Pinch-Drag-Flick vs. Spatial Input: Rethinking Zoom & Pan on Mobile Displays

Martin Spindler
User Interface & Software Engineering Group
University of Magdeburg, Germany
martin.spindler@acm.org

Martin Schuessler
Quality and Usability Lab
Telekom Innovation Laboratories
TU Berlin, Germany
schuesslerm@acm.org

Marcel Martsch
Dep. of Vocational Education & Human Resources Development
University of Magdeburg, Germany
marcel.martsch@ovgu.de

Raimund Dachselt
Interactive Media Lab
Technische Universität Dresden
01062 Dresden, Germany
dachselt@acm.org

ABSTRACT

The multi-touch-based pinch to zoom, drag and flick to pan metaphor has gained wide popularity on mobile displays, where it is the paradigm of choice for navigating 2D documents. But is finger-based navigation really the gold standard? In this paper, we present a comprehensive user study with 40 participants, in which we systematically compare the Pinch-Drag-Flick approach with a technique that relies on spatial manipulation, such as lifting a display up/down to zoom. While we solely considered known techniques, we put considerable effort in implementing both input strategies on popular consumer hardware (iPhone, iPad). Our results show that spatial manipulation can significantly outperform traditional Pinch-Drag-Flick. Given the carefully optimized prototypes, we are confident to have found strong arguments that future generations of mobile devices could rely much more on spatial interaction principles.

Author Keywords

Spatial Input; Spatially Aware Displays; Mobile Displays; Multi-Touch Input; User Study

ACM Classification Keywords

H.5.2. Information interfaces and presentation: User Interfaces – *input devices and strategies*.

INTRODUCTION

The exploration of large 2D information spaces, such as maps, pictures and web documents, is a very common task carried out on mobile displays by millions of users every day. Due to the rather small screen size of the devices, this often involves heavy usage of zoom and pan, usually performed using multi-finger gestures. In this context, the Pinch-Drag-Flick paradigm has proven to be one of the most (commercially) successful gesture sets: pinch to zoom, drag and flick to pan. While these gestures are considered to be easy to learn and perform, there are inherent problems with the approach: fingers occlude virtual items on the screen [29]; virtual travel distances per gesture are short [16]; pinch gestures are difficult to execute if one hand is occupied; and elderly or disabled persons may not possess sufficiently fine motor skills to perform gestures accurately. Besides that, ambiguities of gestures may trigger unintended actions, such as the accidental selection of items, or may require users to explicitly switch between edit and navigation mode, which in turn may induce disorientation [17]. These shortcomings motivated the development of navigation techniques that, for example, employ different finger gestures [2] or extend the interaction to the side [23] or the back of devices [29]. Although such approaches soften some of the issues of Pinch-Drag-Flick, the underlying input strategy still remains surface-constrained and thus continues to rely mostly on fine finger motor skills.

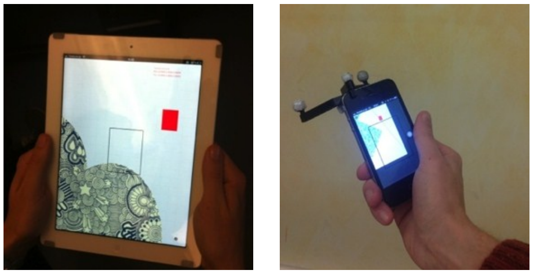

(a) iPad (b) iPhone

Figure 1: We augmented an iPad and iPhone with IR reflective markers for the support of spatial tracking.

In this paper, we study a radically different approach that is based on an alternative input channel: the spatial position and orientation of mobile displays (see Figure 1). Well-known examples of such spatially aware displays are the Chameleon [7] and the Peephole Displays [30]. In contrast to the metaphor of grabbing a document, they build upon the concept of Magic Lenses [1] and thus use the metaphor of moving a viewport (the display) over a virtual information world. For this purpose, distinct motion patterns are mapped to specific navigation tasks, e.g., horizontal movements may change the viewport center (panning), whereas lifting a display up/down may control the zoom factor. As this requires users to move a display through the physical space surrounding them, the motor space is increased considerably (large 3D volume vs. small 2D screen) and a different set of motor skills is addressed (arms vs. fingers).

We see this difference in motor control as a significant opportunity that may help overcome the problems of con-

CHI 2014, April 26 - May 01 2014, Toronto, ON, Canada
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-2473-1/14/04…$15.00.
http://dx.doi.org/10.1145/2556288.2557028

ventional touch-based navigation – not as a superior form of interaction, but as a complimentary one. This implies many advantages including the addition of a more natural way of interaction by addressing principles of spatial manipulation, the support of longer travel distances per gesture, a reduction of item occlusions on the screen, and less mode switches, e.g., by assigning spatial input to navigation and touch input to selection. For these reasons, we consider spatially aware displays an important research topic that has the potential of changing the way we interact with mobile displays – with 2D document navigation being only one possible use case.

Surprisingly little practical work has been done on systematically studying how both navigation approaches perform against each other on mobile displays. Previous attempts either addressed different setups, e.g., involving a wall [15, 20], occupied both hands for spatial input [11], or did not succeed in finding hard evidence in favor of the spatial approach [10, 19] – a gap that we fill with our work. In this paper, we contribute a comprehensive user study with 40 participants that we conducted using state-of-the-art mobile displays (iPhone 4, iPad 3). We found overwhelming proof that spatial input-based navigation does – if designed and implemented properly – outperform Pinch-Drag-Flick for 2D document navigation. We believe that this is due to our design decisions, e.g., regarding the importance of an easy to use clutch and the role of a high quality prototype.

The remainder of this paper is organized as follows. First, we review related work and discuss key design decisions. We then outline the scope of the study and present the method, results, and discussion. This is followed by design recommendations for future generations of mobile displays as well as conclusions and an outlook on future work.

RELATED WORK

Pinch-Drag-Flick-is a well understood and established technique. We will therefore restrict our review of related work to spatial input-based interaction and its evaluation.

One of the first spatially aware mobile displays is the Chameleon presented by Fitzmaurice [7]. Inspired by the notion of see-through interfaces [1], it serves as a “peephole in hand” providing access to a virtual world that can be explored by moving the device around. This concept was later adapted to arm-mounted displays, e.g., the Boom Chameleon [27], or to a tabletop environment, e.g., PaperLens [25]. These systems add further aspects to the overall interaction equation, e.g., the opportunities of multi-display environments or additional input modalities, such as digital pens. While we studied the specifics of mobile displays, we believe that our findings are transferable to such setups.

Evaluating specific spatially aware display systems has been the goal of some research projects. Oh and Hua tested various aspect ratios and sizes of spatially aware peepholes [18]. They concluded that the aspect ratio of a display plays a more important role for smaller screens than for larger ones and that screen sizes are more dominant in impacting the user performance. Spindler et al. [24] tested the specifics of multi-layer interaction above a horizontal reference surface. Two recent projects compared touch- and spatial-based 2D document navigation on handheld devices that share some similarities with our work, yet are based on completely different setups: Kaufmann & Ahlström [15] projected the workspace onto a wall with a Pico projector and Rädle et al. [20] combined a tablet with a wall-sized display. Both projects found advantages of the spatial techniques – particularly in terms of recall performances, which was not our focus. Jones et al. [11] investigated free hand around device input for 2D navigation on a handheld display. They are one of the first to show that spatial input can be as good as touch. In some respects, our spatial technique is a simplified variant of their “1 button simultaneous” condition, as our clutch works differently and our technique does not require a second hand for pointing in mid-air (we use relative device positions instead). These are two likely reasons for why fatigue (guerilla arm effect) was only a negligible problem in our study, as opposed to Jones et al. who let their participants rest every 3 to 5min.

To our knowledge, there is only one previous work that bears considerable resemblance to our work: the Lens Chameleon [19] by Pahud et al., who were driven by similar motivations. They also conducted a series of experiments to compare spatial-based navigation with standard Pinch-Drag-Flick. In contrast to our work, their implementation of spatial-based navigation was significantly slower than Pinch-Drag-Flick. This may be attributed to our design decisions: no use of clutching and a lack of state-of-the-art technology (e.g., cable-bounded device).

DESIGN RATIONALE

While the interaction design specifics for Pinch-Drag-Flick are straightforward (it is the default on most devices), spatial-based navigation is a more complex case. We will now discuss a few design decisions that we made in the process of building the prototype for the spatial technique.

Mapping the Physical to the Virtual World

One key question is how to properly map the physical space to the Space-Scale-Diagram [8]. In pre-tests, we tried various mappings and finally decided on a dynamic mapping that uses the current orientation of the display as the new reference plane for future interpretations of motions. This means that zooming is mapped to movements along the normal of the display (local Z-axis), whereas motions within the display’s XY-plane define panning. Our experiences show that the dynamic mapping supports body-centric usage even better than a spherical mapping [7], which was also recently confirmed in [19]. In addition, it has the benefit of working independently of the user’s position, thus simplifying the interaction design and spatial tracking.

Clutching and Relative Mode

As opposed to [19], where clutching was considered to be of minor relevance (it had performed slightly slower in a pre-test), we argue that mobile devices are moved most of the time without any intention to interact. We therefore think that spatial input should be inactive by default, only to be *enabled on purpose* for a brief moment of interaction – by activating a clutch. With a clutch, the nature of spatial navigation can be changed from absolute to relative mode. We believe that this is a very important and necessary step to support mobile usage. In relative mode, the "volumetric" 2D document (represented by a pyramid Space-Scale-Diagram) travels along with the device like a bubble surrounding it. This enables users to put away the phone, e.g., into the pocket, and to resume navigation later on with the last visited position.

While we fiercely advocate the use of *tactile* clutches (see the discussion in section "A Built-in Tactile Clutch"), we decided on using a touch-based clutch for the user study. This choice was primarily motivated by practical reasons: Existing volume buttons are known to be unsuitable for this purpose [11] and it proved to be challenging to build an adequate alternative in the given time. In our prototype, users can activate the clutch by touching the screen with one or more fingers, e.g., close to the screen bezel in order to prevent occlusion of items in focus. Likewise, removing the finger(s) deactivates the clutch. We believe to have found a close enough approximation of tactile clutches, as this enables users to quickly access the clutch without spending much mental effort on locating it. Hence, we expect that our findings will also apply to the latter case.

Zoom Direction

At first sight, the choice of the proper zoom direction may appear trivial: "If you observe how users react when they can't see something, they always bring the device closer to their eyes" as one of our reviewers wrote. Having implemented both variants, we conducted an informal pre-test with 5 users. All of them preferred the opposite zoom direction, i.e., zoom out when the display gets closer to the user. A look into the literature [19, 30] confirmed this finding. Apparently, the inverted zoom direction would not match the peephole-in-hand metaphor that most users are familiar with (and thus expect), e.g., from using a magnifying glass or looking through a camera. Hence, we decided to conduct the study with the "zoom-out-when-getting-closer" option (the participants could not change this setting).

Zoom Mapping

We tested several mappings and chose one that multiplies each movement along the device normal (Z-axis) by the current zoom factor, thus exponentially speeding up the zoom if the zoom level is getting larger, and vice versa.

Zoom Center

In an early version of the prototype users could dynamically reposition the zoom center via the touch point on the screen (touch-based clutch). We later turned this option down and decided to leave the zoom center in the middle of the screen (out of the user's control) for the following three reasons: First, touching the screen violates the idea of a tactile clutch. Second, it added further complexity as users were forced to keep holding the finger at a particular position on the screen. Third, it caused occlusion of items in focus.

Pan Boundaries

Special care must be taken for handling document boundaries, e.g., to prevent users from traveling into the void [30]. For Drag-Flick-based panning, this may be accomplished by stopping pan motions at the document boundaries (often accompanied by some sense of physicality, e.g., bouncing). For spatial input, we adjusted the boundaries to guarantee that users can align even the document corners to the zoom center (the middle of the screen) as illustrated in Figure 2c.

GOALS AND SCOPE OF THE STUDY

For the reasons discussed above, the main focus of this work is on designing, conducting and evaluating a user study that systematically investigates and compares the two navigation techniques (touch, spatial) on state-of-the-art mobile displays. In particular, we pursued two major goals:

G1 *Efficiency:* We aimed at comparing how fast users perform common navigation tasks with the techniques.

G2 *User satisfaction:* We aimed at investigating how users relate to different usability aspects of the techniques.

Factors of Influence

We considered the following major factors in our study:

Navigation Technique – Our primary attention was, of course, on the two navigation techniques (*touch*, *spatial*).

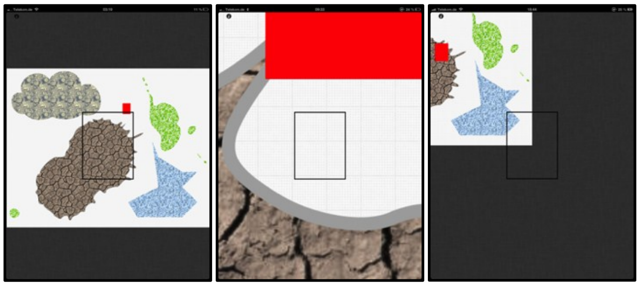
Target Visibility – Zoom and pan are used in diverse application contexts with a variety of intentions that have an influence on the visibility of target items. In our study, we considered *on-screen targets*, i.e., items that are (partially) visible on the display and *off-screen targets,* i.e., items that are not initially visible on the display, e.g., a distant node in a node-link diagram.

Display Size – The screen size is another key factor that can influence the cognitive performance of users [21] and thus the time required for completing navigation tasks [4, 12]. We focused on the two predominant classes of mobile display: *phones* and *tablets*. These do not only differ in screen size, but also in weight, device size, pixel resolution and density that may also affect the navigation performance.

Gender – Previous studies [5, 6] show that women and men differ in their cognitive strategies when performing navigation tasks. In order to compensate for such effects, it is vital to properly incorporate the gender into the study design.

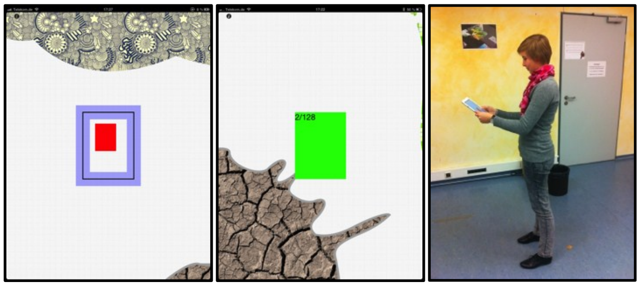
Scope of the Study

Sufficient ecological validity was very important to us. While we believe to have taken into account the most essential factors, there are further variables that may additionally affect the navigation performance. Examples for this are: sitting/standing/walking usage, public/private usage (e.g., shyness and social habits), touch vs. haptic clutches, the mapping from physical to virtual space, and the use of



(a) Entire scene (b) Maximum zoom (c) Pan boundaries

Figure 2: The abstract scenario used in the user study (screenshots taken from iPad prototype).



(a) Tolerance zone (b) Task succeeded (c) During the study

Figure 3: Screenshots of the prototype (a, b) as a participant is performing the navigation tasks (c).

more complex scenarios that involve frequent mode switches (e.g., navigation vs. annotation mode). For practical reasons, we limited our investigations to a user standing in the middle of a free space in an office-like lab (private environment) performing simple navigation-only tasks on a mobile display with a touch-based clutch. We decided to focus on technologically affine users of both genders with advanced multi-touch experiences. This decision was motivated twofold. First, we wanted our baseline (touch) to score very well. Second, we expect that sooner or later the majority of people will acquire similar skills as mobile displays become more widespread.

METHOD

Participants

Forty unpaid students of different departments at the local university participated in the study. Genders were evenly distributed (20 male, 20 female). The average age was 23 years (M = 23.48, SD = 2.27) ranging from 19 to 33 years. All participants (normal eyesight, no colorblindness) were daily users of smart phones or tablets and thus considered themselves as experienced with such devices. This implies that they were confident in performing Pinch-Drag-Flick-based navigation, which we verified in a pre-test.

Study Design

We designed a controlled lab experiment with four independent variables. Our main focus was on the *navigation technique* (touch, spatial), which was the primary independent variable. *Display size* (phone, tablet), *target visibility* (on-screen, off-screen), and *gender* (female, male) were the secondary independent variables. We conducted the user study as a mixed-model design. For *display size* and *gender*, we used an in-between subjects design, i.e., participants were either assigned to work with a phone or a tablet (by balancing out the gender). For *navigation technique* and *target visibility*, we chose a within-subjects design (repeated measures), i.e., each participant performed both techniques exactly once (counter-balanced) using the same task sequence. We used the same task sequence for all users.

Apparatus & Implementation

The experiment was conducted in a quiet lab environment (see Figure 3c). We used popular devices that users were familiar with: iPhone 4 (640×960 pixels @ 51×74mm^2) and iPad 3 (1536×2048 pixels @ 148×198mm^2). This ensured the touch technique to be a strong baseline condition, as these devices provide a high standard of Pinch-Drag-Flick navigation out of the box.

Spatial Tracking: We opted for an optical tracking based on 12 infrared (IR) cameras (Optitrack FLEX:V100R2). This provided precise spatial device positions and orientations at 100Hz with an error of less than 1mm within the tracking volume. Its projected area on the floor was about 3×3m^2. A designated server (running Tracking Tools 2.2) streamed the spatial raw data over a local Wi-Fi network in a standardized form (VRPN). This included time stamps, device IDs, and 4×4 transformation matrices describing the spatial position and orientation in six degrees of freedom (6DoF).

Marker Design: We glued 6 IR-reflective stickers to the iPad's display bezel (see Figure 1a). Only 3 of them needed to be visible at a given time, enabling users to hold it freely in their hands without accidentally interrupting the tracking. Due to the smaller device size, this was not practicable on the iPhone. Here, we built a small, lightweight plastic frame (± 8cm×8cm×3cm^3, 30g) with 4 IR-reflective balls that we plugged into the iPhone's headphone output (see Figure 1b). This guaranteed a robust tracking and ensured enough flexibility for operating the phone with one or both hands.

App Development: We implemented the prototype in Objective C using iOS 6.0 (XCode 4.5). We integrated the touch and spatial technique within a single universal app that runs on both devices. A major problem was the limited RAM of the devices – a problem others [2] faced, too. We solved this by combining several strategies: A zoom pyramid consisting of three layers containing different resolution of the scene, with the most detailed one being built up of tiles that are loaded on demand.

Scenario and Task Design

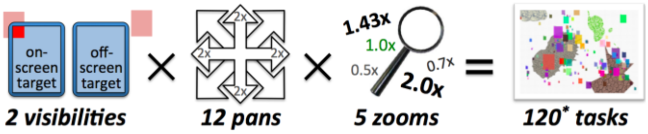
In order to minimize side effects caused by prior knowledge of data, we used an abstract 2D scene for both conditions (touch, spatial). The scene provided visual context to avoid disorientation (Desert Fog) [14]. For this purpose, the scene background featured a thin grid and several distinctively colored and textured shapes (see Figure 2a). The scene had a fixed resolution of 4734 × 3683 pixels (approx. 46×36cm^2

in physical space). We used a maximum zoom factor of 5 (see Figure 2b), which translates to approx. 230×178cm^2.

In the scene, participants completed a pre-defined sequence of 128 navigation tasks using one of the two devices held in portrait mode. They had to match a red rectangular search target with a black reference frame in the middle of the screen (4×5cm^2, see Figure 2). Only one target was visible at a time. If a search target came close to the reference frame, it automatically snapped in and the task was done (Figure 3a illustrates the snapping tolerance). Then, the rectangle turned green and a progress bar was shown (see Figure 3b). After 2 seconds, the next target appeared in the scene. The use of red search targets above a green-bluish background was motivated by the feature integration theory [26], as it reduces the cognitive load (pop-out effect).

Composition of the Task Sequence

To test the user performances depending on the navigation intent (pan/zoom/combined × target visibility), we designed a single sequence of 128 navigation tasks. We wanted this sequence to contain a well-balanced combination of pure and mixed pan/zoom tasks with both on- and off-screen targets. To achieve this, we defined a set of basic composition rules (see Figure 4). We chose *5 zoom factors* (0.5x 0.7x, 1.43x, 2.0x, and no zoom) that we joined with *12 pan directions* (2 × each of the 4 major display sides + 1 × each of the 4 diagonals). To address *target visibility*, we placed the navigation targets either within (on-screen) or outside the display (off-screen). At the beginning of a task, on-screen targets appeared fully visible on the tablet, yet only partially on the phone (due to its smaller screen estate). We used these rules in a script that produced a sequence of 120 navigation tasks (2 *target visibilities* × 5 *zoom factors* × 12 *pan direction)* in a random order. The script also created small randomized local positional offsets. As all tasks of the sequence involved panning, we added 8 extra zoom-only tasks, i.e., 2 × (0.5x, 0.7x, 1.43x, 2.0x). Hence, we obtained a total number of **128 navigation tasks**.



* There were 8 additional zoom-only tasks, i.e., user performed **128 tasks** in total.

Figure 4: Basic composition rules for the task sequence

Procedure

Participants completed the study within 50 to 70 minutes. Before conducting the experiment, we had collected basic demographic information about potential participants via an online form. This included the personal experience with touch screens to sort out applicants with insufficient multi-touch skills. We grouped suitable candidates so that exactly half of the women and men worked with an iPhone or iPad.

(1) Introduction Part

After briefly verifying the personal data of the participant, we explained the goals and procedure of the study by reading out aloud from a sheet of paper. This also included the specific request to complete the tasks as quickly as possible.

(2) Main Part

The main part of the study consisted of three phases that were executed in two runs, once for the touch condition and once for the spatial condition (in counter-balanced order).

Trial Phase: Depending on the group, the participant either started with the touch or spatial technique that we explained and demonstrated using the iPhone or iPad prototype. This also included a brief explanation of the underlying interaction metaphor, e.g., the possibility of clutching (in addition to the standard press-and-hold activation). We then invited the participant to perform a few exercise trials using an example dataset, until he or she felt confident with the technique. In most cases, this took no longer than 5min, even for the spatial condition.

Interaction Phase: In both conditions, participants were asked to walk to the center of the interaction space, marked with a cross on the floor. We enforced a standing usage. Participants were free to move within an area of 2×2m^2.

Assessment Phase: After completing the tasks, we asked the participant to sit down and to fill out a questionnaire. We then conducted a brief interview, where we encouraged the participant to provide additional feedback in form of free comments. Before commencing with the second condition, participants were allowed to remain seated and to rest as long as they wanted (this did not take longer than 7min).

RESULTS

In this section, we present the results of the study, including a detailed analysis of performance and self-report data.

Statistical Methodology

We analysed the data using *three-way repeated measurement ANOVAs*. For all ANOVAs, the in-between factors were *display size* (phone, tablet) and *gender* (male, female).

The repeated-measures factor depended on the analyzed data type. We either used the *navigation technique* (spatial, touch), the *target visibility* (on-screen, off-screen), or their *combination* (spatial on-screen, spatial off-screen, touch on-screen, touch off-screen). All p-values were Greenhouse-Geisser corrected. The alpha level for tests of statistical significance was set to $\alpha = .05$. When effects were significant, we reported Bonferroni adjusted p-values for post hoc comparisons (t-test, two-tailed). For descriptive data, we provided mean values (M) and standard deviations (SD).

Collected Performance Data

For spatial input, the collected raw performance data consists of the spatial position and orientation (6DOF) of the devices over time with a sampling rate of 30 Hz as well as the start/end time of each clutch-cycle. For touch input, we logged relevant events provided by the iOS-framework, e.g., the gesture type, on-screen positions, and start/end times. We also recorded the start and end time of each task.

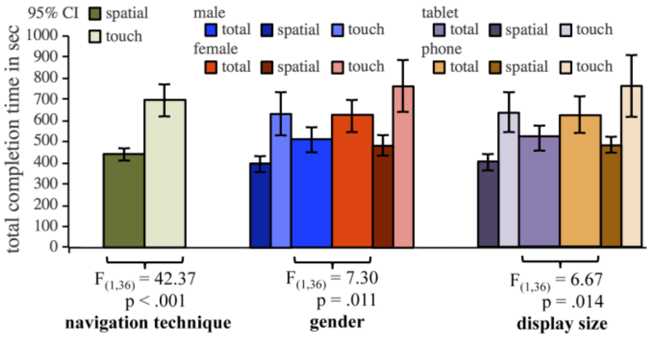


Figure 5: Total completion times broken down by navigation technique, gender, and display size. Error bars denote standard deviations (95% CI).

Completion Times

We used the times that participants spent on completing the tasks as a *measure of performance*. We considered two variants: The *total completion time* is the overall time that participants needed to finish all 128 navigation tasks with either the touch or the spatial condition. In contrast, the *task completion time* is the average time that participants spent on finding on-screen and off-screen targets, respectively. All times are in seconds.

Total Completion Times

The key results concerning total completion times are summarized in Figure 5. We found a main effect of the *navigation technique*, i.e., participants completed the tasks significantly faster with the spatial condition (M = 445.55, SD = 94.96) than with the touch condition (M = 701.15, SD = 259.38). There was also a main effect of the *display size*, i.e., participants worked significantly faster with the tablet (M = 522.46, SD = 121.06) than with the phone (M = 626.93, SD = 168.22), independent of the navigation technique. Beyond that, we found a main effect of *gender*, i.e., males (M = 517.28, SD = 128.90) were significantly faster than female participants (M = 626.62, SD = 158.72).

Task Completion Times

Our analysis confirmed a main effect of *target visibility* (see Figure 6, left). Post-hoc comparisons show that participants reached on-screen targets significantly faster (see Figure 6, middle) with the spatial (M = 2.04, SD = .42) than with the touch condition (M = 3.38, SD = 1.36). The same applies to off-screen targets (see Figure 6, right), where a direct comparison shows that the spatial technique (M = 5.25, SD = 1.25) is faster than touch (M = 8.16, SD = 3.09).

The analysis revealed an interaction effect between *target visibility* and *display size* ($F_{(3,109)}$ = 4.21, p = .038). While the display size had only little influence on the completion time for on-screen targets (see Figure 6, middle), participants reached off-screen targets significantly faster with the tablet than with the phone (see Figure 6, right). The relative performance gain on the tablet was only marginally higher for touch (22.67%) than for spatial input (18.77%).

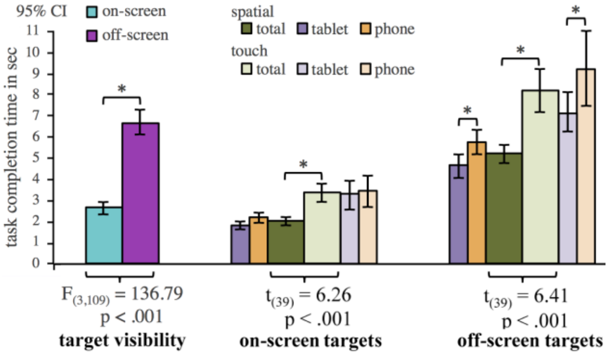


Figure 6: Average task completion times (gender-neutral). Error bars denote standard deviations (95% CI).

We found a main effect of *gender* ($F_{(1,36)}$ = 7.45, p = .01) and *display size* ($F_{(1,36)}$ = 7.27, p = .011). This means that independent of the *target visibility*, males were faster than females and tasks were completed in less time on the tablet than on the phone.

Number of Discrete Actions

We investigated how many discrete actions were initiated for on-screen and off-screen target tasks. We did this by counting the number of clutches (spatial condition) and individual touch gestures (touch condition). We used these numbers as a *measure of handicap*. The rationale behind this is that starting a new action interrupts the navigation and thus negatively affects the overall performance. For example, executing three drag gestures in a row requires the user to lift the finger from the screen two times more if compared to just performing a single continuous drag. The same applies to the spatial condition, where releasing the clutch, e.g., to move the display to a more conformable position, briefly pauses the actual navigation.

Number of Clutches (Spatial Condition)

We found a main effect of *target visibility* (Figure 8, left). Significantly fewer couplings were performed for on-screen (M = 1.18, SD = .36) than for off-screen targets (M = 2.26, SD = 1.41). There was no effect of *display size* or *gender*.

Number of Touch Gestures (Touch Condition)

There was a main effect of *target visibility* (see Figure 8, right), i.e., on-screen targets were reached with significantly less touch gestures (M = 3.93, SD = 1.09) than off-screen targets (M = 10.2, SD = 3.89). We also found an interaction effect between *target visibility* and *display size* ($F_{(1,36)}$ = 17.00, p < .001). For on-screen targets, the number of touch gestures did not vary much between devices. For off-screen targets, in contrast, participants spent four gestures more on the phone (M = 12.01, SD = 4.15) than on the tablet (M = 8.47, SD = 2.74) on average.

Utilized Motor Space (Spatial Condition Only)

The motor space of the spatial technique, i.e., the physical room surrounding the user, is considerably larger than a touch screen. We were interested in how much of the motor space was actually utilized by participants during the spatial condition and whether there were differences between on-

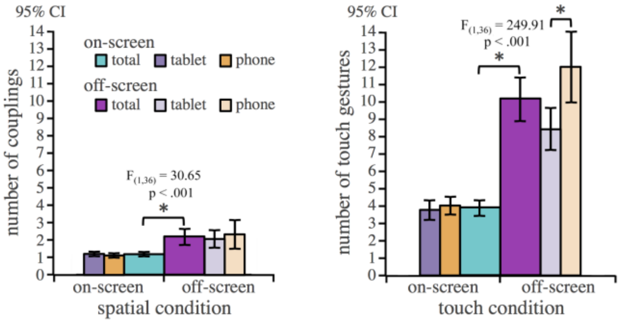


Figure 8: Number of discrete actions (gender neutral). Error bars denote standard deviations (95% CI).

screen and off-screen target tasks. For this purpose, we extracted the 3D bounding box of the physical space that participants used while solving the tasks (i.e., with activated clutch). This was done for each of the 120 tasks (i.e., except zoom-only tasks). We then computed an average bounding box for on-screen and off-screen targets, respectively. We analysed both bounding boxes in terms of the maximum extent along each of the three principle axes (X, Y, Z). We can show that the amount of used motor space significantly depends on the target visibility (see Table 1).

Axis	On-screen targets		Off-screen targets		$F_{(1,36)}$	p
	M in mm	SD	M in mm	SD		
X	57.78	43.81	237.17	102.70	116.66	< .001
Y	49.95	16.06	176.61	65.17	121.61	< .001
Z	78.00	35.75	298.69	69.15	290.41	< .001

Table 1: Size of the utilized motor space for the spatial condition. Mean values (M) and standard deviation (SD) are provided for the physical extent along the main axes.

User Feedback & Observations

Usability Ratings: We compiled a questionnaire with 36 items using a 7-point Likert scale ranging from 1 ("strongly disagree") to 7 ("strongly agree"). These items addressed generic usability aspects [22] as well as specific issues regarding the tested techniques, in particular the perceived influence of zooming and horizontal/vertical panning on the overall performance. To ensure a high degree of validity, we used 3 to 6 items per usability issue (reverse-worded). Both techniques were generally assessed very positively without significant differences, except for *ease of use*, *efficiency to use*, *user experience*, and *zooming* that were rated in favor of the spatial condition (see Figure 7).

Free Comments: Participants gave us very positive feedback about the spatial technique. Some (N = 5) were even a little surprised that completing the 128 tasks with the touch technique *"felt somehow more difficult than with the other one [spatial]"*. One user said that she *"could almost 'see' the map behind and beside the iPad, making it easier to decide where to move the device to next"*.

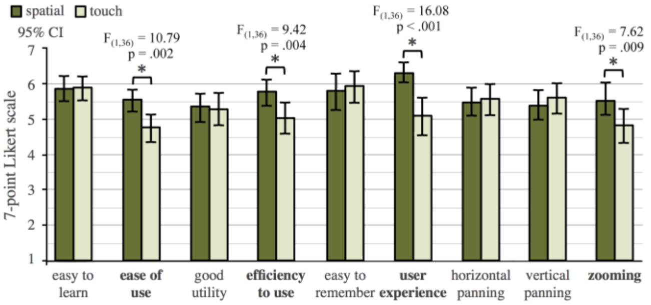


Figure 7: Usability ratings from the questionnaires. Error bars denote standard deviations (95% CI).

Fatigue: All participants completed the tasks without a break. After each condition, participants were asked to choose from a list, which part(s) of their body felt tired. Table 2 summarizes the results that reflect the frequency of fatigue among participants, but not its intensity. Both techniques caused fatigue in the shoulders and the upper arms. The touch technique was more demanding for the fingers, the neck and the forearm, whereas spatial input affected the back and the upper arm more frequently. To our surprise, 10 participants did not experience fatigue for spatial input, yet there were only 4 for the touch condition.

	Spatial	Touch
None	10	4
Shoulders	10	10
Back	6	4
Forearms	3	9
Upper arms	7	6
Fingers	4	12
Wrists	2	9
Neck	5	13
Elbows	1	3

Table 2: Occurrences of fatigue as reported by users

DISCUSSION

Not a single participant completed the 128 navigation tasks faster with the touch than with the spatial condition, even though all of them were multi-touch-experienced and used the spatial technique for the first time. When performing the tasks, several (N = 7) participants already expressed that they liked the spatial technique, even though we had not asked for that yet. While we had predicted a decent performance for the spatial approach, it *outperformed Pinch-Drag-Flick by 36.5%* (M ≈ 445s vs. M ≈ 701s) on overall average – a finding that we had not expected in such clarity.

Influence of Display Size

Independent of the technique, participants benefited most from a larger display for off-screen targets (see Figure 6, right). There are two reasons for that. First, search tasks typically start by zooming out for overview. This phase is usually shorter on devices with a larger screen, because search targets appear earlier on the display. Second, users see more information on a larger display allowing them to come to navigation decisions earlier, e.g., see [9]. We observed that most participants used the phone with one hand, while the tablet was usually taken into both hands.

Influence of Gender

On average, males completed the tasks 17% faster than females, independent of the navigation technique, target visibility, and display size. Hence, the spatial technique is not suitable for compensating gender effects.

Use of Clutching for Repositioning the Device

By design, all participants had to use the clutch in order to activate spatial input. Hence, the minimum number of clutches per task was one. Apart from that, clutching could also be used to move the device to a more convenient position, e.g., closer to the body. In the data, a clutch number larger than one reflects this. For on-screen targets, participants rarely used the clutch for this purpose (M = 1.18). For off-screen targets, however, where the size of the motor space was utilized much more, we counted one or two extra clutches for most participants (M = 2.26). Only the minority of participants (N = 8) did not make use of the technique. Instead, they stretched out their arms farther or did an extra footstep forward. When asked why, common reasons included convenience issues, avoidance of slowing down, or being oblivious of the possibility.

Learning Speed & Other Groups of the Population

All participants learnt how to use the spatial technique very quickly. The majority (N = 31) needed less than 5 minutes for that. Supported by the steep learning curve of the young, healthy and technological-affine users, we expect our findings can be transferred to other groups of the population. One example for this are elderly people or persons having difficulties in precisely controlling their fingers, e.g., due to age-related motor impairment, Gout, or Osteoarthritis. We believe that these groups may particularly benefit from the different motor skills that are relevant for spatial input.

Explaining the Effects

Based on our observations and data analysis, we identified two key factors that we believe are the main reasons for why the spatial navigation technique performed so well.

Motor Space

One key benefit of the spatial navigation technique is the size of its motor space. If we consider the space between hip and chest as the preferable interaction zone, then this space is more than one order of magnitude larger than the average mobile touch screen. This is an important advantage in terms of physical resolution and accuracy. By performing only one continuous gesture, it allows users to cover very long distances within the document – by maintaining a high level of precision at the same time. For touch gestures, travel/zoom distances are considerably smaller per gesture [16], thus forcing users to perform multiple gestures to achieve the same result.

We found clear evidence supporting these claims. Users performed many touch gestures for both on-screen (M ≈ 4) and off-screen target tasks (M ≈ 10). In contrast, clutching was used only marginally in the spatial condition (M < 2.3). Here, participants clearly benefited from the larger motor space that they used more extensively for off-screen targets if compared to on-screen targets (see Table 1).

Motor Skills

Both navigation techniques target different parts of the human muscle system and thus demand different motor skills. The Pinch-Drag-Flick approach primarily addresses fine motor skills of the fingers, usually involving a high physical pointing accuracy within a small (screen) area. In our study, we repeatedly witnessed participants having difficulties with the pinch gesture. As a consequence, many participants found it easier to lift a display up/down for zooming, which is reflected by the user ratings for zooming, as depicted in Figure 7. Apart from that, touch input also requires a high visual attention, e.g., due to little tactile/haptic feedback. In contrast, spatial navigation explicitly supports proprioception, i.e., the sense of relative positions of neighboring body parts. We believe that such kinesthetic cues can reduce the demand of visual attention. As hinted in [7, 28], such cues can also enable users to associate important regions in the document with specific physical positions around their body, making it easier for them to quickly travel within the document. Another important benefit of spatial navigation is simultaneous zooming & panning, which is naturally supported by moving a display diagonally through the space-scale diagram [10, 19]. Participants made use of this very frequently in the study.

Limitations

As standard deviations of completion times indicate, most participants were similarly fast with the spatial technique, but showed diverse performance times for Pinch-Drag-Flick – even though they had prior experiences with the latter technique. One possible reason for this might be that participants were motivated more to succeed in the spatial technique, because it was new to them. Yet, we believe the high performance variations for touch can also be attributed to issues with the touch condition: First, displays were prone to get soiled by a thin film of sweat and grease after working on them for some time. As this affected the touch recognition, we carefully cleaned the displays each time before a participant started to work with them. Second, while holding the device, the (ball of the) thumb of participants occasionally came in contact with the display, thus accidentally interfering with the detection of other touch gestures. Third, we witnessed a few female participants (N = 4) who had problems, caused by their fingernails. Although the nails were not unusually long, these women struggled with a less reliable touch recognition. Many participants reported that they had experienced similar problems before, e.g., when working with their personal phone. Hence, we conclude that these issues do not weaken our findings, but rather reflect the condition of the world outside the lab.

Suggestions on Improving Spatial Navigation

In retrospect, the use of a dynamic spatial mapping based on local device orientations has proven to be a good choice. However, our observations indicate that there is room for improvement. In interviews, several participants (N = 5) asked for a finer mapping, allowing them to move the device less by still covering the same virtual distance in the document. We propose to provide a user setting for this, though it may be worth investigating suitable thresholds that might depend, for example, on the display size. To further improve on that, some participants proposed to adjust the travel speed within the document depending on how fast they were moving the device in physical space.

FUTURE GENERATIONS OF MOBILE DISPLAYS

Current mobile displays are lacking a few technical qualities that hinder a broader success of spatial-based navigation in the mass market. We identified two areas that future generations of mobile devices should address:

Device-Intrinsic Spatial Tracking

One major technical challenge is the support of reliable 6DoF spatial tracking for real mobile usage, where requirements different from those in the lab apply: First, the workspace is not stationary anymore. Therefore, external sensors are unavailable and light conditions are likely to vary considerably. Second, spatial tracking should be relative to users, so they can walk without affecting the interaction (body-centric tracking). Third, the algorithm should be energy efficient to ensure long working times.

Prior to conducting the study, we experimented with alternative sensing approaches that use the built-in sensors of mobile displays, e.g., gyroscopes [15]. Accelerometers are energy efficient and offer low-latency feedback, though cannot detect positional changes of the user (e.g., when walking) and also are prone to induce error drifting. Alternatively, Hansen et al. [10] suggested the tracking of facial landmarks by the front camera for zooming, thus facilitating body-centric tracking. However, this limits the interaction to 3DoF and does not work when the user's face is not within the camera's field of view. To overcome these shortcomings, sensor fusion [13] combines gyroscopic data with face tracking. Yet, this approach still suffers from potential inaccuracies and technical pitfalls, so we found a proper implementation to be too time-consuming. It is also questionable, whether this or similar approaches are already advanced enough in terms of fidelity, spatial range and energy consumption. Nonetheless, we believe that integrating such capabilities into future generations of mobile displays is the one major technical challenge that needs to be solved before spatial interaction becomes more widespread.

A Built-in Tactile Clutch

In our prototype, we used touch input for the de/activation of spatial input, which was primarily due to the lack of alternatives. While this worked generally well in our study (e.g., due to its limited scope), we do not consider touch-based clutches as our preferred solution. There are several reasons for that. First, touching the screen with a finger occludes parts of the viewport. Second, mixing on-screen touch input with spatial input is contrary to the philosophy of hybrid input paradigms, where different input channels should work independently from each other. Third, the clutch must be easily detectable preferably by non-visual cues so users can keep their visual attention on the document. Fourth, users should be provided with eyes-free feedback regarding the current state of a clutch. This is to provide precise control on when and how long the clutch is activated. These requirements may appear trivial, but their influence on the user performance and satisfaction should not be underestimated, e.g., see [11]. This is also supported by our observations and interviews with participants.

Hence, we propose to equip future generations of mobile devices with a clutch that provides some form of tactile feedback. This may be a simple physical button, though it should be larger than the tiny volume controls usually found on mobile phones. Ideally, the clutch would be readily usable independent of the current orientation of the display, for example, by squeezing the display bezel [3].

CONCLUSION AND FUTURE WORK

In this paper, we presented a comprehensive user study that systematically compares the efficiency and user satisfaction of two contrarian input strategies for 2D document navigation on mobile displays: the predominant touch-based Pinch-Drag-Flick approach with a spatial-input-based approach that utilizes positional changes of a mobile display in the physical space surrounding a user. The results surpassed our expectations in various ways. On average, participants were more than 35% faster with the spatial approach, even though all of them were conversant with Pinch-Drag-Flick and used the spatial technique for the first time. This finding was further supported by the questionnaires, where participants rated the spatial approach at least as good as or even better than the touch-based counterpart. To the best of our knowledge, we are the first who provide such clear evidence in favor of spatial input. This was only possible by building high quality prototypes that make use of state-of-the-art mobile devices. Considering the popularity of Pinch-Drag-Flick, our findings could be of interest for future interaction designs of mobile devices – as a complimentary method of interaction, yet not as a complete replacement. Because there are also limitations: social protocols may limit its application, users may perform differently when sitting, and users may prefer to put a display on a desk for certain tasks. However, given the additional advantages of a supplemental input channel, we hope that our findings will help mobile computing embrace spatial interaction principles much more than before.

For future work, we plan to address the technical challenges and design recommendations that we discussed in the previous section, in particular device-intrinsic spatial tracking via sensor fusion and tactile clutches. With this technology, we will then continue our investigations by testing how in-the-wild usage (e.g., when walking) affects performances as well as the accuracy and recall [20]. Beyond that, we intend on studying compound tasks [19] that involve additional tasks, such as selection or annotations, and thus may particularly benefit from combining touch with spatial input.

ACKNOWLEDGEMENTS

We would like to thank all the participants who generously shared their time for the study. We are also grateful to the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

REFERENCES

1. Bier, E.A., Stone, M.C., Pier, K., Buxton, B., and DeRose, T.D. Toolglass and Magic Lenses: The See-Through Interface. In *Proc. SIGGRAPH 1993*, ACM (1993), 445–446.
2. Boring, S., Ledo, D., Chen, X., Marquardt, N., Tang, A., and Greenberg, S. The fat thumb: using the thumb's contact size for single-handed mobile interaction. In *Proc. MobileHCI 2012*, ACM (2012), 207–208.
3. Burstyn, J., Banerjee, A., and Vertegaal, R. FlexView: An evaluation of depth navigation on deformable mobile devices. In *Proc. TEI 2013*, ACM (2013), 193–200.
4. Chae, M. and Kim, J. Do size and structure matter to mobile users? An empirical study of the effects of screen size, information structure, and task complexity on user activities with standard web phones. *Behaviour and Information Technology 23*, 3 (2004), 165–181.
5. Cutmore, T., Hine, T., Maberly, K., Langford, N., and Hawgood, G. Cognitive and gender factors influencing navigation in a virtual environment. *Int. Journal of Human Computer Studies 53*, 2 (2000), 223–249.
6. Dabbs, J., Chang, E., Strong, R., and Milun, R. Spatial ability, navigation strategy, and geographic knowledge among men and women. *Evolution and Human Behavior 19*, 2 (1998), 89–98.
7. Fitzmaurice, G.W., Zhai, S., and Chignell, M. Virtual reality for palmtop computers. *ACM Trans. Inf. Syst. 11*, 3 (1993), 197–218.
8. Furnas, G.W. and Bederson, B.B. Space-Scale Diagrams: Understanding Multiscale Interfaces, In *Proc. CHI 1995*, ACM (1995), 234–241.
9. Gutwin, C. and Fedak, C. Interacting with big interfaces on small screens: a comparison of fisheye, zoom, and panning techniques. In *Proc. Graphics Interface 2004*, Canad. H.-C. Comm. Society (2004), 145–152.
10. Hansen, T.R., Eriksson, E., and Lykke-Olesen, A. Mixed interaction space - Expanding the interaction space with mobile devices. In *Proc. HCI 2006*. Springer London (2006), 365–380.
11. Jones, B., Sodhi, R., Forsyth, D., Bailey, B., and Maciocci, G. Around device interaction for multiscale navigation. In *Proc. MobileHCI 2012*. ACM (2012), 83–92.
12. Jones, M., Marsden, G., Mohd-Nasir, N., Boone, K., Buchanan, G. Improving web interaction on small displays. *Computer Networks 31*, 16 (1999), 1129–1137.
13. Joshi, N., Kar, A., and Cohen, M. Looking at you: fused gyro and face tracking for viewing large imagery on mobile devices. In *Proc. CHI*, ACM (2012), 2211–2220.
14. Jul, S. and Furnas, G.W. Critical zones in desert fog: aids to multiscale navigation. In *Proc. UIST 1998*, ACM (1998), 97–106.
15. Kaufmann, B. and Ahlström, D. Studying spatial memory and map navigation performance on projector phones with peephole interaction. In *Proc. CHI 2013.* ACM (2013), 3173-3176.
16. Malacria, S., Lecolinet, E., and Guiard, Y. Clutch-free panning and integrated pan-zoom control on touch-sensitive surfaces: the cyclostar approach. In *Proc. CHI 2010,* ACM (2010), 2615–2624.
17. Norman, D. The Design of Everyday Things. Basic Books, 2002.
18. Oh, J. and Hua, H. User Evaluations on Form Factors of Tangible Magic Lenses. In *Proc. ISMAR 2006*, ACM (2006), 23–32.
19. Pahud. M., Hinckley, K., Iqbal, S., Sellen, A., and Buxton, B. Toward compound navigation tasks on mobiles via spatial manipulation. In *Proc. Mobile HCI 2013*. ACM (2013), 113–122.
20. Rädle, R., Jetter, H.-C., Butscher, S., and Reiterer, H. The effect of egocentric body movements on users' navigation performance and spatial memory in zoomable user interfaces. In *Proc. ITS 2013*. ACM (2013), 23–32.
21. Reeves, B., Lang, A., Kim, E.Y., and Tatar, D. The effects of screen size and message content on attention and arousal. *Media Psychology 1*, 1 (1999), 49–67.
22. Rogers, Y., Sharp, H., and Preece, J. Interaction Design: Beyond Human-Computer Interaction. John Wiley & Sons, 1st edition, 2002.
23. Spelmezan, D., Appert, C., Chapuis, O., and Pietriga, E. Side Pressure for Bidirectional Navigation on Small Devices. In *Proc. MobileHCI*, ACM (2013), 113–122.
24. Spindler, M., Martsch, M., Dachselt, R. Going Beyond the Surface: Studying Multi-Layer Interaction Above the Tabletop. In *Proc. CHI*, ACM (2012), 1277–1286.
25. Spindler, M., Stellmach, S., and Dachselt, R. Advanced Magic Lens Interaction above the Tabletop. In *Proc. ITS 2009*, ACM (2009), 77–84.
26. Treisman, A., Gelade, G. A Feature Integration Theory of Attention. *Cogn. Psychology 12*, 1 (1980), 97–136.
27. Tsang, M., Fitzmaurice, G., Kurtenbach, G., Khan, A., and Buxton, B. Boom Chameleon: Simultaneous Capture of 3D Viewpoint, Voice and Gesture Annotations on a Spatially-aware Display. In *Proc. UIST 2002*, ACM (2002), 111–120.
28. Ullmer, B., Ishii, H., and Jacob, R. Token + constraint systems for tangible interaction with digital information. *ACM Trans. C. - H. Interaction 12*, 1 (2005), 81–118.
29. Wigdor, D., Forlines, C., Baudisch, P., Barnwell, J., and Shen, C. Lucid touch: a see-through mobile device. In *Proc. UIST* 2007, ACM (2007), 269–278.
30. Yee, K. Peephole Displays: Pen Interaction on Spatially Aware Handheld Computers. In *Proc. CHI 2003*, ACM (2003), 1–8.