Should I Stay or Should I Go? Selecting Between Touch and Mid-air Gestures For Large-Display Interaction

Mikkel R. Jakobsen, Yvonne Jansen, Sebastian Boring, and Kasper Hornbæk

Department of Computer Science, University of Copenhagen, Copenhagen, Denmark
{mikkelrj, yvja, sebastian.boring, kash}@diku.dk

Abstract. Users can interact with large displays in many ways, including touch and mid-air gestures. However, it remains unclear how these ways compare and when users choose one over the other. In a first experiment, we compare touch and mid-air gestures to identify their relative performance for target acquisition. In a second experiment, participants choose freely between touch and mid-air gestures and we artificially require movement to simulate situations where mid-air is considered beneficial. Results from the first experiment show mid-air to be overall slower than touch depending on the task; in the second experiment, participants mostly chose touch in particular for selecting small targets and they rarely switched between mid-air and touch. Results also show that when faced with an increasing cost of using touch in the form of movement, participants chose mid-air over touch; touch remains as fast as mid-air on average.

Keywords: Large display, mid-air, touch, freehand gestures, user study.

1   Introduction

Users can interact with large displays in many ways, including through touch when being close to the display and through mid-air gestures when standing at a distance. Both touch and mid-air gestures leverage our basic human ability to point with our hands at objects of interest. Each of them has been researched in isolation [31,38] and they have been researched in combination [24,26,39], but they have rarely been compared (except for specific public display scenarios [20]).

Touch and mid-air gestures seem appropriate for different tasks or situations. For working with detailed information up close with a large high-resolution display, touch requires direct interaction through physical contact, which may be faster and preferred over indirect input (e.g., using a mouse [35]). In contrast, users may want to view large displays from a distance to gain an overview. At a distance, mid-air gestures allow users to interact with targets anywhere on the display [40]. Both interaction styles may be combined to support large-display interaction, allowing users to transition between them. However, an important question remains to be answered: When do users choose one over the other? This paper aims at answering this question.

Users may have different reasons for choosing to interact through touch or mid-air. First, the relative performance of touch and mid-air may influence users' choice.

There are several reasons for expecting touch and mid-air to perform differently for common tasks (e.g., target acquisition): Display space and input space are unified in touch but decoupled in mid-air; touch gives tactile support, but incurs friction while dragging, in contrast to mid-air movement; mid-air gestures can be performed at a distance, but distance affects accuracy; touch is limited to display parts within arms' reach and extensive movement is required to interact with remote parts; and touch onset naturally delimits gestures, whereas mid-air gestures need an explicit delimiter (e.g., pinching). Empirical studies are needed to help understand these differences.

However, users choose to interact through touch or mid-air not only based on their relative performance. For instance, mid-air gestures might be used for a task because they require less effort, even though they might be slower or less accurate than touch. This choice may depend on the task or result out of convenience (e.g., in order to avoid repetitive locomotion). Investigating when users choose one over the other and how they switch between touch and mid-air gestures is important for understanding how and when they might be combined.

We present two experiments: (1) a controlled experiment that compares the relative performance of touch and mid-air gestures for different target acquisition tasks in which we vary target size, distance, and whether target locations are known; and (2) an experiment in which users can freely choose between touch and mid-air gestures, but are required to step away from the display at different intervals (i.e., simulating conditions that benefit from mid-air input and impose a cost on using touch). The experiments present the first empirical data on users' choice between touch and mid-air, which may help better take advantage of both types of input for wall-displays.

2 Related Work

Touch and mid-air gestures are particularly interesting input modes for interacting with large displays: they allow free movement in front of the display, can be used without a dedicated input device, and can therefore be used straight away and by several users at a time. Other input options that allow freedom of movement have been researched, including gyroscopic mice, handheld devices (e.g., smart phones [6,25]), and tangibles [18]. However, in this paper we mainly discuss direct touch and mid-air gestures that use free hand movement. Also, while many types of gestures have been researched (e.g., for moving objects or executing commands), we focus on selection.

In the following, based on a review of literature, we discuss factors that may influence the use of, and choice between, touch and mid-air gestures. We also review research about the combination of touch input and mid-air gestures.

2.1 Touch-based Interaction

Touch is familiar to many people and simplifies interaction on large displays (e.g., by allowing direct pointing to an object instead of moving a mouse pointer [38]). Yet, it introduces new challenges: First, finger occlusion makes accurate pointing at small targets difficult; this has been addressed through novel interaction techniques [41].

Second, touch requires users to be within reach of the point of interaction. When people want to interact with content further away on a large display, they must physically move there. Techniques such as Frisbee [19] and Drag-and-pop [4] provide access to distant content with less movement. Nacenta et al. [28] compared different techniques for reaching distant targets located on multiple displays. They found a control-display ratio of 1:1 to be preferable to amplified touch movements.

Third, content further away is not always visible when standing close to the display; close proximity makes it difficult to search the display. Although users can step back in order to get an overview, and have been found to do so [16], additional effort is required to go back to the display in order to interact. Several researchers have explored distant touch interaction that allow for overview at a distance (e.g., Touch Projector [6] or ARCPad [25]), but they require the use of a handheld device.

2.2 Mid-air Gestures

Mid-air gestures have the advantage that users can directly point to an object, similar to touch, except that users can do so from a distance [38]. Most prominent are mid-air techniques using ray casting, which extend a finger or object with an imaginary line to determine the point of contact with the display. Early research used laser pointers to interact with distant content [28]. Later work has investigated freehand pointing [40]. Research on mid-air interaction has addressed several challenges.

First, for techniques that continuously track the user's hand, there is no differentiation between action (i.e., selecting an object) and movement (i.e., moving towards a target). Naturally, techniques that require a dedicated device can have buttons to perform selection operations. For freehand pointing, several gestures have been proposed to trigger a selection: the most often used is the pinch gesture where users pinch together index finger and thumb to trigger an action [22,34,42]; other techniques such as AirTap [40], ThumbTrigger [40], or SideTrigger [3] use different gestures. Techniques differ in how fingers used for ray casting (often the index finger) move during the selection gesture, which affects pointing.

Second, mid-air pointing generally suffers from low accuracy. A common cause is the natural hand tremor, which is particularly problematic for small targets at far distance [27]. Vogel et al. compensate for these problems by switching between relative pointing and absolute pointing [40]. Nancel et al. used different regions on a mobile device for different control-display ratios [29]. Relative pointing techniques can thus improve pointing accuracy, but require recalibration or clutching.

Third, users become less accurate without visual feedback even with direct pointing through ray-casting [9]. Users have to relate hand movements to on-screen cursor movement because the input space is separated from the output space [14].

2.3 Combination of Touch and Mid-air Gestures

Having both touch and mid-air input available at the same time is feasible and earlier work has emphasized that the techniques may be integrated (e.g., [24]). Directly related to our work is the study of Schick et al. that compared a touch-and-point condition,

in which participants could point at an object and hold their arm still for .25s to select it, to a touch-only condition [37]. Touch-and-point was faster and preferred, requiring less physical effort. However, the study involved moving rather large 300 × 300 pixel blocks (on a 25ppi display) and did not control for target distance. Vogel and Balakrishnan developed an ambient display that supported transitions from implicit interaction at a distance to explicit interaction through touch [39]; others have explored such transitions [23,26]. Touch and mid-air interaction has also been unified for tabletop displays: the continuous interaction space allows moving from touch to gestures above the surface [24]; Hilliges et al. supported picking up objects for mid-air manipulation [13]. Pointable, which augments touch with in-air pointing to allow bimanual interaction with distant content, was found to perform comparable to multi-touch [3].

In sum, empirical comparisons of touch and mid-air gestures for wall-display interaction are rare. Several factors might affect how users would choose between the two modalities if both were available at the same time, but it is unclear how users choose; we have found no research on this. Thus, with this paper we aim to contribute empirical data to help understand when users choose between touch and mid-air gestures.

3 Experiment 1: Touch vs. Mid-Air

We first conducted an experiment comparing touch and mid-air for target acquisition tasks. The purpose was to obtain empirical data on user performance and satisfaction for touch and mid-air gestures that would allow us to hypothesize about when users would choose one or the other for different tasks. Our aim was not to conclude on the relative performance of touch and mid-air in general; the results do not necessarily generalize to other implementations. The experiment focused on interaction with wall-sized displays, on which mid-air gestures have been primarily used [22,30].

3.1 Interfaces

Participants used two interfaces that implement pointing and selection using either touch or mid-air. With the *Touch* interface, a touch cursor appears when a participant's finger touches the surface, and the touch is registered as a selection. With the *Mid-air* interface, participants move a cursor on the display using ray casting similar to Vogel and Balakrishnan's technique [40]: the cursor is placed at the point where the ray cast from the tip of the user's index finger intersects the display plane. We chose ray casting because it is the "canonical pointing technique" (according to Bowman et al. [7], p. 82) and it is straightforward to use. Participants make a selection using a SideTrigger gesture [3]: while pointing, they move their thumb towards the middle finger, which is curled toward the palm (see **Fig. 1**–C). "Clicking" the middle finger provides kinesthetic feedback while minimizing involuntary movements of the index finger during selection [3]. We considered using the other hand to make selections, but decided against it, since touch is also (in the present experiment) a single-handed technique. We used the €1-filter [8] to compensate for jitter in cursor movements.

4

3.2 Apparatus

We conducted the experiment on a wall-sized display (see **Fig. 1**) that consists of 12 HD projectors with a total of 7680 × 3240 pixels and a resolution of around 68 pixels per inch. Touch on the display is detected through camera-based tracking. Input from six cameras, each capturing 640 × 480 pixels at 30 frames per second (fps), are processed by Community Core Vision. A custom program written in Java multiplexes the tracked touch points. The overall touch resolution is around 17 pixels per inch.
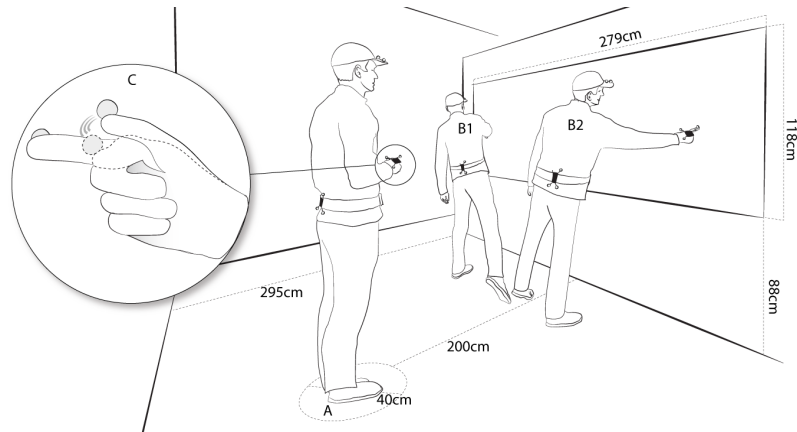

Fig. 1. Experimental setup: A–starting position for all tasks; B–maximum distance to cover in the touch condition; C–thumb trigger gesture used to make selections in the mid-air condition.

Participants were tracked using an OptiTrack motion capture system (.5 mm tracking error, 50 fps). Participants wore a baseball cap, a belt, and a glove with markers attached. This helped quantify head turning and body movement, and gave position and orientation of the hand and the position of the tips of the index finger and thumb.

3.3 Tasks

We used two types of target selection task: *Varied* and *Fixed*. Both consisted of consecutive selections of targets, typical for evaluations of input devices. Previous studies have typically chosen to either (1) vary the size and distance of targets [e.g., 21] or (2) keep the size and distance constant within a sequence (e.g., the reciprocal task [35]).

We designed our tasks to manipulate participants' ability to anticipate target locations, which may influence the relative performance of Touch and Mid-air. While participants can anticipate the next target location in a reciprocal selection task, mixing combinations of size and distance in a sequence requires them to visually search for targets, and they cannot anticipate the direction in which to move for selecting the next target. As Touch requires close proximity to the display, visual search is harder due to the limited field of view. We further expect larger anticipation effects for Touch where larger body movements are required to reach distant targets. Using both types of task helps us investigate these differences. The two tasks are as follows:

- *Varied*. Participants select 13 targets shown (as a red circle) one at a time. When they successfully select a target, the next target appears at a random location, but at a given distance from the previous target. All combinations of size and distance occur once in the sequence. The first target is always 128 pixels in diameter and placed in the center of the display.
- *Fixed*. Participants perform nine alternate selections of two targets of a fixed size, spaced a given distance apart. The current target is shown as a red circle, while the other target is represented as a gray circle. Upon successful selection of the current target, the other target turns red.

The selection of the first target marks the beginning of both tasks; only data from the following selections were used.

We also wanted to understand how the distance and size of targets influence the relative performance of touch and mid-air and thus varied both (see **Table 1**). We varied target size from 32 pixels to 512 pixels (1.2cm to 18.8cm on the display): smaller targets were deemed impractical for both Touch (due to occlusion) and Mid-air (due to limits in pointing accuracy). The corresponding visual angle was between 1.3° and 21° when standing 50cm from the display. The visual angle of a target varies proportionally to viewing distance: A 128px target has the same visual angle at 50cm distance (~ comfortable touching distance) as a 512px target at 2m.

We varied target distance from 768 pixels to 6144 pixels, which is 10% to 80% of the display width, and 31° to 132° visual angle viewed from 50cm distance.

Participants were allowed to move around freely in both interface conditions. We considered restricting movement in the Mid-air condition, but since movement is required for Touch, we allowed movement so as to make the conditions more similar. However, participants started each task from a fixed position (**Fig. 1**–A). For Mid-air participants could thus move in order to point more accurately; moving changes the control-display ratio, which depends on both the viewing distance and viewing angle.

3.4 Participants

We recruited 19 volunteers (14 male), 19–36 years old ($M = 26$), to participate; all but two were right handed. Participants received an equivalent of €25 as compensation.

3.5 Experimental Design

The experiment used a within-subjects design with interface (Mid-air, Touch), task type (Varied, Fixed), target size (3 levels), and target distance (4 levels) as factors. For each interface, participants performed a series of tasks for both task types. The

Table 1. Target sizes and distances. Visual angle is at a 50cm distance to the display.

	Target sizes			Distances between targets			
pixels	32	128	512	768	1536	3072	6144
cm	1.2cm	4.7cm	18.8cm	28cm	56cm	112cm	224cm
visual angle	1.3˚	5.3˚	21˚	31˚	59˚	96˚	132˚

order of interface was counterbalanced across participants to compensate for learning and fatigue. For both task types, participants performed 8 repetitions for each of the 12 combinations of size and distance. Participants thus performed 8 Varied tasks (8 × 12 = 96 timed targets) and 12 Fixed tasks (12 × 8 = 96 timed targets). Altogether, the experiment gave data from 19 participants × 2 (interfaces) × 2 (task types) × 3 (target sizes) × 4 (target distances) × 8 (repetitions) = 7296 target selections.

3.6   Dependent Variables and Data Collection

As dependent variables we measured accuracy, target selection time, subjective satisfaction, and preference. We also collected data on participants' physical movement in order to describe how participants performed the tasks using the two interfaces.

- *Accuracy*: We calculated the error rate as percentage of targets that were not selected on the first attempt; outside-target selections do not cause the next target to appear, only a correct selection does.
- *Target Selection Time*: We split the elapsed time into a pointing phase (time spent approaching the target) and a selection phase (time spent touching or making a selection gesture on target). For Mid-air, we determined when the cursor had first entered the target; as a proxy for the cursor in the Touch condition, we orthogonally projected participants' index finger onto the display plane.
- *Subjective Satisfaction*: We used 12 questions from the ISO 9241-9 device assessment questionnaire [11] including questions on fatigue. We changed the anchors of questions from "too low" / "too high" to "appropriate" / "inappropriate" as we believe that the original anchors were confusing (i.e., what is too low fatigue?).
- *Physical movement*: We quantified participants' locomotion (from belt position), head turning, and hand movements necessary for selecting targets. Our measures of movement were calculated from our tracking data, which we filtered using the Douglas-Peucker algorithm (1cm tolerance) to compensate for jitter.

3.7   Procedure

We first introduced participants to the experiment and calibrated the system. For calibration, we asked participants to raise their hand in a pointing gesture with their thumb touching the knuckle of their curled middle finger (**Fig. 1**–C), and repeat this gesture a number of times. This was captured to build a template for the selection gesture. Participants then did five practice tasks with each interface. Participants operated both interfaces using their preferred hand; we automatically verified that touch events were produced by the gloved hand. The introduction took around 15 minutes.

Before each task, we asked participants to stand at the starting position 2m away from the center of the display (**Fig. 1**–A); an on-screen indication helped them find the position; when in position, the first target was shown. Participants selected the first target to begin the task. Participants were asked to select targets as quickly as possible, while maintaining high accuracy. Participants could rest after each task. Once they completed all tasks with one interface, they were handed the questionnaire.

After completing all tasks, participants were asked to explain which interface they preferred. The experiment lasted around an hour on average for each participant.

3.8   Hypotheses

We expected that the different control-display ratios of touch and mid-air would result in a speed-accuracy tradeoff. Moreover, mid-air should generally be slower, as input space and output space are decoupled and users therefore must relate movements to the visual feedback. We hypothesized the following:

- Touch is faster and less error-prone than mid-air for subsequent targets close to each other on the display. Compared to mid-air, direct coupling of motor space and display space gives users direct feedback on interaction as it occurs.
- Mid-air is faster for distant targets as users can cover any distance to a target solely through arm/hand/finger movements. For touch, in contrast, distant targets require extensive body movements, which are slower.
- Mid-air is slower for small targets, because the higher control-display ratio makes pointing more difficult; users may need to move closer to point more accurately.
- Touch performs relatively worse for Varied tasks, especially with large distances and small targets, because visual search is harder due to the limited field of view.

3.9   Results

We report results based on the estimation approach [10], that is, as effect sizes with confidence intervals following the latest recommendations from the APA [1]. We report geometric means, as they predict population means of completion times more reliably than other metrics [36], and 95% confidence intervals. Note that geometric means may lead to asymmetric confidence intervals.

Accuracy with Touch and Mid-Air. We observed a high error rate ($M = 25\%$) affecting both Touch ($M = 16\%$) and Mid-air ($M = 34\%$). As shown in **Table 2**, the error rate depends on task type, target distance, and target size. Small targets were particularly difficult to select and produced high error rates with both interfaces (Touch: $M = 39\%$, Mid-air: $M = 59\%$); these error rates are consistent with previous

Table 2. Error rate across target size (rows) and distance (columns) for the two interfaces and task types.

	Mid-air					Touch				
Fixed	768	1536	3072	6144	M	768	1536	3072	6144	M
32	41%	51%	55%	68%	54%	27%	30%	32%	41%	32%
128	16%	26%	34%	42%	29%	1%	2%	2%	7%	3%
512	9%	11%	14%	23%	14%	1%	1%	1%	1%	1%
Varied										20%
32	60%	66%	65%	68%	65%	36%	45%	48%	51%	45%
128	36%	34%	33%	41%	36%	8%	6%	11%	17%	10%
512	12%	15%	20%	27%	19%	2%	7%	4%	9%	6%

studies of touch (29% for 1.26cm targets, 19-57cm distances [33]) and mid-air (56% for 1.6cm targets, 134-402cm distances [40]). However, error rates for larger sizes are higher than expected. We identified 3% of the errors as due to participants making selections far from the target or where a "double-selection" was made within 200ms after a successful selection. When we compare selection times below, we only analyze trials where targets were successfully selected in the first attempt ($N = 5450$).

Target selection time with Touch and Mid-Air. Interface has a clear effect on selection time: Participants spent 40%, CI [29%, 52%] more time selecting targets with Mid-air ($M = 1698$ms, CI [1570ms, 1863ms]) than Touch ($M = 1214$ms, CI [1111ms, 1327ms]). This is in line with our expectation of mid-air being generally slower because of the decoupled input and output spaces.

We see from **Fig. 2** that the mean selection time is higher for Mid-air for both task types, but that the difference is larger for Fixed tasks (1.6 times, CI [1.45, 1.77]), where target placement was predictable, than for Varied tasks (1.22 times, CI [1.1, 1.35]), where targets appeared in random locations. This difference in ratios is likely because searching for randomly appearing targets is easier when using Mid-air at a distance from the display.

One reason for the relatively poor performance of Mid-air is difficulties with the thumb trigger gesture. We occasionally had to recalibrate the trigger gesture during the experiment (mostly due to a shifted glove) in order to ensure correct recognition of selections. For Mid-air, a relatively large amount of time ($M = 26$%, CI [24, 30]) is spent selecting the target after having pointed at the target.
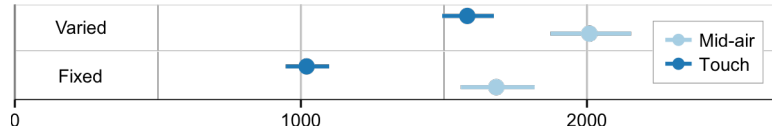

Fig. 2. Average target selection times showing main effects for interface and task type.

Effects of distance and size. We expected the relative performance between Touch and Mid-air to depend on target size and distance. **Fig. 3** shows selection times as the ratio of Mid-air to Touch; a ratio larger than 1 means that Mid-air is slower. The figure shows that the main effect of interface holds for most of the tested conditions: ratios are larger than 1 for 21 out of 24 (task type × size × distance) combinations.

The advantage of Touch diminishes with increasing distances, in particular when targets cannot easily be reached without much body movement (cf. **Fig. 1**–B1/B2). Mid-air even performs better than Touch for the combination of largest targets at the farthest distance, although the ratio is relatively small (0.89 times, CI [0.82, 0.97]).

Contrary to our expectations, there seems to be less variation in task completion times for randomly placed targets (the Varied task) than for reciprocal placements (the Fixed task). For small targets, the results are less reliable due to high error rates for these targets. Still, the overall trend is clear that Touch performs well for selections across short distances, which require little or no locomotion.
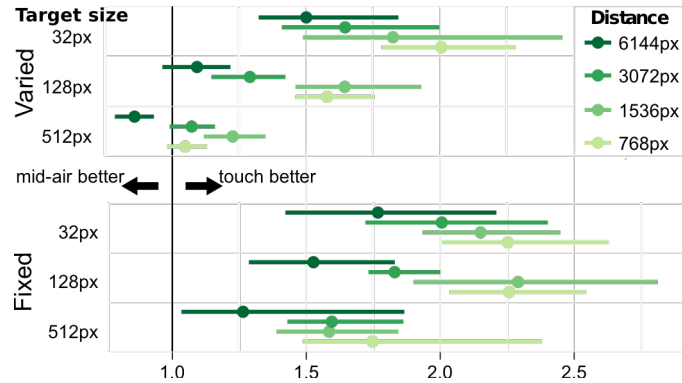
Fig. 3. Selection times for the two interfaces as the ratio of Mid-air to Touch (x-axis) for all combinations of target size (y-axis) and distance (color). A lower ratio means that Mid-air is better. Error bars show the 95% confidence interval for bootstrapped effect sizes [9] .

Physical movement. As expected, more movement was required for Touch than for Mid-air (see **Fig. 4**). For Touch, participants naturally have to move their body to bring their hand within physical reach of the target on display: they moved their hand 237cm on average to reach targets at the largest distance of 224cm. For Mid-air, targets can be selected from a distance with arm and hand movements only.

Movement increased with larger distances for both interfaces (Fig. 4, top), but in particular for Touch ($M = 115$cm vs. $M = 33$cm for Mid-air, at the largest distance) as participants moved their whole body in order to get in a position to better reach the target. On average, participants moved more and approached the display more for small targets (Fig. 4, bottom). Also, participants moved sideways in order to gain a better visual angle, and therefore had to move more if they were closer to the display.

Participants also turned their head much more for Touch ($M = 201°$) than Mid-air ($M = 92°$). The field of view is limited when being close to the display and visual search for targets is likely more time consuming. This impacts only the Varied task, which explains the difference in relative performance between the two interfaces for the two task types.
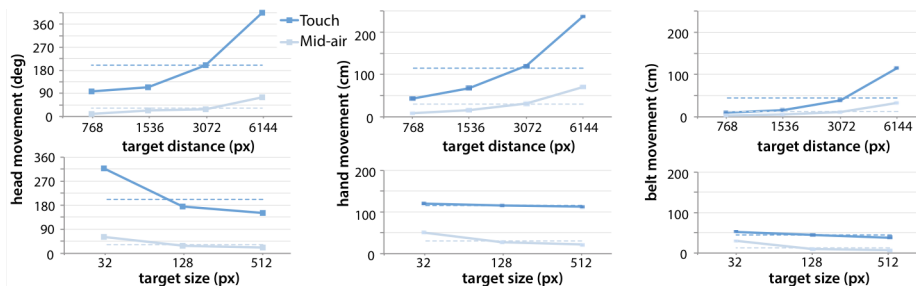

Fig. 4. Mean physical movement for Touch (dark blue) and Mid-air (light blue) across different target distances (top) and sizes (bottom).

Subjective satisfaction and preference. Participants gave Touch more positive scores on 6 out of 12 questions about subjective satisfaction (see **Fig. 5**). Interestingly, participants reported higher wrist and finger fatigue for Mid-air than for Touch, which contradicts movement data. Holding the hand and fingers in a static mid-air pointing gesture seems to be more straining than more dynamic movements for touch input.

Twelve participants preferred Touch, seven preferred Mid-air (not significant by X^2-test). Participants hinted at the reasoning behind their preferences in their comments. Participants explained that Touch was accurate and reliable (9 participants), but that the requirement for moving was taxing (4). Mid-air was thought to give an overview and made it "easy to see targets" (5) and required less walking (4), but accuracy was low particularly for small targets (10).
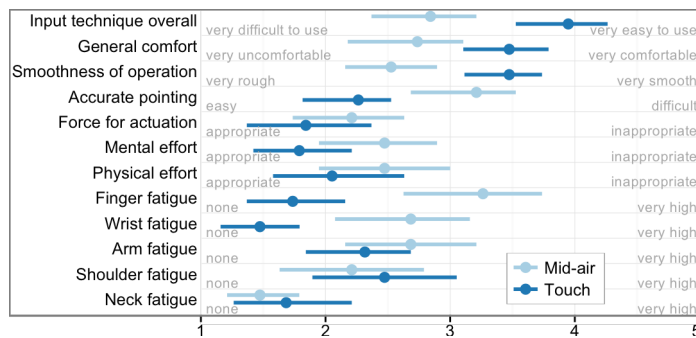

Fig. 5. Bootstrapped confidence intervals for ratings on questions about subjective satisfaction.

4   Experiment 2: Movement Costs and User Choice

Our second experiment investigated which input modality users choose when both mid-air and touch are available. We were interested in how the introduction of conditions under which mid-air input is thought to be beneficial (e.g., backing away to overview a display [30] or use a keyboard [22]) affects performance, preference, and choice of interface: we simulated these conditions by artificially requiring movement.

4.1   Hypotheses

Our first hypothesis was that participants overall would choose touch over mid-air. This hypothesis is based on the results of Experiment 1, which showed that touch performs the best except for one distance/size combination. However, about a third of the participants preferred mid-air input in Experiment 1, which suggests that they might choose mid-air interaction. Due to the cost of switching and incurred performance degradation for some targets, we still expect touch to be chosen overall.

Our second hypothesis for the experiment was exploratory. We investigated how manipulating the cost of location-dependent input (such as touch) changes perfor-

mance and preference. We introduced a backing-up request, requiring participants to move to a particular place in the room. This request abstracts situations where users have to move during or in-between interaction (e.g., to type on a keyboard, write on paper, consult with peers, or get an overview); earlier work has thought that mid-air might be beneficial in such situations (e.g., [22,30]). Our expectation was that—with an increasing cost associated with location-dependent input (i.e., having to walk back and forth between touch and backing-up requests)—mid-air would be chosen more often, perform better, and be more preferred.

4.2   Interface and apparatus

Our interface combines the two input techniques (Touch and Mid-air) used in Experiment 1: The experimental interface allows participants to either touch or, at a distance, point in mid-air. We found this to be simple to understand and thus decided against attempting to integrate the two techniques. When the participant's hand or index fingertip is more than 20cm from the display the ray-pointing cursor is shown. As the participant's finger approaches to touch the display (<15cm distance, using hysteresis tolerance) that cursor disappears. This was done to avoid confusion about the cursor being shown while interacting with the display through touch.

4.3   Tasks

We used only the Fixed task from Experiment 1, for which there was the greatest performance benefit for touch input, in order to reduce the length of the experiment and the risk of tiring participants. We used the same target sizes (3 levels) and target distances (4 levels) as in Experiment 1.

The *backing-up request* required participants to move to a 40cm-wide circular area located 2m away from the display (see **Fig. 1**–A). The request abstracts situations in large-display interaction where users have to move away from the display, for example to gain an overview or to access a keyboard in a particular location. We considered asking participants to type on a keyboard, but since we were only interested in the consequent effects of having to move away from the display, we decided against introducing an arbitrary task. The request was signaled by a message on the display asking participants to move to the location: this message was removed when the participants had stayed in the area for 500ms (as determined by the tracked position of the head). We varied the frequency of backing-up requests as follows: *Absent* (no requests, corresponding to Experiment 1), *Infrequent* (a third of the trials), and *Frequent* (half of the trials). Requests were made after randomly determined trials.

4.4   Participants

We recruited 10 volunteers (5 female), 18–47 years old (M=24), to participate; all were right handed. Participants received an equivalent of €25 as compensation.

4.5 Experimental design

We varied backing-up requests within participants; size and distance were also varied within participants as in Experiment 1. We varied the order of levels of backing-up request across participants using a Latin square. Participants performed 8 repetitions for each combination of size and distance, for a total of 96 timed targets. Altogether, the experiment gave data from 10 participants × 3 (backing-up request frequencies) × 3 (target sizes) × 4 (target distances) × 8 (repetitions) = 2880 target selections.

4.6 Dependent variables and data collection

We collected task time, error rate, data on whether selections were done with touch or mid-air gestures, and participants' preference for either touch or mid-air gestures.

4.7 Procedure

The calibration and instructions were as in Experiment 1. Participants first performed four practice tasks with each input type to familiarize themselves with them; then they performed eight practice tasks (four with infrequent and frequent backing-up requests, respectively) where they could freely choose and switch between touch and mid-air. Participants selected the first target to begin a task. In order to avoid bias against touch input, participants did not have to stand 2m away from the display to start each task, as was required in Experiment 1. After completing a task, they could rest and move freely in order to use either touch or mid-air gestures to begin the next task. The backing-up request required participants to move to the location 2m away from the center of the display (see **Fig. 1**–A), as described above. After standing there for 500ms participants could select the next target. After completing all tasks, participants stated which interface they preferred. The experiment lasted 45 minutes on average.

4.8 Results

Choice of input. Overall, participants completed 978 trials with touch (34%) and 1902 trials with mid-air gestures (66%). Touch was the most often used when backing-up requests were absent, see **Fig. 6** (top row). This supports our first hypothesis.

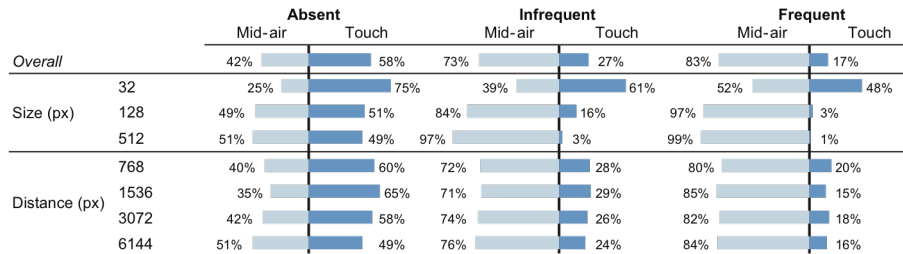|		Absent		Infrequent		Frequent	
		Mid-air	Touch	Mid-air	Touch	Mid-air	Touch
Overall		42%	58%	73%	27%	83%	17%
Size (px)	32	25%	75%	39%	61%	52%	48%
	128	49%	51%	84%	16%	97%	3%
	512	51%	49%	97%	3%	99%	1%
Distance (px)	768	40%	60%	72%	28%	80%	20%
	1536	35%	65%	71%	29%	85%	15%
	3072	42%	58%	74%	26%	82%	18%
	6144	51%	49%	76%	24%	84%	16%

Fig. 6. Frequency of trials made with each input mode for each condition.

However, we had not expected mid-air to be chosen so often (42% of trials). We had hypothesized that the cost of using touch, which was imposed by the backing-up requests, would lead mid-air to be chosen more often. There was a significant association between backing-up requests and input used, $X^2(2) = 393.8$, $p < .001$. Indeed, with more frequent requests, mid-air was chosen more often (83% of trials for frequent requests).

The question then is whether target size and distance had an effect on choice? **Fig. 6** (middle rows) shows how often touch and mid-air were chosen for different target sizes. It seems choice of input depends on target size. For medium-sized and large targets, touch and mid-air were used equally often when requests were absent, whereas almost all selections were done using mid-air when requests were frequent. Participants chose to use mid-air surprisingly often for selecting small targets, which are particularly challenging with mid-air, even in absence of backing-up requests. **Fig. 6** (bottom rows) suggests that distance had less effect on participants' choice of input.

Switching input: Staying or going. Participants switched between using touch and mid-air for 158 (out of 1902) target selections, altogether, of which 67 times were associated with a backing-up request. After backing up, they switched to using mid-air 70% of the times; the remaining times they walked back to use touch (85% of which were for small targets). Considering that 840 backing-up requests were made, participants were prone to stay at a distance from the display. Also, 32 switches to using mid-air were not associated with a request. We saw no instances of alternating between using touch and mid-air for consecutive target selections.

The frequency of switches depends on the frequency of backing-up requests in the task: Participants switched more often (.61 times on average) when requests were infrequent than when they were absent or frequent (.32 and .39 times, respectively), which suggests that participants were more challenged in making the tradeoff between staying and using mid-air or going back to the display in order to use touch.

Task time. We hypothesized that mid-air would perform better with increasing cost of location-dependent touch input: the backing-up requests penalize touch because participants must spend time walking *back* to the display (in the following analysis of target selection times we exclude time spent backing up). Generally, it takes time to switch between the two modes of input, which impacts both mid-air and touch. As can be seen in **Table 3** (rightmost column), the overall mean selection time (which includes selections with errors) is comparable for Mid-air and Touch. However, selection times depend much on whether participants switched from another input mode.

Preference. Seven out of ten participants preferred mid-air gestures (cf. only seven out of 19 in Experiment 1). It seems that with an increasing cost of touch, by way of movement induced by the backing-up requests, mid-air becomes preferable. As benefits of mid-air, seven participants mentioned the lack of a need to move (e.g., "little movement required") and the ease of selecting distant targets (e.g., "much easier to click dots that are far apart"). Five participants liked touch for being precise.

14

Table 3. Mean target selection times after having used the same mode of input as for the previous selection (i.e., not switched) and having switched from another mode of input.

Size (px)	Not switched			Switched			*M*
	32	128	512	32	128	512	
Midair	5341	2323	1362	7914	2714	2581	**2570**
Touch	2786	1138	885	8725	6806	10799	**2565**
M	**3805**	**2055**	**1276**	**8500**	**3981**	**2897**	**2568**
N	870	918	934	90	42	26	2880

5 Discussion

Summary of results. Touch was between 22% and 60% faster than mid-air in Experiment 1. Selection with touch was uncomplicated and had lower error rates. Touch also scored higher than mid-air on several aspects of satisfaction. Touch performance suffered when a target's size and position could not be anticipated; participants turned their head much more, presumably searching for targets. In Experiment 2, touch was as fast as mid-air on average, even with the requirement to do additional movement.

Mid-air was slow and error-prone in Experiment 1. In Experiment 2 mid-air was also slow compared to touch, but users chose it frequently, especially when they were asked to back away, and almost exclusively for selecting medium-sized and large targets (97%-99%). Preferences also shift between the two experiments: 7/10 preferred mid-air with backing-up requests in Experiment 2 versus 7/19 in Experiment 1. Mid-air required less movement and was therefore liked.

Interpretations of Results. The results can be interpreted in several ways. First, the results suggest a place for mid-air interaction. While touch is hard to compete with, mid-air seems to work well and to be chosen by users in situations where earlier work has suggested that it is beneficial (e.g., walking to type on a keyboard [22]). Further, users might choose to manipulate even small targets from a distance when they do not need to inspect them in detail up close: Participants in Experiment 2 chose mid-air for over half of the smallest targets with frequent backing-up requests. This calls for accurate mid-air pointing techniques. These are key implications of the present study.

Second, the results seem to present a new case of performance-preference dissociation; several studies in usability research have shown that people are not necessarily performing best with the interfaces they prefer [15,31]. Mid-air might benefit from the principle of least effort: users prefer not to move, even if small targets are hard to select at a distance. Similarly, a study found users largely preferring virtual navigation over locomotion for a classification task using a gyroscopic mouse, despite possible performance benefits of locomotion [17]. Public display research has also presented subjective feedback that suggests users might minimize physical effort [20].

Third, the viewing angle and distance to the display has played an important role in earlier work when users need to overview [2] or make visual comparisons [5] of data on large displays. Here, we show it is also important for choice of input mode. The

benefit of mid-air, in part, comes from the lower need to visually scan or to move in order to point at far-between targets, when users stand at a distance.

Limitations and Future Work. Several limitations of the study and avenues for future work are clear. First, the high error rates of the study are a concern. Even if comparable to earlier studies (e.g., [40]), investigating interaction techniques that may reduce them is crucial. Many such techniques exist [12,41] that could be adapted and tested for mid-air. Improving the trigger implementation could also reduce error rates.

Second, we studied just one task, pointing, but other tasks also need studying. Users' performance with and choice between touch and mid-air gestures may look quite different for other types of task (steering, manipulation of data, etc.), and for collaborative tasks in particular, which is an important use case for large displays.

Third, we artificially manipulated participants to move to a distant location. A next step for research is to study both realistic tasks and cognitively demanding tasks that benefit from using the display from a distance (e.g., overview of information) and from off-loading cognitive effort into physical movement. Such studies might see users choose differently between touch and mid-air gestures.

6   References

1. American Psychological Association. *The Publication manual of the American psychological association (6th ed.).* Washington, DC, 2010.
2. Ball, R., North, C., and Bowman, D.A. Move to improve: promoting physical navigation to increase user performance with large displays. *Proc. CHI*, ACM (2007), 191–200.
3. Banerjee, A., Burstyn, J., Girouard, A., and Vertegaal, R. Pointable: An In-air Pointing Technique to Manipulate Out-of-reach Targets on Tabletops. *Proc. ITS*, ACM (2011), 11–20.
4. Baudisch, P., Cutrell, E., Robbins, D., et al. Drag-and-pop and drag-and-pick: Techniques for accessing remote screen content on touch-and pen-operated systems. *Proc. INTERACT*, (2003), 57–64.
5. Bezerianos, A. and Isenberg, P. Perception of Visual Variables on Tiled Wall-Sized Displays for Information Visualization Applications. *IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis) 18*, 12 (2012).
6. Boring, S., Baur, D., Butz, A., Gustafson, S., and Baudisch, P. Touch Projector: Mobile Interaction Through Video. *Proc. CHI*, ACM (2010), 2287–2296.
7. Bowman, D.A., McMahan, R.P., and Ragan, E.D. Questioning naturalism in 3D user interfaces. *Commun. ACM 55*, 9 (2012), 78–88.
8. Casiez, G., Roussel, N., Vogel, D. 1 € Filter: A Simple Speed-based Low-pass Filter for Noisy Input in Interactive Systems. *Proc. CHI*, (2012), 2527–2530.

9. Cockburn, A., Quinn, P., Gutwin, C., Ramos, G., and Looser, J. Air pointing: Design and evaluation of spatial target acquisition with and without visual feedback. *International Journal of Human-Computer Studies 69*, 6 (2011), 401–414.

10. Cumming, G. The New Statistics: Why and How. *Psychological science 25*, 1 (2014), 7–29.

11. Douglas, S.A., Kirkpatrick, A.E., and MacKenzie, I.S. Testing Pointing Device Performance and User Assessment with the ISO 9241, Part 9 Standard. *Proc. CHI*, ACM (1999), 215–222.

12. Grossman, T. and Balakrishnan, R. The bubble cursor: enhancing target acquisition by dynamic resizing of the cursor's activation area. *Proc. CHI*, ACM (2005), 281–290.

13. 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. *Proc. UIST*, ACM (2009), 139–148.

14. Hinckley, K. and Wigdor, D. Input Technologies and Techniques. In A. Sears and J.A. Jacko, eds., *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*. CRC Press, 2011.

15. Hornbæk, K. and Law, E.L.-C. Meta-analysis of correlations among usability measures. *Proc. CHI,* ACM Press (2007), 617-626.

16. Jakobsen, M.R. and Hornbæk, K. Up Close and Personal: Collaborative Work on a High-resolution Multitouch Wall Display. *ACM Trans. Comput.-Hum. Interact. 21*, 2 (2014), 11:1–11:34.

17. Jakobsen, M.R. and Hornbæk, K. Is Moving Improving? Some Effects of Locomotion in Wall-Display Interaction. *Proc. CHI*, ACM (2015), 4169-4178.

18. Jansen, Y., Dragicevic, P., and Fekete, J.-D. Tangible remote controllers for wall-size displays. *Proc. CHI*, ACM (2012), 2865–2874.

19. Khan, A., Fitzmaurice, G., Almeida, D., Burtnyk, N., and Kurtenbach, G. A Remote Control Interface for Large Displays. *Proc. UIST*, ACM (2004), 127–136.

20. Kurdyukova, E., Obaid, M., and André, E. Direct, Bodily or Mobile Interaction?: Comparing Interaction Techniques for Personalized Public Displays. *Proc. MUM*, ACM (2012), 44:1–44:9.

21. Mahyar, N., Sarvghad, A., and Tory, M. A closer look at note taking in the co-located collaborative visual analytics process. *IEEE VAST*, (2010), 171–178.

22. Markussen, A., Jakobsen, M.R., and Hornbæk, K. Vulture: A Mid-air Word-gesture Keyboard. *Proc. CHI*, ACM (2014), 1073–1082.

23. Marquardt, N., Ballendat, T., Boring, S., Greenberg, S., and Hinckley, K. Gradual engagement: facilitating information exchange between digital devices as a function of proximity. *Proc. ITS*, ACM (2012), 31–40.

24. 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. In *Proc. INTERACT*, Springer (2011), 461–476.

25. McCallum, D.C. and Irani, P. ARC-Pad: Absolute+Relative Cursor Positioning for Large Displays with a Mobile Touchscreen. *Proc. UIST*, ACM (2009), 153–156.

26. Müller, J., Bailly, G., Bossuyt, T., and Hillgren, N. MirrorTouch: Combining Touch and Mid-air Gestures for Public Displays. *Proc. MobileHCI*, ACM (2014), 319–328.

27. Myers, B.A., Bhatnagar, R., Nichols, J., et al. Interacting at a Distance: Measuring the Performance of Laser Pointers and Other Devices. *Proc. CHI*, ACM (2002), 33–40.

28. Nacenta, M.A., Gutwin, C., Aliakseyeu, D., and Subramanian, S. There and Back Again: Cross-Display Object Movement in Multi-Display Environments. *Human–Computer Interaction 24*, 1-2 (2009), 170–229.

29. Nancel, M., Chapuis, O., Pietriga, E., Yang, X.-D., Irani, P.P., and Beaudouin-Lafon, M. High-precision Pointing on Large Wall Displays Using Small Handheld Devices. *Proc. CHI*, ACM (2013), 831–840.

30. Nancel, M., Wagner, J., Pietriga, E., Chapuis, O., and Mackay, W. Mid-air pan-and-zoom on wall-sized displays. *Proc. CHI*, (2011), 177–186.

31. Nielsen, J. and Levy, J. Measuring usability: preference vs. performance. *Communications of the ACM 37*, 4 (1994), 66–75.

32. Olsen, D.R., Jr. and Nielsen, T. Laser Pointer Interaction. *Proc. CHI*, ACM (2001), 17–22.

33. Pedersen, E.W. and Hornbæk, K. An Experimental Comparison of Touch Interaction on Vertical and Horizontal Surfaces. *Proc. NordiCHI*, ACM (2012), 370–379.

34. Pierce, J.S., Forsberg, A.S., Conway, M.J., Hong, S., Zeleznik, R.C., and Mine, M.R. Image Plane Interaction Techniques in 3D Immersive Environments. *Proc. I3D*, ACM (1997), 39–ff.

35. Sasangohar, F., MacKenzie, I.S., and Scott, S.D. Evaluation of Mouse and Touch Input for a Tabletop Display Using Fitts' Reciprocal Tapping Task. *Proc. HFES 53*, 12 (2009), 839–843.

36. Sauro, J. and Lewis, J.R. Average task times in usability tests: what to report? *Proc. CHI*, ACM (2010), 2347–2350.

37. Schick, A., van de Camp, F., Ijsselmuiden, J., and Stiefelhagen, R. Extending Touch: Towards Interaction with Large-scale Surfaces. *Proc. ITS*, ACM (2009), 117–124.

38. Swaminathan, K. and Sato, S. Interaction design for large displays. *interactions 4*, 1 (1997), 15–24.

39. Vogel, D. and Balakrishnan, R. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. *UIST '04: Proc. UIST*, ACM (2004), 137–146.

40. Vogel, D. and Balakrishnan, R. Distant Freehand Pointing and Clicking on Very Large, High Resolution Displays. *Proc. UIST*, ACM (2005), 33–42.

41. Vogel, D. and Baudisch, P. Shift: a technique for operating pen-based interfaces using touch. *Proc. CHI*, (2007), 657–666.

42. Wilson, A.D. Robust Computer Vision-based Detection of Pinching for One and Two-handed Gesture Input. *Proc. UIST*, ACM (2006), 255–258.

43. Wilson, A. and Shafer, S. XWand: UI for Intelligent Spaces. *Proc CHI*, ACM (2003), 545–552.