Going Beyond the Surface: Studying Multi-Layer Interaction Above the Tabletop

Martin Spindler[1], Marcel Martsch[2], Raimund Dachselt[1]
[1]User Interface & Software Engineering Group
[2]Department of Vocational Education and Human Resources Development
University of Magdeburg, Germany
{spindler, marcel.martsch, dachselt}@ovgu.de

ABSTRACT

Lightweight spatially aware displays (Tangible Magic Lenses) are an effective approach for exploring complex information spaces within a tabletop environment. One way of using the 3D space above a horizontal surface is to divide it into discrete parallel layers stacked upon each other. Horizontal and vertical lens movements are essential tasks for the style of *multi-layer interaction* associated with it. We conducted a comprehensive user study with 18 participants investigating fundamental issues such as optimal number of layers and their thickness, movement and holding accuracies, and physical boundaries of the interaction volume. Findings include a rather limited overall interaction height (44 cm), a different minimal layer thickness for vertical and horizontal search tasks (1 cm/4 cm), a reasonable maximum number of layers depending on the primary task, and a convenience zone in the middle for horizontal search. Derived from that, design guidelines are also presented.

Author Keywords

Multi-Layer Interaction; Spatially Aware Displays; Above the Tabletop; Tangible Magic Lens; User Study.

ACM Classification Keywords

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

INTRODUCTION

With ever increasing complex information spaces, technology and user interfaces are required to explore them and to focus attention to the relevant information. Magic lenses, first introduced in 1993 for graphical user interfaces [1], have been proven to be one promising solution. Since then, they were made tangible as additional active or passive displays being used on tabletops [9, 20], above tabletops [8, 9, 15, 16], in the air [4] or as Mixed Reality lenses [2, 3]. Typically, these setups include one or more larger (interactive) surfaces. In this work, we will focus on tangible magic lenses in combination with horizontal displays (tabletops).

With spatially aware tangible magic lenses, the physical interaction space is no longer limited to an interactive two-dimensional (2D) surface, but extended to the third dimension (3D). For this purpose, the 3D space above the work surface can be divided into physically separated horizontal layers [15]. They can be explored by moving spatially aware lightweight displays through the "air" that constantly provide immediate visual feedback regarding current layers.

Until now, many of the aforementioned magic lens approaches present "cool" technologies and interesting interaction opportunities, whereas a careful investigation of the actual physical characteristics and limitations is still missing. Few studies have addressed multi-layer interaction above tabletops, with [17] being a rare exception and focusing on pen interaction only. Although multi-layer tangible lens interaction has proven to be a powerful tool for a broad spectrum of application domains, such as the exploration of spatial information spaces [15], information visualization [16], and collaborative face-to-face sketching and brainstorming [14], only little is known about appropriate boundaries for the physical interaction space, adequate amounts of layers, or minimum layer thicknesses – especially with respect to typical interaction tasks with layers.

This paper aims at filling this gap by presenting a comprehensive study which we conducted to find answers regarding these questions. In particular, we were interested in:

- How accurate do users actually interact in the space above the tabletop (with respect to a particular task)?
- How to design the interaction space in terms of layer thicknesses, number of layers, and optimal lower and upper heights for the physical interaction volume?

We conducted a comparative user study with 18 participants. For this purpose, a simple tangible magic lens system with limited features was implemented. As the participants accomplished several tasks with the system, more than five hours of performance data was collected. Beyond that, we gathered detailed self-reported data directly after each task. Our findings address valuable insight into physical precision as well as practical guidelines on how to extend a (tabletop) system with multi-layer interaction techniques.

The remainder of this paper is structured as follows. After presenting background/related work, we describe the goals and the scope of the study in more detail. This is followed

CHI'12, May 5–10, 2012, Austin, Texas, USA.

by a detailed account of the method, results, and discussion in the respective sections, followed by conclusions.

BACKGROUND: TANGIBLE DISPLAYS

In this section, a brief overview of common tangible mobile display approaches is presented and set into context of the used reference surface. This is done in order to highlight technical environments and potential usage scenarios that are likely to benefit from multi-layer exploration techniques as will be discussed in the next section.

Tangible Display Interaction (With Respect to a Surface)

Merging the digital world with the physical world is the vision of ubiquitous computing, as defined by Weiser [21]. This concept was adapted by Ishii's and Ullmer's tangible user interfaces (TUIs) [7], where interaction with digital information is provided through physical manipulation of real-world objects. Inspired by the notion of see-through interfaces [1], these objects can also be spatially aware physical displays (e.g., mobile phones) that serve as tangible magic lenses into the virtual world. One of the first mobile displays for ubiquitous usage has been proposed by Fitzmaurice, who presented a spatially aware palmtop computer for exploring 3D-situated information spaces for revealing virtual information associated with arbitrary objects in an office environment [4].

Horizontal Reference Surfaces. Contrary to Fitzmaurice's approach, the metaDESK project by Ullmer and Ishii [20] makes use of a horizontal reference surface only: a tabletop. Here, users can freely navigate through polygonal 3D models by moving an arm-mounted LCD display through the space above the tabletop that is also responsible for providing contextual graphical information. Hirota and Saeki [5] as well as Konieczny et al. [10] provide technical solutions for 3D volume slicers using tangible magic lenses, but even though both do not make use of a tabletop, they use the (horizontal) floor as reference.

Vertical Reference Surfaces. In terms of vertical reference surfaces, Sanneblad and Holmquist [13] used mobile phones to magnify details of a larger contextual image that was shown on a wall display, but without using the space in front of the wall for interaction purposes, such as controlling the level of detail. Besides providing a technical solution for projecting video content on spatially aware projection screens, Lee et al. [11] suggested using spatially aware displays to explore data sets with multiple information layers in front of a vertical display.

User's Body as Reference Surface. The peephole displays by Yee [22] explore the virtual information space wrapped around a user by taking the distance between user and PDA into account. This is used for zooming and layering operations in the context of desktop applications, such as calendar, web browsing or geographical maps.

Lightweight Tangible Display Solutions

In various cases, active display solutions, such as LCD panels, do not provide suitable form factors for a seamless integration into a tangible magic lens system. This is because active displays are often too heavy, too thick, too big, too rigid, and sometimes even too expensive if many devices are needed. This lack of technology motivated researchers to come up with a variety of lightweight display solutions, with many of them being designed for horizontal table(top)s. Most of these solutions use a passive approach, i.e., image content is projected onto spatially aware projection screens in one or another form. Lee *et al.* [11], Holman *et al.* [6], and Spindler *et al.* [15] presented such systems that all use ceiling-mounted projectors for projecting dynamic video content onto optically tracked paper-like displays. In contrast, SecondLight by Izadi *et al.* [8] is technically more complex in that it is based on electronically switchable diffusers. It supports dual back-projections on both a tabletop surface and tangible lenses above it. For UlteriorScape, Kakehi and Naemura [9] use a special projection foil that changes its translucency depending on the projection angle and can be used to simultaneously back-project different image content onto a tabletop surface and a tangible lens, respectively. Another promising technological development are organic LEDs (OLED) that one day might unify the advantages of active and passive displays.

BACKGROUND: MULTI-LAYER INTERACTION

As demonstrated in the previous section, a vivid research community is working in the field of tangible displays. Although the aforementioned systems show that a broad spectrum of technical solutions and interaction techniques has been created for tangible displays, only few of these solutions truly utilize the 3D space above a reference surface for the purpose of *multi-layer interaction*.

A multi-layer space is a batch of discrete horizontal layers that are stacked one upon each other with each layer consuming a distinct height in physical 3D space. In this way, the multi-layer stack features a physical extent into the third dimension that can be used as a "habitat" for spatially aware tangible displays. Multi-layer stacks have been proven to be an effective interaction metaphor with a broad field of applications, e.g., geometric zooming of large images, semantic zooming of node-link diagrams, exploration of space-time data, and layered geo maps – such as demonstrated by our PaperLens [15] project. This is a lightweight paper-based display solution that enables users to explore various types of information spaces by moving paper-like screens through the physical space above the tabletop (see Figure 1). For this purpose, layers of a virtual data space are mapped onto physical layers of a multi-layer stack that remain parallel to a horizontal table surface. In such setups, three basic tasks play a dominant role (see Figure 2):

- **Holding** (keeping a display at a fixed position)
- **Vertical Search** (moving a display upwards/downwards)
- **Horizontal Search** (moving a display at the same height)

One example for information spaces that can be explored with these basic interaction tasks are temporal data sets (e.g., videos) that users can browse through by lifting and lowering a mobile display (vertical search) in fast or slow

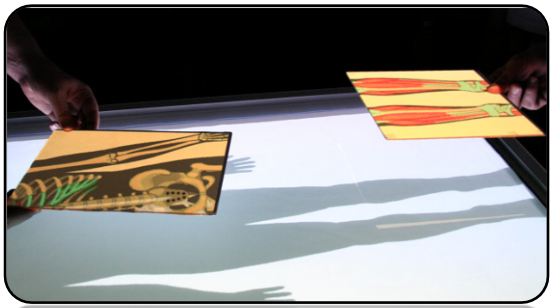


Figure 1: Exploring a multiple layer information space by lifting and lowering a tangible magic lens above a digital table that serves as an additional visual context.

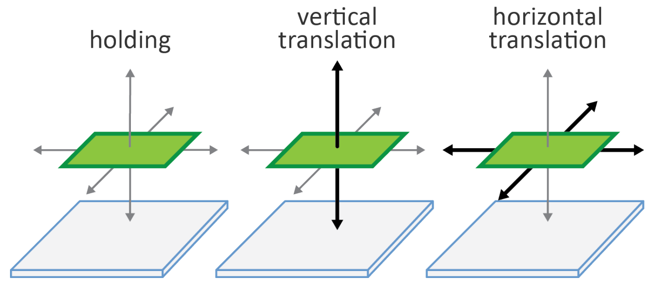


Figure 2: Three basic tasks for tangible magic lens-based multi-layer interaction above a horizontal surface that were tested in terms of physical accuracy in our study.

motion, or they can look at a particular still image by holding the lens at a particular height (holding task). Another example is the *layered information space* [15] that enables users to explore several 2D information layers (see Figure 1), e.g., anatomic images by moving a display horizontally above the table surface (horizontal search). Here, selecting a layer is accomplished by vertical search. For all these types of information spaces, the tabletop usually serves as a contextual reference that provides a graphical overview to the detail that is displayed on a mobile display.

In a concluding formative user study [15], we confirmed that their exploration techniques are easy to use and intuitive to work with, i.e., users were able to instantly accomplish given tasks without being taught how to actually interact with them, making these techniques ideal not only for public installations. We believe that such techniques will play a vital role in future tabletop research.

Evaluating Multi-Layer Tangible Display Interaction

Evaluating specific tangible display systems has been the goal of a few research projects only. Besides our own work [15], a notable example is Oh and Hua's evaluation on form factors of tangible magic lenses [12]. In a comparative user study, they tested various lens aspect ratios and sizes. They came to the conclusion that the aspect ratio of a lens plays a more important role for smaller lenses than for larger ones and that lens sizes are more dominant in impacting the user performance. However, to the best of our knowledge, no evaluation of the particular needs of the *multi-layer interaction aspect* of such systems has been done before.

Multi-Layer Pen Interaction above Digital Tables

In [17], Subramanian *et al.* presented and evaluated pen interaction above a digital table for the purpose of multi-layer interaction. Although this is similar to the previously introduced PaperLens concept, it does not couple input and output within the same physical device. Instead, Subramanian *et al.* designed and tested a set of 3D pen gestures that can be used to interact with a multi-layer stack. In a pilot study with 5 users (all sitting) they estimated a minimal layer thickness of 4 cm to prevent frustration due to accidentally changing layers, which is close to our findings. In order to reduce fatigue, they argued for limiting the maximum height above the work surface to about 16 cm, thus, resulting in a maximum number of 4 layers. They also noted that when users only navigated through layers (similar to our "vertical search") these could be made thinner than layers where selection gestures were performed (similar to our "horizontal search"). At the same time, layers closer to the work surface could be made thinner than others.

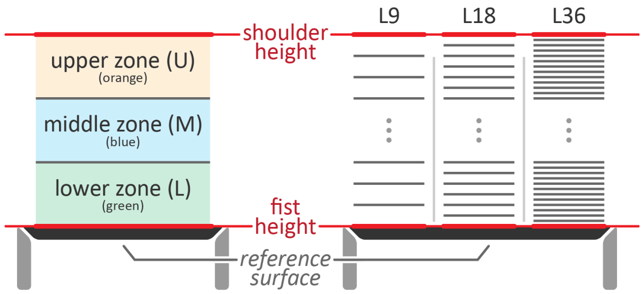
GOALS AND SCOPE OF THE STUDY

While the long-term goal of our work is to fully understand general affordances of tangible magic lens interaction above the tabletop, the focus of this study is the thorough investigation of the special requirements of multi-layer interaction and its physical boundaries. In particular, we are interested in the three basic interaction tasks discussed in the previous section: holding, vertical search, and horizontal search. These operations shall be examined in terms of the accuracy at which users can execute them in physical space. In addition, design principles concerning layer thickness, quantity and convenience volumes for multi-layer interaction shall be derived as a sound foundation for designing future magic lens systems.

Hence, the main focus of this work is on designing, conducting and evaluating a user study that addresses the following issues:

- Finding possible physical lower and upper thresholds that vertically limit the interaction space
- Finding exact measures on how accurately users accomplish vertical & horizontal search and hold operations
- Gaining an understanding about layout details for a multi-layer stack and its dynamic exploration

Many possible variables impact the interaction with tangible magic lenses, e.g., two- vs. one-handed use, tilting, display sizes and weights, device thicknesses, and display frame widths. In order to maintain a manageable study design while still providing ecological validity, we needed to make some tradeoffs. We therefore restricted our investigations to the case of a user standing in front of a horizontal tabletop and a two-handed use (as opposed to a seated setup and one-handed use). We also decided to disregard lens tilting. This was because our main focus was on stacked multi-layer spaces, where lens tilting often plays a minor



(a) Interaction Zones (b) Layer Subdivisions

Figure 3: Conceptual design of interaction zones (a) and layer subdivisions (b) as used for the study.

role (e.g., for the ease of comfort) or is used as additional modality that only adds further degrees of freedom. In informal pre-tests we observed no practical influence of display sizes for a two-handed use (we tested paper formats of A6 to A4), so we skipped this variable. Another example is display weights. We assume that in the near future a new generation of lightweight displays (e.g., based on OLEDs) will dramatically improve form factors of tablets, smart phones and alike, so that display weights are likely to come closer to the weight we used in our study (about 190g). A high priority was the avoidance of additional disturbing effects, such as a visual context displayed on the table surface. In order to diminish cognitive load from participants, mainly stimuli were considered that address low-level processing. Based on the feature integration theory [19], we designed single-feature search tasks with targets that are clearly visually separated from distractors. This is achieved by using discriminative features (pop-out effect), for instance strong contrasts of light, shade and color, which can be preattentively processed.

METHOD

Participants

Eighteen students and staff members from our department (4 female, 14 male) participated in the user study. Their age ranged from 20 to 32 (M = 26.9, SD = 3.2). The average body height was 178.9 cm (SD = 9.4). All were daily users of computers and had advanced knowledge in the fields of computer graphics, simulation or image processing.

Design & Tasks

In order to test the accuracy at which users perform the three basic interaction tasks (holding, vertical search and horizontal search), a minimalistic layered information space was designed that consists of vertically stacked randomized integer numbers (between "1" and "99") representing the search targets unknown to participants. The independent variable was the number of layers that directly correlates with layer thicknesses. As pre-tests suggested, we expected this parameter to impact the accuracy of task performances. Three different levels of layer subdivision were used: 9 layers (L9), 18 layers (L18) and 36 layers (L36), as illustrated in Figure 3b. The use of multiples of "3" was mainly motivated by the three interaction zones as explained later (see Figure 3a). We used a within-subject design.

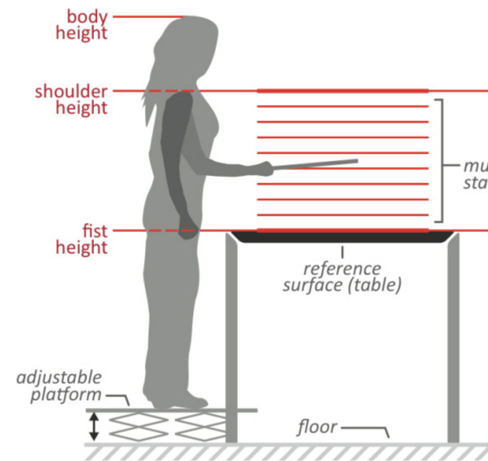


Figure 4: Principle setup of the study. The platform's height was adjusted in order to match fist height and table height.

Compensating Body Sizes

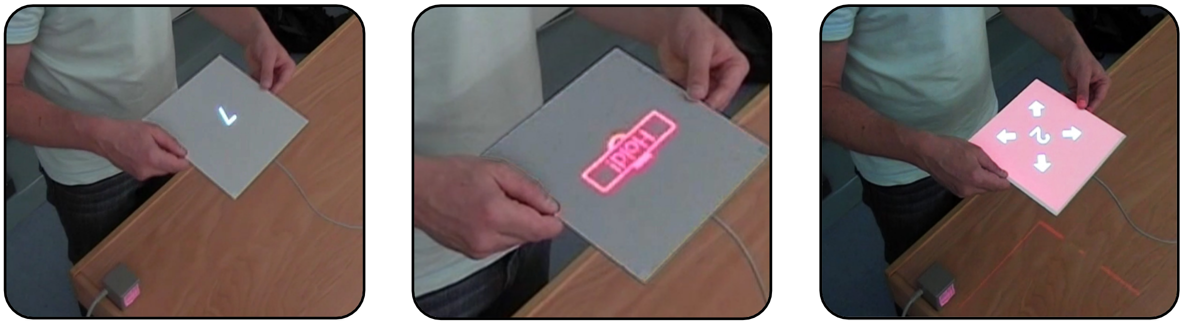
To compensate for different body sizes of participants (e.g., longer arm lengths), we decided to use the *fist height* and the *shoulder height* as a priori limits for the lower and upper boundaries of the interaction volume, with the fist height being the distance of fist to floor when the arm hangs loosely. These values could be adjusted for each participant by using a platform (see Figure 4). The use of fist and shoulder heights was motivated by ergonomics literature – in our case [18] – and the outcome of a small pre-test: a table surface that is too low forces users to bend the upper body, whereas holding lenses higher than the shoulder quickly leads to fatigue. We used the shoulder-fist distance to define the overall height of the interaction volume for each participant (M = 691.11 mm, SD = 54.33). Due to constant subdivision of layers, the layer thicknesses varied relative to individual heights of the physical interaction volume, with L9 (M = 76.80 mm, SD = 6.04), L18 (M = 38.40 mm, SD = 3.02) and L36 (M = 19.2, SD = 1.51).

Vertical Search: Distributing Targets within the Layer Stack

For vertical search tasks, a sequence of targets was defined that was distributed over the overall volume height. For this purpose, we distinguished between three equispaced interaction zones: the lower (L), middle (M) and upper (U) zone (see Figure 3a). We chose three random layers from each zone as a search target. This allowed us to design six different combinations of vertical search tasks: L-L, M-M, U-U, L-M, L-U, and M-U, which was further multiplied by the direction, e.g., L-M vs. M-L or L_1-L_2 vs. L_2-L_1. Since at least two variants were to be tested for each combination, we finally came up with 6 × 2 × 2 = 24 vertical search tasks. For all participants and layer subdivisions (L9, L18, L36) the same sequence was used.

Horizontal Search: Distributing Targets within a Single Layer

For horizontal search tasks (panning), two search targets were randomly placed in each horizontal layer. To avoid



(a) Vertical Search Task (b) Holding Task (c) Horizontal Search Task

Figure 5: Apparatus and three interaction tasks. Lens content appears brighter than during the user study, e.g., for holding tasks the background was usually deep black. This is due to light conditions during taking photographs. Interaction was two-handed.

that participants have to bend their backs, we chose a conservative working radius of 40 cm (150 cm tall women have a shoulder-finger reach of about 60 cm, see [18]).

Overall Task Cycle

According to this design, for each level of layer subdivision (L9, L18, and L36) a randomized sequence of 24 holding tasks (3 sec each), 24 vertical search tasks, and 9 horizontal search tasks was defined. The same sequence was used for each participant. See Figure 6 for an example height plot.

Procedure

For each participant, the following order of test parts was maintained: (1) introduction part, (2) interaction part, and (3) assessment part. Users spent on average 40 minutes performing all parts.

(1) Introduction Part

After participants completed a brief questionnaire soliciting demographics and computer usage information, the body, shoulder and fist heights were measured. Thereon, the platform's height was adjusted to match the table height with the fist height, see section "Compensating Body Sizes" above. Then, the main task was explained verbally in a standardized way by reading out aloud from a sheet of paper. To ensure that all participants had perfectly understood all relevant aspects, they were invited to perform a few exercise trials without collecting data until they felt confident in handling the task (never longer than two minutes).

(2) Interaction Part

Participants were asked to complete the interaction part in three cycles – one for each level (L9, L18 and L36). The order of levels was counter-balanced. For each cycle, participants were to hold the tangible display in front of their body in order to explore a stack of white random integer numbers by vertically lifting and lowering it (*vertical search task*, see Figure 5a) until they found the search target (a single red-colored number). Users were instructed to read out aloud the red number[1]. After a "beep" sound and a visual feedback on the display (red "Hold!" label, see Figure 5b) they had to hold the tangible lens stable at the very same layer (*holding task*) for 3 seconds until the "Hold!" label disappeared and another "beep" sound indicated that the participant should continue searching for the next red number. From time to time (at 9 different layers), users had to perform a sequence of *horizontal search tasks*. For this purpose, the red number was replaced by a slightly smaller white number above a red background with white arrows hinting that the participant should now continue searching horizontally until two randomized search targets (white numbers) within the same layer were found that had to be read out aloud (see Figure 5c). Whenever participants accidentally left the layer, they got immediate visual feedback (the lens turned dark). In such cases, users had to manually find back to the layer by lifting/lowering the lens until it turned red again. Neighboring layers did not contain any numbers. This prevented reports of false numbers. By reaching the starting position again (white number with arrows, see Figure 5c), the horizontal search task ended, followed by the next vertical search task.

(3) Assessment Part

Self-Report. At the end of each cycle in the interaction part, participants were asked to rate their agreement on a 7-point Likert scale from "1" ("do not agree at all") to "7" ("completely agree") to several items in a questionnaire. These items addressed the usability (effectiveness, efficiency and user satisfaction) regarding holding tasks, vertical search tasks and horizontal search tasks.

Perceived Interaction Zones. After all tasks of the interaction part had been completed, participants were asked to fill out another questionnaire that was very similar to the one from the interaction part, but this time with respect to how they perceived the interaction in each of the three interaction zones (see Figure 3a). In order to utilize spatial memory, participants could play around with the tangible lens once again while the lens was tinted according to the scheme depicted in Figure 3a (no targets were displayed).

Preferred Boundaries. In the final and additional part, participants were asked to define – from their point of view – the ideal vertical lower and upper boundaries of the interac-

[1] Filtering a single red-colored number out of many white-colored ones is a preattentive process. Although this does not apply to reading two-digit numbers, we still consider it to be fast enough to not significantly affect our experiment.

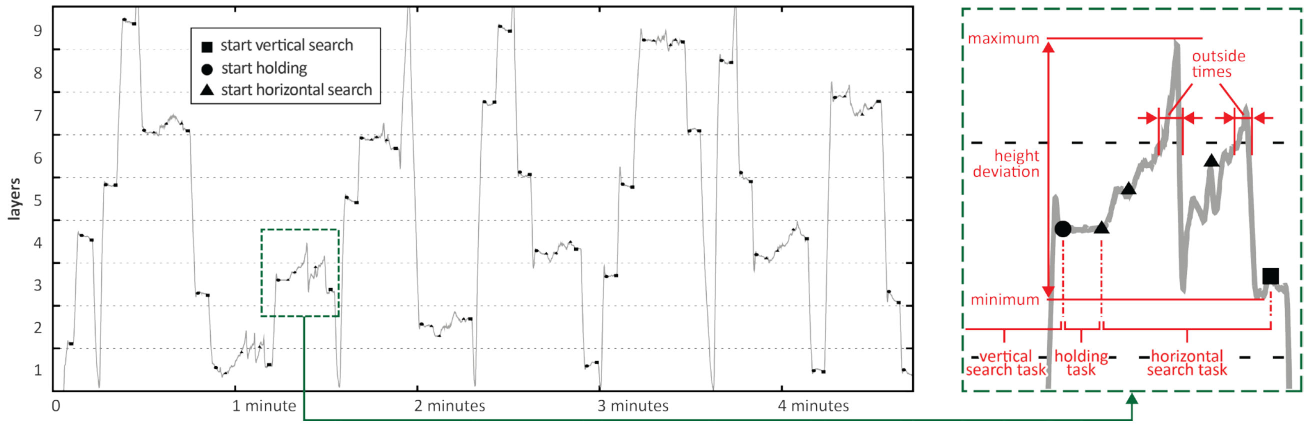


Figure 6: A typical plot of a cycle for level L9 showing height measures of the tangible lens over time. In the detailed view at the right side, the derivation for "outside times" and "height deviations" (as used in the data analysis) is illustrated.

tion volume with respect to the table surface. This was accomplished by holding the lens at the favored lower and upper heights above the table and then saying "okay".

Apparatus

The user study was conducted in a dark and quiet lab environment. The technical setup used for the experiments consisted of a horizontal table (reference surface), a piece of a rectangular cardboard of size 21.5 × 21.5 cm (tangible display), and a ceiling-mounted video projector directly above the table. In order to guarantee a high degree of spatial precision we opted for a magnetic-based tracking approach (Polhemus Fastrak) that enabled us to limit the spatial error to 0.3 mm within a working volume of 70 × 70 × 70 cm³. We decided against projecting additional visual context onto the table. Instead, the table surface primarily served as horizontal spatial reference.

On the software side, our system was implemented with C#. We opted for a client/server model with the server being responsible for lens tracking and the client being responsible for displaying and application (user study tasks). Communication between client and server was achieved by a simple self-tailored UDP/IP-based protocol. It allows for sending and receiving a stream of 4 × 4 transformation matrices that describe the current 3D position and orientation of the tangible lens' center.

Collected Data

For later evaluation, the stream of transformation matrices was continuously logged into a file at a rate of about 33 samples per second. Along with this, additional marker timestamps were recorded (e.g., "start vertical search", see Figure 6). For this purpose, the study leader had to manually press the ENTER key every time a number was read out aloud to trigger the next task. To ensure equal times for the holding task, the system automatically generated a "beep" sound after 3 seconds of holding.

RESULTS

In this section, results from the experiment are presented in three parts. First, we report on the basic analysis of performance data. Second, we provide further analysis with respect to the three interaction zones. And third, we investigate data from questionnaires.

Performance Data

In order to analyze collected performance raw data, the following types of derived data have been extracted:

- *Task completion time* (in sec) that we used as a distinct measure of performance for vertical and horizontal search tasks. For holding tasks this time is constant (24 × 3 sec = 72 sec) and thus not relevant.
- *Total completion time* (in sec) is an aggregated measure of performance that sums up the times spent on horizontal and vertical search tasks.
- *Height deviation* (in mm) is the minimum and maximum displacement for horizontal search and holding tasks used as a measure of accuracy (see Figure 6). By using MIN/MAX operators, we decided on a conservative but simple statistical tool. Choosing other tools, such as MEAN/VARIANCE, would only further narrow our findings.
- *Outside time* (in sec), i.e., amount of time that participants unintentionally spent outside of a layer while performing a holding or horizontal search task (see Figure 6). We used this as a measure of error.

We also analyzed tilting of lenses. However, in contrast to our assumptions, we did not find any significant effect and thus will omit the discussion about it.

Statistical Methodology

Collected data (performance measures and self-report data) was analyzed with a *repeated measurement ANOVA*. For all ANOVAs, p-values were Greenhouse–Geisser corrected. When main effects were significant, Bonferroni corrected p-values are reported for *post hoc* comparisons (t-test, two-tailed). For testing the relationship between layer subdivision and accuracy of horizontal search tasks, correlation coefficients (Pearson r) were computed. For statistical tests, the a priori threshold of $\alpha = .05$ was used. If not stated otherwise, results are in sec; "ns" stands for "not significant".

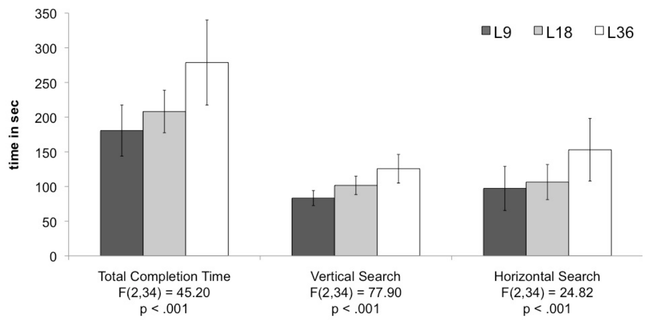


Figure 7: Task completion times of horizontal and vertical search tasks for L9, L18 and L36. Both times sum up to the total completion time. Error bars show standard deviations.

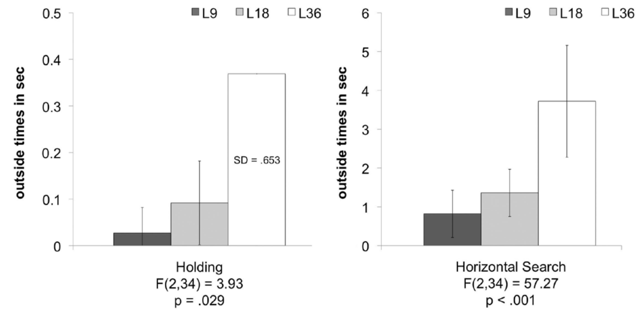


Figure 8: Outside times of holding tasks and horizontal search tasks for L9, L18 and L36. Error bars represent standard deviations.

Basic Analysis of Performance Data

In the following, a basic analysis of performance data is presented. See Figure 7 for completion times.

Total Completion Times

Regarding total completion times, we found a significant main effect of layer subdivisions ($F_{(2,34)} = 45.2$, $p < .001$). Completion times for L9 were significantly shorter than for L18 ($t_{(17)} = 3.95$, $p = .001$). The same holds for the contrast L18 vs. L36 ($t_{(17)} = 12.73$, $p < .001$).

In summary, total completion times show that with decreasing thickness of layers participants needed significantly more time for completing horizontal and vertical search tasks. In the following, further results are presented for each task.

Holding Task

The analysis of height deviations revealed no significant main effect of layer subdivisions ($F_{(2,34)} = 2.41$, ns). However, the results of height deviations depict that the precision of holding tasks generally improves slightly from L9 (M = 7.18 mm, SD = 2.33) over L18 (M = 7.00 mm, SD = 2.30) to L36 (M = 6.23 mm, SD = 1.44).

Analysis of outside times (see Figure 8, left) revealed a significant main effect of layer subdivisions ($F_{(2,34)} = 3.93$, $p = .029$). Pairwise comparisons showed that participants accidentally left target layers significantly more frequently ($t_{(17)} = 3.00$, $p < .024$) for L18 than for L9. Most layer crossings happened for L36, but neither differed significantly from L9 ($t_{(17)} = 3.00$, ns) nor from L18 ($t_{(17)} = 1.72$, ns), which is due to a high standard deviation.

Vertical Search Task

In terms of the completion time, we observed a significant main effect of layer subdivisions for vertical search tasks ($F_{(2,34)} = 77.9$, $p < .001$). Participants needed significantly less time ($t_{(17)} = 6.33$, $p < .001$) for L9 when compared with L18. In the same way, differences of completion times between L18 and L36 are significant ($t_{(17)} = 6.14$, $p < .001$).

Horizontal Search Task

For horizontal search tasks we found a significant main effect of completion times of layer subdivisions ($F_{(2,34)} = 24.82$, $p < .001$). Although pairwise comparisons show that there is no significant effect ($t_{(17)} = 1.56$, ns) for L9 vs. L18, participants needed significantly ($t_{(17)} = 4.90$, $p < .001$) more time for L36 than for L18.

We observed no main effect of layer subdivisions for height deviations ($F_{(2,34)} = 2.25$, ns). For outside times we found a significant main effect of layer subdivisions ($F_{(2,34)} = 57.27$, $p < .001$). The participants crossed layer boundaries less often for L9 than for L18 and L36, see Figure 8 (right).

To test the interrelation of total completion times (TCT) against height deviations (HD) and outside times (OT), several correlation coefficients were computed that show a significant relation of TCT × HD (L9: r = .526, p < .001; L18: r = .495, p < .001; L36: r = .643, p < .001) and TCT × OT (L9: r = .682, p < .001; L18 = .871, p < .001; L36: r = .903, p < .001). This shows for all layer subdivisions that participants who accomplished tasks accurately were the ones with a good overall performance and vice versa.

Further Analysis Regarding Interaction Zones

In this section, further analysis of performance data is presented with regards to the lower, middle and upper interaction zone (see Figure 3a). For this purpose, we rearranged already analyzed performance data with respect to these zones. In particular, we looked into *height deviations* and *outside times* for holding and horizontal search tasks separately for each layer subdivision (within-subject factor).

Holding Task

We did not find any significant main effect of the three interaction zones for height deviations or for outside times.

Horizontal Search Task

For L9 and L18 we observed no significant main effects with regard to the three interaction zones, neither for task completion times and height deviations, nor for outside times. In contrast, for L36 we revealed significant effects of task completion times ($F_{(2,34)} = 12.08$, $p = .001$), height deviations ($F_{(2,34)} = 4.10$, $p = .049$) and outside times ($F_{(2,34)} = 10.45$, $p = .003$). For the following discussion, see Figure 9.

For task completion times, pairwise comparisons show that the middle zone is less demanding than the lower ($t_{(17)} = 4.56$, $p = .001$) and upper zone ($t_{(17)} = 4.13$, $p = .002$).

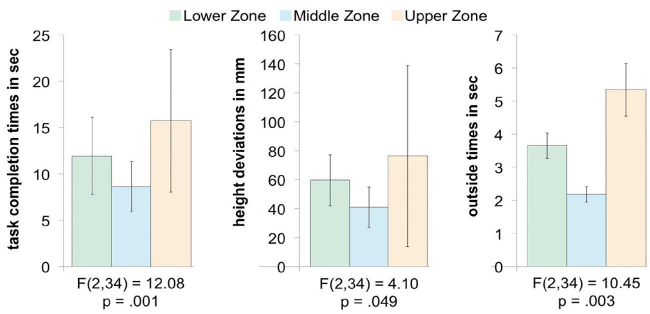


Figure 9: Performance data for the horizontal search task (only L36) broken down by the lower, middle and upper interaction zone (see Figure 3). Error bars represent standard deviations.

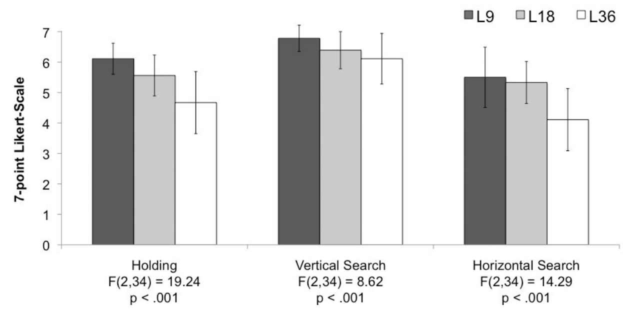


Figure 10: Usability ratings for holding task, vertical search task and horizontal search task with respect to single layer subdivisions (L9, L18, and L36). Error bars represent standard deviations.

Height deviations of the middle zone are slightly (but not significantly) smaller than the ones of the lower and upper zone. This indicates that the middle zone takes on a special role for horizontal search tasks. This is also supported by outside times that are significantly longer for the upper zone, when compared to the lower ($t_{(17)} = 4.63$, $p = .001$) and middle zone ($t_{(17)} = 3.96$, $p = .003$).

Questionnaires & User Preferences

In the following, the results of user ratings concerning usability are presented as agreement values on a 7-point Likert scale. For a brief summary, see Figure 10.

Holding Task

Ratings on holding tasks differ depending on layer subdivisions ($F_{(2,34)} = 19.24$, $p < .001$). Pairwise comparisons show that L9 is rated best when compared with L18 and L36. Individual comparison shows that contrasts are significant for all combinations.

Vertical Search Task

With respect to the ratings for vertical search, we found a significant main effect of layer subdivisions ($F_{(2,34)} = 8.62$, $p = .001$). Pairwise comparisons show that L9 is assessed as significantly easier than both L18 and L36. For L18 vs. L36, this effect is not significant.

Horizontal Search Task

Ratings for horizontal search tasks have a significant main effect of layer subdivisions ($F_{(2,34)} = 14.29$, $p < .001$). Pairwise comparisons reveal that L36 is rated as significantly more difficult than L9 and L18.

Perceived Interaction Zone

We observed a significant main effect of the three interaction zones ($F_{(2,34)} = 25.20$, $p < .001$). The usability of interaction for the middle zone (M = 6.46, SD = .62) was rated significantly better when compared with the lower (M = 6.04, SD = .62) and upper zone (M = 4.47, SD = 1.38).

Preferred Boundaries

The analysis of subjective boundaries of the interaction volume results in a preferred lower limit (M = 68.28 mm, SD = 51.01) and a preferred upper limit (M = 508.77 mm, SD = 78.45) above the table surface.

DISCUSSION

In this section, we will reflect on our results and provide answers to the questions of minimal layer thicknesses, vertical boundaries for the physical interaction zone, and maximum number of layers. Based on these findings, we will derive design guidelines that will be presented at the end of this section.

Layer Thicknesses & Accuracy

In general, we can show that a diminishing layer thickness correlates with decreasing test performances. This applied in particular to the vertical and horizontal search task. In contrast, the holding task remained mostly unaffected.

Holding Task. Compared to other tasks, holding was performed most accurately. On average, the height deviation for holding tasks was approx. 7 mm. By taking into account a standard deviation of approx. 2 mm, this leads to a rather conservative **minimum layer thickness** of approx. 9 mm or **about 1 cm**. This value is even substantially lower than the average thickness of L36 (M ≈ 19 mm), which is most likely the reason for why the control variable "layer subdivision" only marginally affected the error measure "outside times" for holding tasks. Interestingly, participants performed the holding tasks most accurate with L36 (M ≈ 6 mm), as opposed to L9 (M ≈ 7 mm). This effect was significant and indicates that **visual feedback helps to further improve the accuracy** at which holding tasks are performed – simply because users can adapt to errors quicker (visual correction impulse). This interpretation is also supported by slightly longer "outside times" for L36 (M ≈ 0.4 sec), which correlates with more frequent visual feedback due to thinner layers. Surprisingly, these findings are somehow contradicted by self-reports. Here, users had the subjective impression that they would perform significantly worse for L36 (M ≈ 4.7) compared to L9 (M ≈ 6.1), which was just not true. One reason for such ratings could be disappointment due to an assumed bad performance that was associated with leaving a layer.

Vertical Search Task. We found significant main effects for task completion times between all combinations of layer subdivisions. Although the search space was doubled each time, the completion times for vertically searching it did not. More precisely, for L9-L18, there were 9 added layers with 18 sec of longer completion times that resulted in a delay of 2 sec per extra layer. In contrast, for L18-L36, there were 18 added layers with only 25 sec penalty producing a delay of less than 1.4 sec per extra layer. This indicates that the vertical search task is an **efficient interaction technique** for single-feature searches (when pop-out effects are being utilized) that **performs better than linear** with a growing search space. These insights are also supported by self-reports.

Unfortunately, data gathered during our experiments made it difficult to directly derive a minimum layer thickness. This is because "outside times" and "height deviations" are not available for vertical search. However, since search tasks are usually accompanied by holding tasks, we can borrow from our findings there. Thus, our conservative estimate for the **minimum layer thickness** for vertical search is **roughly 1 cm**.

Horizontal Search Task. For horizontal search, task completion times and outside times indicate that the threshold for a minimum layer thickness lies somewhere between L18 and L36. Thus, a conservative measure is found by using the average layer thickness of L18 (M ≈ 3.8 cm) as the **minimum layer thickness** that is **about 4 cm**. A less conservative measure would be slightly smaller but should not get too close to L36 (M ≈ 1.9 cm). A follow-up study with finer layer subdivisions might help further narrowing down this value. However, conducting another user study is probably not worth the effort.

Physical Interaction Space & Number of Layers

Holding tasks were performed **equally well** in all of the three interaction zones. **Horizontal search** tasks, in contrast, were best accomplished in the **middle zone** (with respect to height deviations and outside times) and not as we anticipated in the lower zone. This was a surprising insight and somehow the precursor of another unexpected finding:

Lower and Upper Boundaries. When asked for their preferred lower interaction boundary, most participants did not choose the table surface. Instead, they opted for a slightly higher value that was roughly **7 cm** (SD ≈ 5 cm) above it. On the one hand, this was because most participants did not consider letting the lens loose in order to put it on the table surface. On the other hand, this means that the initially used "fist height" was a rather weak estimate for the table height. This was probably due to the frequent bending of the upper part of the body that was involved, for instance, with the horizontal search task. We therefore propose the "**wrist height**" as a more adequate **measure for the ideal tabletop height**. This insight can also be useful for conventional interactive table displays, in particular whenever a standing usage is intended.

In terms of a physical upper boundary, participants preferred a height of roughly **51 cm** above the table surface with a standard deviation of about 8 cm, which closely matches the standard deviation of the body heights (SD ≈ 9 cm). By compensating for the preferred lower boundary, we obtain a **corrected upper boundary of about 44 cm** (i.e., 51 cm - 7 cm) above the adjusted table surface.

Number of Layers. A reasonable threshold for the maximum number of layers is found by mapping minimal layer thicknesses (1cm / 4cm) onto the absolute height of the physical interaction volume (44 cm). Depending on the task, this results in a number of **44 layers** (holding and vertical search tasks) or **11 layers** (horizontal search tasks).

Design Guidelines & Further Observations

When designing multi-layer applications, the general rule of thumb is to **use as few layers as necessary** with the exact number depending on the primary interaction goal. For instance, for the exploration of a multi-layer geo-referenced map, where panning (horizontal search) is the dominant task, at most eleven layers should be adopted. In addition to that, most relevant information layers should be assigned to the **middle ("comfort") zone**, followed by the lower zone.

One important outcome of our study was that **vertical search** was clearly **favored over horizontal search**. Thus, it should be the first choice whenever possible. For example, for the exploration of temporal data sets (e.g., surveillance videos), vertical search should be reserved for the dominant goal of time-browsing, whereas horizontal search should be used for secondary goals, such as selecting a video.

Although our study mostly addressed single-feature search tasks in discrete (non-continual) layers, many of our findings also apply to continuous layer setups. Examples are the measures for lower and upper boundaries of the physical interaction space, the minimum layer thickness for holding tasks, and the recommendation for restricting the interaction to the middle and lower zones.

A simple way to improve the accuracy of holding and horizontal search tasks is to **provide instant visual feedback regarding layer borders.** Another strategy is to **adjust the center of the layer** to match the current display height **after a certain dwell time**. This prevents the problem of losing the layer when the tangible lens was originally too close to one of its boundaries – an effect that we frequently encountered during our studies.

Another noticeable effect was over- and undershooting during vertical search, especially whenever an unknown search target was encountered (pop-out effect). We only gathered limited evidence regarding fatigue (participants spent only approx. 30 min with the tangible display). Nevertheless, a multi-layer application should always provide users with the ability to rest from time to time, e.g., by freezing particular views and putting them down on the table (cf. [16]).

Beyond that, in many scenarios the required number of physical layers can easily extend a reasonable amount. In such cases, better layer subdivision strategies must be found, such as distorting the physical interaction space for horizontal search tasks by increasing the thickness of a particular layer (fisheye effect).

CONCLUSIONS & FUTURE WORK

In this paper, a class of interaction techniques that extends the design space of horizontal digital tables to the physical third dimension was systematically investigated with respect to exploring a layered information space. With a comprehensive user study, we studied the accuracy at which three fundamental tasks (vertical search, horizontal search and holding) are performed with tangible magic lenses. This also considered other fundamental issues such as the optimal number of layers and their thickness. Our findings include considerable differences in vertical and horizontal search tasks, e.g., a minimal layer thickness of 1 cm vs. 4 cm or a resulting maximum number of layers (11 vs. 44). Derived from these findings, design recommendations were also presented. Given the generic character of interaction tasks in stacks of discrete layers and their universal applicability for many systems and domains, we hope that our findings will help the research community to design and develop a new generation of applications using tangible magic lenses.

For future work, we intend to investigate further aspects of tangible multi-layer interaction, such as rotation tasks, the dynamic non-linear arrangement of layers, perceptual and cognitive issues as well as novel application domains.

ACKNOWLEDGEMENTS

This work was funded by the German Ministry of Education and Science (BMBF) project ViERforES-II (01IM1000 2B). We thank Ricardo Langner for his artwork and video editing.

REFERENCES

1. E. A. Bier, M. C. Stone, K. Pier, W. Buxton, and T. D. DeRose. Toolglass and Magic Lenses: The See-Through Interface. *Proc. SIGGRAPH 1993*, ACM Press (1993), 445-446.
2. L. D. Brown, and H. Hua, Magic Lenses for Augmented Virtual Environments. *IEEE Computer Graphics and Applications.* 26, 4 (July 2006), 64-73.
3. L. K. Y. Chan and H. Y. K. Lau. Poster: A Tangible User Interface Using Spatial Augmented Reality. *Proc. IEEE 3DUI*, IEEE Comp. Society, 2010, 137-138.
4. G. W. Fitzmaurice. Situated Information Spaces and Spatially Aware Palmtop Computers, *Communications of ACM*, 36, 7 (1993), 39-49.
5. K. Hirota and Y. Saeki. Cross-section Projector: Interactive and Intuitive Presentation of 3D Volume Data using a Handheld Screen. *Proc. 3DUI 2007*, IEEE Computer Society Press (2007), 57-63.
6. D. Holman, R. Vertegaal, M. Altosaar, N. Troje, and D. Johns. Paper Windows: Interaction Techniques for Digital Paper. *Proc. CHI '05*, ACM Press (2005), 591-599.
7. H. Ishii and B. Ullmer. Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. *Proc. CHI 1997*, ACM Press (1997), 234-241.
8. S. Izadi, S. Hodges, S. Taylor, D. Rosenfeld, N. Villar, A. Butler, and J. Westhues. Going Beyond the Display: A Surface Technology with an Electronically Switchable Diffuser. *Proc. UIST '08*, ACM Press (2008), 269-278.
9. Y. Kakehi and T. Naemura. UlteriorScape: Interactive Optical Superimposition on a View-Dependent Tabletop Display. *Proc. TABLETOP '08*. IEEE Computer Society Press (2008), 189-192.
10. J. Konieczny, C. Shimizu, G. W. Meyer, and D. A. Colucci. Handheld Flexible Display System. *Proc. VIS '05*, IEEE Computer Society Press (2005), 75-81.
11. J. C. Lee, S. E. Hudson, J. W. Summet, and P. H. Dietz. Moveable Interactive Projected Displays Using Projector Based Tracking. *Proc. UIST '05*, ACM Press (2005), 63-72.
12. J. Oh and H. Hua. User Evaluations on Form Factors of Tangible Magic Lenses. *Proc. ISMAR '06*, ACM Press (2006), 23-32.
13. J. Sanneblad and L.E. Holmquist. Ubiquitous Graphics: Combining Hand-held and Wall-size Displays to Interact with Large Images. *Proc. AVI '06*, ACM Press (2006), 373-377.
14. M. Spindler and R. Dachselt. Poster: Towards Making Graphical User Interface Palettes Tangible. *Proc. ITS '10*, ACM Press (2010), 291-292.
15. M. Spindler, S. Stellmach, and R. Dachselt. Advanced Magic Lens Interaction above the Tabletop. *Proc. ITS '09*, ACM Press (2009), 77-84.
16. M. Spindler, C. Tominski, H. Schumann, and R. Dachselt. Tangible Views for Information Visualization. *Proc. ITS '10*, ACM Press (2010), 157-166.
17. S. Subramanian, D. Aliakseyeu, and A. Lucero. Multi-Layer Interaction for Digital Tables. *Proc. UIST '06*, ACM Press (2006), 269-272.
18. A.R.Tilley. The Measure of Man and Woman: Human Factors in Design. John Wiley & Sons, 2nd ed., 2002.
19. A. Treisman and G. Gelade. A Feature Integration Theory of Attention. *Cognitive Psychology*. 12, (1980), 97–136.
20. B. Ullmer and H. Ishii. The metaDESK: Models and Prototypes for Tangible User Interfaces. *Proc. UIST '97*, ACM Press (1997), 223-232.
21. M. Weiser. The Computer for the 21st Century. *Scientific American*. 265, 3 (1991), 66-75.
22. K. Yee. Peephole Displays: Pen Interaction on Spatially Aware Handheld Computers. *Proc. CHI '03*, ACM Press (2003), 1-8.