Does (Multi-)Touch Aid Users' Spatial Memory and Navigation in 'Panning' and in 'Zooming & Panning' UIs?

Hans-Christian Jetter[1]   Svenja Leifert[1]   Jens Gerken[2]   Sören Schubert[1]   Harald Reiterer[1]

[1]HCI Group, University of Konstanz
Universitätsstraße 10, Box 73
78457 Konstanz, Germany

{jetter, leifert, schubert, reiterer}
@inf.uni-konstanz.de

[2]ICT AG
Erscheckweg 1
72664 Kohlberg, Germany

j.gerken@ict.de

ABSTRACT

Recent findings from *Embodied Cognition* reveal strong effects of arm and hand movement on spatial memory. This suggests that input devices may have a far greater influence on users' cognition and users' ability to master a system than we typically believe – especially for spatial *panning* or *zooming & panning* user interfaces. We conducted two experiments to observe whether multi-touch instead of mouse input improves users' spatial memory and navigation performance for such UIs. We observed increased performances for *panning* UIs but not for *zooming & panning* UIs. We present our results, provide initial explanations and discuss opportunities and pitfalls for interaction designers.

Categories and Subject Descriptors

H.5.m [**Information interfaces and presentation**]: Miscellaneous

General Terms

Experimentation, Human Factors, Measurement, Theory.

Keywords

Touch, Multi-Touch, Zooming, Panning, ZUI, Spatial Memory, Navigation, User Study, Embodied Cognition.

1. INTRODUCTION

User interfaces (UIs) with a *panning* or a *zooming & panning* style of navigation have become ubiquitous in present-day computing. They are employed whenever only a limited amount of physical display space can be used for navigating a larger virtual canvas of spatially distributed data. Examples for *panning* UIs are the home screens of smartphones or tablets where panning is used to scroll through many pages of apps and icons. Examples for *zooming & panning* UIs are InfoVis or mapping applications that show vast amounts of geographical or visual data at different scales and levels of detail. Today, *panning* and *zooming & panning* navigation have become an important part of the design

guidelines of novel natural user interfaces (NUIs) for touch and gesture. For example, Wigdor & Wixon propose "the spatial NUI" that uses a 2D planar space instead of hierarchical navigation to leverage users' spatial memory [30]. One reason for this increasing popularity of *panning* and *zooming & panning* NUIs is their adequacy for devices with multi-touch input – in particular pads, tablet PCs and tabletops: The famous 'zoom' or 'pinching' multi-touch gesture is shown in almost every TV advertisement for multi-touch devices to demonstrate the naturalness of zooming into maps, web pages or photos by touch. Another frequently advertised interaction is sliding a finger on a screen to move a larger page or document, so that the desired section can be panned into the display's focus.

Despite this popularity, HCI still knows surprisingly little about the effects that this shift from mouse to touch input has on the users' spatial cognition and their ability to master *panning* and *zooming & panning* UIs. Recent findings from *Embodied Cognition* [5, 9, 14] provide plausible arguments why replacing the mouse with touch might have strong effects that go far beyond the already documented (dis)advantages of touch such as speed and accuracy of target acquisition (see [7, 19]). For example, various studies demonstrate that the ability to transform mental images is linked to motor processes, so that rotating one's hands in the direction opposite to the required mental rotation slows down the speed of mental rotation [9, 29]. Others report that hand and arm gestures facilitate the maintenance of spatial representations in working memory [28]. Therefore, it seems very plausible that an UI's input modality and its proprioceptive and kinesthetic feedback may also have an effect on users' spatial memory. In HCI, such an effect could have a great impact on the usability of *panning* and *zooming & panning* UIs where spatial memory is needed for recalling visual landmarks and their spatial relation to enable an effective and efficient navigation. For this reason, we have conducted two experiments to reveal if touch instead of mouse input aids users' spatial memory and improves navigation performance for such user interfaces.

2. RELATED WORK

Embodied Cognition renders the traditional Cartesian dualism with a mind-body separation as obsolete. In this novel embodied view, the body and cognitive skills coevolved from the beginning as a unit: "(…) it is the entire system of muscles, joints, and proprioceptive and kinesthetic functions and appropriate parts of the brain that evolve and function together in a unitary way" [13]. Accordingly, recent experiments have observed strong effects that

hand or arm movement has on mental tasks, images, speech or working memory [4, 9, 20, 28, 29]. For example, Wesp et al. [28] and De Ruiter [4] argue that gestures may help maintain spatial images in the visuo-spatial scratchpad of Baddeley's model of working memory [1]. Although *Embodied Cognition* is increasingly applied in HCI [5, 11, 14], the particular role of mouse vs. touch for spatial cognition has not been researched yet. Traditional comparisons of input devices only focus on Fitts' law and speed/accuracy tradeoffs [7, 16, 18, 19]. While they are an essential part of using any kind of UI, they are not subject to our research here. Instead, we focus on aspects of spatial cognition that have a similar or even greater impact on the usability of a *panning* or *zooming & panning* UI: *spatial memory* and *navigation*. Both are essential for knowing where to move and how to move to desired locations in virtual space and thus are critical for any effective and efficient usage. They demand cognitive abilities such as the development and use of a mental spatial representation of the virtual canvas, deciding on a path towards a currently invisible destination, and following this path using the input device until the destination becomes visible. While we do not know about any prior work addressing this issue, we identified four categories of related work from HCI:

1.) Comparing accuracy and speed of mouse vs. touch – Meyer et al. reported that mouse outperformed touch in terms of speed for a desktop display [18]. Micire et al. studied selection tasks on a tabletop and concluded that task completion was faster with touch and the error rate was comparable to the mouse for target sizes above 30mm [19]. Forlines et al. showed that users of tabletops may be better off when using a mouse for unimanual input and their fingers for bimanual input [7]. They also suggested that other design considerations such as spatial memory should play a role and further investigations into such qualities are needed.

2.) Fitts' law for 2D tasks and zooming – The original Fitts' law models one-dimensional target acquisition as a speed vs. accuracy tradeoff. Mackenzie et al. applied this to two-dimensional tasks [16] and provided the basis for above-mentioned comparative studies [7, 19]. Guiard & Beaudouin-Lafon extended Fitts' original pointing paradigm to multiscale UIs or zoomable user interfaces (ZUI) and showed that Fitts' law does also apply to such UIs [10]. However, this does not help us to directly address our research questions concerning spatial memory and navigation.

3.) Touch gestures for zooming and panning – Today's dominant design is the aforementioned two-finger 'zoom' or 'pinching' gesture. An alternative to this 'pinching' design was proposed by Malacria et al. [17]: CycloPan and CycloZoom are clutch-free single-touch gestures based on oscillatory motions. However, embodied views of spatial memory or navigation were not a part of their design rationale or evaluation. Other work on touch gestures is concerned with the design of entire multi-touch gesture sets, e.g. [25, 31]. However, they also were not evaluated with respect to their effect on spatial memory or navigation.

4.) Spatial memory – In the past, HCI studies of spatial memory for bookmark or window management, e.g., Data Mountain [23], Task Gallery [24], only focused on the visuo-spatial metaphors on the screen and how they affect the user. The input method with mouse and keyboard was not a subject to research. More recent research on spatial and content memory completely removed user input from the study and focused solely on the effect of visual grids [15]. Ebert et al. explored the combination of spatial and kinesthetic memory in front of a wall-sized 3D vision display with real depth perception [6]. Users performed a memory task in

a 3D scene with mouse vs. physical navigation, i.e., walking in front of the screen. However, Ebert et al.'s results were inconclusive and cannot be applied to 2D *panning* UIs or *ZUIs* without depth perception and only hand and arm movement.

In conclusion, we are only aware of a single study in HCI that is closely related to our work. In 2002, Tan et al. conducted a user study on the effect of kinesthetic cues on spatial memory [26]. They compared mouse vs. touch input for a memorization task on a vertical 18.1" screen during which 28 users had to memorize objects by dragging them into given locations in an 11 by 7 invisible grid. In contrast to our research, the UI was neither a *panning* nor a *zooming & panning* UI. The grid remained constantly at the same absolute screen position. Tan et al. reported a significant 19% improvement of the accuracy of remembered locations for touch input.

3. FOUNDATIONS AND DEFINITIONS

In the following, we first provide foundations for and definitions of the core concepts of our study to enable a precise discussion:

A *zooming & panning* UI or Zoomable User Interface (ZUI) [22] relates two parts: a *viewport* and a *canvas*. The *viewport* serves as a window through which the user can view and interactively explore the spatially distributed data that is contained in the *canvas*. The part of the *canvas* that is currently visible inside the *viewport* is called the *view*. The current *view* can be specified using three coordinates: The top-left corner of the *view* in *canvas*-coordinates X and Y and the scale-factor S [8]. S determines the size of the *view* in relation to the size of the *canvas*: The greater S, the smaller is the part of the canvas that the *view* covers, but the greater is the rendered level of detail. $S = 1.0$ typically means that the *view* covers the entire *canvas*. To explore the *canvas*, users can employ two interaction primitives: *panning* and *zooming*. *Panning* means changing X and Y without changing S. To the user, *panning* feels like moving sideways. *Zooming* means changing S without changing X and Y. To the user, *zooming* feels like approaching or stepping back from the *canvas*. A *panning UI* is simply a ZUI where only panning is allowed and $S = const.$ with $S > 1.0$.

Spatial memory is an essential cognitive process that humans use to encode the space around them [27]. It develops and uses a mental representation or cognitive map in our mind [3]. Different parts of the brain provide us with this ability to remember the positions and identities of objects that we have seen [21]. For the purpose of our study, we define *spatial memory* as the users' mental representation of the virtual canvas of a ZUI. It enables them to recall locations of objects based on their identity. The *spatial memory performance* can be measured by analyzing the accuracy of the results of an explicit reproduction task. This task resembles a retrospective map drawing exercise during which users try to assign objects to their original location in the canvas.

Navigation is the aggregate task of wayfinding and motoric motion [3]. For our purposes, we define *navigation* as a user activity during which a directed movement takes place from a home position in the virtual canvas to a destination object. Since the destination object is not visible from the home position, successful *navigation* makes use of *spatial memory*. Good *navigation performance* means that the path used to move to the destination is short and ideally the shortest possible. Thus, the *navigation performance* can be measured by calculating the spatial length of the executed navigation path. For our purposes, we consider *navigation performance* as a measure of length and

not of time or motor costs since they would introduce noise that is not related to memory (e.g. user-specific experience with using multi-touch, individual preferences for mouse velocity). We also measure *spatial memory performance* and *navigation performance* individually to identify potential commonalities and differences between them. For example, physical ergonomics or motor memory [14] might affect *navigation performance* but not *spatial memory performance*.

4. EXPERIMENTS

We conducted two experiments (E1, E2) with the hypothesis that touch instead of mouse input would result in better *spatial memory performance* and *navigation performance*. This assumption was loosely based on existing studies on the effect of hand and arm movement [28, 29] but in particular on the work of Tan et al. [26]. Both experiments simulated real-world UIs using an abstract UI design with a spatial layout of objects that was greater than the visible screen size. E1 compared the impact of touch vs. mouse on a pure *panning* UI that resembled a large home screen (e.g. Windows 8) with many apps to switch between. E2 mimicked a classic *zooming & panning* UI such as Pad++ [22] or Google Earth with semantic zooming where objects at different locations and scales reveal their details only after zooming in. Both experiments took place at different points in time and with different participants.

During the experiments participants performed navigation tasks in which frequent switching between a home position and destination objects was necessary. This kind of repeated switching and navigation between objects is typical for many real-world applications. Memorization of object identities and locations becomes a by-product of normal use and makes them last in spatial memory for at least several minutes. This is decisive for the UI's usability, since the time for navigation and visual search is reduced drastically when the destination location can be recalled.

Figure 1. Physical setup of the experiments using a tabletop.

4.1 Experiment 1 – Panning UI

The goal of E1 was to observe whether the improved spatial memory performance with touch that Tan et al. reported can also be observed in a *panning* UI where the objects do not keep their absolute screen positions, but move on the screen following the users' panning operations. 20 participants (7 female, 13 male) were recruited from the campus of our university. Participants ranged in age from 18 to 28 years and were paid 7 EUR in compensation for their time.

4.1.1 Apparatus

We used a horizontal interactive tabletop (a Microsoft Surface) with a diameter of 30" and a resolution of 1024x768. The device was used in its original "coffee-table" configuration without further elevation (Figure 1). For touch input we used the built-in touch tracking of the Surface. For mouse input we used a wireless Logitech Anywhere Mouse MX. The mouse was operated by the participants on top of a small rolling table next to the tabletop that had the same height as the Surface.

For our study, a trade-off had to be made between a natural use and orientation of the mouse and keeping the users' visual frame of reference constant, i.e., perceived screen size, viewing angle, relative position to Surface. As the latter is more critical for spatial memory measurement, we decided to keep the visual setting constant and avoided occlusion of the screen with arm or mouse by providing a rolling table that users could place as desired (Figure 1). Since we did not measure time or motor activity, but judged a device by how precisely positions are remembered after using it and how short the user's navigation paths were, this potentially unfamiliar and physically demanding mouse position could not bias our measurement too strongly. Also, the proprioception of relative movements is not compromised by the mouse orientation on the rolling table. The mouse sensitivity was set to the default values of the Logitech mouse driver and was kept constant in both experiments and for all participants. These values enabled participants to operate the mouse without extensive clutching while still maintaining a high level of precision and thereby providing a realistic setting.
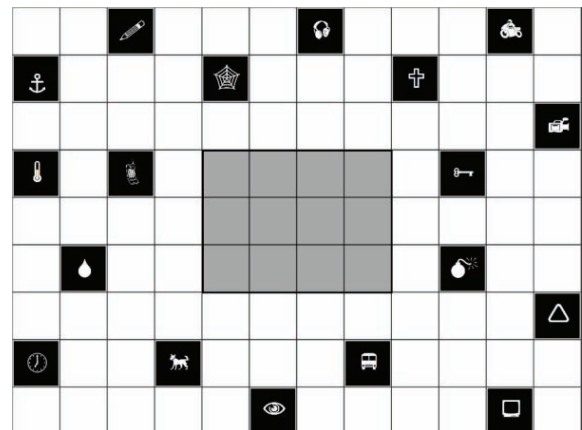

Figure 2. A configuration from E1. The position and size of the view at the home position is highlighted in grey.

At the tabletop, participants navigated a canvas using panning operations. The canvas contained a 12 by 9 grid with a spatial configuration of 18 items (Figure 2). At no time the entire grid was visible, since the screen always showed only a 4 by 3 section of the grid. The empty space in the center served as a home position without visible items. Each item was of similar size and color. The positions of the items were initially random and manually altered to avoid that participants can easily apply obvious memorization strategies, e.g., "all living things are at the bottom".

4.2 Conditions

The design of the mouse condition emulated popular *panning* UIs: The mouse cursor could be moved freely over the canvas. By pressing the mouse button over an arbitrary location in the canvas,

the canvas could be grabbed and dragged around. In the touch condition, the same principle was applied but participants used their fingers to touch & slide for panning. In both conditions clutching was necessary for long-distance panning, since the mouse cursor or the finger could not be used outside the screen's physical boundaries.

4.2.1 Navigation Task
In the *navigation task*, the system started at a home position in the empty center of the canvas where no item was visible. Then a destination item that had to be found in the canvas was shown in a transparent overlay in the center of the screen. Participants had to find this item in the canvas and pan it into the screen's center with a tolerance of 100 pixels around the center. After this, the item was considered as found and the task continued at the home position with a different item.

A sequence of 8 of the 18 items in the canvas was presented one after another as destination items. The same sequence of items was then repeated 8 times (8 blocks), resulting in 64 search trials. The first attempts in a navigation task were always bound to fail, since right after the start the configuration was still completely unknown to the participant. However, with each block, the participant saw and memorized more sections of the canvas, the items, and item locations. Therefore, we could observe the development of an increasingly accurate mental representation and increasingly efficient navigation paths.

During the navigation task, all panning operations and their *XYS*-coordinates were continuously logged to calculate the navigation performance. We excluded the data from the first block from data analysis to take into account the randomness of the first navigation attempts.

4.2.2 Spatial Memory Task / Reproduction Task
In the *spatial memory task* (or *reproduction task*), we asked participants to reproduce the learned item configuration, similar to a map drawing exercise. A random sequence of the 8 items that participants had to navigate to during the navigation task was shown and for one item after another, they were requested to put the item into the grid at the exact location where it had been during the navigation task. Participants did not receive feedback on the accuracy of their placements and for each item of the sequence they had to start over with an empty grid. To eliminate potential influences by the unconscious use of motor or kinesthetic memory, the reconstruction task was not performed with mouse, touch or pointing at locations on the tabletop. Instead, participants used the cursor keys of a wireless keyboard to move the object inside the grid. During the spatial memory task we used the Euclidean distance as measurement and asked users to optimize for accuracy.

4.2.3 Procedure
We used a counterbalanced within-subject design for our experiment with input modality as the independent variable. Therefore, for each participant, the test consisted of two main parts, one for each input device. Each part was divided into three phases, in each of which one navigation task and one spatial memory task had to be completed. In phase 1, the subjects were introduced to the usage of the input device and the nature of the test tasks by a simple example and a short introduction by the experimenter. In phase 2, the practice phase, the participants were given the opportunity to get used to the input device and try out different tactics. While they still had the opportunity to ask questions, they were encouraged to constantly increase the pace

during the navigation task and to achieve a high accuracy during the spatial memory task. During phase 3, the actual data collection phase, each user performed the entire navigation and reproduction task as described above. Like in the second phase, users were asked to focus on achieving a high speed respectively accuracy in the spatial memory task. Different item configurations were used in the practice and data collection phases leading to a total amount of four item configurations (two in the mouse and two in the touch condition) which were counterbalanced between the participants.

4.2.4 Results for Spatial Memory Performance
The comparison of the results of the spatial memory task with mouse vs. touch shows a significantly better spatial memory performance for the touch condition. While the mean error in grid units was 1.092 (SD = 0.157) in the mouse condition, touch led to more accurate results with a smaller mean error of 0.795 grid units (SD = 0.096). This difference is statistically significant with $F_{1,19} = 5.724$, $p < 0.05$. Thus, similar to Tan et al.'s 19% improved performance with touch for a *non-panning* UI, we also observed a 37% improvement for a *panning* UI. Compared to the mouse, touch input does apparently facilitate the encoding of object locations in the users' mental representation.

Tan et al. attribute this to "kinesthetic cues" when using touch, but do not further elaborate on this or explain how they work. We also believe in such an effect based on proprioceptive and kinesthetic feedback and provide a first hypothesis about its nature in the discussion section. In both conditions, we were surprised by the absolute spatial memory performance that exceeded our expectations. We therefore increased the difficulty of the spatial memory task in the second experiment to avoid ceiling effects.
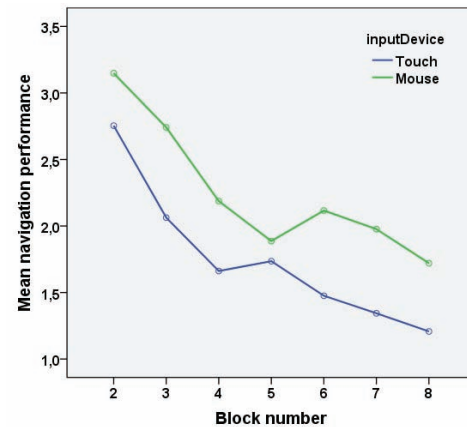

Figure 3. Navigation performance based on mean panning distance (E1).

4.2.5 Results for Navigation Performance
To determine the navigation performance for the conditions, we used all panning distances of the navigation task except block 1 and divided each of them by the optimal, shortest possible distance. Therefore, the closer this measure is to 1.0, the better. With touch the mean was 1.749 (SD = 0.127) and significantly smaller than the mean with mouse of 2.254 (SD = 0.252) with $F_{1,19} = 8.703$, $p < 0.05$.

As discussed before, this 29% improvement in navigation performance is not necessarily based solely on the improved spatial memory performance. Other effects could come into play

here. However, since the result resonates with the increased spatial memory performance, it suggests that touch input not only results in a more accurate mental representation, but that it also enables users to effectively apply this during navigation tasks.

Regarding the development of the performance over time (Figure 3), we observed the expected learning progress that our study design was based on. Interestingly, in block 5 (touch) and block 6 (mouse) there was a sudden increase in panning distance and a decline in navigation performance. However, during the following blocks, participants returned to their previous performance and improved further. We can only speculate on the reason for this, but could imagine that this was due to a temporary decrease in attention after participants noticed the repetitive nature of the task and temporarily lost interest or motivation.

4.3 Experiment 2 – Zooming & Panning UI

E2 was very similar to E1, so that we only describe the differences between them. The most prominent difference was the use of a *zooming & panning* UI instead of a *panning* UI. Based on E1, we assumed that zooming & panning on a large horizontal touch screen would lead to better performances for spatial memory and navigation. In particular, we assumed that the employed two-finger "pinching"-style for zooming with multi-touch could have a strong effect: Similar to the mouse, "pinching" enables users to control panning and zooming simultaneously. However, by touching two points in the canvas and simultaneously dragging them, the user cannot only specify the direction of changing S (e.g. *zooming in*), but also the precise absolute amount by which S is changed (e.g. *zooming in by factor 4*). For example, if users intend to zoom into a small object on the screen, they can touch the object at opposite corners and slide their fingers apart to zoom in until the object is displayed exactly in the desired size and screen position. This is different from the less direct control of S with a mouse wheel by using incremental finger movements. To still achieve a precise control of S, we immediately multiplied or divided S by a small factor $\Delta s = 0.95$ whenever a wheel changed event was fired. This resulted in a smooth and continuous zooming with many updates of the screen content per second.

To ensure that participants make intensive use of zooming, the initial home position was changed, so that the participants had an overview of the entire canvas from the beginning. However, the items only became visible after zooming in until S exceeded a threshold. When S was below this threshold, the items' identity was not visible and only their locations in the grid were indicated by empty black boxes (Figure 4). Furthermore, the *XY*-distance between the items was increased compared to E1. This lead to a stronger "desert fog" phenomenon [12] when users tried to pan large distances at high scale factors without zooming out first.

To avoid the potential ceiling effects for spatial memory that we discussed for the first experiment, the width and height of the grid was quadrupled to a total of 48 by 36 grid cells. This increase in grid size did not change the nature of the tasks or the interaction, but only increased the measuring resolution of distances by factor 4 (similar to using feet instead of yard). Also, all grid lines or other visual landmarks except the items were removed to avoid confounding variables. Due to these alterations, participants made full use of zooming and panning as expected.

16 participants (8 female, 8 male) were recruited from the campus of our university. Participants ranged in age from 20 to 35 years and were paid 10 EUR in compensation for their time.

4.3.1 Results for Spatial Memory Performance

Unlike in E1, we could not observe an improvement of spatial memory performance with touch in E2. For the mouse, the mean error in grid units was 2.628 (SD = 0.260). For (multi-)touch, the mean error was 2.612 (SD = 0.308). This difference was statistically not significant ($F_{1,15} = 0.003$, p = 0.960).
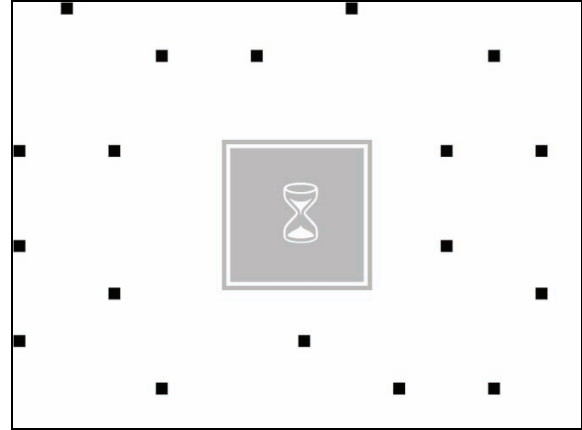
Figure 4. Start of ZUI navigation task. The destination item to navigate to is indicated in the center. Participants can zoom & pan into the black boxes to look at the contained items.

4.3.2 Results for Navigation Performance

Calculating path lengths in a ZUI is not trivial. Therefore, we used a ZUI navigation cost metric for our analysis of navigation performance which is based on the space-scale diagrams of Furnas & Bederson and their discussion of the "joint pan-zoom problem" [8]. This metric takes into account that the length of a path in a ZUI is highly dependent on the movement in S and not only in X and Y: "A vast distance may be traversed by first zooming out to a scale where the old position and new target destination are close together, then making a small pan from one to the other, and finally zooming back in (…). Since zoom is naturally logarithmic, the vast separation can be shrunk much faster than it can be directly traversed, with exponential savings in the limit" [8].

Following Furnas & Bederson's suggestions, the "cost" of navigation in a ZUI depends on the amount of visual information that has to be added into a viewport while zooming and panning and thus can also be seen as a measure for the additional visual information that has to be processed by the user. During panning, entirely new information moves into the viewport and that is "expensive". However, zooming always keeps parts of the current visual information and is therefore "less expensive". Due to this different nature of panning and zooming, the cost of a pure pan is linear in the distance panned, and the cost of a pure zoom is logarithmic with change of scale. Both costs are related using a constant that is determined by the number of pixels in the viewport [8]. We can therefore formulate the ZUI navigation cost metric as follows:

A user executes a navigation step of Δx, Δy, Δs in a ZUI where $\Delta x = |x_n - x_{n-1}|$, $\Delta y = |y_n - y_{n-1}|$ and $\Delta s = S_n / S_{n-1}$. Here, x_n, x_{n-1}, y_n, y_{n-1} are screen coordinates that can be calculated from the canvas coordinates X, Y. Given a ZUI with a viewport size of $W \times H$, e.g. 1024 x 768 pixels, the cost c of the navigation step can then be formulated as:

$$c = H \Delta x + W \Delta y - \Delta x \Delta y + WH |\log(\Delta s)|$$

$H \Delta x + W \Delta y - \Delta x \Delta y$ is the amount of pixels that have to be updated due to the panning motion. $H \Delta x$ and $W \Delta y$ are the amounts of pixels that are panned into the viewport. Since horizontal and vertical panning happens in parallel, there is an overlap of $\Delta x \Delta y$ pixels that do not have to be updated.

$WH |\log(\Delta s)|$ is the amount of pixels that have to be updated because of zooming in or out. The unit of the ZUI navigation cost metric is *square pixels* which may appear surprising, but is simply because we use *pixels* as a unit for length while the resulting cost is a screen area.

The navigation performance can then be determined by using this ZUI navigation cost metric to calculate the total cost of all zooming and panning operations during the navigation task (Figure 5): When applying this metric for navigation performance, the resulting mean for mouse condition is 2,988,269 (SD = 249,037) and 3,820,188 (SD = 377,600) for touch. This difference is statistically significant with $F_{1,15} = 4.727$, $p < 0.05$.

Similar to the spatial memory performance, we could not observe the expected increase of navigation performance in the touch condition. Instead, the navigation cost in the mouse condition was 28% smaller than with touch. This is noteworthy, since other empirical data is in favor of touch: The mean task completion time for touch was 540 seconds (SD = 19.2) and 633 seconds (SD = 20.1) for mouse, so that touch was 17% faster. The difference in the mean is statistically significant with $F_{1,15} = 15.012$, $p = 0.001$. Furthermore, all participants preferred touch over mouse input when asked after the experiment.
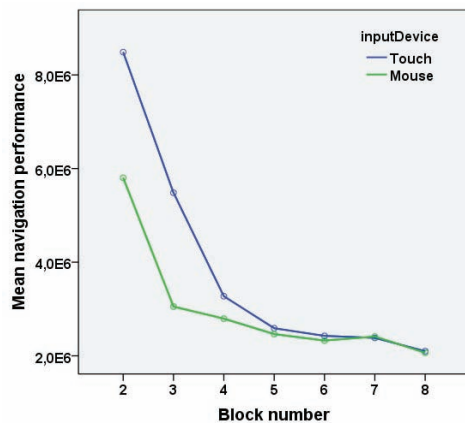

Figure 5. Navigation performance in E2 based on the ZUI navigation cost metric.

4.3.3 Non-Memory Effects
During the experiments, we made an observation that could explain why in E2 the navigation performance with touch was below that of the mouse, although participants generally preferred touch and were faster. Figure 6 visualizes this using two heat maps that show the density of interaction points for all users. They contain all screen locations where the mouse cursor or fingers were used for panning or zooming. Using touch, participants had a tendency to focus their touches to a region below the center of the tabletop. Apparently this region was convenient to reach and touches closer to the edges were avoided, since they required greater motor activity. Also, when the target was close to the screen's edges, zooming in with two fingers seemed too cumbersome because there was not enough room for sliding two fingers apart. Most participants did not use the

possibility to zoom into targets by putting one finger close to the edge and sliding only the other finger towards the center.

Using mouse, there was a similar central region where most interaction happened, but on the whole, there was a wider spread of activity across the screen. Also, participants often moved the mouse cursor against the screen's edges and used the wheel to conveniently zoom into targets along these edges. In terms of ZUI navigation cost, this approach is highly efficient and comes close to an optimal zoom-pan-trajectory. The mouse heat map in Figure 6 reveals the wider spread of interaction points across the screen and the higher density of interactions along its edges.

5. DISCUSSION
In the following, we suggest an initial explanation of our observations that can serve as a basis for future models of the effect of the input modality on spatial memory and navigation when using (multi-)touch in a *panning* or *zooming & panning* UI. Although it has to be considered only as a first step towards a model, we believe it as an important contribution to guide further research. It also enables us to formulate first implications for design.

5.1 Touch Input for Panning UIs
Based on the results of our first experiment and previous work of Tan et al., we can conclude that there is an effect of touch vs. mouse input for *non-panning* and *panning* UIs. We assume that the greater intensity of proprioceptive and kinesthetic feedback of touch facilitates the encoding of object locations in the users' mental representation. Although the nature of E1 was different from that of Tan et al., we see commonalities in the results that we can explain as follows:

In a *panning* UI the absolute positions of objects on the screen constantly change with the user's panning operations. During panning, the user never sees the entire canvas but watches it through a window that travels in the canvas and only exposes temporary views on small sections. This has two consequences: First, the absolute positions on the screen are ephemeral and constantly change in time. Second, to develop a permanent mental representation of the entire canvas with global locations, users have to integrate these many ephemeral views with their temporary screen positions over time. We believe that this is not a purely visual process but can also be aided by proprioceptive and kinesthetic feedback from hand and arm movement. Our assumption is that there is a greater intensity of proprioceptive and kinesthetic feedback when using touch on a tabletop instead of a mouse next to a tabletop. More precisely, we assume that the greater perceived amplitude (or length) of movements of the hand in the horizontal plane, which is sensed by stretch receptors of wrist, elbow, arm and glenohumeral joints and muscles, amplifies the encoding of locations in spatial memory. This feedback provides an additional motor perception of the panned distances that aids the process of relating the current visual input with its local positions to the existing mental representation of the canvas with its global locations. Users combine their visual perception of the panning distance in display space with the proprioceptive and kinesthetic sensation of the panning distance from motor space to form a multi-modal sense of distance. This multi-modal sense facilitates the process of integration which seems plausible, since similar effects exist for path integration and geographic orientation in 3D virtual environments [2]. Orientation based on visual flow alone proved to be more inaccurate and unreliable.
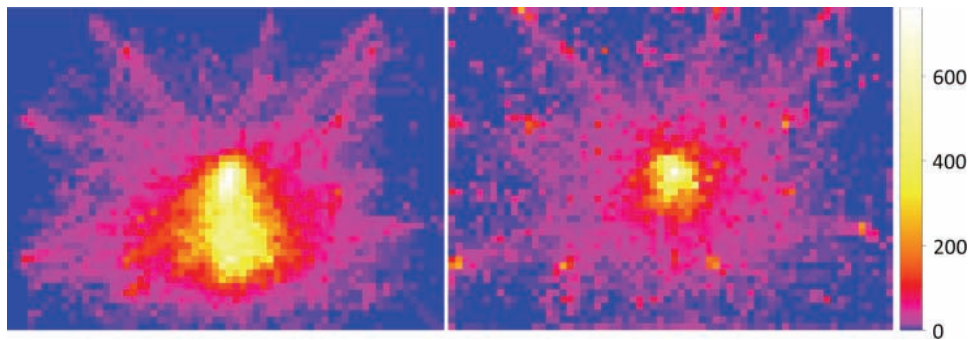
Figure 6. Heat maps for touch (left) and mouse (right). The color map indicates the number of logged events/region.

While proprioceptive and kinesthetic feedback is present for mouse and touch, there are two important differences: First, for touch screens larger than a typical mouse pad the intensity and precision of proprioceptive and kinesthetic feedback is greater. Second, on a touch screen the ratio between visual and motor space is always 1, so that the perceived visual and motor distances always match. There is no difference between the perceived local distances and the global distances in the mental representation of the canvas. Since Tan et al.'s *non-panning* UI is essentially a trivial case of a *panning* UI, an improved spatial memory performance can be observed for both. We believe this improved spatial memory performance also facilitates navigation tasks and leads to the observed increase in navigation performance for *panning* UIs from experiment 1.

5.2 Touch Navigation in ZUIs

In a ZUI, not only the absolute positions of objects change on the screen, but also the visual and motor distances between them: When a user zooms in by increasing the scale factor, the visual and motor distance between two objects on the touch screen increases. When a user zooms out, the visual and motor distance decreases. Thus, unlike in a *panning* UI, the changing scale factors in a ZUI lead to a changing ratio between the currently visible view with its local distances and the mental representation of the canvas with its global distances. Thus, the integration of the local view into the global representation has to take into account the current scale factor.

To overcome this gap during the integration process, users must rely on their visual perception. Although "pinching" or two-finger zoom gives users a proprioceptive and kinesthetic feedback on the current scale, we believe this feedback has not enough longevity to provide a reliable motor sense of the current scale factor. Instead, scale can only be inferred from the relative size of visual objects and landmarks, e.g., items or grids, or by consciously recalling previous zooming operations. Therefore, the benefit of a multi-modal sensation of distance cannot come into play. Accordingly, we did not observe any differences in spatial memory performance with touch vs. mouse in ZUIs.

6. IMPLICATIONS FOR DESIGN

1.) Designers can expect that users benefit from using touch instead of mouse input for *non-panning* and *panning* UIs. Users are able to remember the locations of items in the canvas more precisely and they need shorter navigation paths to move to currently invisible items. We observed this for a 30" touch screen and [26] indicates that this also true for a 18" touch screen. However, it remains open if this also true for smaller touch screens. If the screen size comes closer to that of a typical mouse pad, this effect could diminish with the lower intensity of proprioceptive and kinesthetic feedback. A bivariate study with mouse vs. touch where screen size is introduced as a second independent variable could provide interesting insights. Ideally, such a study could also serve to better quantify the effect and contribute to the formulation of a predictive quantitative model.

2.) For *zooming & panning* UIs, designers cannot expect that users benefit from touch instead of mouse in terms of spatial memory and shorter navigation paths. While we witnessed significantly smaller task completion times and users unanimously preferred touch, we could not observe more accurate spatial memory or better navigation performance for ZUIs. If designers intend to create UIs that are superior in these terms, they should consider if the amount of spatial data to provide is small enough to employ a *panning* UI instead.

3.) The dominant design of touch ZUIs with "pinching" or two-finger zoom does not come without cost. The closer the zoom target gets to the edges of the screen, the more difficulties users have with using two-finger zooming efficiently. Even after a considerable practice phase, our participants mostly preferred to pan the target into the center before zooming in. In terms of the length of ZUI navigation paths this is inefficient and the mouse performs significantly better. A potential solution could be ZUIs where the touch-sensitive area is greater than the viewport. Such a touch-sensitive rim could possibly afford more efficient two-finger zooming at the viewport's edges. It could also be worthwhile to search for new solutions for the problem of losing track of scale during zooming by introducing better visual cues or landmarks.

7. CONCLUSION & FUTURE WORK

In this paper, we have presented results from two experiments that enabled us to observe whether multi-touch instead of mouse input improves users' spatial memory and navigation performance in *panning* UIs and ZUIs. For panning UIs with touch input, we observed a 37% increase of spatial memory performance and a 29% increase in navigation performance. For ZUIs, we observed an unanimous user preference and a 17% improvement in task completion times for touch, but we could not observe better spatial memory and navigation performance.

We provided an explanation for this effect based on a multi-modal sensation of distance. While *panning* UIs can benefit from this sensation, this is not the case for ZUIs due to their changing scale factors. We also observed a 28% better navigation performance in terms of path lengths in a ZUI with mouse. We can attribute this to frequent problems that participants had with two-finger or "pinching" gestures for zooming into targets that were close to the screen's edges. Based on these results, we suggested first implications for design to guide interaction designers.

As future work, we suggest bivariate studies with mouse vs. touch and different screen sizes as independent variables to contribute to the formulation of a predictive quantitative model of the observed effects. In particular, further research on smaller screen sizes or touch pads would help to validate our findings.

More generally, we would like to motivate other researchers to conduct studies based on *Embodied Cognition* in HCI. Cognitive science provides strong evidence for the effects that body movement inevitably has on cognitive functions such as memory but also language use, social behavior and emotional attachment to other people or things [9]. HCI could strongly benefit from understanding and exploiting these effects when designing future user interfaces and gesture sets for gestural and touch interaction.

8. ACKNOWLEDGMENTS

9. REFERENCES

1. Baddeley, A. D. 1986. *Working Memory*. Oxford University Press, Oxford.

2. Bakker, N. H., Werkhoven, P. J. and Passenier, P. O. 1999. The Effects of Proprioceptive and Visual Feedback on Geographical Orientation in Virtual Environments. *Presence: Teleoper. Virtual Environ. 8*, 1, 36-53.

3. Darken, R. P. and Peterson, B. 2001. Spatial orientation, wayfinding, and representation. In *Handbook of Virtual Environment Technology*, 1-21.

4. De Ruiter, J.-P. 1998. *Gesture and speech production*. PhD Thesis, MPI for Psycholinguistics.

5. Dourish, P. 2001. *Where the Action Is: The Foundations of Embodied Interaction*, MIT Press.

6. Ebert, A., Deller, M., Steffen, D. and Heintz, M. 2009. "Where Did I Put That?" – Effectiveness of Kinesthetic Memory in Immersive Virtual Environments. In *Universal Access in Human-Computer Interaction*. Springer.

7. Forlines, C., Wigdor, D., Shen, C. and Balakrishnan, R. 2007. Direct-touch vs. mouse input for tabletop displays. In *Proc. CHI ' 07*, ACM, 647-656.

8. Furnas, G. W. and Bederson, B. B. 1995. Space-scale diagrams. In *Proc. CHI '95*, ACM, 234-241.

9. Gibbs, R. W. 2006. *Embodiment and Cognitive Science*. Cambridge University Press.

10. Guiard, Y. and Beaudouin-Lafon, M. 2004. Target Acquisition in Multiscale Electronic Worlds. *International Journal of Human-Computer Studies 61*, 6, 875-905.

11. Imaz, M. and Benyon, D. 2007. *Designing with Blends: Conceptual Foundations of Human-Computer Interaction and Software Engineering*. MIT Press.

12. Jul, S. and Furnas, G. W. 1998. Critical zones in desert fog: aids to multiscale navigation. In *Proc. UIST '98*, ACM, 97-106.

13. Kelso, J. A. S. 1995. *Dynamic Patterns : The Self-Organization of Brain and Behavior*. MIT Press, Cambridge, MA, USA.

14. Klemmer, S. R., Hartmann, B. and Takayama, L. 2006. How bodies matter: five themes for interaction design. In *Proc. DIS '06*, ACM, 140-149.

15. Leifert, S. 2011. The influence of grids on spatial and content memory. In *CHI EA '11*, ACM, 941-946.

16. MacKenzie, I. S. and Buxton, W. 1992. Extending Fitts' law to two-dimensional tasks. In *Proc. CHI '92*, ACM.

17. Malacria, S., Lecolinet, E. and Guiard, Y. 2010. Clutch-free panning and integrated pan-zoom control on touch-sensitive surfaces: the cyclostar approach. In *Proc. CHI '10*, ACM, 2615-2624.

18. Meyer, S., Cohen, O. and Nilsen, E. 1994. Device comparisons for goal-directed drawing tasks. In *CHI EA '94*, ACM, 251-252.

19. Micire, M., Schedlbauer, M. and Yanco, H. 2007. Horizontal Selection: An Evaluation of a Digital Tabletop Input Device. In *Proc. AMCIS 2007*, AIS.

20. Morsella, E. and Krauss, R. M. 2004. The Role of Gestures in Spatial Working Memory and Speech. *The American Journal of Psychology 117*, 3, 411-424.

21. Moscovitch, M., Kapur, S., Kohler, S. and Houle, S. 1995. Distinct neural correlates of visual long-term memory for spatial location and object identity. *Proc Natl Acad Sci U S A. 92*, 9, 3721-3725.

22. Perlin, K. and Fox, D. 1993. Pad: an alternative approach to the computer interface. In *Proc. SIGGRAPH '93*. ACM, 57-64.

23. Robertson, G., Czerwinski, M., Larson, K., et al. 1998. Data Mountain : Using Spatial Memory for Document Management. In *Proc. UIST '98*, ACM, 153-162.

24. Robertson, G., Dantzich, M. v., Robbins, D., Czerwinski, M. et al. 2000. The Task Gallery: a 3D window manager. In *Proc. CHI '00*. ACM, 494-501.

25. Schmidt, S., Nacenta, M. A., Dachselt, R. and Carpendale, S. 2010. A set of multi-touch graph interaction techniques. In *Proc. ITS '10*, ACM, 113-116.

26. Tan, D. S., Pausch, R., Stefanucci, J. K. and Proffitt, D. R. 2002. Kinesthetic cues aid spatial memory. In *CHI EA '02*, ACM, 806-807.

27. van Asselen, M. 2005. *The neurocognitive basis of spatial memory*. PhD Thesis, Helmholtz Institute, Utrecht University, Utrecht**.

28. Wesp, R., Hesse, J., Keutmann, D. and Wheaton, K. 2001. Gestures Maintain Spatial Imagery. *The American Journal of Psychology 114*, 4, 591-600.

29. Wexler, M., Kosslyn, S. and Berthoz, A. 1998. Motor processes in mental rotation. *Cognition 68*, 77-94.

30. Wigdor, D. and Wixon, D. 2011. *Brave NUI World : designing natural user interfaces for touch and gesture*. Morgan Kaufmann.

31. Wobbrock, J. O., Morris, M. R. and Wilson, A. D. 2009. User-defined gestures for surface computing. In *Proc. CHI '09*, ACM, 1083-1092.