StrikeAPose: Revealing Mid-Air Gestures on Public Displays

Robert Walter *
robert.walter@tu-berlin.de

Gilles Bailly *
gbailly@free.fr

Jörg Müller * ‡
joerg.mueller@tu-berlin.de

* Quality and Usability Lab
Telekom Innovation Laboratories
TU Berlin, Germany

‡ University of the Arts
Berlin, Germany

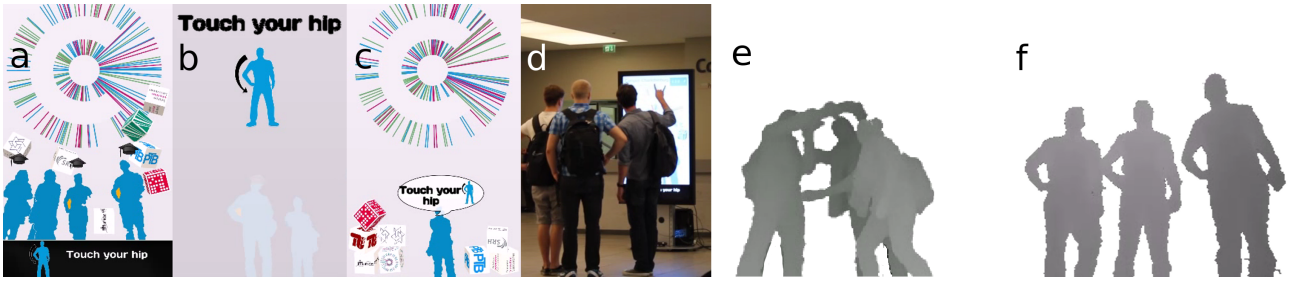

Figure 1. a,b,c) Three strategies for revealing an initial mid-air gesture on public displays: a) spatial division, b) temporal division, c) integration; d,e,f) examples of findings from our field study: d) the Teapot Gesture is fluently integrated with other gestures, e) users explore a potential gesture vocabulary, f) users often imitate other users' gestures.

ABSTRACT

We investigate how to reveal an initial mid-air gesture on interactive public displays. This initial gesture can serve as gesture registration for advanced operations. We propose three strategies to reveal the initial gesture: spatial division, temporal division, and integration. Spatial division permanently shows the gesture on a dedicated screen area. Temporal division interrupts the application to reveal the gesture. Integration embeds gesture hints directly in the application. We also propose a novel initial gesture called *Teapot* to illustrate our strategies. Our main findings from a laboratory and field study are: A large percentage of all users execute the gesture, especially with spatial division (56%). Users intuitively discover a gesture vocabulary by exploring variations of the Teapot gesture by themselves, as well as by imitating and extending other users' variations.

Author Keywords

Public Displays; Initial Gesture; Revelation; Field Study

ACM Classification Keywords

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

CHI 2013, April 27–May 2, 2013, Paris, France.

INTRODUCTION

Touch-based interaction is the common modality for public displays. However, distant interaction through mid-air gestures has several advantages for public display interaction. (1) It does not require to touch public installations which may be inappropriate for hygienic reasons. (2) Users do not need to come close to the screen to interact. (3) It can help noticing interactivity of public displays because passers-by can interact inadvertently [26]. (4) It may favor performative interaction, e.g., expressive and highly visible gestures [31].

Several interaction techniques [4, 13, 14, 22] have been proposed to guide the execution of gestures in the context of pen-based or touch interfaces. However these techniques have never been applied to mid-air gestures and assume that users already know how to *register* the gesture (e.g. how to define the beginning of the interaction [33]). Usually executing a gesture requires to initiate the gesture by pressing a button or touching an interactive surface. Triggering a help system or menu usually requires to touch or press and wait for one second [4, 13]. For mid-air gestures on public displays, a registration or *initial gesture* to define the beginning of advanced interaction is not yet established. Due to the novelty of the interaction technique it is unclear to users how to initiate the interaction.

The context of public displays introduces additional challenges for revealing the initial gesture, especially for mid-air gestures: First, many users approach the device for the first time.They are unaware that the system can capture mid-air gestures, which gestures are available and how to execute them. Second, users interact with the system for a short time [26] and thus the system only has a couple of seconds/minutes for communicating the initial gesture. Third, a public display

generally has only "one shot": Users may give up if they do not immediately succeed with their interaction [23].

In this paper we investigate the question of how to *reveal* an initial mid-air gesture on public displays to enable advanced interactions such as navigating through a menu. We propose three strategies to reveal gestures: spatial division, temporal division and integration. *Spatial division* permanently shows the gesture on a dedicated screen area (Figure 1a at the bottom). For *temporal division* the running application is interrupted to reveal the gesture in full screen (Figure 1b). For *integration* hints are directly embedded into the application, similar to product placement techniques in movies (Figure 1c).

We also propose the Teapot gesture, a novel initial gesture for mid-air gestural interaction with public displays. Users touch their hip to enclose an inner area with their body and their arm in their contour image (Figure 1f). We show that the Teapot gesture is easy to recognize by the system, comfortable for the user, socially acceptable, and easy to understand. We use the Teapot gesture to illustrate our gesture revelation strategies in a laboratory and a field study.

The laboratory study shows that (1) users do not notice the hint with spatial division. (2) Users notice the hint with integration but have difficulties to understand it. (3) Users both notice and understand how to perform the initial gesture with temporal division. From these observations we derived improved versions of the most promising technique for each strategy to compare them against each other in a field study.

The main findings of our field study are: (1) A large percentage of all users execute the gesture, especially with spatial division (56%). This is a surprisingly high number for an in-the-wild study, especially since users are free to do what they want, are not instructed by experimenters, and the game alone is already fun to play. (2) Users intuitively discover a gesture vocabulary by exploring variations of the Teapot gesture. This provides us with a user-defined gesture set. (3) Users discover this gesture vocabulary by imitating other users or by trying to find more interesting gestures than the other members of the group. (4) Users discover the gesture inadvertently while doing unrelated movements.

RELATED WORK

Gestures

In this paper we refer to the definition of gestures of Kurtenbach and Hulteen [20, 9]: "A gesture is a motion of the body that contains information". Several classifications or taxonomies [19, 24, 35, 10] have been proposed to categorize gestures. For instance Cadoz [10] proposes three types of gestures depending on their function: Semiotic (to communicate meaningful information), ergotic (to manipulate the physical world), and epistemic (to learn from the environment). While ergotic gestures are usually used for direct manipulation of virtual objects, semiotic gestures are used for the execution of commands. Semiotic gestures can be further subdivided into symbolic, deictic, iconic, and pantomimic gestures. Symbolic gestures signify gestures that iconify a certain meaning (such as drawing a question mark), but also gestures without an immediately obvious meaning would be classified as symbolic (such as touching a certain body part).

Mid-air Gestures and Public Displays

Only a few public displays support mid-air gesture interaction. While [26, 31] investigate direct manipulation through ergotic gestures, [27, 3] and [32] investigate the use of symbolic gestures for the execution of commands. Still, no field studies have investigated the revelation of symbolic mid-air gestures in the field. The closest to our work is [16], who investigates touch gestures for a public multi-touch table in a field study. They find that gestures are integrated into a continuous flow of gestures and the choice of gesture is influenced by previous gestures and social context. However, these results can not be transfered to mid-air gestures because prolonged interaction with an interactive table differs from playful interaction with a vertical display.

Mid-air gestures in front of public displays can also be described as performative interaction [28]. This concept proposes that users are simultaneously in three different relationships: (1) the interaction with the public display, (2) the perception of themselves within the the situation and (3) acting about a role for others to observe [12]. Important concepts for performative interaction are *manipulations* and *effects* [28] because they impact social learning and the honeypot effect [26, 31]. Manipulations refer to the performer's gestures while effects refer to the visible result of the gestures on the display.

Gesture Registration

Gestures can be described in three phases [6, 13, 36]: (1) *registration* that clearly marks the beginning of the gesture, (2) *continuation* which is the dynamic part and (3) *termination* that marks the end of the gesture. In the case of a touch screen, these phases could be (1) touch the screen, (2) swipe finger and (3) release finger. Especially for mid-air gestures, the registration and termination phases appear less obvious, since there is no explicit delimiter that marks the beginning and the end of a gesture.

Wigdor [33] proposes three possible delimiters: (1) *Multimodality* could be applied to make use of additional channels (e.g. speech, button, etc.) to communicate a delimiter. For example a user could say "put that ..." while pointing at an object, then point at another location saying "... there!" [7] to move an object. However some modalities may be unavailable or inappropriate on interactive public displays. Moreover, discovering additional modalities itself introduces new problems. (2) *Reserved actions* (such as drawing a pigtail [15], or drawing a corner with the pen [14]) can define that the previous or next action should be interpreted as a command. (3) *Clutching* provides a certain state in which gestures are recognized. A possible clutching mechanisms for mid-air gestures may be a virtual and invisible curtain that the user's hand needs to penetrate in order to initiate the gesture tracking. Still it is not clear how this surface should be shaped and positioned. If it is too close it may generate false positive- and if it is too far away it is prone to false negative detection.

An initial gesture can be defined as a reserved action or as clutching. While reserved actions can define either the registration or termination, clutching has the advantage to define both. We now discuss techniques for revealing this initial gesture to novice users.

REVELATION: FROM TOUCH TO MID-AIR GESTURES

Kurtenbach [22] introduced the concepts of self-revelation, guidance and rehearsal for gestures. Several techniques have been proposed for guidance or rehearsal in the context of pen-based or touch interaction. These techniques include *Marking menus* [22] and its variants [1, 2, 3, 37, 38], *HoverWidgets* [14], as well as *Octopocus* [4] and its variants for multi-touch surfaces [5, 13]. Only *LightGuide* [30] has been proposed in the context of mid-air gestures by projecting guidance hints directly onto the user's hands.

In contrast, very few techniques have been proposed for *revelation* [8, 17], although it is an essential issue for all gesture-based systems, especially in public space. We now detail three approaches to reveal gestures on touch surfaces: guessability, interaction techniques, and crib-sheets. We discuss their adequacy for mid-air gestures on public displays.

Guessability: The design of *guessable* gestures [35] appears not very promising for public displays because generally users are not aware which commands are available. However this is one major prerequisite for guessability. Besides users of public displays usually do not have a specific goal or a command to execute in mind.

Interaction Techniques for Revelation: To the best of our knowledge, only three techniques focus on revelation of gestures in the context of mouse [8] and touch [17, 13] interfaces. Firstly, *GestureBar* [8] is a technique for integrating gestures into conventional WIMP interfaces. It uses an advanced toolbar which, instead of executing the command when the corresponding icon is clicked, displays a video of how to execute the command via a mouse gesture. Secondly, Hofmeester recently investigated the revelation of a single gesture in the context of Tablet PCs [17]. In the project a *slide to select* gesture to launch applications on Windows 8 [17] is taught to the user. A tutorial is not used to avoid impairing the user experience. The authors found that visual cues that raise curiosity are an important factor to improve the discoverability of gestures. Finally, *ShadowGuides* [13] displays various hand poses for gesture registration, once users have touched the display and dwelled for one second. *ShadowGuides* also guides the gesture continuation after the user has executed the registration gesture.

As these projects [8, 13, 17], we aim at improving the discoverability of gestures. However, our approach differs in several aspects as we focus on public displays.

First, these systems assume that users already know how to interact in a first modality (mouse and toolbar for *GestureBar*; touch and dwell for *ShadowGuides*; touch for Windows 8). This prior knowledge about the first modality is then used to reveal gestural interaction as a second modality.

Second, these systems have been designed for a context where users want to achieve a goal. In this scenario users are aware of available commands and explore the system for them. In contrast, users of public displays often do not have a specific goal [26]. The interaction is spontaneous and initiated by curiosity or playfulness.

Third, for *GestureBar* and *ShadowGuides* users were already instructed that they should operate a gestural interface. They were aware of "the concept of gestural commands and how to use them" [8]. In contrast, passers-by are generally not aware that public displays are interactive, how to interact with them and whether gestural interaction is supported [26].

In consequence, passers-by should understand that gesture-based interaction is possible and how gestures are invoked, both in a very short time as passing-by interaction is generally quite short (a couple of seconds/minutes) [26].

Crib-sheets: Another alternative is the use of crib-sheets [21]. Most of them are displayed *on demand* by pressing a help button. In Tivoli [21], users *press and hold* to get information about commands and gestures. But this technique is not compatible with immediate usability of public displays. Another strategy may be to *always* display the crib-sheet on the screen. For traditional platforms, permanent crib-sheets are often criticized because they require a lot of space, especially for large gesture sets.

The spatial division techniques presented in this paper are similar to permanent crib-sheets. A major difference to our approach is that not *all* the available gestures are shown, but only *one*: the initial gesture. This single gesture would serve as a registration for advanced gestures, to access a larger set of gestures, or perform other interactions. We believe that presenting several gestures will confuse or overload users by displaying too much information simultaneously. Finally, while different kinds of labels have been used in crib-sheets (text, icons, or animations), they have not been evaluated or compared in the context of distant interaction with public displays.

STRIKEAPOSE

StrikeAPose is an interactive public display game to investigate how an *initial mid-air gesture* can be revealed to users (see Figure 1 a,b,c and 2).

Game

Inspired by [11, 26, 31], we designed a simple but engaging game based on physics simulation to motivate passers-by to interact. Passers-by see their mirror image on the screen and can use it to play with virtual cubes (Figure 1). Users can toss them into a specific target to collect points. They can also perform an *initial gesture* to enable an advanced operation. While the focus is only on revealing this initial gesture and in order to keep the experiments as simple as possible, the advanced operation only consists of adding a funny bunny mask (laboratory study, Figure 2) or doctoral hat (field study, Figure 1a) to the users' contour.

Teapot Gesture

We propose the Teapot gesture as a novel initial gesture for mid-air gestural interaction on public displays. The gesture can be described as a full-body version of the pinch gesture [34], where users touch their hip with their arm to enclose a distinct inner area in their contour image (Figure 1). The Teapot gesture overcomes two limitations of the pinch gesture in the context of public displays: (1) The inner area is large enough to be easily detected by the system, even if users are positioned a couple of meters away from the screen. (2) The area is naturally oriented towards the sensor as the interacting user is facing the screen. The Teapot gesture is also well suited as a gesture registration, because it can clearly indicate the beginning and the end of a gesture or interaction. The first implementation of the recognizer was based on the skeletal tracking capabilities of the *OpenNI / NiTE* framework. Though thick clothes might cause reliability issues with the contour-based recognition, it turns out to be more robust than the skeleton-based recognizer. As a side effect, any gesture besides the Teapot gesture, that generates such an inner area would trigger the recognizer. In the field study described below we observed, that people naturally tend to explore this set of possible gesture variations. A Pilot study also shows that the Teapot gestures is well accepted by users [29].

Revelation Strategies

We propose three strategies inspired by advertisement placement to reveal an initial gesture:

- *Spatial division:* The screen is split into two areas: the game and a ribbon below explaining the gesture. This strategy is for instance implemented as banner ads on Youtube videos.
- *Temporal division:* The game is temporally interrupted to reveal the gesture in full screen. This strategy is similar to classical television ads.
- *Integration:* Visual cues are integrated into the game itself. This strategy is similar to product placement, where certain products (e.g., cars, drinks) are placed in a movie.

These three strategies suggest *when* and *where* to reveal the initial gesture. However, it is not yet defined *how* to explain it.

Labels: Text can provide precise descriptions but its intelligibility depends on the user's language skills. Icons do not have limitations related to readability but can be ambiguous or insufficient for complex or dynamic gestures. Videos are ideal for dynamic gestures but users have to observe and memorize the entire sequence, which may require too much time and cognitive load. Videos can also highlight the link between the manipulation (gesture) and the effect of this gesture. Icon- and video labels can also be combined with text. To reduce the number of conditions to be evaluated in laboratory and field studies, we ran a pilot study to determine the three most promising cues among the five label variations: text; icon; video; text+icon; text+video. Results show that the conditions that include text were more effective in triggering users to execute the gesture than those without text. Based on these results, we decided to test text, text+icon, and text+video in a laboratory study.

Integration Cues: We propose three examples of cues that can be used for the integration strategy.

- *Hip Button:* As an example technique for affording a specific action or movement a button positioned at the hip is added to the user's contour. The Hip Button is supposed to afford users to "touch their hip" (Figure 2c).
- *Voodoo User:* The user is temporally dispossessed of control over his own mirror image. Instead of mirroring the user's movement, the mirror image would perform the gesture (Figure 2c).
- *Fake User:* An additional virtual user is added to the game. For the Fake User an actor has been previously video-recorded passing by the display, stopping, and executing the initial gesture. The Fake User enters every 30 seconds for four seconds. As people tend to imitate behavior of other people, we expect that the real user would imitate the gesture of the Fake User.

To test a reasonable number of conditions in the laboratory study, we conducted a second pilot study to determine the two most promising techniques. Results show that the Fake User performed relatively poorly. While 8/10 users noticed the Fake User, no one actually imitated the presented gesture. Users reported that "The bunny guy comes in to distract me from the game!" or as "a reward" for good performance. Apparently users paid most attention to their own mirror image and have only perceived the effect but not the manipulation of the gesture from the Fake User. Based on the results of this pre-study, we decided to retain the Voodoo User and the Hip Button for a laboratory study.

LABORATORY STUDY

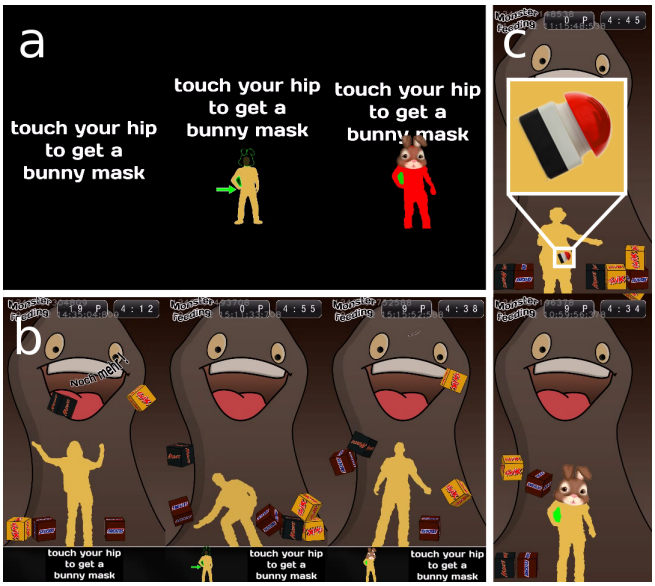

Figure 2. Screenshots of all Conditions: a) Temporal Division (Text, Text+Icon and Text+Video); b) Spatial Division (Text, Text+Icon and Text+Video); c) Integration (Hip Button and Voodoo User)

We conducted a laboratory study to (1) reduce the design space of techniques for revealing gestures, (2) identify the most promising techniques for each strategy, and to (3) optimize the strategies before evaluating them in a field study.

Experimental Design

Label: Three different labels (Figure 2) were used for the temporal and spatial division strategies. A text label "Touch your hip to get a bunny mask" explains both the manipulation (gesture) and the effect (the result of the manipulation). An iconic label shows the static pose, the highlighted inner area of the Teapot gesture, and the bunny mask. A video shows an actor performing the gesture and receiving the bunny mask. To avoid that users confuse the actor in the video with themselves, the actor's contour was rendered in a different color and position.

For the temporal strategy the label was displayed in the center of the screen. It was presented every 30 seconds for four seconds. Four seconds were sufficient to present a video of a user executing the gesture and show the effect, but short enough not to make users wait too long before being able to continue playing the game. For the spatial strategy, the label was shown permanently in the lower part of the screen.

Apparatus and Participants: The system was installed in a room 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 166 passers-by, aged between 13 and 72 years (mean=26.5; sd=9.8) participated in the test. They received candies for their participation. For the entire time of the interaction the system logged the raw depth video, a screen capture, and various events to a text file.

Instructions and Task: After introducing the interactive public display to the participants, we asked them to play with it as they would do in a public place. Participants did not receive any further instructions. In particular they were not instructed about the gestural interaction, the initial gesture, the principle of the game, or the bunny mask. The experiment was aborted as soon as the participant successfully performed the initial gesture (manipulation) to trigger the effect, or after a maximum time of two minutes, an approximation of the maximal usage time for casual interaction in the field [26].

After the test we interviewed all of the 166 participants of the laboratory study individually for approximately 3 minutes. The interviews were semi-structured and included questions like whether or not participants Q1) noticed the hint, Q2) tried to perform the Teapot gesture, Q3) understood the manipulation before trying and Q4) understood the effect before trying. The most important interview results are summarized in figure 3. In addition to the mentioned questions, we gathered general data like age, gender, and occupational background from participants.

Design: We used a between subjects design as we were particularly interested in first-time users.

Results

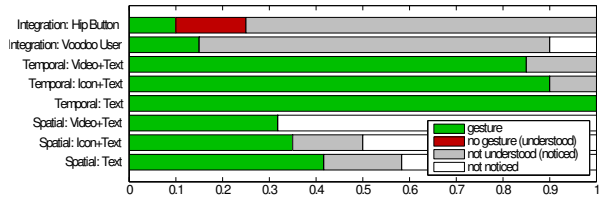

Figure 3. Percentage of participants that performed the gesture (green); did not perform the gesture though they noticed and understood the hint (red); noticed but did not understand the hint (gray); and did not notice the hint (white)

Conversion Rate

The conversion rate is defined as the percentage of users that execute the gesture. It was derived from the results of question Q2 of the interview.

Techniques: A Kruskal-Wallis test reveals no effect for spatial- and temporal division labels on the conversion rate. However, it reveals a significant effect between the two integration techniques ($\chi^2 = 7, p < 0.01$). Voodoo User (15%) triggers significantly more gestures than Hip Button (10%). For the Hip Button technique, we observed that a few users did not perform the gesture correctly. They recognized the button, understood that they had to push it but did not do it in the intended way: Instead of touching their hip by pushing the button from the side, they decided to hit it *in front* of their body. Ultimately, the Hip Button triggered more actions than the Voodoo Users, but in contrast to the Hip Button, the Voodoo User appears to be less ambiguous.

Strategies: A Kruskal-Wallis test reveals an effect for strategies on the conversion rate ($\chi^2 = 51.5, p < 0.0001$). Temporal division (92%) triggers significantly more gestures than spatial division (36%) which triggers significantly more gestures than integration (12%).

Comprehensibility Rate

The comprehensibility rate is defined as the percentage of users that understood the manipulation of the technique. It was derived from the results of question Q4 of the interview.

Techniques: A Kruskal-Wallis test reveals neither an effect for spatial and temporal divisions labels on the comprehensibility rate nor for the integration techniques.

Strategies: A Kruskal-Wallis test reveals an effect for strategies on the comprehensibility rate ($\chi^2 = 45, p < 0.001$). Users understood temporal division (85%) significantly more often than spatial division (35%) and integration (15%). For spatial division, people are distracted from the game and thus may notice but disregard the hint. This was also reported by some participants in the interview.

Noticeability Rate

The noticeability rate is defined as the percentage of users that notice the hint. The data were gathered from question Q1 of the user interviews.

Techniques: A Kruskal-Wallis test reveals neither an effect for the spatial and temporal division strategies on the noticeability rate nor for the integration techniques.

Strategies: A Kruskal-Wallis test reveals an effect for strategies on the noticeability rate ($\chi^2 = 31.0, p < 0.0001$). The hint is noticed significantly more often for temporal division (100%) and integration (95%) than for spatial division (47%).

Finally, we observed that 96% of the participants that understand the manipulation would actually perform the gesture. This is independent of whether or not people did also understand the effect of the gesture.

In summary, results show that (1) The temporal division strategy (92%) triggers a high conversion rate. (2) There are no differences between the different spatial and temporal division labels on the conversion rate. (3) Users that understand the manipulation are likely to perform the gesture.

Optimizing Strategies

This user study highlighted weaknesses for each of the strategies: For the integration techniques, almost all users notice the cue, but it was too subtle to be understood. In contrast, a lot of participants did not notice the hint for spatial division. Finally, for the temporal division strategy, we observed that users noticed the hint and executed the gesture. However, as they executed the gesture mostly during the inserts, it did not show any effect.

almost only while the cue was shown, such that they could not see the effect.

Label: As we did not observe differences between labels, we decided to use the text+icon label for the temporal and spatial division, because it shows both accurate textual description as well as a language independent iconic description of the gesture.

Highlighting Spatial Division: Since the hint was sometimes unnoticed by users for the spatial division, we decided to highlight the hint occasionally using looming stimuli [25]: the cue jumps repeatedly towards the user to capture attention. This highlight appears every 30 seconds for 4 seconds.

Feedback for Temporal Division: To allow users to observe the effect of the gesture during the hint in the temporal division, we decided to fade the mirror image while the hint was blended in (compare Figure 1b to 2a).

Comprehensibility for Integration: We built on the Hip Button of the affordance technique as it is noticed better than the Voodoo User. But as this subtle cue was not understood by the users, we decided to make the hint more explicit. We propose to attach a *Speech Bubble* to the mirror image of the user as shown in Figure 1c. The Speech Bubble uses the same text+icon label as proposed for temporal and spatial division and appears every 30 seconds for four seconds.

Communicating Manipulation Only: For all techniques, we decided not to communicate the effect of the gesture but only the manipulation as it did not seem to effect whether users executed the gesture.

FIELD STUDY

In order to evaluate the gesture revelation techniques in an ecologically valid setting, we conducted a field study. We deployed *StrikeAPose* for five working days in the entrance hall of a university cafeteria. The screen was oriented sideways along the main walking path.

Conditions: We tested the three techniques derived from the laboratory study (see Figure 1 a-c). These conditions were counterbalanced and automatically switched every 10 minutes to minimize the influence of time of day on the results. To avoid interruptions of user interactions, the switch was delayed until no users were detected or after 15 minutes at the latest. All hints were shown (time division and integration) or highlighted (space division) every 30 seconds for 4 seconds.

System: We used the same hardware and software as in the lab study, but we updated the game assets to reflect the deployment location (compare Figure 2 / 1 a-c).

Data Analysis: We collected both qualitative and quantitative data. As quantitative data, we recorded a screen capture as well as the raw depth video from the sensor for the entire time of the deployment. Qualitative data was gathered from observations, interviews and from analyzing manual video recordings.

We interviewed a total number of 46 users in 20 semi-structured group interviews. They were randomly picked, regardless of whether or not they executed the Teapot gesture. Typical questions include whether they R1) already knew the screen, R2) noticed the hint, R3) were annoyed by the hint and R4) tried to perform the Teapot gesture. However, quantitative results of the field study were only derived from the video annotations. Besides that, we collected general comments and opinions of users on the system. Interviews were usually shorter than five minutes and were rewarded with candies.

Observations were conducted from an inconspicuous location on a nearby bench without interfering with the interaction. The depth videos were manually annotated for gesture execution (conversion rate), gesture variations, disengagement, and temporal relation to the hint.

Quantitative Results

During the five days of deployment 558 individual users interacted with the screen while 274 of them performed the gesture at least once.

Conversion Rate: A Chi-squared test revealed that spatial division (56%) triggers significantly more gesture execution than integration (39%) ($\chi^2_{1,N=384} = 10.8, p < 0.05$). The conversion rate for temporal division is 47%.

Timing: We annotated when the gesture was performed within five seconds after the hint appeared or was highlighted. Assuming that people would perform the gesture randomly, this value would be 5 / 30 = 16.7%. We observed that integration (53.7%) and temporal division (56.1%) generates a high probability that users perform the gesture during the appearance or highlighting of the cue. In contrast, for spatial

division only 21.0% of users performed the gesture within that time.

Throughout all three conditions people interacted with the screen for about 41 seconds in average (std=45.7).

Disengagement: The hint appears or is highlighted every 30 seconds. Assuming the hypothesis that the hint would not trigger disengagement, the expected random disengagement rate would be only 16.7%. However, we observed that 27.4% of the users of the temporal division leave within five seconds after the hint appears while 13.8% of the users do so for spatial division and 18.8% for integration.

Discussion: Overall, with 56% for spatial division, a surprisingly large percentage of all users executed the gesture. This shows that gesture revelation works very well for public displays. It seems that the lack of attention observed in the laboratory study was resolved by the periodic highlighting. Temporal division also performs well (47%) but seems to make a large percentage of users leave while the cue is shown. This relates to the finding of Huang [18], who observed that when people look at a film on a public display, they will mostly leave when the film ends or there is an interruption.

Design Recommendations: In order to communicate an initial gesture on a public display, it is recommendable to use a gesture revelation strategy like spatial division. A large percentage of users can be expected to execute the gesture.

All revelation strategies work well, but have different benefits and drawbacks. One should not assume that users will casually play with a screen before executing a symbolic gesture. Almost one quarter of all users executed the gesture before actually start playing. There is also a growing body of evidence that interruptions on public displays make people leave. Thus, interruptions should be avoided or used very carefully.

Qualitative Results

In this section we report our three main qualitative findings: Flow of gestures, exploration of gesture variations, and imitation of gestures. For each of these behaviors, we report, compare, and discuss our observations in the context of related work. Finally, we provide design recommendations.

Flow of gestures

Users did not perform the Teapot gesture in isolation. They were engaged in a constant flow of gestures, be it symbolic, deictic, iconic, pantomimic, or ergotic gestures. For example, when approaching the screen, some users did a symbolic waving gesture towards the screen. Others pulled their friends by the arm to make them abort interaction. They were also pointing at the screen (deictic) while talking to others. One user iconified the depth sensor with both hands while talking to the friends, and another user pantomimed the walking style of his friend with crutches. The biggest category of gestures, however, were ergotic gestures, movements to manipulate the environment. Users play with cubes on the screen, push their friend, or grab objects. A variety of gestures were also executed simultaneously with the Teapot gesture, using every part of the body. Some users were scratching their head while executing the Teapot gesture. Others executed the Teapot gesture with one hand (symbolic) and then tried to lift their virtual doctoral hat with the other hand or to continue to play the game (ergotic). If they executed the Teapot gesture with both arms or held an object in the other hand, they became more creative. They continued playing using their head, their shoulders, and their legs. Some users even punched each other in the face in the mirror image (while standing at different distances) while executing the gesture (Figure 6).

Users also engaged in very *expressive* behavior. For example they wildly swung their hip, posed as on a catwalk, or performed skillful dances while performing the Teapot gesture (Figure 4).

Users sometimes discover the Teapot gesture *inadvertently* through a flow of gestures. For example, one user performed the gesture while putting a cigarette behind his ear. Another user pulled a wallet from his back pocket, inadvertently executing the gesture.

Discussion: The fact that gestures are not performed in isolation, but are rather linked into interwoven sequences, was also observed by Hinrichs and Carpendale [16] for a multi-touch table. They describe that users perform different gestures depending on their previous motion for physical ease (e.g., keeping hand posture) but also social functions. This interweaving of gestures is much more pronounced for mid-air gestures. For multi-touch users, there is a separation between touch gestures for manipulating the screen and mid-air gestures for communicating and manipulating the world. For mid-air gestures, this separation almost disappears, and gestures of various kinds for various purposes melt into a continuous flow, with the same gesture often fulfilling multiple purposes (like posing when executing the Teapot).

Design recommendations: The initial gesture will be woven into a continuous flow of gestures, and this needs to be supported. For example, the Kinect *Guide Gesture* requires users to stand still while stretching their arm at 45° for about two seconds. Users are not allowed to move their other arm or legs during this time. This would be counter-productive in public settings, as it interrupts the natural flow of gestures. Similarly, forcing users to use a particular hand, as it is done for the Kinect, would be unsuitable for public situations. A large proportion of users carry objects, like coffee, bags, or jackets in one hand, so that they would prefer to use the other hand for executing the gesture.

Some users may execute gestures inadvertently, which may be beneficial for gesture discovery. However, this did not occur very often in our case and a separate gesture revelation technique may still be necessary.

As it was the case for the Teapot gesture, the gesture should be easy to avoid if desired.

Finally, because gestures are interwoven into a gesture flow, it should be easy to recognize beginning and end of gesture for system and user (e.g., through tactile feedback when touching the own body).

Figure 4. Teapot gesture performed in an expressive Way

Exploration of Gesture Variations

Once users discovered the Teapot gesture, they performed many variations of it. These variations include the modification of the location or the size of the inner area as shown in Figure 5. *Location* was the most frequent modification, and includes switching arm (13%) or using both arms (28%). User also touch their head to define the inner contour image area with the left (7%), right (2%), or both (2%) arms. Finally, some users also used their legs to span an area.

Users also explored different *sizes* for the area. Minimal size included pinch with one hand (Figure 5f) while maximal size included the formation of a circle with the arms in mid-air above the head (2%), on the left (2%) or on the right side (2%) as shown in Figure 5. Users also explored size by coming closer or using various objects (bags, umbrellas, and even scooters).

Additionally, *multi user* gestures as shown on Figure 1 or 5 were explored. This may be either by active collaboration where people try to define very large areas by holding their hands together or passively by using the body of another user as a border for the enclosed area. Some users freely interpreted the instruction "touch your hip" by touching the hip of their friends (Figure 5).

Discussion: While not actively encouraged users, users explored variations of the Teapot gesture. They not only try to execute the gesture, but try to discover other gestures and additionally identify the limits of the system. We believe that several users consider the Teapot gesture as part of a *gesture vocabulary*. Indeed users rarely tried arbitrary gestures but restrict their exploration to some variations of the Teapot (mainly by spanning an area at different locations). This observation supports our main hypothesis: revealing the initial gesture is necessary to discover it but it is also sufficient for several users to discover advanced gestures. These novel gesture variations form a whole that can for instance be used to organize commands in a gesture-based menu.

Variations of gestures have been observed in the context of multi-touch surfaces [16]. However, it is interesting to notice that these variations occur at two different levels of abstraction. In [16] authors observe that users adopt different strategies to perform the same gesture (e.g the use of 2, 3, 4, or 5 fingers to rotate an object). Authors conclude that designers should provide a variety of ways to perform a gesture. In that study users do not intentionally *explore* these variations. In contrast, our observed variations occur at a different level of abstraction and seek to to determine a gesture vocabulary: It is *not* a variation of the way to execute the gesture, but a variation of the gesture itself leading to different gestures. Interestingly, the Teapot gesture supports these two levels of variations. At a low level, users are not forced to *precisely* touch their hip (they can also touch their waist, adopt different hand or body pose, etc.). At a high level, it raises curiosity to explore a potential gesture vocabulary, an important factor to improve the discoverability of other gestures [17].

Design recommendations: Public displays can rely on users exploring the gesture vocabulary if the initial gesture has been chosen carefully. To do so, designers should design an initial gesture which is part of a gesture vocabulary (e.g., one that encloses an inner area in the users' contour image). The initial gesture should also raise curiosity to support the discoverability of the other gestures. The Teapot gesture is one example of an initial gesture supporting these two properties.

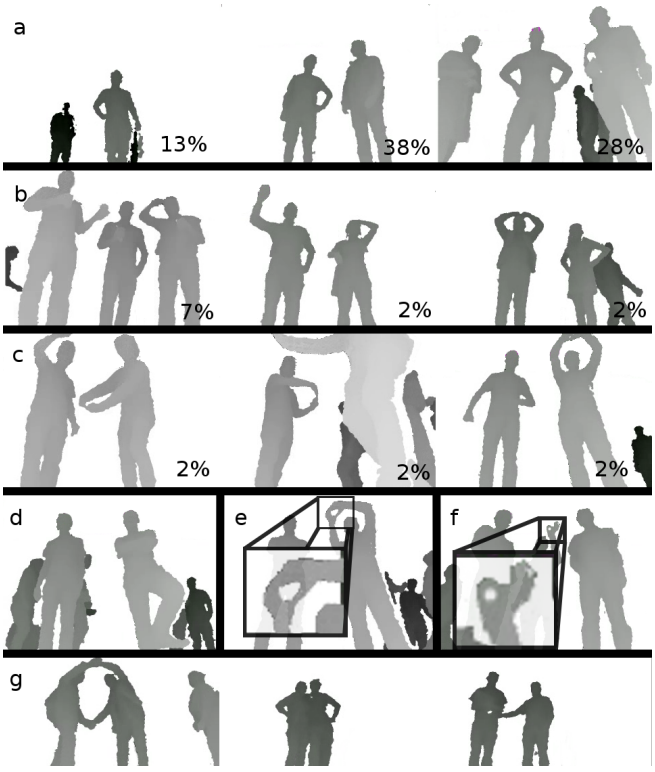

Figure 5. Teapot Gesture Derivations at different Locations: a) Hip (left / right / both) b) Head (left / right / both) c) Hands (Body left / Body right / above Head) d) Feet e) Fingers f) Pinch g) Multi-user

Figure 6. Users hold the Teapot gesture while playing the game

Imitation of Gestures

We were surprised by the number of imitations occurring between users, mostly within groups, but also between groups (see Figure 7). As soon as one user in a group performed the gesture, there was a high probability that within a few seconds others would perform the gesture, too. While for some cases they may all have seen the cue at the same time, there

was also a substantial number of cases where this happened and no cue was shown.

We distinguish direct and indirect imitation. The first imitation occurs when people *directly* observe a user performing the gesture. In contrast, some people who played together in the same group rarely looked at each other but at the screen. Therefore, they probably *indirectly* copy the gestures from the mirror images on the screen. Finally, spectators in the environment seemed to position themselves so that they can see both the users and the screen, although sometimes the screen was occluded. They seemed to copy gestures more from directly observing the bodies of other users.

Users did not only copy the Teapot gesture, but also variations of it as well as other gestures. For example, in Figure 7a, one user discovers that he can execute the gesture by touching his head and this is copied by another user (Figure 7b).

Interestingly, users did not only try to simply imitate gesture variations. Instead, when they saw someone performing a variation of the gesture, they tried to build on this and find different, more interesting gesture variations. This sometimes led to a kind of *competition*, where each creative variation of the Teapot gesture by one user was answered by a more unconventional variation by another user.

Discussion: [16] reports on two cases of imitative behavior (operating a multi-touch table with sleeves and hoarding objects). In these cases, the observation of the manipulation and the effect (on the screen) is merged. In contrast, in our case there is a difference between users in the same group, imitating from seeing the effect (mirror image), and in other groups, rather imitating from seeing the manipulation. In our case, imitation behavior seems to be an important factor for the revelation of the gesture, in particular when someone in the same group has already performed the gesture. Groups also seemed to quickly explore the gesture vocabulary by building on each others gesture in a competitive game.

Design recommendations: Imitation is an important part of gesture revelation. In particular, a registration gesture should be easy to recognize not only for the system, but also for bystanders, and easy to imitate. Because players concentrate on the screen, it is recommendable if the gesture is imitable by only seeing its effect on the screen, for example through a mirror image. In addition, a large set of interesting gesture variations should be supported, in order to enable groups of users to discover the vocabulary by competitively trying to find more interesting gestures.

Figure 7. Imitation between Groups (a → b) and within Groups (c → d)

GENERALIZABILITY AND LIMITATIONS

In this paper, we investigated how to reveal mid-air gesture on public displays. In this section, we discuss the generalizability of our results as well as potential limitations.

Initial gesture. In this paper, we proposed and used *Teapot* as an initial gestures. The focus of this paper is not to find the optimal initial gesture, and we do not claim that the Teapot gesture should be used for all mid-air gesture interfaces on public displays. However, our studies reveal that the Teapot gesture has several advantages: It is easy to recognize for the system [34], apparently comfortable and socially acceptable, easy to understand even with very short description, encourages exploration of the gesture vocabulary, is easy to imitate, and can be held while performing other gestures simultaneously (asobserved in the field study). Further investigations are advised to compare Teapot to other gestures (e.g., to the guide gesture of the Kinect).

Revelation strategies. We tested several revelation strategies both in laboratory and field studies. Several questions remain: First, do these techniques provide similar results if applied to other initial gestures? We believe that the temporal and spatial division strategies are quite independent of the chosen gesture and interaction paradigm (like mirror image interaction). In contrast, we believe that the integration strategy can provide different results depending on the adequacy of the visual cues to the gesture. Additionally, in a different interaction paradigm (e.g., pointer interaction), integrated cues would need to look differently. Second, some decisions have been taken for the timing of strategies. Currently, all cues are highlighted or shown every 30 seconds for four seconds. Further investigations are necessary to determine the optimal timing. Finally, we were surprised by the low performance of our integrated techniques. These visualization techniques have a lot of parameters (color, shape, metaphor, etc.) that need to be tested in a more systematic way to investigate their impact on noticing, understanding, and executing gestures.

Game. Our user studies are based on a simple but engaging game to motivate people to interact. Different scenarios (such as getting news, entering tweets, etc.) should be investigated in the future in order to generalize our results to a larger variety of applications. More generally, it would be useful to investigate interaction between the context (size of the screen, location of the screen, walking path, etc.) and techniques on the number of passers-by executing gestures.

CONCLUSION

In this paper we investigated mid-air gesture revelation strategies for public displays. We proposed three strategies that have been shown to be efficient to make users execute the gesture. For spatial division, 56% of all interacting users executed the gesture, followed by temporal division (47%) and integration (39%). Spatial division is very effective, does not interrupt the content, and does not cause people to leave while the cue is shown. However, it constantly consumes screen space. Temporal division does not have this problem, but interrupts the content and causes users to leave while the cue is shown. Integration seems to be less effective, but can show different cues to different users. Our findings also revealed

that once users execute the gesture, they will explore the gesture vocabulary. They do not only imitate gestures from other users, but try to go beyond other users' gestures in a kind of competition. Finally, the Teapot seems to be a promising initial gesture. We hope that our work can provide a foundation for the investigation of initial mid-air gestures on public displays.

ACKNOWLEDGMENTS

This work was supported by the European Institute of Innovation and Technology and the Alexander von Humboldt Foundation. We thank Uta Hinrichs, Dieter Eberle, Ines Ben Said, Thor Bossuyt, Niklas Hillgren and Christina Dicke.

REFERENCES

1. Bailly, G., Lecolinet, E., and Nigay, L. Wave menus: improving the novice mode of hierarchical marking menus. In *Springer-Verlag INTERACT'07* (2007), 475–488.
2. Bailly, G., Lecolinet, E., and Nigay, L. Flower menus: a new type of marking menu with large menu breadth, within groups and efficient expert mode memorization. In *ACM AVI '08* (2008), 15–22.
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 *Springer-Verlag INTERACT'11* (2011), 248–262.
4. Bau, O., and Mackay, W. E. Octopocus: a dynamic guide for learning gesture-based command sets. In *ACM UIST '08* (2008), 37–46.
5. Bau, O., Ghomi, E., Mackay, W. Arpege: Design and learning of multi-finger chord gestures. In *Technical Report 1533, LRI. February 2010.* (2010).
6. Baudel, T., and Beaudouin-Lafon, M. Charade: remote control of objects using free-hand gestures. *Commun. ACM 36*, 7 (July 1993), 28–35.
7. Bolt, R. A. Put-that-there: Voice and gesture at the graphics interface. In *ACM SIGGRAPH '80* (1980), 262–270.
8. Bragdon, A., Zeleznik, R., Williamson, B., Miller, T., and LaViola, Jr., J. J. Gesturebar: improving the approachability of gesture-based interfaces. In *ACM CHI '09* (2009), 2269–2278.
9. Buxton, B. Gesture Based Interaction (chapter 14), 24 August, 2011. **`http://www.billbuxton.com/input14.Gesture.pdf`**.
10. Cadoz, C. Le geste canal de communication homme-machine. La communication "instrumentale". *Techniques et Sciences Informatiques* (1994), 31–61.
11. Coutrix, C., Kuikkaniemi, K., Kurvinen, E., Jacucci, G., Avdouevski, I., and Mäkelä, R. Fizzyvis: designing for playful information browsing on a multitouch public display. In *ACM DPPI '11* (2011), 27:1–27:8.
12. Dalsgaard, P., and Hansen, L. K. Performing perception - staging aesthetics of interaction. *ACM Trans. Comput.-Hum. Interact. 15*, 3 (Dec. 2008), 13:1–13:33.
13. Freeman, D., Benko, H., Morris, M. R., and Wigdor, D. Shadowguides: visualizations for in-situ learning of multi-touch and whole-hand gestures (2009). 165–172.
14. Grossman, T., Hinckley, K., Baudisch, P., Agrawala, M., and Balakrishnan, R. Hover widgets: using the tracking state to extend the capabilities of pen-operated devices. In *ACM CHI '06* (2006), 861–870.
15. Hinckley, K., Baudisch, P., Ramos, G., and Guimbretiere, F. Design and analysis of delimiters for selection-action pen gesture phrases in scriboli. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '05, ACM (2005), 451–460.
16. Hinrichs, U., and Carpendale, S. Gestures in the wild: studying multi-touch gesture sequences on interactive tabletop exhibits. In *ACM CHI'11* (2011), 3023–3032.
17. Hofmeester, K., and Wolfe, J. Self-revealing gestures: teaching new touch interactions in windows 8. In *ACM CHI EA '12* (2012), 815–828.
18. Huang, E., Koster, A., and Borchers, J. Overcoming assumptions and uncovering practices: When does the public really look at public displays? In *Pervasive Computing*, vol. 5013 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, (2008), 228–243.
19. Karam, M., and Schraefel, C. A taxonomy of gestures in human computer interactions". *Technical report, Electronics and Computer Science, University of Southampton* (2005).
20. Kurtenbach, G., and Hulteen, E. Gestures in human-computer communication. *The Art of Human-Computer Interface Design* (1990), 309–317.
21. Kurtenbach, G., Moran, T. P., and Buxton, W. Contextual animation of gestural commands. In *Computer Graphics Forum* (1994), 305–314.
22. Kurtenbach, G. P. *The design and evaluation of marking menus*. PhD thesis, Toronto, Ont., Canada, Canada, 1993. UMI Order No. GAXNN-82896.
23. 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. CHI '11, 3033–3042.
24. Mulder, A. Hand gestures for HCI. *Tech report 96-I.* (1996).
25. Müller, J., Alt, F., Michelis, D., and Schmidt, A. Requirements and design space for interactive public displays. In *ACM MM '10* (2010), 1285–1294.
26. Müller, J., Walter, R., Bailly, G., Nischt, M., and Alt, F. Looking glass: a field study on noticing interactivity of a shop window. In *ACM CHI '12* (2012), 297–306.
27. Perry, M., Beckett, S., O'Hara, K., and Subramanian, S. Wavewindow: public, performative gestural interaction. In *ACM ITS '10* (2010), 109–112.
28. Reeves, S., Benford, S., O'Malley, C., and Fraser, M. Designing the spectator experience. In *ACM CHI '05* (2005), 741–750.
29. Rico, J., and Brewster, S. Gestures all around us: user differences in social acceptability perceptions of gesture based interfaces. In *MobileHCI '09*, ACM (New York, NY, USA, 2009), 64:1–64:2.
30. Sodhi, R., Benko, H., and Wilson, A. Lightguide: projected visualizations for hand movement guidance. In *ACM CHI '12* (2012), 179–188.
31. 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 *ACM CHI '12* (2012), 317–326.
32. 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.
33. Wigdor, D., and Wixon, D. *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. Morgan Kaufmann. Elsevier Science, (2011).
34. Wilson, A. D. Robust computer vision-based detection of pinching for one and two-handed gesture input. In *ACM UIST '06* (2006), 255–258.
35. Wobbrock, J. O., Morris, M. R., and Wilson, A. D. User-defined gestures for surface computing. In *ACM CHI '09* (2009), 1083–1092.
36. Wu, M., Shen, C., Ryall, K., Forlines, C., and Balakrishnan, R. Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. TABLETOP '06, IEEE Computer Society (2006), 185–192.
37. Zhao, S., Agrawala, M., and Hinckley, K. Zone and polygon menus: using relative position to increase the breadth of multi-stroke marking menus. In *ACM CHI'06* (2006), 1077–1086.
38. Zhao, S., and Balakrishnan, R. Simple vs. compound mark hierarchical marking menus. In *ACM UIST '04* (2004), 33–42.