MirrorTouch: Combining Touch and Mid-air Gestures for Public Displays

Jörg Müller[1,2] **Gilles Bailly**[2,3] **Thor Bossuyt**[2] **Niklas Hillgren**[2]

[1]Alexander von Humboldt Institute for Internet and Society, Berlin, Germany
[2]Quality and Usability Lab, Telekom Innovation Laboratories, TU Berlin, Germany
[3]TELECOM ParisTech, Paris, France

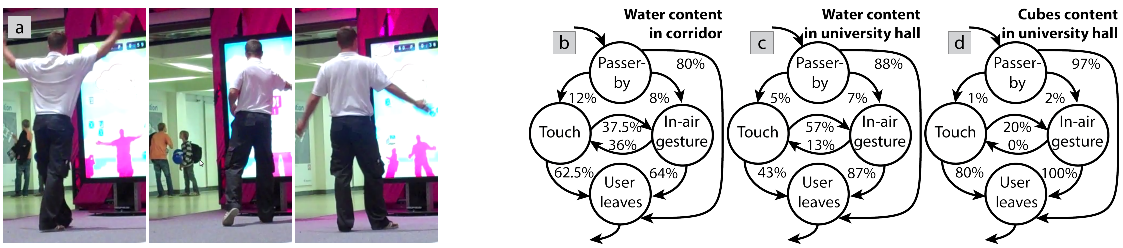

Figure 1. a) A user performs a transition from mid-air gestures to touch and back to mid-air gestures again. b), c) and d) Conversion Diagrams for transitions in two locations (corridor vs. hall) and for two contents (water vs. cubes), each one based on around 350 passers-by.

ABSTRACT

In this paper we present a series of three field studies on the integration of multiple modalities (touch and mid-air gestures) in a public display. We analyze our field studies using Conversion Diagrams, an approach to model and evaluate usage of multimodal public displays. Conversion diagrams highlight the transitions inherent in a multimodal system and provide a systematic approach to investigate which factors affect them and how. We present a semi-automatic annotation technique to obtain Conversion Diagrams. We use Conversion Diagrams to evaluate interaction in the three field studies. We found that 1) clear affordances for touch were necessary when mid-air gestures were present. A call-to-action caused significantly more users to touch than a button (+200%), 2) the order of modality usage was different from what we designed for, and the location impacted which modality was used first, and 3) small variations in the application did lead to considerable user increase (+290%).

Author Keywords

Public displays; field studies; touch; mid-air gestures

ACM Classification Keywords

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

http://dx.doi.org/10.1145/2628363.2628379

INTRODUCTION

The majority of interactive public displays support either touch or mid-air gestures. Recently, Jurmu et al. [13] presented a study on a public display that combined touch and mid-air gestures. They found that people who already interact in one modality are reluctant to switch to the other modality. Because in their study, interaction was supervised, they propose that future studies should investigate unsupervised interaction with multimodal public displays. They argue that the current lack of models and metrics, as well as variety of objectives hamper research in multimodal public displays. In particular, the order of modality usage, the conditions of modality switching and the impact of the environment and application are unsolved questions. In this paper we address these issues through a series of in-the-wild studies of multimodal public displays. In a prestudy, we observe the same effect as [13] that people are reluctant to switch modalities once they started to interact in one modality.

To address the lack of models and metrics for multimodal public displays, we use Conversion Diagrams. In Conversion Diagrams, the percentage of users converting from usage of one modality to other modalities is explicitly given. They can be used 1) as a conscise description of actual usage of a multimodal public display, 2) as a design guideline to make implicit assumptions about how a display will be used explicit and compare these assumptions to how it is actually used in the wild, and 3) to derive hypotheses regarding the influence of independent variables on transitions.

While data for Conversion Diagrams can be obtained by simple manual counting, we present a semi-automatic coding technique. We capture a depth video of user interactions, as well as interaction logs (e.g., touches). In a first analysis round, video parts with at least one user in front of the screen are extracted. These parts are then manually annotated.

We designed MirrorTouch, a multimodal public display to investigate usage of multimodal public displays through three field experiments. MirrorTouch integrates touch interaction and mid-air gestures.

From the three field studies, we learned that

- Clear affordances for touch were necessary when mid-air gestures were present, with a call-to-action being more effective than a button (+200%).
- Although the displays were designed for using mid-air gestures before touch, many users used touched first, depending on the location.
- Small variations in the application did lead to considerable user increase (+290%).

While these findings need to be further evaluated and validate using different locations and applications, they shed first light on the usage of multimodal public displays in the wild. Further, Conversion Diagrams, the semi-automated coding technique and our field studies are intended as tools for researchers investigating usage of multimodal public displays in field studies.

MOTIVATION

Touch is known to afford highly accurate interaction, which is beneficial for fine-grained interaction like text entry, selection and manipulation of interface objects, selection from complex menus etc. In addition, touch provides a natural delimiter (i.e. when touching the surface). It also provides tactile feedback for the delimiter and structural support for the users' hands, thereby reducing strain. A major drawback of touch is that users may not know whether a surface is a touchscreen or not.

When a mirror image of passers-by is shown on a displays, passers-by may interact inadvertently with the display, thereby capturing their attention [19]. Mid-air gestures can also be a lot of fun, and users do not have to make a detour to approach the display. They also work when the display is out of reach, and have less hygienic issues. On the downside, delimiters for mid-air gestures are difficult to design and communicate [27].

As we see, the benefits of mid-air gestures and touch are largely complementary. In order to highlight the mutual benefits that can arise by combining these two modalities, we provide three scenarios.

S1. *Alice is a tourist in an unknown city. Strolling through a street, she notices her mirror image on a display that she would otherwise not have noticed. She sees that she had inadvertently kicked a few virtual objects, and plays with them a bit. She notices that the display provides a map of the city center with information what is currently going on there. She approaches and touches the display and browses through the events, until finding an art event to attend.*

S2. *Bob goes window shopping on a sunday. He passes by a shop window and sees in it his mirror image augmented with some cool clothes. This attracts his attention, and he approaches the window to touch it and browse through a catalogue. He finds some nice new clothes and orders them to be delivered to his home.*

S3. *Lisa is waiting at a bus stop and sees her contour mirrored on a large display where she can play with a virtual ball. See sees some other people from a different city on the same screen and plays a bit with them. After some time, she approaches the screen where she can use a soft keyboard to chat with them. Finally, she makes some new friends and exchanges email addresses with them.*

RELATED WORK ON MULTIMODAL PUBLIC DISPLAYS

In this section we review related work on public displays combining touch and mid-air gestures and supporting different user phases.

Combining Touch & Mid-air Gestures

While traditionally most interactive devices support either touch or mid-air gestures, more recently the combination of touch and mid-air gestures is receiving increasing attention. Schick [24] proposes a technique to use the pointing direction as an extension to direct touch for very large displays. Hilliges [6] extended a multitouch tabletop through gestural interaction above and enabled to lift virtual objects to put them into containers.

Lightspace [29] uses projectors and depth cameras to enable touch interaction on arbitrary surfaces in a room as well as in mid-air. For example, virtual objects can be picked up from surfaces and carried through the room to other surfaces. Marquardt [16] propose a design concept of a continuous interaction space above interactive surfaces. Bailly proposes fingercount [1], a gesture set and menu technique that can be used in touch as well as mid-air gestures.

For public displays, the combination of touch and mid-air gestures has been proposed by the Interactive Public Ambient Display [26] and the Proxemic Peddler [28]. A first field study of a public display combining touch and mid-air gestures has recently been presented by Jurmu [13]. They deployed a display at a science fair and asked passers-by to interact. One observation was that once users started to interact using one modality, they were reluctant to switch to the other modality. Our paper investigates this issue in more detail in an unsupervised in-the-wild setting.

Kukka et al. [15] present an extensive in-the-wild study investigating which factors entice passers-by to interact with a touchscreen. They found that text is more effective than icons, color more than greyscale, and static more than animated. In comparison, our study focusses on the switching between the touch and mid-air gesture modalities. Our first field study can be interpreted as and extension of the findings of [15]. We compare a call-to-action (icon and text, static and color) to a conventional button (icon and text, static and color). We find that the call-to-action outperforms the button by +200%.

System Design Models for Different User Phases

A number of public display systems have been developed which support different *phases* of user interaction. They mostly provide interaction with different modalities in these phases. The Hello.Wall [21] proposes to recognize whether the user is in an ambient zone, notification zone, or cell interaction zone, and adapt its behavior accordingly. Similarly, the Interactive Public Ambient Display [26] can be in ambient mode, implicit interaction, subtle interaction, or personal interaction mode based on the users' distance and head direction. More recently, the Proxemic Peddler [28] initially scrolls the product list at fast speed, slowing down when the user orients towards the display, and shows more product details when the user walks closer. In contrast to previous models, the Peddler actively tries to move a user from one phase to the next, e.g., to recapture attention once the user turns away.

It is important to distinguish these models for system design from models like Conversion Diagrams that describe actual user behavior based on observations. These models describe how a system should or does react when the user shows a certain behavior, and are not used to evaluate a public display, or to describe observations from a field study.

However, such system models could clearly benefit from descriptive models like Conversion Diagrams. For example, while the Proxemic Peddler actively tries to move the user to a specific phase, it is unclear which techniques (e.g., moving content) can achieve this. Conversion Diagrams allow to systematically investigate the impact of various techniques on conversion probabilities. A system could then use this to predict what a user would do next with which probability when the system applies a certain technique (like moving content). It could then decide to use the technique that maximizes the probability that the user does a certain action (e.g., touch the screen).

EVALUATION METHODS FOR PUBLIC DISPLAYS

In this section we analyze which kinds of field studies have been used for evaluating public displays. We classify the different methods that have been used (see Table 1) and illustrate them with examples of findings. Finally, we extract the metrics that have been used in quantitative field experiments (see Table 2). We show how Conversion Diagrams complement these existing methods. For practitioners, this section provides an overview of evaluation methods for public displays. For researchers, the benefit is to outline the state of the art in public display evaluation and provide them a simple framework (study types and metrics) to select the right study for their problem.

For the evaluation of public displays, almost exclusively field studies have been used. Because the actual situation in-the-wild is hard to predict, it is difficult to design laboratory studies where the results actually transfer to real settings. For example, Marshall et al. [17] have shown that actual multi-user behavior in the wild differs widely from what is usually simulated in laboratory studies for multitouch tables.

Descriptive Field Studies

The majority of field studies of public displays are descriptive. Most often, a single prototype is deployed in the wild, and usage is being described, without variations of the prototype.

Descriptive Qualitative	[23][3][2][10][5]
Descriptive Quantitative	[4][20][11][17][18]
Descriptive Existing Systems	[8][9]
Comparative Qualitative	[25]
Comparative Quantitative	[14][12][19]

Table 1. Categories of Field Studies for Public Displays

Qualitative Descriptive Studies

Blueboard [23] is an early example of large touchscreens in semi-public settings. It was deployed in a longitudinal setting, usage was captured on video and analyzed. Among others, it was found that social learning plays a large role, e.g., people observe others to learn how to use the system. The Vision Kiosk [3] was a touchscreen kiosk with an on-screen avatar that followed passers-by with the eyes. Among others, it was found that the perception of the quality of the system depended strongly on the quality of the content.

Opinionizer [2] was a large screen that enabled users to leave their opinion via a keyboard at a party. Usage was videotaped and analyzed. One of the major findings was that once someone already interacted with the display, also others would start to interact (the Honeypot effect). Dynamo [10] was a multi-user display that enabled users to collaborate and share data that was installed in a school. Usage was also videotaped and analyzed. One finding was that the display was used extensively for performances in front of audiences. Finally, Fischer [5] conducted observations and interviews of usage of the SMS Slingshot. From these observations, he constructed a model of the space around the displays.

Quantitative Descriptive Studies

All studies which collected quantitative data also performed a qualitative analysis of user behavior. Arguably, one of the first quantitative studies of public display usage is the Plasma Poster Network [4]. For a network of large touch-enabled screens, the number of clicks and posted content was logged. For example, it was found that posting behavior was skewed in the population and 8 users were responsible for 75.2% of the posted materials on the screen.

The CityWall [20] was a large multitouch screen installed in downtown Helsinki. Usage was captured with cameras and manually annotated. In particular, the duration of interaction sessions, number of simultaneous active users, and number of bystanders were annotated. One result was that very few passers-by interacted when alone - 72% of all users were in pairs.

Worlds of information [11] extended this system to a 3D interface that was investigated during a trade fair. Usage was logged on video and manually annotated, and in addition questionnaires were collected. In particular, the number of simultaneous users, the duration of interactions, and the number of "interaction spots" was annotated. Among the findings was that almost all interactions started with using only one finger. Because most of the interaction relied on multi-finger gestures, users had difficulties in discovering them.

Marshall et al. [17] observed interaction with a multitouch table in a tourist information office. They conducted observations and captured the durations of interaction sessions as well as the number of users per session. A major finding was that groups of tourists did not approach the table simultaneously. In contrast, one member started interaction, maybe a second joined some time later, the first member left etc. Michelis [18] conducted observations of gesture enabled public displays. They propose the Audience Funnel model, where the number of passers-by showing a certain behavior (e.g., stopping and interacting with the screen) is observed.

Descriptive Field Studies of Existing Systems
A number of studies investigate systems that were already available and were not a research prototype. All of these studies obtained some quantitative data, but, probably because it is difficult to change conditions on a commercial deployment, all of them were of descriptive nature.

Hornecker [8] observed usage of a multitouch table in a museum. Besides a large number of qualitative observations, she counted about 200 passers-by. 40 of them did not look at the table, while 60 looked but did not touch. 40 touched, but only played with the interface, and 70 touched and read some content. Of these, 35 read more than one content item. Huang [9] observed whether passers-by of large screens looked at them. She found that most displays only receive very few glances.

Comparative Field Studies
Only a handful of comparative field studies of public displays have been conducted.

Qualitative Comparative Studies
Very few comparative studies use only qualitative data. Chained Displays [25] was a deployment of a space-invaders game set up on a concave, flat, and convex large display. Usage was observed by nearby researchers. Among others, it was found that for group interactions, the flat arrangement was best. In order to maintain group coherence, it was much more important that the group could see the same thing on the screen (shared focus) than that they could see each other (as with the convex setting).

Quantitative Comparative Studies
Comparative field studys where at least two conditions of an independent variable are compared, a quantitative dependent variable is measured, and subjects are assigned to the conditions randomly, will also be called *field experiments* throughout the paper.

In the Ubidisplays deployment [14], a shortlist menu for applications (the wizard) was compared to a general menu regarding the number of application launches. A single display was used to compare application launches within three non-consecutive phases of 60 days each. It was found that applications were started significantly more often when available in the wizard compared to when available from the menu.

Rogers [22] investigated whether ambient displays (LEDs in the floor) can make more people take the stairs compared to the elevator. They determined the ratio of people using the stairs compared to the elevator in a 6 month period before deployment compared to 8 weeks after deployment, as measured by pressure mats. They could show that the ambient displays could indeed increase the percentage of people taking the stairs significantly.

Absolute Number of Users	[8][9][19]
Percentage of Users	[18][12][22][7]
Number of Interactions	[14]
Duration of Interactions	[20][11][17]
Number of Simultaneous Users	[20][11][17]

Table 2. Metrics for Field Studies for Public Displays

Ju [12] conducted a comparative field study to determine whether an animatronic hand can attract more users to touch a kiosk compared to a projected hand. She also compared the animatronic hand to an animatronic arrow. Conditions were switched every 30 minutes. She manually counted the number of passers-by and how many looked at and touched the kiosk. She found that the physical motion attracted significantly more views and touches, but the hand did not attract more compared to the arrow.

Looking Glass [19] was a field experiment of the effectiveness of different interactivity cues on whether passers-by start to interact. Conditions were switched every 30 minutes, and anonymous depth video of interactions was recorded and manually annotated for number of interaction sessions (with a new session starting when nobody was in front of the screen for at least 2 seconds). It was found that inadvertent interaction through a colored mirror image of the user attracts significant more interactions than other visualizations or a classical call-to-action.

Metrics
The metrics used for quantitative evaluations of public displays can broadly be categorized into five categories (see Table 2).

The *absolute number of users* was used in descriptive studies, for example to determine the number of interactions [8] or views towards displays [9]. These metrics are heavily influenced by variations in numbers of passers-by over time. When used for experiments, conditions need to be switched often to reduce the impact of varying numbers of passers-by on the dependent variable.

The *percentage of users* showing a certain behavior is used in descriptive as well as experimental studies. This percentage can be based on the total number of passers-by [18], sometimes restricted to people walking into a certain direction [12]. In other cases, the percentage is based on the number of users showing a certain behavior. For example, those going upstairs [22] or those using a certain gesture variation to execute a specific command [7].

The absolute *number of interactions* can often be determined easily from automated system logs. For example, [14] used this metric to determine how often an application was started.

Finally, the *duration of interactions* and *number of simultaneous users* or bystanders were used in a descriptive way by [20, 11, 17]. They are usually used to round up the description of usage.

While a number of metrics have been proposed for describing interaction with a public display, none of them are especially suited for *multimodal* public displays. For this reason, in this paper we use Conversion Diagrams provide a conscise model of the entire interaction process with a multimodal public display, from passing-by to leaving.

USING CONVERSION DIAGRAMS

Conversion Diagrams (see Figure 2) contain states that can describe the current behavior of the user. For example, a user may be passing by, interact via touch or mid-air gestures, or leave. The frequency with which users transition from one state to another is annotated on the edges and can serve as an estimation of the probability of one particular user transitioning to the other state. Conversion Diagrams reveal how users transition between the modalities or states, which modality they use first, and which modality they use last before leaving. We present different use cases that benefit from the use of Conversion Diagrams.

Conversion Diagrams as a Descriptive Tool

Conversion Diagrams enable the quantitative comparison of usage of different applications, times and locations. In particular, unlike the number of interactions, they are very robust against variations in footfall (number of passers-by) caused by variations in time or location. Importantly, all transitions of all users observed are annotated on the edges. Thus, a conversion diagram is a maximum likelihood estimator of user transitions.

As one consequence, one user transitioning very often can heavily skew the transition probabilities between modalities (but not those entering and exiting the diagram). However, they are a very compact representation of usage. Conversion diagrams also omit important information, like how long each modality is being used, multiuser effects (e.g., whether users are influenced by other simultaneous users), or what users exactly do while they use a certain modality. For these reasons, they should always be augmented with other metrics (see section on enriching Conversion Diagrams).

Conversion Diagrams as Design Guideline

For both researchers and practitioners, Conversion Diagrams are also intended as a design guideline. During the early design phases of a multimodal public display, we recommend creating Conversion Diagrams with the intended transition probabilities annotated (Figure 2 a). This can serve as a tool to more explicitly think about and communicate how the display is intended to be used.

During the design process (field observations, early prototypes, prestudies in the field, and field studies of final prototypes), the Conversion Diagram should then gradually be refined to reflect actual usage of the system. When discrepancies between actual and intended usage are discovered (Figure 2 b), systematic experiments can be conducted regarding

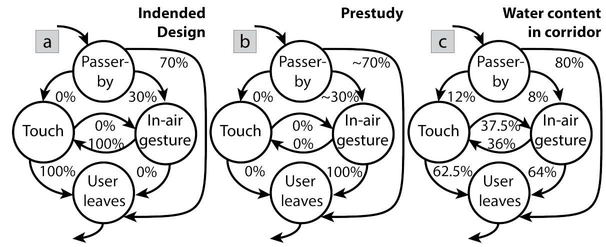

Figure 2. Conversion Diagrams are useful in making implicit assumptions explicit during the design process, and systematically comparing them to reality. For example, we implicitly assumed that people would use mid-air gestures before touching (a). During a first prestudy, we noticed, however, that not a single user touched the screen at all (b). After adding an effective touch affordance, in one location more users touched first than gesturing first (c).

the effectiveness of design changes on particular transition probabilities (Figure 2 c).

Conversion Diagrams for Deriving Experiments

Conversion Diagrams can support the generation of hypotheses and experiments, in particular regarding the impact of various variables on conversions. Regarding the formulation of hypotheses, two equivalent perspectives can be taken. In the state perspective, all users entering or exiting a certain state or set of states can be investigated. In the transition perspective, all users making a certain transition or set of transitions can be investigated.

Conversion Diagrams depict the absolute number of *transitions*, thereby enabling to predict the next action of a user. For experiments, one would generally be more interested in how many *users* make a transition (and only consider each user once). Still, conversion diagrams can help in generating these hypotheses. We will demonstrate which kinds of hypotheses can be derived from Conversion Diagrams during the field studies.

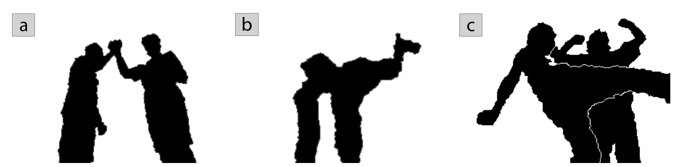

Figure 3. Examples from log files: a) two users perform a high five after reaching the highscore b) two users mime a camel in the university hall c) two children roughhouse the screen in the university corridor.

Obtaining Conversion Diagrams

Conversion Diagrams are useful in particular if it is efficient to generate them. A simple method to obtain Conversion Diagrams would be to manually observe the deployment and keep a list of transitions. Because interactions may be rare, and when they happen, multiple interactions may happen simultaneously, this may be very tedious. Instead, we present a semi-automatic annotation technique.

The presence of a user can be determined using background subtraction. Our prototype (see next section) uses the mirror image of the user to support gesture-based interaction with

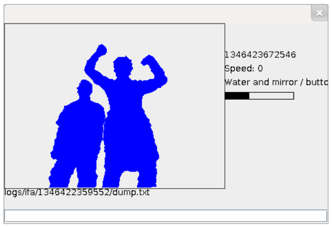

Figure 4. The logging tool used for creating conversion diagrams. When a touch is detected, the contour of the closest user turns red. Playback speed is adjustable, and only situations with at least one user in front of the display or a touch detected are shown. Transitions are annotated manually into a spreadsheet program.

virtual objects. We log the anonymous contours of users (see Figure 3). A custom annotation software (Figure 4) is used to play back situations where at least one user was in front of the screen. The annotating researcher could fast-forward, rewind and pause on sequences of interest. When a pause was made, the current timeframe was automatically copied to the clipboard. This way, it could be pasted into an excel sheet for later reference. When observing an entire day in detail, the timestamps of users entering/leaving the scene were noted this way, as well as the sequence of interactions made by that person. This way it was easy to go back from the excel sheet to the corresponding point in the video. The system also logs touch events, and during a touch, the user closest to the screen is highlighted. Finally, the annotations are compiled into Conversion Diagrams.

Enriching Conversion Diagrams

Conversion Diagrams can easily be extended with other metrics and qualitative observations to provide a broader perspective. For example, it always makes sense to annotate the absolute number of users observed. It would also be useful to annotate the average duration of interactions and number of simultaneous users at the phases.

On a qualitative side, they can be extended through data from observations (either manual or automatic) and interviews. Although we keep a complete anonymous interaction video log for later analysis, the narrow field of view of the camera means that usually not the entire situation is captured. Therefore, we use researchers present at the experiment site at all times to perform manual observations. Exemplary cases of how Conversion Diagrams can be extended through observations are provided in the three field studies in the following sections.

MIRRORTOUCH

Conversion Diagrams can be used for any kind of multimodal public displays. We think that the integration of touch and mid-air gestures is a very interesting case, since 1) both modalities are well established, 2) they have complementary benefits, as mid-air gestures are good for attracting users and touch is good for enabling highly accurate extended interactions, and 3) to our knowledge, no field studies investigating public displays that combine touch and mid-air gestures have been done yet. For these reasons, we designed, implemented and deployed MirrorTouch to illustrate the use of Conversion Diagrams for the specific case of a public display that integrates touch and mid-air gestures.

MirrorTouch (Figure 5) is a public display that combines mid-air gestures and touch interaction. People see their own mirror image as a contour and can play with virtual objects. The central idea behind the design was to attract users through mid-air gestures and then enable detailed and accurate interaction (e.g., looking up information) through touch. Gestural interaction was designed so that passers-by would discover the fact that the display was interactive inadvertently while passing-by (as in [19]). Because we were not especially interested in the details of touch interaction, and only in the fact whether people touch at all, the touch interaction was rather simple (looking up information or spawning ducks).

Water and Cubes Designs

We developed two variations of MirrorTouch. In the *water* variant (see Figure 5 b), people can play with water particles that are spawned from a virtual faucet. When they touch the screen, a duck appears that floats on the water. In the *cubes* variant (see Figure 5 a), people can play with virtual cubes, and throw them into a cloud, where they disappear. When they touch the display, an informational screen is shown and the game is interrupted.

Implementation

MirrorTouch uses a Samsung 650TS 65” touchscreen in portrait format. Users are tracked through an Asus Xtion Pro camera. The user contour is extracted using OpenNI. The water variant is 2D and uses Box2D as physics engine and Processing for rendering. The cubes variant is 3D, where the interaction is constrained to a 2D plane. It uses Bullet as physics engine and Java OpenGL for rendering.

Prestudy

As a prestudy we deployed the water prototype for one evening during a university fair. We let passers-by discover interaction by themselves but stood in the vicinity to observe and answer questions. Our system was designed to attract people through mid-air gestures and then enable detailed interaction through touch. Surprisingly, this approach did not

Figure 5. Two types of content used for MirrorTouch: a) cubes, b) water.

Figure 6. During the prestudy, not a single user touched the screen.

work at all. We did not observe a single user (among several hundreds who used mid-air gestures) who touched the screen (see Figure 6). Apparently, once users started to interact through mid-air gestures, they were caught in that interaction modality and did not consider that other modalities were also available. While our implicit design assumption was that users would transition from mid-air gestures to touch, this simple study already showed that the actual Conversion Diagram was very different. The same behavior was also observed by Jurmu et al. [13]. Consequently, we started our field studies with finding efficient affordances for touch.

FIELD EXPERIMENT 1: AFFORDANCES FOR TOUCH

We conducted a field experiment to compare two affordances for touch, a (more classical) button design and a call-to-action "touch here" in combination with a hand icon (see Figures 8 a and 8 b). This study can be interpreted as an extension of the touch affordances study by Kukka et al. [15], who found that static colored text works best. Both our affordances use static colored text, but we compare a button design to a call-to-action.

Method

As apparatus for this field experiment the water prototype was used (see Figure 5 b). We deployed the screen for one day in a corridor leading to a university cafeteria (see Figure 7 a). Participants were mostly students, faculty, and visitors to the university. Most users would come from the elevators, walk past the display towards the cafeteria, or come from the cafeteria and walk towards the elevators. We sat nearby in a student working room to inconspicously observe interaction with the display. The two conditions were switched every 5 minutes. We captured the contours of passers-by as well as touch events. Later all events with at least one user in front of the camera were automatically extracted. These events were then manually annotated regarding the number of users who passed-by and the number who touched the display. The entire data was annotated by two independent coders. Inter-rater reliability was very good (Cohens Kappa=0.82).

Location	University corridor (a)	University hall (b)
Space	narrow	spacious
Audience	students	students
Content	water	water & cubes

Figure 7. Deployment locations: a) the university corridor b) the university hall

Figure 8. Two types of touch affordances: a) standard button, b) call-to-action.

Results

We counted 697 passers-by of whom 66 touched the screen. A Chi-square test revealed that the percentage of passers-by utilizing the touch screen significantly differed with the affordance used ($\chi^2_{1,N=697} = 20.9, p < .05$). The call-to-action affordance had a 15% touch conversion rate, while the button had 5%. To our surprise, however, people did not use the modalities in the order we intended. Instead of gesturing first before touching, some users touched the screens directly without gesturing first.

We also observed some interesting details about user behavior. The display was installed in a distance of about 10m from the elevators. Quite often we observed people leaving the cafeteria, passing by the display without interacting and walking toward the elevators. When they saw that the elevators had not arrived yet, they walked back to the display and interacted. When the elevator arrived, it played a ringing tone, and people hurried towards the elevator, although some missed it. This demonstrates that waiting zones are important even when the display itself is not installed immediately within the waiting zone. It may thus be benefitial to install displays in well-frequented areas (e.g., hallways) but close to this kind of waiting zone. These qualitative observations illustrate that some external factors (like elevators) can have large effects on the transitions, and should be reported. In particular, the duration of interactions in this case may be determined less by the display itself, but rather by the arrival time of the elevator.

Discussion

This study demonstrates that different design of affordances can have large differences in effectiveness (in our case +200%). We learned that the order of modalities was different from what we designed for. Many users touched without gesturing first. The study also demonstrates how to systematically investigate the impact of design decisions on the Conversion Diagram. Although the call-to-action achieved a much higher transition probability than the button, in order to derive a general recommendation, this result should be

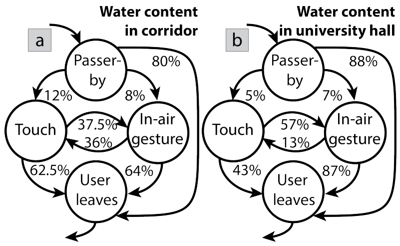

Figure 9. Conversion Diagrams for Field Experiment 2. The water content was deployed in the university corridor (a) and the university hall (b).

validated in different user populations, locations, and applications. For the following study, we decided to investigate in more detail which modality is used first. We proceeded with using the call-to-action as affordance.

FIELD EXPERIMENT 2: IMPACT OF THE LOCATION

After having identified an effective touch affordance (call-to-action) we conducted a second field study to determine whether there would be differences in the Conversion Diagrams between the two locations.

Method

As apparatus, the water content (see Figure 5 b) was used in the university hall (see Figure 7 b) and the university corridor (see Figure 7 a). Both locations had similar participants, being mostly students, but also faculty and visitors to the university. We tested each location for one day. Otherwise, the methodology used was the same as in Field Experiment 1.

Results

During the two days of deployment we observed a total of 117 users and 718 passers-by. A Chi-square test revealed that the first used modality significantly differed with location ($\chi^2_{1,N=117} = 3.9, p < .05$). Passers-by mainly used mid-air gestures as a first modality in the hall (7% vs. 5%) while they mainly used touch first in the corridor (12% vs. 8%). The two Conversion Diagrams are illustrated in Figure 9.

We also observed what seems to be a very interesting interaction effect between playful mid-air gestures and touch interaction. Some people were roughhousing the screen during touch interaction after engaging in intense mid-air gesture interaction. In particular, on one occasion some children were playing with the screen in the university corridor (see Figure 10).

They started with rather modest motion playing with the water. Over time, the motion became increasingly expressive and even agressive. When they then started touching the screen, they started to push it harder and harder (see Figure 3 c). The screen started shaking considerably, and we had to intervene (entering from our concealed position) to prevent damage or accident.

We observed a very similar behavior in the university hall. One employee of the cafeteria also became more and more engaged over time playing with the water through mid-air gestures. When she started touching the screen, she began to shake it violently so we had to intervene. These qualitative observations illustrate that the details of usage of one modality may differ by which other modality was used before. Such unexpected effects can only be discovered by qualitative observations.

Discussion

From this study we learned that not only many users used the modalities in a different order than intended, but the order in which they used the modalities also depended on the location. We do not know the reason why they touched first more often in the corridor, but the corridor was more narrow than the hall. It is possible that the fact that people were closer to the screen when passing by influenced in which order they used the modalities. It would be, however, necessary to validate this hypothesis in different locations and with different applications. The experiment serves to illustrate how to evaluate the ratio between two transitions in a field experiment. It also indicates that mid-air gestures can lead to more violent touches and roughhousing. When validated in further experiments focussing on this aspect, it could mean that touchscreens that also support (playful and engaging) mid-air gestures would need to be very physically robust.

FIELD EXPERIMENT 3: IMPACT OF THE CONTENT

Our third field study aimed at comparing two different contents and keeping the location (university hall) constant. We hypothesized that there would be no differences between the Conversion Diagrams between the cubes and water content (Figure 5).

Method

As apparatus, we compared the two applications described in the MirrorTouch section (Figure 5). Both applications were installed in a university hall close to a major cafeteria (see Figure 7 b). Paricipants were mostly students, faculty, and visitors to the university. People would pass by the display either on their way to the cafeteria or the way back. We used the same data analysis technique as for Field Experiment 1.

Results

The prototypes were deployed during one day each, for a total of 53 users and 700 passers-by. A chi-square test revealed that

Figure 10. Children after roughhousing the screen.

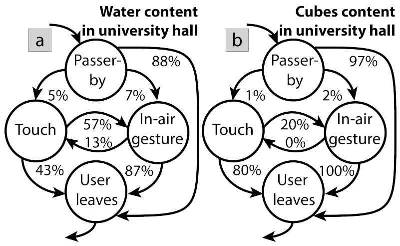

Figure 11. Conversion Diagrams for Field Experiment 3. The water content (a) and the cubes content (b) were deployed in the university hall.

the percentage of passers-by utilizing the public display significantly differed with content ($\chi^2_{1,N=700} = 20.5, p < .05$). The water content increased the number of users by +290%, compared to the cubes content. Furthermore, the content did not have a significant effect on which modality people chose to use first, nor did it affect which one they used last. The Conversion Diagrams are shown in Figure 11.

Mid-air gestures seemed to cause more playful behavior and cause a stronger Honeypot effect (see Figure 12) than touch. We also observed that some users who used touch first never noticed their mirror image (the feedback for mid-air gestures). Because they were very close to the screen, they failed to see their mirror image displayed in the lower area of the display. These observations illustrate that one modality (mid-air gestures) may not even be discovered if another modality (touch) is used before, augmenting the information represented in Conversion Diagrams.

Discussion

We learned two things from this study. First, a variation of content did lead to a huge difference in the number of transitions. We found this surprising as the two applications were quite similar: both were 1) a game with 2) physics simulation, 3) they also had the same affordances and 4) the same modalities (touch and mid-air gestures). Although it is unclear what exactly caused these large differences, it would be interesting to investigate this phenomenon further in different contents and with different locations and user populations.

Second, it may be difficult to combine feedback for touch and mid-air gestures in such a way that users still perceive the feedback when they are touching the screen. As they see only a very small part of the screen, and feedback for mid-air gestures usually requires considerable space, it is a challenge to show the feedback for mid-air gestures in such a way that touching users still perceive it.

CONCLUSION

In this paper we presented a series of three field studies on the integration of multiple modalities (touch and mid-air gestures) in a public display. We used Conversion Diagrams as an approach to evaluate multimodal public displays. We also presented MirrorTouch, a public display combining touch and mid-air gestures. We learned that the actual usage of such displays in the wild can be very different from what was intended. Small variations in affordance, location and application had considerable influence on actual usage. For example, in our case the call-to-action was three times more effective than a traditional button. Contrary to what we designed for, many people touched before mid-air gesturing, and the order of modalities depended on the location. Finally, small variations in the content increased the number of users by 290%. Conversion Diagrams provide a systematic way to address these issues. In particular, they enable us to narrow the (often huge) gap between what we believe how our devices are or should be used and how they are actually used in the wild.

Figure 12. Mid-air gestures seemed to cause stronger Honeypot effect than touch.

ACKNOWLEDGEMENTS

We want to thank Robert Walter for help during the implementation and field studies. This work was supported by the European Institute of Innovation and Technology.

REFERENCES

1. Bailly, G., Müller, J., and Lecolinet, E. Design and evaluation of finger-count interaction: Combining multitouch gestures and menus. *International Journal of Human-Computer Studies* (2012).

2. Brignull, H., and Rogers, Y. Enticing people to interact with large public displays in public spaces. In *INTERACT*, M. Rauterberg, M. Menozzi, and J. Wesson, Eds., IOS Press (2003).

3. Christian, A. D., and Avery, B. L. Speak out and annoy someone: experience with intelligent kiosks. In *ACM CHI'00* (2000), 313–320.

4. Churchill, E. F., Nelson, L., Denoue, L., and Girgensohn, A. The plasma poster network: Posting multimedia content in public places. In *INTERACT'03*, vol. 3 (2003), 599–606.

5. Fischer, P. T., and Hornecker, E. Urban hci: spatial aspects in the design of shared encounters for media facades. In *ACM CHI'12* (2012), 307–316.

6. Hilliges, O., Izadi, S., Wilson, A. D., Hodges, S., Garcia-Mendoza, A., and Butz, A. Interactions in the air: adding further depth to interactive tabletops. In *ACM UIST '09* (2009), 139–148.

7. 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.

8. Hornecker, E. I dont understand it either, but it is cool - visitor interactions with a multi-touch table in a museum. In *IEEE TABLETOP'08* (2008), 113–120.

9. 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.

10. Izadi, S., Brignull, H., Rodden, T., Rogers, Y., and Underwood, M. Dynamo: a public interactive surface supporting the cooperative sharing and exchange of media. In *ACM UIST'03* (2003), 159–168.

11. 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 *ACM CHI'10* (2010), 2267–2276.

12. Ju, W., and Sirkin, D. Animate objects: How physical motion encourages public interaction. In *Persuasive Technology*. Springer Berlin Heidelberg, 2010, 40–51.

13. Jurmu, M., Ogawa, M., Boring, S., Riekki, J., and Tokuda, H. Waving to a touch interface: Descriptive field study of a multipurpose multimodal public display. In *ACM PerDis '13* (2013), 7–12.

14. 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 (March-April), 56–63.

15. Kukka, H., Oja, H., Kostakos, V., Gonçalves, J., and Ojala, T. What makes you click: Exploring visual signals to entice interaction on public displays. In *ACM CHI'13* (2013), 1699–1708.

16. Marquardt, N., Jota, R., Greenberg, S., and Jorge, J. A. The continuous interaction space: interaction techniques unifying touch and gesture on and above a digital surface. INTERACT'11, Springer-Verlag (2011), 461–476.

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 *ACM CHI'11*, ACM (2011), 3033–3042.

18. Michelis, D., and Müller, J. The audience funnel: Observations of gesture based interaction with multiple large displays in a city center. *Int. J. Hum. Comput. Interaction 27*, 6 (2011), 562–579.

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 *ACM CHI '12* (2012), 297–306.

20. 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 *ACM CHI '08* (2008), 1285–1294.

21. Prante, T., Röcker, C., Streitz, N., Stenzel, R., Magerkurth, C., Van Alphen, D., and Plewe, D. Hello. wall–beyond ambient displays. In *Adjunct Proceedings of Ubicomp* (2003), 277–278.

22. Rogers, Y., Hazlewood, W. R., Marshall, P., Dalton, N., and Hertrich, S. Ambient influence: can twinkly lights lure and abstract representations trigger behavioral change? In *ACM Ubicomp'10* (2010), 261–270.

23. Russell, D., Drews, C., and Sue, A. Social aspects of using large public interactive displays for collaboration. *ACM UbiComp'02* (2002), 663–670.

24. Schick, A., van de Camp, F., Ijsselmuiden, J., and Stiefelhagen, R. Extending touch: towards interaction with large-scale surfaces. In *ACM ITS'09* (2009), 117–124.

25. 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.

26. 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.

27. Walter, R., Bailly, G., and Müller, J. Strikeapose: Revealing mid-air gestures on public displays. In *ACM CHI '13* (2013), 841–850.

28. Wang, M., Boring, S., and Greenberg, S. Proxemic peddler: a public advertising display that captures and preserves the attention of a passerby. In *ACM ISPD'12* (2012), 3.

29. Wilson, A. D., and Benko, H. Combining multiple depth cameras and projectors for interactions on, above and between surfaces. In *ACM UIST '10* (2010), 273–282.