka, H., Goncalves, J., Tselios, N., and Ojala, T. Multipurpose public displays: How shortcut menus affect usage. *Computer Graphics and Applications, IEEE 33*, 2 (2013), 56–63.
13. Krueger, M. *Artificial reality II*. Addison-Wesley, 1991.
14. Kules, B., Kang, H., Plaisant, C., Rose, A., and Shneiderman, B. Immediate usability: a case study of public access design for a community photo library. *Interacting with Computers 16*, 6 (2004), 1171–1193.
15. Kurtenbach, G., and Buxton, W. User learning and performance with marking menus. In *CHI'94*, ACM (1994), 258–264.
16. Lepinski, G. J., Grossman, T., and Fitzmaurice, G. The design and evaluation of multitouch marking menus. In *CHI'10*, ACM (2010), 2233–2242.
17. Marshall, P., Morris, R., Rogers, Y., Kreitmayer, S., and Davies, M. Rethinking 'multi-user': an in-the-wild study of how groups approach a walk-up-and-use tabletop interface. In *CHI'11*, ACM (New York, NY, USA, 2011), 3033–3042.
18. Müller, J., Alt, F., Michelis, D., and Schmidt, A. Requirements and design space for interactive public displays. In *MM'10*, ACM (2010), 1285–1294.
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 *CHI'12*, ACM (2012), 297–306.
20. Müller, J., Wilmsmann, D., Exeler, J., Buzeck, M., Schmidt, A., Jay, T., and Krüger, A. Display blindness: The effect of expectations on attention towards digital signage. In *Pervasive'09*, Springer (2009), 1–8.
21. Nilsen, E. L. *Perceptual-motor control in human-computer interaction*. PhD thesis, Ann Arbor, MI, USA, 1991. UMI Order No. GAX91-35659.
22. Peltonen, P., Kurvinen, E., Salovaara, A., Jacucci, G., Ilmonen, T., Evans, J., Oulasvirta, A., and Saarikko, P. It's mine, don't touch!: interactions at a large multi-touch display in a city centre. In *CHI'08*, ACM (2008), 1285–1294.
23. Samp, K. Designing graphical menus for novices and experts: connecting design characteristics with design goals. In *CHI'13*, ACM (2013), 3159–3168.
24. Shoemaker, G., Tsukitani, T., Kitamura, Y., and Booth, K. S. Body-centric interaction techniques for very large wall displays. In *NordiCHI'10*, ACM (2010), 463–472.
25. Song, H., Grossman, T., Fitzmaurice, G., Guimbretiere, F., Khan, A., Attar, R., and Kurtenbach, G. Penlight: combining a mobile projector and a digital pen for dynamic visual overlay. In *CHI'09*, ACM (2009), 143–152.
26. 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 *CHI'12*, ACM (2012), 317–326.
27. Valkanova, N., Walter, R., and Müller, J. Myposition: A public polling projection display designed to spark civic participation.
28. Vogel, D., and Balakrishnan, R. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *ACM UIST'04* (2004), 137–146.
29. Walter, R., Bailly, G., and Müller, J. Strikeapose: revealing mid-air gestures on public displays. In *CHI'13*, ACM (2013), 841–850.
30. Wu, M., Shen, C., Ryall, K., Forlines, C., and Balakrishnan, R. Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. In *TABLETOP'06*, IEEE (2006), 185–192.
31. Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R., and Baudisch, P. Earpod: eyes-free menu selection using touch input and reactive audio feedback. In *CHI'07*, ACM (2007), 1395–1404.