Imaginary Interfaces: Spatial Interaction with Empty Hands and without Visual Feedback

Sean Gustafson, Daniel Bierwirth and Patrick Baudisch
Hasso Plattner Institute
Potsdam, Germany
{sean.gustafson, patrick.baudisch} @ hpi.uni-potsdam.de
daniel.bierwirth @ student.hpi.uni-potsdam.de

ABSTRACT

Screen-less wearable devices allow for the smallest form factor and thus the maximum mobility. However, current screen-less devices only support buttons and gestures. Pointing is not supported because users have nothing to point at. However, we challenge the notion that spatial interaction requires a screen and propose a method for bringing spatial interaction to screen-less devices.

We present *Imaginary Interfaces*, screen-less devices that allow users to perform spatial interaction with empty hands and without visual feedback. Unlike projection-based solutions, such as Sixth Sense, all visual "feedback" takes place in the user's imagination. Users define the origin of an imaginary space by forming an L-shaped coordinate cross with their non-dominant hand. Users then point and draw with their dominant hand in the resulting space.

With three user studies we investigate the question: To what extent can users interact spatially with a user interface that exists only in their imagination? Participants created simple drawings, annotated existing drawings, and pointed at locations described in imaginary space. Our findings suggest that users' visual short-term memory can, in part, replace the feedback conventionally displayed on a screen.

Author Keywords: mobile, wearable, spatial, screen-less, memory, gesture, bimanual, computer vision

ACM Classification Keywords: H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

General Terms: Design, Human Factors, Experimentation

INTRODUCTION

Today's desktop computer interfaces are mostly based on *spatial* interaction: objects are located at a specific location; in order to interact with an object, users select it by pointing at its location. A key component in this interaction model is the screen, as it provides the reference frame in which pointing takes place.

In mobile interaction, however, screens need to be much smaller than on the desktop. In order to achieve ultimate mobility, some researchers have abandoned screens altogether [23]. These wearable devices are operated with hand gestures, because there is no screen to point to anymore. Unfortunately, the gestures are generally categorical, rather than spatial: Gesture Pendant [25], for example, allows users to perform a series of commands out of a finite vocabulary, such as "open door".

Virtual Shelves [20] extends this concept by allowing users to point at set of virtual positions without visual feedback. This result is interesting because it suggests that users might be able to interact with arbitrary spatial interfaces without seeing what they are interacting with.

In this paper, we investigate the question: To what extent can users interact spatially with a user interface that exists only in their imagination?

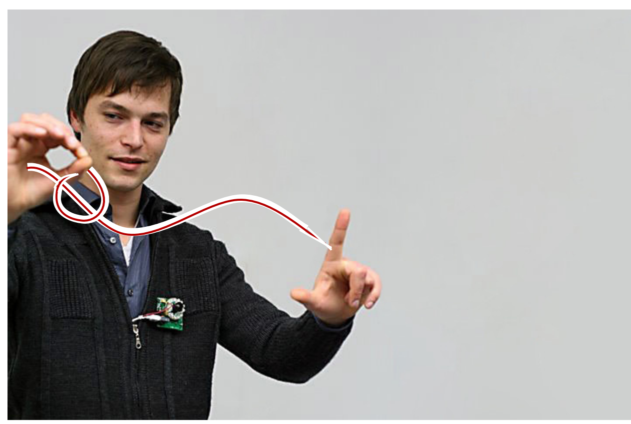

Figure 1: User sketching a stock curve using an imaginary interface. The curve exists only in the user's imagination. Unlike *Sixth Sense* [22] there is no projector or display—users only see their bare hands.

IMAGINARY INTERFACES

Imaginary Interfaces allow users to create and interact with objects located in an invisible 2D space (Figure 1). As with conventional computer systems, users select objects by pointing at them and they manipulate objects spatially. Unlike these interfaces, however, with Imaginary Interfaces there is no screen. All visual "feedback" takes place in the user's imagination.

Figure 2 illustrates an envisioned scenario. Karl has called his stock broker using a Bluetooth earpiece to discuss his investment portfolio. (a) He invokes an imaginary interface

UIST'10, October 3–6, 2010, New York, New York, USA.

by forming an ‘L’ with his left hand. The shared whiteboard app is invoked automatically, where his index finger and thumb specify an imaginary interaction plane. (b) “The stock was as high as thirteen dollars last week.” He moves his right hand to the top of his index finger and *pinches* to start the imaginary ink, sketching the stock price curve. “Then it dipped down to less than six dollars.” (c) The broker follows the sketch plotted in real time on his computer screen as his client explains. Karl completes the stroke by releasing the pinch.

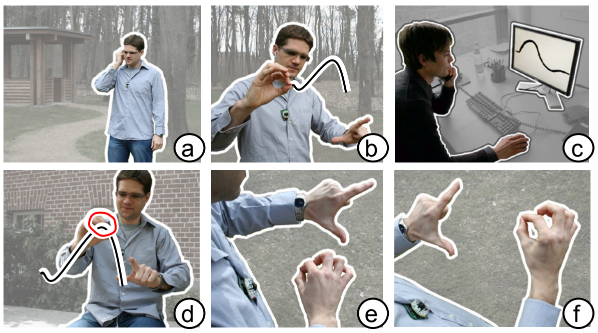

Figure 2: This user has invoked an Imaginary Interface by forming an “L” with his left hand. As he is talking to his stock broker he sketches a stock curve to explain his point and then adds an appointment to his timetable.

Karl sits down on a park bench, (d) reestablishes the imaginary interaction plane using the ‘L’ gesture and adds, “You should have sold at the peak.” Karl changes his imaginary ink color to red and circles the crest of the curve. Despite not being able to see what he has sketched before Karl is easily able to, with the help of the left reference hand, recall the precise spatial location of the curve’s features and add annotations. The broker wants to satisfy his client and suggests they meet for lunch tomorrow to discuss it further. Karl agrees and (e) launches his appointment book by pinching the space directly below his thumb (a location he has reserved for this function). He then (f) selects mid-day (aligned with his index finger tip) to reserve a slot for lunch.

As demonstrated in this scenario, Imaginary Interfaces can be used to clarify spatial information during a conversation, such as to explain where a player was located when he or she scored the goal, where apartment keys were deposited, how to get to the lab, or how to layout an article.

On the other hand, parsing the user’s input programmatically can allow using Imaginary Interfaces to enable real-time interaction, such as to operate virtual buttons, scrub a virtual audio slider, pick a color from a 2D color widget, or to operate a user interface a user has sketched in the Imaginary Interface itself.

Benefits and Contribution

The main benefit of Imaginary Interfaces is that they bring 2D spatial interaction to screen-less devices. While screen-less devices allow for ultimate miniaturization [23], these devices currently offer *no* spatial interaction. Imaginary Interfaces enable basic spatial interaction and, by leveraging visual short term memory, they allow users to extend, annotate or edit a drawing after it was initially created.

Since Imaginary Interfaces leave the user’s hands empty, invoking an imaginary interface is fast—there is no need to retrieve a device. This makes them particularly useful for microinteractions [1], i.e., interactions too short to merit whipping out a more interactive device. This allows Imaginary Interfaces to integrate well into conversations, just as users gesture when talking with a person face-to-face.

RELATED WORK

Imaginary Interfaces build on wearable computing, gestural input, mobile computing and spatial interaction.

Wearable Computing

Imaginary Interfaces derive in spirit from wearable computing. We share the goal of creating always available, nearly invisible and completely integrated personal computing technology to support day to day life.

Fukomoto’s Body-Coupled FingerRing [10] is particularly inspiring because of its ability to turn any surface into an input device. Similarly GestureWrist and GesturePad [24] are unobtrusive input devices that exploit physical movement as input.

With Disappearing Mobile [23], Ni and Baudisch explored extremely small wearable input devices. They tested text input using a single point scanning interface mounted on the user’s wrist. They point out that one of the main challenges for spatial interaction without visual feedback is to *connect* strokes (*relative position features* [23]). As illustrated by Figure 3, participants’ ‘D’ Graffiti characters [11] were often misrecognized as ‘P’, because participants failed to connect the end point of the stroke to the starting point due to the absence of visual control.

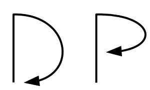

Figure 3: In the study reported by Ni and Baudisch, Graffiti ‘D’s were often misrecognized as ‘P’s [23].

Gestures

Beyond some wearable devices, gestures have become a popular input mode for all sorts of Reality-Based Interaction [18]. As computers become more adept at interpreting the language of people (i.e., reality) they necessarily resort to input modalities akin to gestures.

For years, immersive 3D environments have used freehand gesture input (i.e., whole hand interaction [27]) as their primary interaction mode. Projects such as Charade [3] have investigated how humans can interact within virtual environments as they do in person. The Gesture Pendant [25] senses free-hand gestures in front of the user’s body from a camera worn around the user’s neck.

Researchers have also transferred this interaction style to the desktop. Wilson and Oliver’s GWindows [29] is a desktop gestural interface based on robust stereo vision techniques.

Often though, gestural input is limited to a series of categorical gestures. That is, one posture of the hand results in one corresponding action. However, in normal social interaction gestures support a conversation by adding a secondary stream of information that complements speech with spatial relationships that would be difficult to express otherwise [8].

Spatial Interaction

Billinghurst et al. [5] classify spatial interfaces as head, body or world-stabilized. Our system is, in one sense, body-stabilized as the camera is mounted to the user's chest and tracks the location of both hands with respect to the body. However, we take a cue for bimanual interaction research, such as [12], and use the non-dominant hand to coarsely set the frame of interaction, while the dominant hand performs fine input within that space. This effectively creates a fourth style of stabilization (beyond Billinghurst et al.'s three): a non-dominant hand stabilized interface.

Hinkley and Pausch [16] investigated using the non-dominant hand as a reference when performing spatial interaction. They found that two hands interacting together provide enough context to maintain a frame of reference without visual feedback.

Mobile Spatial Input

Recent research in mobile computing has led to a number of input techniques that allow for spatial input beyond a touchscreen. For example, SideSight [6], Abracadabra [13], Minput [14], Skinput [15], and HoverFlow [19] all provide a level of spatial input in a mobile setting.

Peephole displays [31] and the original concept from Cameleon [9] offer spatial input that maps physical movement of the device to movement in a virtual world. One can revisit a virtual location by returning to the same physical position.

Sixth Sense [22], a projector and camera-based wearable computer, provides spatial gesture input like Imaginary Interfaces but relies heavily on the projector to provide visual feedback. Similarly, Brainy Hand [28] is an ear-worn camera and projector system.

Spatial Interaction without a Screen

The main reason why mobile devices have not continued to be miniaturized is because the screen must remain a certain size to be usable. Projects such as nanotouch [4] and Ridgepad [17] have increased touch accuracy on very small screens but real miniaturization begins when the screen can be eliminated. For example, Stratchan et al.'s BodySpace [26] assigns positions on the body to specific functions.

Imaginary Interfaces build on Virtual Shelves [20] in that both allow for feedback-free spatial interaction. Virtual Shelves allows picking an item out of 11 locations spread out along the phi and theta planes of a hemisphere. Imaginary Interfaces, in contrast support continuous input in 2D Euclidean space and higher precision because of the *visible* and *re-definable* reference frame.

Visual Memory

Baddeley and Hitch's [2] model of working memory contains a subsystem called the *visuospatial sketchpad* that is responsible for maintaining short term visual memories and spatial relationships. Inspired by this and to avoid complicated physiological models of memory, throughout this paper we will refer to people's short term capacity to recall the location of objects in space as *visuospatial memory.*

USER STUDIES

We conducted three user studies. All three studies investigated users' ability to interact spatially using Imaginary Interfaces. Our main goal across the studies was to determine to what extent users' visuospatial memory could replace visual feedback. That is, can users define an invisible interaction space and successfully manipulate and annotate objects located in that space.

In the following experiments we use sensing hardware (an optical tracker installation) which is not appropriate for a real life deployment. Using a highly reliable and accurate platform allowed us to test the underlying limits of user performance with the interaction style instead of the quality of our implementation.

The first study investigated participants' visuospatial memory while drawing single stroke characters and simple sketches. Participants completed the single stroke characters with high (94.5%) recognition rate. The quality of multi-segment drawings, however, decreased with the number of strokes as visuospatial memory fades over time.

The second study investigated in how far user motion impacts visuospatial memory. Participants' performance in annotating an imaginary drawing dropped when they had to turn around between drawing and annotating. The use of the left hand as a spatial reference, however, alleviated the effect in part.

The third and final study tested participants' ability to point to a location specified in Euclidian coordinates, more specifically vertical units of index fingers and horizontal units of thumbs. We found targeting error to correlate with the Manhattan distance of the target from the user's fingertips, which serve as visual landmarks.

USER STUDY 1: BASIC SHAPES AND DRAWINGS

The purpose of this study was to test if Imaginary Interfaces are subject to the difficulties in connecting stroked reported by Ni and Baudisch [23]. Participants' task was to reproduce a series of predefined sketches. Each sketch required participants to align or connect a stroke to a stroke drawn earlier.

Tasks and Procedure

For each trial, participants were shown a letter-size sheet of paper with one of the stimulus drawings reproduced in Figure 5.

When the participant indicated they were ready, the drawing was hidden and participants replicated it in the space in front of them as illustrated by Figure 4a. All participants created the 'L' posture with their left hand and sketched with their right hand. The trial was complete when participants indicated so.

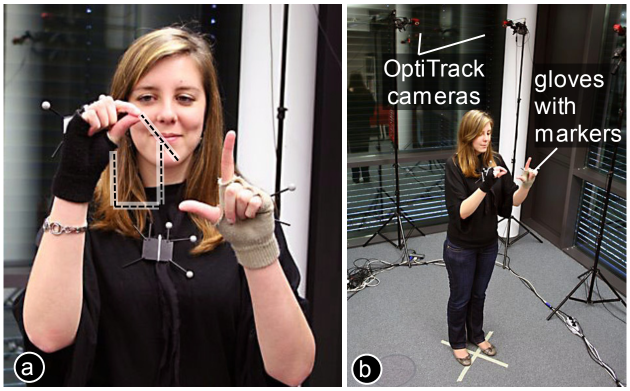

Figure 4: (a) For each trial the participant replicated a simple sketch in imaginary space. (b) An optical tracking system tracked hand positions over time.

The trials were divided into three tasks:

1. Graffiti. The participants drew seven Graffiti characters. The characters, shown in Figure 5a, were selected from [23] because they were especially difficult to complete eyes-free. The challenge was to connect/align the stroke end to obtain proper recognition.

2. Repeated drawing. In this task participants drew a simple square and triangle (stimulus is shown in Figure 5b) repeatedly five times in a row without stopping—the next always drawn over the previous.

3. Multi-stroke drawing. In this task participants drew a set of five simple sketches (see Figure 5c), each of which involved multiple drawing strokes.

Each participant completed all three tasks, i.e., drew all the sketches shown in Figure 5. Each participant completed the experiment session within ten minutes.

Figure 5: Stimulus drawings for the participants to reproduce. (a) Task 1: single-stroke Graffiti character. (b) Task 2: simple shapes drawn repeatedly. (c) Task 3: more complex shapes involving multiple strokes. Arrows show how to complete the shape.

Hypotheses

We only had one quantitative hypothesis, namely that participants would perform fewer Graffiti recognition errors than reported by Ni and Baudisch [23]. Even though the sketches were imaginary, we expected that participants to build up visuospatial memory by watching their hands act and successfully complete the shapes.

The purpose of the *repeated drawing task* was to allow us to obtain a measure of the lower limit of error on connecting (the vertices of the shapes) by using strokes with very simple movements and very short time periods. The purpose of the *multi-stroke task* was to explore how stroke connection accuracy decreases with increasing number of strokes.

Apparatus

The apparatus is shown Figure 4a. In order to obtain full 3D position and rotation information we used an optical tracking system (an 8-camera *OptiTrack*). Participants wore a marker set on the back of each of their hands. A third marker set on the participant's sternum allowed us rotate the collected 3D position data to a common orientation for all users. The system tracked the marker sets with 1mm accuracy.

Participants

12 participants (5 female) were recruited at our institution to take part in the study. They were between the ages of 20 and 30 (mean=24.2, s=2.95). All participants set the reference frame with their left hand and drew with their right. One participant was left-handed (participant #1) but was proficient with using a mouse in his right hand. Participants were given a small gratuity for their time.

Data Processing

We converted the 3D positions/rotations to 2D coordinate space by rotating them as if the participants were directly facing the XY-plane, i.e., their gaze followed the Z-axis. The Z coordinate was then discarded, leaving the remaining X and Y coordinates to specify the location of both hands on a perfectly vertical interaction plane. The final drawing position was taken relative to the left hand.

The apparatus allowed us to track the position of participants' hands, but not the pinch gestures. We manually marked up the beginning and end of each stroke post-hoc. Most participants paused briefly at the start and end of each drawing motion, which simplified the classification.

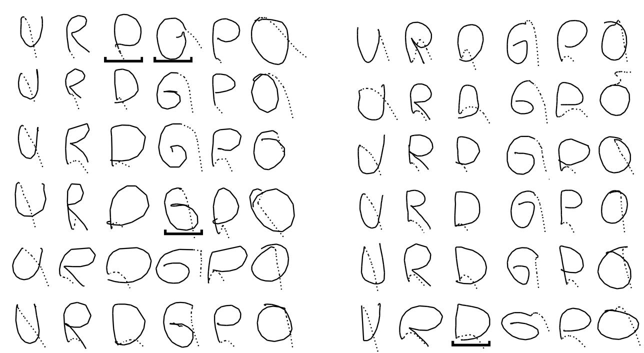

Figure 6: (a) All characters participants drew for the *graffiti task*, resized to match. Each half row is data from one participant. Misrecognized characters are underlined.

We analyzed the *graffiti task* by running the captured drawings through a Graffiti recognizer [7]. The *repeated drawing task* was analyzed by measuring the average distance per vertex of the triangle/diamond. *Multi-stroke drawings* were not formally analyzed.

Results

Graffiti task: Figure 6 shows the complete set of drawings created by the participants for the *graffiti task*. Overall only 5.5% of the gestures were unsuccessfully recognized versus 15.0% for the same subset from [23].

Figure 7 compares error rates to Ni and Baudisch. The error rate of the Imaginary Interface condition is comparable to Graffiti text entry on pen devices (2.9% for the same subset of characters by first time users with 5 minutes of training [21]). While this is not enough data for statistical analysis, the data suggests that the *relative position feature* issue pointed out by Ni and Baudisch does not apply to Imaginary Interfaces, at least not to the same extent. This supports our hypothesis that visuospatial memory helps fill in the blanks.

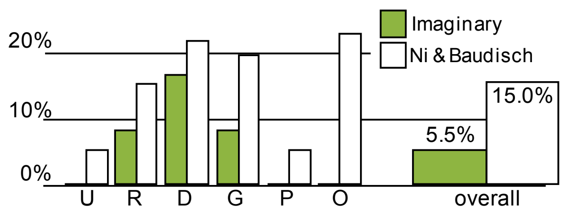

Figure 7: Graffiti recognition error rates compared to Ni and Baudisch's feedback-less gesture input.

Repeated drawing task: Figure 8 shows the complete set of drawings created by the participants for this task.

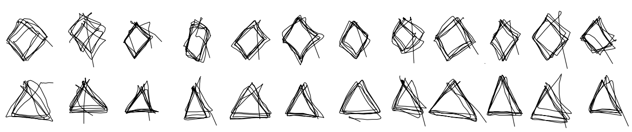

Figure 8: All drawings made by all participants for task 2. Each column is data from one participant.

Figure 9 shows the average distance between all vertexes for (a) each repetition from the first and (b) each repetition from the previous. Overall the average error from previous for the diamond was 2.20cm (s=0.90cm) and for the triangle was 3.25cm (s=2.31cm). The decreasing error distance from the previous suggests that participants built up visuospatial memory of the shape with repetition and the fact that users kept drifting away from the first suggests that later loops overwrite earlier ones.

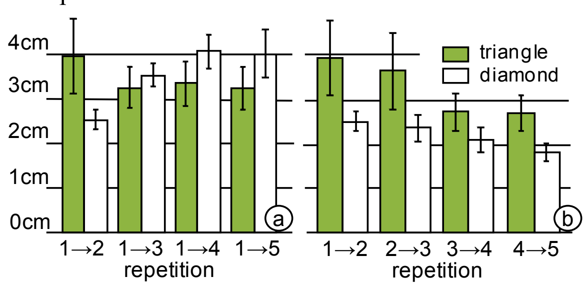

Figure 9: Graph showing all error rates by repetition (+/- standard error of the mean): (a) error rate relative to first drawing and (b) error rate relative to previous drawing.

Multi-stroke drawing task: Figure 10 shows all sketches created by the participants. We observed several interesting points: (1) Alignment between strokes seems to decrease with the complexity of the sketch. (2) The individual letters of the ABC string are well drawn; participants, however, condensed whitespace, causing letters to overlap. (3) Relative scale appears reasonably correct. Misalignment appears to be caused mostly by translation errors derived from choosing the wrong starting point.

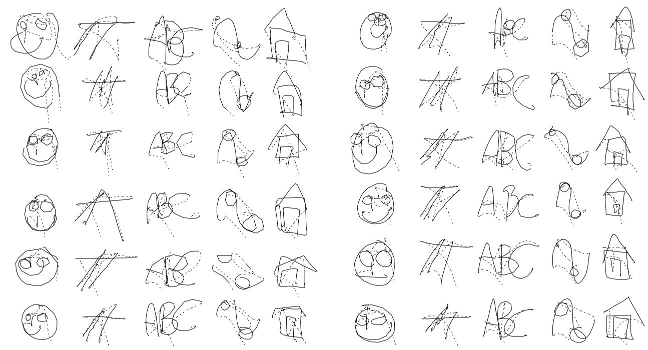

Figure 10: The drawings made by all 12 participants in multi-stroke drawing task. Each half row is one participant, resized for comparison.

Discussion

Participants created the basic characters and sketches despite the lack of visual feedback. Unlike the finding of Ni & Baudisch, problems with closing shapes were minimal; we think this is because users built up visuospatial memory by watching their hands throughout the interaction.

While alignment within a stroke was good, aligning strokes in multi-stroke sketches seemed to challenge participants.

However, it remains unclear under what conditions visuospatial memory begins to fade. We examined this in the next user study.

USER STUDY 2: RETURNING TO A DRAWN FEATURE

The purpose of this study was to understand what causes visuospatial memory to fade when using Imaginary Interfaces. Participants' task was to draw a simple shape, then go back and point to one of the vertices of what they just drew. In one condition, participants pointed right away; in another condition participants had to turn around between drawing and pointing. Our main hypothesis was that intermitted actions affect visual memory and thus reduce accuracy.

Task

Participants began each trial (Figure 11a) with their hands at their side and were shown a simple glyph on a computer screen. As shown in Figure 12, all glyphs were constructed from four strokes.

Participants pressed the footswitch to begin the trial and, depending on condition, raised their hands performing the 'L' gesture with their left hand. Now participants replicated the glyph with the right hand (Figure 11b) and pressed the footswitch again. In half of the trials (see below), participants rotated their bodies by 90° at this point.

Figure 11: (a) At the start of each trial the participant stood in front of one of the footswitches and views the glyph on a monitor. Each trial consists of two phases: (b) drawing the glyph and (c) selecting a corner of the drawn glyph.

Now a computer generated voice announced "select 1", "select 2" or "select 3" and started the timer. Participants selected the respective corner from the glyph they just had drawn (Figure 12b) and committed by pressing the foot switch, which stopped the timer.

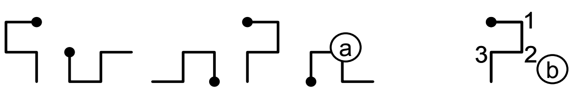

Figure 12: During each trial, participant (a) drew one of these glyphs and then (b) pointed to one of its three corners.

Participants dropped their arms to get ready for the next trial and if they had not rotated earlier in this trial they rotated now.

The system recorded the participant's rendition of the glyph and the selected position. Error was defined as the Euclidean distance between the selected position and the location of the corner drawn earlier. Task time was the duration of corner selection activity.

Independent Variables

There were four conditions defined by the two-level independent variable *body rotation* and the two-level independent variable *reference system*.

Body rotation: In the *rotate* conditions, users rotated their entire body by 90° between drawing the glyph and acquiring the point on it. In the *stay* conditions they did not.

Reference system: In the *hand* condition, participants drew the glyph with their reference hand in the air and position was computed relative to this hand. In the *none* condition, they drew with their left hand along their side and position was computed relative to their torso.

Hypotheses

We had two hypotheses.

H1: We expect lower error in the *stay* condition than in the *rotate* condition. Participants should be able to fully use their visuospatial memory in the *stay* condition, while the body rotation would impair it.

H2: In the rotate conditions, we expect lower error in the *hand* condition than in the *none* condition. In the *hand* condition the left hand should in part fill in for the lack of the visual reference frame eliminated by the body rotation.

Although we had no expectation for a difference in performance between *hand* and *none* conditions when the participant did not rotate, we were interested in seeing if an effect was noticeable. We are also curious if the *hand* and *none* conditions would differ in task time.

Apparatus

The apparatus for this study was identical to that used in User Study 1 except that we added two monitors on the floor and four footswitches arranged in a circle around the participant as shown in Figure 11a.

Participants

A new set of participants, 10 males and 2 females, were recruited from our institution and community to participate in this study. They were aged between 22 and 31 years old (mean=24.2, s=3.3). All were right handed. Each received a small gift in exchange for their time.

Experiment Design

The experiment was a 2 *body rotation* (*rotate*, *stay*) × 2 *reference system* (*hand, none*) within subjects design. Each condition consisted of 15 trials. Each trial within a condition used one unique combination of glyph (see Figure 12) and corner number. The presentation order within a block was randomized and the blocks were counterbalanced using a balanced latin square.

With 12 participants, 4 conditions and 15 trials per condition, a total of 720 trials were completed in the study. Due to tracking errors with the 3D tracking system, 16 of those trials (2.2%) were discarded and not included in the subsequent analysis.

Participants were trained on the operation of the experimental apparatus and the task until they indicated they were comfortable and understood what was required of them. All participants completed the study in 30 minutes or less.

Results

We performed a multivariate 2×2 repeated measures ANOVA on error and task completion time. Figure 13 contains a summary of the results.

The *hand* condition had an average error of 5.4cm (s=0.4cm) and the *none* condition had an average of 6.3cm (s=0.3cm). The ANOVA indicated this to be significantly different ($F_{1,11}$=16.007, p=0.002).

Also, the *stay* condition (mean=5.1cm, s=0.3cm) had significantly less ($F_{1,11}$=16.007, p=0.002) average error than the *rotate* condition (mean=6.6cm, s=0.4cm).

There was no interaction between the *reference system* and *body rotation* conditions.

To investigate closer we performed four post-hoc paired samples t-tests (and subsequently controlled for inflation of Type I error by using an adjusted α of 0.05÷4=0.0125). The first two post-hoc tests compared the *hand* and *none* conditions separately in both *body rotation* conditions. When the participant did not rotate (i.e., in the *stay* condition) the test did not show a significant difference but in the *rotate* condition it did (t_{11}=4.621, p=0.001).

Similarly, the last two post-hoc tests compared the *stay* and *rotate* conditions separately in each *reference system* conditions. When the reference system was *hand* the test did not show a significant difference but it did (t_{11}=3.978, p=0.002) in the *none* condition.

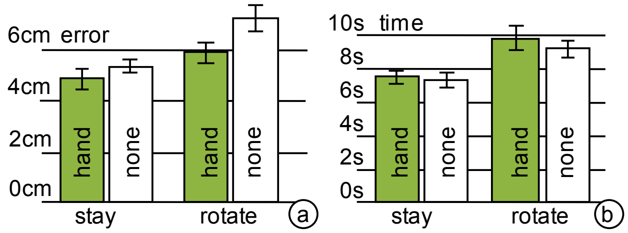

Figure 13: (a) Error and (b) time by body rotation & reference system (+/- standard error of the mean).

With respect to task completion time, the ANOVA indicates a significant main effect for the *body rotation* factor ($F_{1,11}$=42.135, p<0.001). This effect was completely expected as it took time for participants to perform the rotation.

More interestingly, the ANOVA also showed a significant main effect for the *reference system* factor ($F_{1,11}$=5.657, p<0.038). Although the difference was small (8.3s compared to 8.7s) the hand condition was significantly slower than without the hand.

Discussion

Causing a disruption of visuospatial memory, in this case by forcing the participant to make a quarter turn, led to significantly higher error rates. However, this effect was significantly lessened by using the left hand as a reference.

Imaginary Interfaces in part rely on well-maintained visuospatial memory. Using a visible reference point, such as the left hand, improves performance to a point, but we must still deal with situations when visuospatial memory has completely degraded or was never present in the first place. That is subject of the following sections.

IMAGINARY POINTING BY COORDINATES

The previous two studies explored the use of Imaginary Interfaces in situations where users either sketch vaguely or recall a sketch from visuospatial memory. While we think of this as being the most common application scenario, there are situations where users might want to take *factual coordinate data* into account when drawing or pointing. A user might construct a stock curve that passes through certain specific points or users might want to point to a coordinate pair received over an audio connection (the audio manual says "the green button is located at…").

Imaginary Interfaces offer a very basic mechanism to support this. As illustrated by Figure 14, the 'L' shape forms a coordinate system with the user's thumb forming the unit vector in *x* and the index finger forming the unit vector in *y*. This reference allows referring to locations in imaginary space using coordinate pairs, such as (2,1) in the shown example. Negative coordinates can be used to refer to locations left of the index finger and below the thumb.

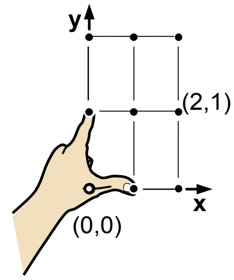

Figure 14: The user's thumb and index finger span a coordinate system that gives each point on the plane a unique address. Here the point "two thumbs right, one index finger up" is labeled (2,1).

To understand the capabilities and limitations of coordinate-based imaginary pointing we conducted a third user study.

USER STUDY 3: POINTING BASED ON COORDINATES

The purpose of this study was to measure the performance of coordinate-based imaginary pointing. Participants acquired targets given to them as coordinate pairs in (thumb, index) length units. We measured error as the Euclidean distance from the target. We hypothesized that error would grow with the distance from the tips of index finger and thumb, which participants use would as visual landmarks.

Task

For each trial, participants started in a neutral position with their hands held loosely at their sides. Participants now received the target location as two digits via audio and displayed on a monitor. They pressed a footswitch which started the timer and started the trial.

Participants now raised their hands, formed an 'L' gesture with their left hand and acquired the respective target by pinching with their right hand (Figure 15a). Participants committed the acquisition by pressing a footswitch again. This recorded the 3D location and rotation of the body and both hands, played a confirmation sound and completed the trial.

In preparation for the next trial, participants dropped their arms and rotated approximately 90° to simulate mobile use.

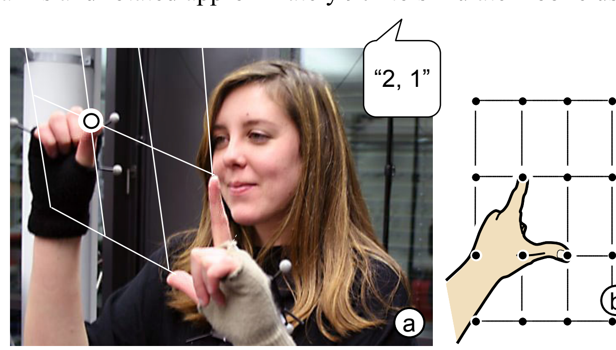

Figure 15: (a) Participant selected a target at the coordinate announced by the system. (b) Participants acquired 16 positions from (-1,-1) to (2,2).

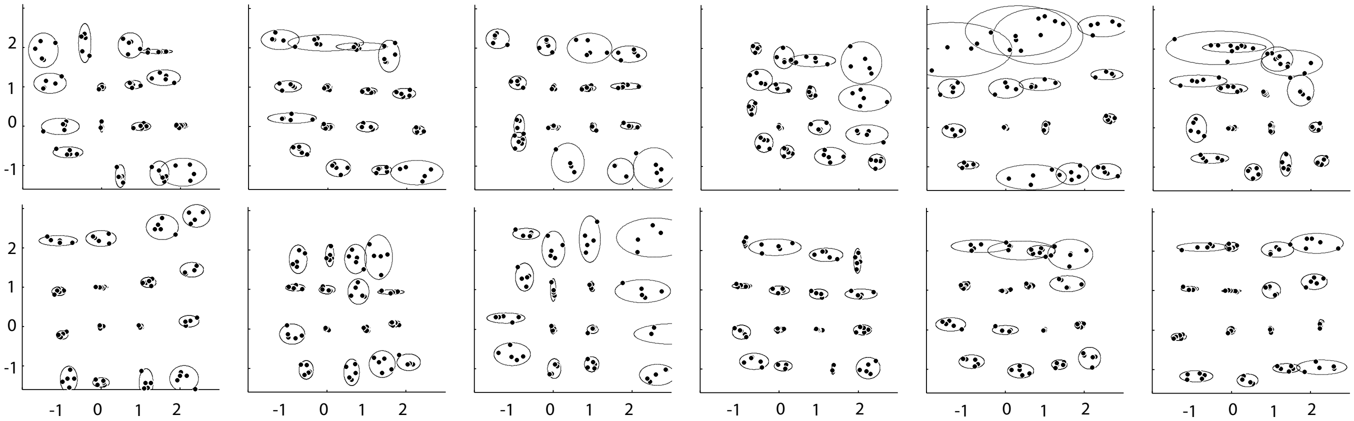

Figure 16: Error per target position per participant. Each oval approximates 95% of selections for that target.

Experiment Design

The participants selected from all positions on a 4×4 grid (Figure 15b) 5 times each in random order for a total of 80 trials. Together, 12 participants completed 960 trials.

All participants completed the trials in 30 minutes or less.

Apparatus

We used the same marker-based tracking system as in User Study 2.

Participants

We recruited a new set of 12 participants (5 female) from our institution and community. Two participants were left-handed and both used the mouse in their right hand. Participants were between 21 and 27 years old (mean=23.0, s=2.0). Each received a small gift for their time.

Hypotheses

We expected that error would increase with distance from the two main landmarks, i.e., the fingertips of the 'L' hand. In particular:

H1: pointing accuracy will be highest at the fingertips.

H2: pointing accuracy will decrease as the distance from the nearest fingertip increases.

Data Preparation

To get comparable numbers we corrected for difference in participants' interaction planes and hand sizes.

First we corrected for participants interacting in planes of different tilt. For each user, we determined the interaction plane using a linear planar regression based on all 3D positions. We then rotated the plane into the XY plane and projected onto that plane, discarding the Z coordinate. All further processing was done with the resulting 2D data.

Next we corrected for differences in finger sizes. Since the tracking system tracked only entire hands, we reconstructed finger tips from the data for the (0,0), (0,1) and (1,0) targets (this step was only required for the marker-based tracking setup in the study, our camera-based prototype "sees" finger lengths).

Next we corrected for the rotated and skewed interaction planes based on the three points gathered in the last step. The coordinate space was transformed until the tip of the index finger was located at (0,1), the tip of the thumb at (1,0) and the origin at (0,0).

We removed 16 outlier trials (1.7%) from the data. We defined an outlier trial where either the X or Y position was more than 2 standard deviations from the overall mean for that position after the above corrections. Many of the outliers appeared to be the participants mistakenly swapping the X and Y coordinates.

Results

Figure 16 shows the error distribution for each location tested in this study for all users. Each oval encodes two standard deviations around the mean position (i.e., 95% of all targeting positions) for that location.

Figure 17 shows another view of the same data. It graphs error spread (i.e., four times the standard deviation per user per location) by the Manhattan distance from the nearest visual landmark (i.e., the index finger tip or thumb tip). The figure shows that locations with a Manhattan distance of 0 (i.e., the finger tips) were indeed the most accurate.

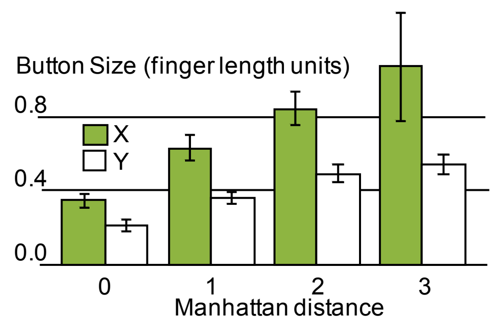

Figure 17: Error spread by Manhattan distance from the nearest finger tip (+/- standard error of the mean). The button sizes would capture 95% of select events per user.

Discussion

In this study we quantified the positioning error for 16 locations surrounding the left hand and calculated the smallest button sizes that remain selectable at these locations.

As expected the fingertips were the most accurate locations, with button sizes of 0.35 thumbs wide and 0.21 index fingers high. As the Manhattan distance increased from the nearest finger tip error increased linearly as shown in Figure 17.

Overall, the size of the targeting areas differs greatly within the range tested. Upon inspection of Figure 16 you can see that some participants had a lot of trouble with the more distant targets while others did not. We expect that with better training all users could reasonably target distant imaginary targets.

Despite complaints from the participants, who stated they did not like the locations in the negative quadrants, these locations offered similar targeting ability as those in the positive quadrant.

IMPLICATIONS FOR DESIGN

Based on the three studies presented above we have gained first insights into the design space of Imaginary Interfaces. Here is a summary of the implications for designers of systems based on the Imaginary Interfaces concept:

1. To benefit most from visuospatial memory, encourage users to create annotations right away while memory is still fresh.
2. When users are mobile using a reference hand helps to maintain visuospatial memory. This is beneficial when annotating drawings created earlier.
3. Exploit the features (such as finger length) of the reference hand to provide directions to imaginary interface elements. For example, the OK button is located at position 1,2.
4. Pointing accuracy decreases as we move away from the reference hand. To improve pointing accuracy, draw and point close to the reference hand. To use imaginary space effectively, complement small virtual buttons close to the reference hand with large virtual buttons at a distance.

IMPLEMENTATION

A wide variety of mechanisms can track a user's hands and can thus be used to implement Imaginary Interfaces. For the experiments above we used an optical tracker system to create our prototype. This setup, however, requires a fixed location and bulky markers to be placed on the user.

To investigate more practical usage scenarios we also built a second prototype, shown in Figure 18. Derived from Gesture Pendant [25], it uses computer vision to track the hands of the user when held in front of their chest. The device consists of a *Fire-i* black-and-white camera with a 107° view-angle (www.unibrain.com) and an infrared-pass filter, as well as a ring of infrared LEDs (shown in Figure 18e). It measures 5.25×5.25cm and uses a computer connected via Firewire 400 as its computational backend.

We obtain good separation of the user's hands from the background by illuminating the space with infrared light (Figure 18a,b). Note, however, that this approach requires controlled lighting and, for example, does not work in sunlight. From the raw thresholded image, we find the contours (Figure 18c). Next, as shown in Figure 18d, we identify the left and right hands, determining the tip of the thumb and index finger and compute the intersection between the lines defined by the two. These three points define the interaction plane.

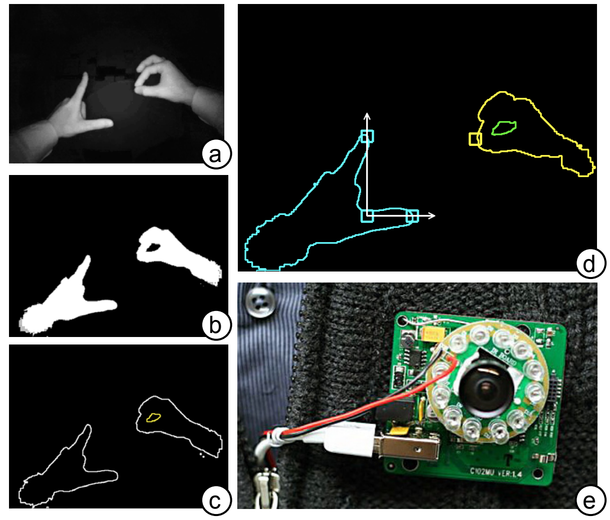

Figure 18: (a) Raw image from camera. (b) Threshold applied. (c) The extracted contours. (d) Three points from left hand captured and the pinch point from the right hand. (e) The device itself.

The device then recognizes pinch gestures of the right hand by using a connected components algorithm, as introduced by Wilson [30]. The location of the pinch is transformed to fit the coordinate system defined by the left hand and passed to the currently running application.

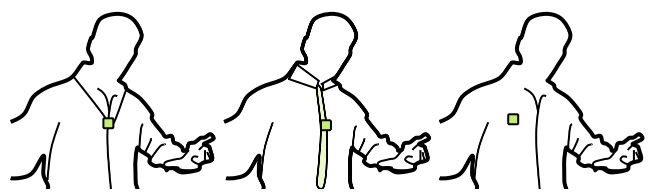

Figure 19: We envision future versions to be a small self-contained clip with cellular radio that can be worn as a brooch, pendant or clipped onto clothing.

As cameras, processors, and wireless communication components continue to shrink and vision-based sensing continues to get more powerful, we envision future versions of the device to shrink to the size of a button or brooch, so that they could be worn almost invisibly as part of the user's clothing (Figure 19).

CONCLUSIONS

We presented *Imaginary Interfaces*, user interfaces that allow users to interact spatially without visual feedback. While traditional interfaces are either spatial (e.g., touch screens) or non-visual (gestures), Imaginary Interfaces combine both aspects.

The main application of Imaginary Interfaces is to bring spatial interaction to screen-less mobile devices. Such devices have the highest potential for miniaturization. The ability to create and share simple sketches on the fly opens up a range of new application scenarios, such as sharing sketches with driving directions as part of a phone call or even, when sufficiently learned, interacting with any conventional spatial interface that has been loaded onto the user's imaginary interaction plane.

Promising directions for future work include investigating methods of learning an imaginary interface, extending Imaginary Interfaces to allow annotation of interactions with speech, adding auditory cues as a feedback channel that allows users to explore an imaginary space and extending Imaginary Interfaces to 3D.

ACKNOWLEDGEMENTS

We thank the members of our lab, in particular Anne Roudaut, Christian Loclair and Christian Holz for many useful discussions and help with the paper, video and implementation. We also thank the study participants for their time and the reviewers for their valuable comments.

REFERENCES

1. Ashbrook, D. *Enabling Mobile Microinteractions*. PhD Thesis (2010).
2. Baddeley, A.D., and Hitch, G. Working memory. In G.H. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory 8*, (1974), 47–89.
3. Baudel, T. and Beaudouin-Lafon, M. Charade: remote control of objects using free-hand gestures. *CACM 36*, 7 (1993), 28–35.
4. Baudisch, P and Chu, G. Back-of-device interaction allows creating very small touch devices. In *Proc. CHI,* (2009), 1923–1932.
5. Billinghurst M., Bowskill J., Dyer N., and Morphett J., Spatial information displays on a wearable computer, *IEEE CG&A 18*, (1998), 24–30.
6. Butler, A., Izadi, S., and Hodges, S. SideSight: multi-"touch" interaction around small devices. In *Proc. UIST*, (2008), 201–204.
7. Couvreur, J. *Pen Stroke Recognition.* Graffiti code download at http://blog.monstuff.com/archives/000012.html
8. Emmorey, K., Tversky, B., and Taylor, H.A. Using space to describe space: perspective in speech, sign, and gesture. *Spatial Cognition and Computation 2*, 3 (2000), 157–180.
9. Fitzmaurice, G.W. Situated information spaces and spatially aware palmtop computers. *CACM 36*, (1993), 39–49.
10. Fukumoto, M. and Tonomura, Y. Body-coupled FingerRing: wireless wearable keyboard. In *Proc. CHI,* (1997), 147–154.
11. Graffiti. http://en.wikipedia.org/wiki/Graffiti_(Palm_OS)
12. Guiard, Y. Asymmetric division of labor in human skilled bimanual action: the kinematic chain as a model. *Journal of Motor Behaviour 19*, 4 (1987), 486–517.
13. Harrison, C. and Hudson, S.E. Abracadabra: wireless, high-precision, and unpowered finger input for very small mobile devices. In *Proc. UIST*, (2009), 121–124.
14. Harrision, C. and Hudson, S. Minput: enabling interaction on small mobile devices with high-precision, low-cost, multipoint optical tracking. In *Proc. CHI*, (2010), 1661–1664.
15. Harrison, C., Tan, D., and Morris, D. Skinput: appropriating the body as an input surface. In *Proc. CHI*, (2010), 453–462.
16. Hinckley, K., Pausch, R., and Proffitt, D. Attention and visual feedback: the bimanual frame of reference. *Proc. Symposium on Interactive 3D Graphics*, (1997), 121–126.
17. Holz, C. and Baudisch, P. The generalized perceived input point model and how to double touch accuracy by extracting fingerprints. In *Proc. CHI*, (2010), 581–590.
18. Jacob, R.J., Girouard, A., Hirshfield, L.M., Horn, M.S., Shaer, O., Solovey, E.T. and Zigelbaum, J, Reality-based interaction: a framework for post-WIMP interfaces. In *Proc. CHI*, (2008), 201–210.
19. Kratz, S. and Rohs, M. HoverFlow: expanding the design space of around-device interaction. In *Proc. MobileHCI*, (2009), 1–8.
20. Li, F.C.Y., Dearman, D., and Truong, K.N. Virtual Shelves: interactions with orientation aware devices. In *Proc. UIST*, (2009), 125–128.
21. MacKenzie, I.S. and Zhang, S.X. The immediate usability of Graffiti.In *Proc. GI*, (1997), 129–137.
22. Mistry, P., Maes, P., and Chang, L. WUW - wear Ur world: a wearable gestural interface. In *CHI Extended Abstracts*, (2009), 4111–4116.
23. Ni, T. and Baudisch, P. Disappearing mobile devices. In *Proc. UIST*, (2009), 101–110.
24. Rekimoto, J. GestureWrist and GesturePad: Unobtrusive Wearable Interaction Devices. In *Proc. ISWC*, (2001), 21–27.
25. Starner, T., Auxier, J., Ashbrook, D., and Gandy, M. The Gesture Pendant: a self-illuminating, wearable, infrared computer vision system for home automation control and medical monitoring. In *Proc. ISWC*, (2000), 87–94.
26. Strachan, S., Murray-Smith, R., and O'Modhrain, S. BodySpace: inferring body pose for natural control of a music player. In *CHI Extended Abstracts*, (2007), 2001–2006.
27. Sturman, D.J. and Zeltzer, D. A design method for "whole-hand" human-computer interaction. *TOIS 11*, 3 (1993), 219–238.
28. Tamaki, E., Miyaki, T., and Rekimoto, J. Brainy hand: an ear-worn hand gesture interaction device. In *CHI Extended Abstracts*, (2009), 4255–4260.
29. Wilson, A.D. and Oliver, N. GWindows: robust stereo vision for gesture-based control of windows. In *Proc. ICMI*, (2003), 211–218.
30. Wilson, A.D. Robust computer vision-based detection of pinching for one and two-handed gesture input. In *Proc. UIST*, (2006), 255–258.
31. Yee, K. Peephole Displays: Pen Interaction on Spatially Aware Handheld Computers. In *Proc. CHI*, (2003). 1–8.