To Touch or not to Touch? Comparing 2D Touch and 3D Mid-Air Interaction on Stereoscopic Tabletop Surfaces

Gerd Bruder
Department of Computer Science
University of Würzburg
gerd.bruder@uni-wuerzburg.de

Frank Steinicke
Department of Computer Science
University of Würzburg
frank.steinicke@uni-wuerzburg.de

Wolfgang Stürzlinger
Department of Computer Science and Engineering
York University
wolfgang@cse.yorku.ca

ABSTRACT

Recent developments in touch and display technologies have laid the groundwork to combine touch-sensitive display systems with stereoscopic three-dimensional (3D) display. Although this combination provides a compelling user experience, interaction with objects stereoscopically displayed in front of the screen poses some fundamental challenges: Traditionally, touch-sensitive surfaces capture only direct contacts such that the user has to penetrate the visually perceived object to touch the 2D surface behind the object. Conversely, recent technologies support capturing finger positions in front of the display, enabling users to interact with intangible objects in mid-air 3D space. In this paper we perform a comparison between such 2D touch and 3D mid-air interactions in a Fitts' Law experiment for objects with varying stereoscopical parallax. The results show that the 2D touch technique is more efficient close to the screen, whereas for targets further away from the screen, 3D selection outperforms 2D touch. Based on the results, we present implications for the design and development of future touch-sensitive interfaces for stereoscopic displays.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – Input Devices and Strategies, Evaluation / Methodology.

Keywords

Touch-sensitive systems, stereoscopic displays, 3D interaction.

1. MOTIVATION

Two different technologies dominated recent exhibitions and the entertainment market: (multi-)touch-sensitive surfaces and 3D stereoscopic displays. These technologies have the potential to provide more intuitive and natural interaction setups for a wide range of areas, including geo-spatial applications, urban planning, architectural design, or collaborative tabletops. These two technologies are orthogonal, as (multi-)touch is about *input* and 3D stereoscopic visualization about *output*. First commercial hardware systems have recently been launched (e. g., [4]), and interdisciplinary research projects explore interaction with stereoscopic content on 2D touch surfaces (e. g., [1, 2]). Moreover, an increasing number of hardware solutions provide the means to sense hand and finger poses and gestures in 3D space without input devices or instrumentation (e. g., Leap Motion [3]). The combination of these novel technologies provides enormous potential for a variety of new interaction concepts.

Until recently, research in the area of (multi-)touch interaction was mostly focused on monoscopically displayed data. There, the ability to directly touch elements has been shown to be very appealing for novice as well as expert users. Also, passive haptics and multi-touch capabilities have both shown their potential to improve the user experience [7]. Touch surfaces build a consistent and pervasive illusion in perceptual and motor space that two-dimensional graphical elements on the surface can be touched. Yet, three-dimensional data limits this illusion of place and plausibility [31]. 3D data sets are either displayed monoscopically, which has been shown to impair spatial perception in common 3D tasks, or stereoscopically, which can enrich the experience and interaction, but causes objects to appear detached from the touch surface [26, 30].

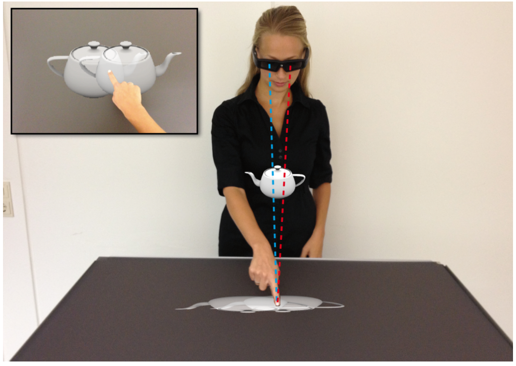
Stereoscopic display technology has been known for decades. It has recently been revived in the rise of 3D cinema and 3D televisions. With stereoscopic displays, each eye sees a different perspective of the same scene through appropriate technology. This requires showing two distinct images on the display. Objects may be displayed with *negative*, *zero*, or *positive* parallax, corresponding to in front, at, or behind the screen. Objects with centroid at *zero parallax* appear attached to the screen and are perfectly suited for touch interaction. In contrast, it is more difficult to apply direct-touch interaction techniques to objects that appear in front of or behind the screen [18, 27, 29]. In this paper we focus on the major challenge in this context, namely objects that appear in front of the screen such as a virtual object floating above the surface within the user's personal interaction space [12]. Teather and Stuerzlinger [34] provide a review of interaction techniques for distant objects behind the screen.

Two methodologies can be used for interacting with stereoscopic objects in front of a tabletop display:

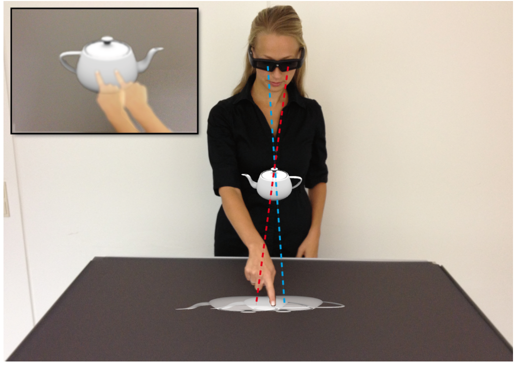
1. If the touch-sensitive surface captures only direct contacts, the user has to penetrate the visually perceived object to touch the 2D surface behind the object [36, 37].

2. Alternatively, if finger poses in front of the screen can be captured, the user can directly interact with the intangible object in 3D space [3].

SUI'13, July 20–21, 2013, Los Angeles, California, USA.



(a)



(b)

Figure 1: Illustration of the main problem of 2D touch interaction with stereoscopically displayed 3D data: The user is either focused (a) on her finger, which makes the selection ambiguous, or (b) on the object, which disturbs the visual perception of the finger.

Due to the discrepancy between perceptual and motor space and missing haptic feedback, both approaches provide natural feedback only for objects rendered with zero parallax. One question posed by this issue is where users "touch" a stereoscopically displayed intangible object in 3D space, considering the misperception of distances in virtual 3D scenes [22]. Conversely, it also brings up the issue where users "touch" a stereoscopically displayed object on a 2D display surface, considering that there are two distinct projections for each eye [36]. If the user penetrates the object while focusing on her finger, the stereoscopic impression of the object is disturbed, since the user's eyes are not accommodated and converged to the display surface. Thus, the left/right image pairs of the object appear blurred and can potentially not be merged (Figure 1(a)). Yet, focusing on the virtual object causes a disturbance of the stereoscopic perception of the user's finger, since her eyes are converged on the object's 3D position (Figure 1(b)). When the user selects an object in 3D space, by holding her finger in front of the screen, she can see a stereoscopic image while converging to her finger. However, due the vergence-accomodation conflict, the virtual object will appear blurred in comparison to the real finger (Figure 2).

In this paper we address the challenge of how to interact with stereoscopic content in front of a touch-sensitive tabletop surface. We evaluate interaction with touch-sensitive screens to select a 3D object, and compare this approach to systems where the user's finger is tracked in 3D space. We use a Fitts' Law experimental design to determine differences in 3D object selection performance for varying object parallax in front of the screen. The results of this experiment provide guidelines for the choice of touch technologies, as well as the optimal placement and parallax of interactive elements in stereoscopic touch environments.

Our contributions are:

- A direct comparison of the performance of 2D touch and 3D mid-air selection for different spatial configurations of interactive 3D objects.
- Guidelines for designing user interfaces for stereoscopic touch-sensitive tabletop setups.

The remainder of this paper is structured as follows. Section 2 summarizes background information on touch interaction and stereoscopic display. Section 3 describes the experiment we conducted to evaluate and compare 2D/3D interaction performance. Section 4 presents the results, which are discussed in Section 5. Section 6 concludes the paper.

2. BACKGROUND

Recently, many approaches for extending multi-touch interaction techniques to 3D applications with *monoscopic* display have been proposed [18, 25, 28, 29, 41]. In order to extend interaction possibilities with monoscopic 2D surfaces, Hancock et al. [18] presented approaches for 3D interaction within a limited range above the surface. Yet, interaction with stereoscopically displayed scenes introduces new challenges [30], since the displayed objects can float in front of or behind the interactive display surface.

2.1 Interaction with Stereoscopic Objects

In this section we describe work related to interaction with stereoscopically displayed objects. In particular, we discuss 2D touch and 3D mid-air selection techniques.

2.1.1 3D Mid-Air Interaction Techniques

To enable selection of stereoscopically displayed 3D objects in space, 3D tracking technologies capture a user's hand or finger motions in front of the display surface. The kinematics of point and grasp gestures in 3D space and the underlying cognitive functions have been studied [16, 23, 39]. For instance, it has been shown that the arm movement during grasping consists of two distinct phases: (1) an initial, *ballistic phase* during which the user's attention is focused on the object to be grasped (or touched). The motion is essentially controlled by proprioception, and (2) a *correction phase* that reflects refinement and error-correction of the movement, incorporating visual feedback in order to minimize the error between the hand or finger and the target [21]. MacKenzie et al. [23] investigated real time kinematics of limb movements in a Fitts' task and showed that, while Fitts' Law holds for the total limb-movement time, humans decelerate the motion sooner, if the target seems to require more precision in the end phase. The changes of the kinematics and control for reaching tasks within virtual environments have been investigated [14, 38].

Hilliges et al. [19] investigated extending the interaction space beyond the touch surface. They tested two depth-sensing approaches to enrich multi-touch interaction on a tabletop with monoscopic display. Although 3D "mid-air" interaction provides an intuitive technique, it has been shown that touching an intangible object, i. e., *touching the void* [11], leads to confusion and a significant number of overshoot errors. This is due to the fact that depth perception is less accurate in virtual scenes compared to the real world, as well as the introduced double vision and vergence-accommodation conflicts. Bruder et al. [10] investigated the effects

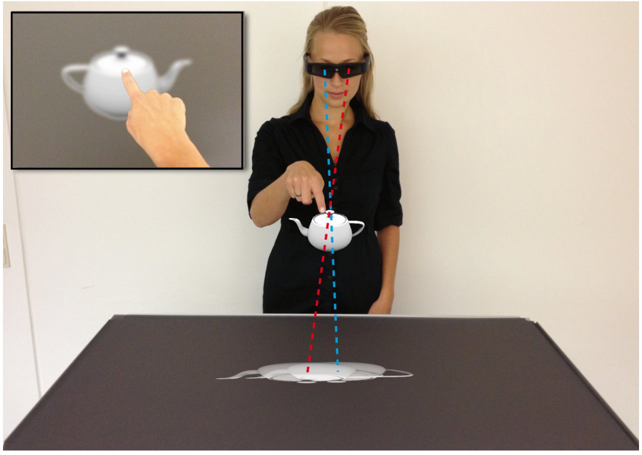


Figure 2: Illustration of the main problem of 3D mid-air interaction with stereoscopically displayed 3D data: The user sees a stereoscopic image while converging to her finger, but due to the vergence-accommodation conflict, the virtual object appears blurred in comparison to the finger.

of visual conflicts on 3D selection performance with stereoscopic tabletop displays. Some devices, such as the CyberGrasp, support haptic feedback when touching objects in space, but require extensive user instrumentation. Other approaches are based on the user moving tangible surfaces in 3D space to align with floating objects, e. g., through transparent props [11], or on controlling the 3D position of a cursor through multiple touch points [5, 32]. Toucheo uses 2D projections to define widget for interaction with objects presented stereoscopically above a multi-touch display [17]. Yet, the projection direction for Toucheo is straight down towards the display surface. This paradigm does not work well for objects that are stacked one above the other, as their projections then conflict.

2.1.2 2D Touch Techniques

Recently, multi-touch devices with non-planar surfaces, such as cubic [13] or spherical [6], were proposed. These can specify 3D axes or points for indirect object manipulation. Interaction with objects with negative parallax on a multi-touch tabletop setup was addressed by Benko et al.'s balloon selection [5], as well as Strothoff et al.'s triangle cursor [32], which use 2D touch gestures to specify height above the surface.

Valkov et al. [36] performed a user study, in which they displayed 3D objects stereoscopically in front or behind a large vertical projection screen. They instructed users to touch the virtual 3D objects by touching *through* the objects until their finger hit the display surface and recorded user behavior. This study found that users tended to touch between the projections for the two eyes with an offset towards the projection for the dominant eye. Bruder et al. [9] further analyzed stereoscopic 2D touch interaction and identified three distinct user behaviors (see Figure 3): users consistently touched either towards the dominant eye projection, the non-dominant one, or the midpoint between the projections. While these three behaviours varied between subjects, they found little within-subjects variation.

In a different study, Valkov et al. [37] showed that users are, within some range, insensitive to small misalignments between visually perceived stereoscopic positions and the sensed haptic feedback when touching a virtual object. Moreover, users are less sensitive to discrepancies between visual and tactile feedback for objects with negative parallax. They proposed to manipulate the stereoscopically displayed scene so that objects are moved towards the screen when the user reaches for them [35, 37]. This only works for objects displayed close (approximately 5cm) to the surface. Yet, the problem is that objects have to be shifted in space, which leads to a disturbed perception of the virtual scene for larger manipulations.

So far, no comparative analysis exists for 2D touch and 3D mid-air interaction in stereoscopic tabletop setups. Thus, it remains unclear if 2D touch is a viable alternative to 3D mid-air selection.

2.2 Fitts' Law and Selection

Fitts' Law [15] is a well-known empirical model for user performance in selection tasks. The model predicts the movement time MT for a given target distance D and size W by $MT = a + b \times \log_2(D/W + 1)$; where a and b are empirically derived. The log term is the *index of difficulty (ID)* and indicates overall task difficulty. This implies that the smaller and farther a target, the more difficult it is to select accurately. A valuable extension supported by an international standard [20] is the use of "effective" measures. This post-experiment correction adjusts the error rate to 4% by re-sizing targets to their effective width (W_e). This enables the computation of effective throughput, a measure that incorporates both speed and accuracy, by "normalizing" the accuracy as effective scores. This *throughput* is computed as $TP = \log_2(D_e/W_e + 1)/MT$, where D_e is the effective distance (average of measured movement distances), and W_e the effective width (standard deviation of error distances multiplied by 4.1333 [24]). Previous 3D research [34] suggests that one should use the point closest to the target along the ray to compute an accurate representation of the effective width W_e, as using the actual 3D cursor position would artificially inflate the effective measure. In essence, this suggestion projects the 3D task into 2D before computing throughput for touch-based interaction techniques. Even more recent work [33] reveals that the distortion due to perspective also has an effect. This work recommends the use of the 2D projections of sizes and distances to compute a screen-projected throughput for all *remote-pointing* techniques, such as ray-pointing.

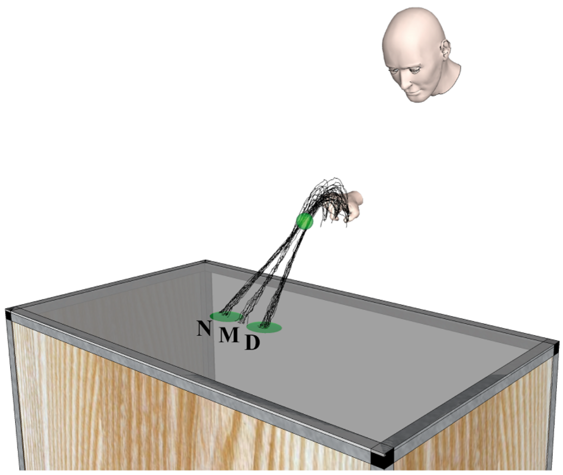


Figure 3: Illustration of finger movement trails for user groups touching towards the dominant eye projection (D), non-dominant eye projection (N), or towards the midpoint (M) using the 2D touch technique [9]. The trails have been normalized and are displayed here for a right-eye dominant user.

3. EXPERIMENTS

Here we describe our experiments to compare the performance of 2D touch and 3D mid-air interaction. We used a Fitts' Law selection task on a tabletop setup with 3D targets displayed on the surface or at different heights above the surface, i. e., with different negative stereoscopic parallax.

3.1 Experimental Setup

For the experiment we used a 62×112cm active stereoscopic multi-touch tabletop setup. The system is shown in Figure 4. The setup uses a matte diffusing screen with a gain of 1.6. For stereoscopic back projection screen we use a 1280 × 800 Optoma GT720 projector at 120Hz. The active DLP-based shutter glasses are driven by the projector at 60Hz per eye. We use an optical WorldViz Precision Position Tracking X4 system with sub-millimeter precision and accuracy to track the subject's finger and head for view-dependent rendering. For this, we attached wireless markers to the shutter glasses and another diffused IR LED on the tip of the index finger of the subject's dominant hand. We tracked and logged both head and fingertip movements during the experiment. The view of the 3D scene was rendered stereoscopically using off-axis projections. We measured an end-to-end latency of approximately 55ms between physical movements and a visual response.

The visual stimulus used in the experiment is a 3D scene in a 30cm deep box, fit to the horizontal dimensions of the physical tabletop setup (see Figure 4). We matched the look of the scene to the visual stimuli used in [9, 10, 33, 34] for improved comparability. The targets in the experiment were represented by spheres, arranged in a circle (Figure 4). A circle consisted of 11 spheres rendered in white, with the active target sphere highlighted in blue. The targets highlighted in the order specified by ISO 9241-9 [20]. The center of each target sphere indicated the exact position where subjects were instructed to touch with their dominant hand in order to select a sphere. Subjects indicated target selection using a Razer Nostromo keypad with their non-dominant hand. The target spheres highlighted green when the finger of the user was within the target to provide subjects with feedback about successful selection, to minimize systematic errors in Fitts' Law experiments [23]. Head-tracked off-axis stereoscopic display was active in all conditions. The size, distance, and height of target spheres were constant within circles, but varied between circles. In other words, targets were at a constant height for each circle of targets. Target height was measured upwards from the level screen surface. All target spheres were presented with positive height, i. e., in front of the screen. The virtual environment was rendered on an Intel Core i7 computer with 3.40GHz processors, 8GB of main memory, and an Nvidia Quadro 4000 graphics card.

3.2 Methods

The experiment used a 2 × 5 × 2 × 2 within-subjects design with the method of constant stimuli. The independent variables were selection technique (2D touch vs. 3D mid-air interaction), target height (0cm to 20cm, in steps of 5cm), as well as distances between targets (16cm and 25cm) and size (2cm and 3cm). Each circle represented a different index of difficulty with combinations of 2 distances and 2 sizes. This yielded four uniformly distributed IDs ranging from approximately 2.85bps to 3.75bps, representing an ecologically valuable range of Fitts' Law task difficulties for a touch screen setup. Each circle used one of 5 different target height, between 0cm and 20cm in steps of 5cm. Distances between targets, sizes and heights were not related from one circle to the next, but presented randomly and uniformly distributed. The dependent variables were movement time, error distance, error rate (percentage of targets missed), and effective throughput.

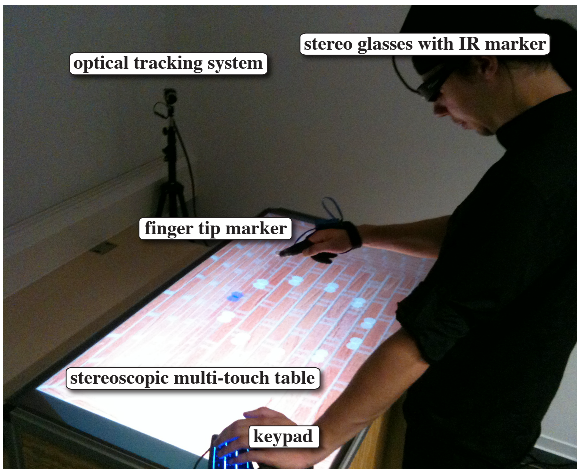
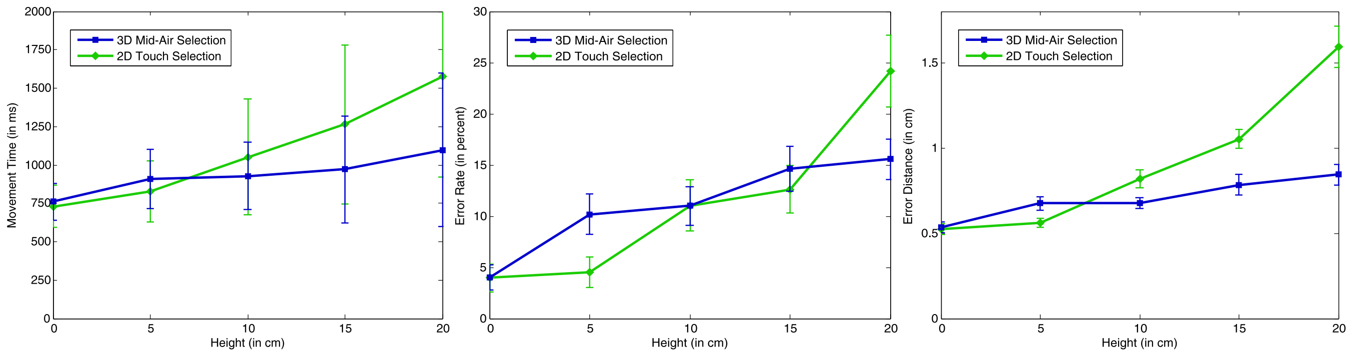


Figure 4: Experimental setup: photo of a subject during the experiment with illustrations.

The experiment trials were divided into two blocks: one for 2D touch selections and one for 3D mid-air selections. We randomized their order between subjects. At the beginning of each block, subjects were positioned standing in an upright posture in front of the tabletop surface (Figure 4). To remove a potential confound in terms of target visibility and view angle, we compensated for the different heights of subjects by adjusting the height of a floor mat below the subject's feet, resulting in an eye height of about 185cm for all subjects during the experiment. The experiment started with task descriptions, which were presented via slides on the projection surface in order to reduce potential experimenter biases. Subjects had to complete 5 to 15 training trials for both techniques to minimize later training effects. These training trials were excluded from the analysis. In order to compensate for misperceptions of the targets, we performed a calibration phase based on Bruder et al. [9]. During this calibration, subjects were instructed to touch the center of the target spheres as accurately as possible with 2D touch as well as 3D mid-air selection. Subjects had as much time as needed and they were free to place their index finger in the real world where they perceived the virtual target to be. We used the resulting calibrated positions to define the target centers in the Fitts' Law trials for each subject as described in [9, 10].

After the calibration, subjects were instructed to select the targets as quickly and accurately as possible, a common instruction in Fitts' Law experiments [33, 34]. Subjects received visual feedback when their finger was inside a target, by targets turning green. Then, subjects indicated selection by pressing a key with their non-dominant hand. If subjects pressed the key while the target sphere was not green, we recorded this as a selection error and advanced the trial state. We computed the distance of the position of the tip of the index finger to the calibrated sphere center. A valid 3D selection occurred if this distance was less than the sphere radius for 3D mid-air interactions. For 2D touch interactions, we computed the projected 3D target position and size on the 2D touch surface (see Figure 3). Then we judged a 2D touch selection to be valid if the finger position was within the projected circle (cf. [36]). There were 11 recorded target selections per circle. Circles were shown twice to each participant in randomized order for each configuration of independent variables. Thus, each participant completed a total of 80 circles, with a total of 880 recorded target selections.



(a) (b) (c)

Figure 5: Results for Fitts' Law trials with target object height on the horizontal axis and pooled for (a) movement time, (b) error rate, and (c) error distance, on the vertical axis. The error bars show the standard error.

Questionnaires.

In addition to the performance data collected in the Fitts' Law trials, we also asked subjects to judge various characteristics of the techniques through subjective questionnaires. Before and after the 2D/3D interaction conditions, subjects were asked to complete a Simulator Sickness Questionnaire (SSQ). Moreover, asthenopia, visual discomfort symptoms, were measured with a questionnaire about blurred vision, ocular soreness, itching of the eyes, increased blinking, heaviness of the eyes, and double vision on 4-point scales (0=none, 1=slight, 2=moderate, 3=severe), i. e., analogous to the SSQ sickness symptoms. After each technique, subjects were asked to complete a Slater-Usoh-Steed (SUS) presence questionnaire, a NASA TLX mental workload questionnaire, as well as a general usability questionnaire, in which we asked subjects to judge the technique according to the criteria learnability, efficiency, memorability, errors, and satisfaction on 5-point Likert scales.

3.3 Participants

10 male and 5 female subjects (ages 20-35, M=27.1) participated in the experiment. Subjects were students or members of the local university. 3 subjects received class credit for participating in the experiment. All subjects were right-handed. All subjects had normal or corrected to normal vision. 1 subject wore glasses and 4 subjects wore contact lenses during the experiment. None of the subjects reported known eye disorders, such as color weaknesses, amblyopia or known stereopsis disruptions. We verified the ability for stereoscopic vision of all subjects. We measured the interpupillary distance (IPD) of each subject before the experiment [40], which revealed IPDs between 5.8cm and 7.0cm (M=6.4cm). We used each individual's IPD for stereoscopic display in the experiment. 14 subjects reported experience with stereoscopic 3D cinema, 14 with touch screens, and 8 had previously participated in a study involving touch surfaces. Subjects were naïve to the experimental conditions. Subjects were allowed to take a break at any time between trials to minimize effects of exhaustion or lack of concentration. The total time per subject was about 1.5 hours.

4. RESULTS

Here we summarize the results from the experiment. We had to exclude two subjects from the analysis who misunderstood the task (i. e., showed 100% incorrect selections). All other trials have been included in the analysis. As stated above, we used for each subject the calibrated target positions as valid target centers. Results were normally distributed according to a Shapiro-Wilk test at the 5% level. We analyzed the results with a repeated measure ANOVA and Tukey multiple comparisons at the 5% significance level (with Bonferonni correction). Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity when Mauchly's test indicated that the assumption of sphericity had been violated.

4.1 Movement Time

The results for the movement time are illustrated in Figure 5(a). We found no significant main effect of technique ($F(1,12)$=3.870, p>.05, η_p^2=.244) on movement time. The average movement time during the experiment was M=1090ms (SD=521ms) for 2D touch, while 3D selection had M=934ms (SD=324ms).

The results show that the movement time for heights differs significantly ($F(1.272, 15.265)$=27.127, p<.001, η_p^2=.693). Post hoc tests revealed that the movement time was significantly increased when objects were displayed with heights of 15cm (p<.05) or 20cm (p<.001) in comparison to 0cm. As expected, we found a significant main effect of the ID on movement time ($F(1.220, 14.635)$=23.061, p<.001, η_p^2=.658).

We found a significant two-way interaction effect between technique and height ($F(1.360, 16.319)$=9.453, p<.01, η_p^2=.441). Post hoc tests revealed that subjects took significantly longer with 2D touch than 3D selection when objects were displayed with a height of 20cm (p<.05). We found no significant difference between the techniques for lower heights.

4.2 Error Rate

The results for error rate are illustrated in Figure 5(b). We found no significant main effect of technique ($F(1,12)$=0.009, p>.05, η_p^2=.001) on error rate. The average error rate during the experiment was M=11.6% (SD=18.5%) for 2D touch, while 3D selection had M=11.3% (SD=14.1%).

The results show that the error rate for heights differs significantly ($F(1.848, 22.172)$=17.186, p<.001, η_p^2=.589). Post hoc tests revealed that the error rate was significantly increased when objects were displayed with a height of 20cm (p<.05) in comparison to 0cm. As expected, we found a significant main effect of the ID on error rate ($F(3,36)$=15.359, p<.001, η_p^2=.561).

We found no significant two-way interaction effect between technique and height ($F(1.798, 21.570)$=2.685, p>.05, η_p^2=.183).

4.3 Error Distance

The results for the error distances, between the center of each sphere and the finger position during selection, are illustrated in Figure 5(c). We found a significant main effect of

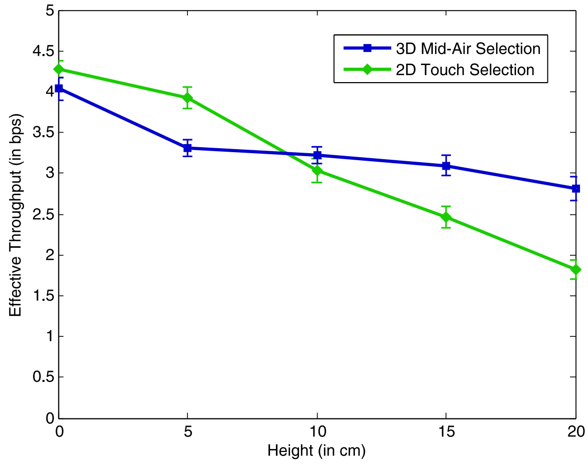


Figure 6: Effective throughput metric combining errors and movement time: The horizontal axis shows the target height, and the vertical axis shows the movement time. Higher throughput is better. The error bars show the standard error.

technique ($F(1, 12)$=5.115, p<.05, η_p^2=.299) on the error distance. Subjects made significantly larger errors when using 2D touch (M=0.91cm, SD=0.62cm) in comparison to 3D selection (M=0.70cm, SD=0.35cm).

The results show that the error distance for the height differs significantly ($F(1.419, 17.032)$=34.99, p<.001, η_p^2=.745). Post hoc tests revealed that subjects made significantly larger errors when objects were displayed with heights of 15cm (p<.05) or 20cm (p<.001) in comparison to 0cm. As expected, we found a significant main effect of the ID on error distance ($F(1.28, 15.361)$=5.669, p<.03, η_p^2=.321).

We found a significant two-way interaction effect between technique and height ($F(1.427, 17.120)$=11.293, p<.002, η_p^2=.485). Post hoc tests revealed that subjects made significantly larger errors with 2D touch than 3D selection when objects were displayed with a height of 20cm (p<.01). We found no significant difference between the techniques for lower heights.

4.4 Effective Throughput

The results for the effective throughput are shown in Figure 6. We found no significant main effect of technique ($F(1, 12)$=1.658, p>.05, η_p^2=.121) on throughput. The average throughput during the experiment was M=3.11bps (SD=1.29bps) for 2D touch, while 3D selection had M=3.30bps (SD=0.98bps).

The results show that the throughput for heights differs significantly ($F(1.696, 20.358)$=71.995, p<.001, η_p^2=.857). Post hoc tests revealed that throughput was significantly reduced when objects were displayed with heights of 10cm (p<.05), 15cm (p<.001) or 20cm (p<.001) in comparison to 0cm. As expected, we found a significant main effect of the ID on throughput ($F(3, 36)$=8.083, p<.001, η_p^2=.402).

We found a significant two-way interaction effect between technique and height ($F(2.408, 28.898)$=23.979, p<.001, η_p^2=.666). Post hoc tests revealed that throughput was significantly higher with 3D selection than 2D touch when objects were displayed with a height of 20cm (p<.05). In addition, we found a trend that the throughput was also higher with 3D selection for objects displayed with a height of 15cm (p<.08). In contrast, we found the inverse trend for objects displayed with a height of 5cm (p<.07). Here, throughput for 2D selection was higher. We found no significant difference between the techniques for lower heights.

4.5 Modeling

Fitts' Law can also be used as a predictive model, by regressing movement time on index of difficulty. We performed this analysis for both techniques at the five different heights. The regression lines for movement time are presented in Figure 7. The predictive quality of the model (as expressed by χ^2 values) is very high for 2D touch (for heights 0cm to 20cm χ^2=0.18, 0.06, 0.006, 0.04, and 0.037) and for 3D selection (for height 0cm to 20cm χ^2=0.10, 0.06, 0.08, 0.24, and 0.01).

4.6 Questionnaires

Also the results were normally distributed according to a Shapiro-Wilk test at the 5% level. Before and after each of the 2D touch and 3D selection conditions, we asked subjects to judge their level of simulator sickness and visual discomfort. Results were analyzed using paired-samples t-tests. For simulator sickness, we found a significant difference between the two conditions ($t(13)$=2.86, p<.02), with an average increase of mean SSQ-scores of 5.61 (SD=16.15) for the 2D touch technique, and 12.16 (SD=12.77) for 3D selections, which may be explained by missing physical support during 3D selections (cf. [8]). We found no significant difference ($t(13)$=0.16, p>.05) for the asthenopia questionnaire between the two techniques, but we observed a general before-after increase in visual discomfort for both 2D touch (M=0.18, SD=0.37) and 3D selection (M=0.19, SD=0.33). Again, the results do not exceed typical effects in stereoscopic display environments. For the reported sense of feeling present in the virtual scene, we did not observe a significant difference ($t(13)$=0.60, p>.05) for mean SUS-scores for 2D touch (M=3.92, SD=1.15) and 3D selection (M=4.08, SD=1.14). Both scores indicate a high sense of presence. We did not find a significant difference ($t(13)$=0.15, p=.88) between 2D touch (M=2.85, SD=0.43) and 3D selection (M=2.92, SD=0.56) on the mean five general usability criteria scores learnability, efficiency, memorability, errors, and satisfaction. Individual usability scores for 2D touch respectively 3D selection were (M=3.15 & M=3.00) for learnability, (M=3.54 & M=3.29) efficiency, (M=3.08 & M=3.43) memorability, (M=2.31 & M=2.71) errors, and (M=2.46 & M=2.00) for satisfaction. We could not find any significant differences between 2D touch and 3D mid-air selection for these metrics. We found no significant difference ($t(13)$=0.46, p>0.05) between 2D touch (M=10.44, SD=3.27) and 3D selection (M=9.91, SD=3.07) for the NASA TLX mental workload questionnaire scores.

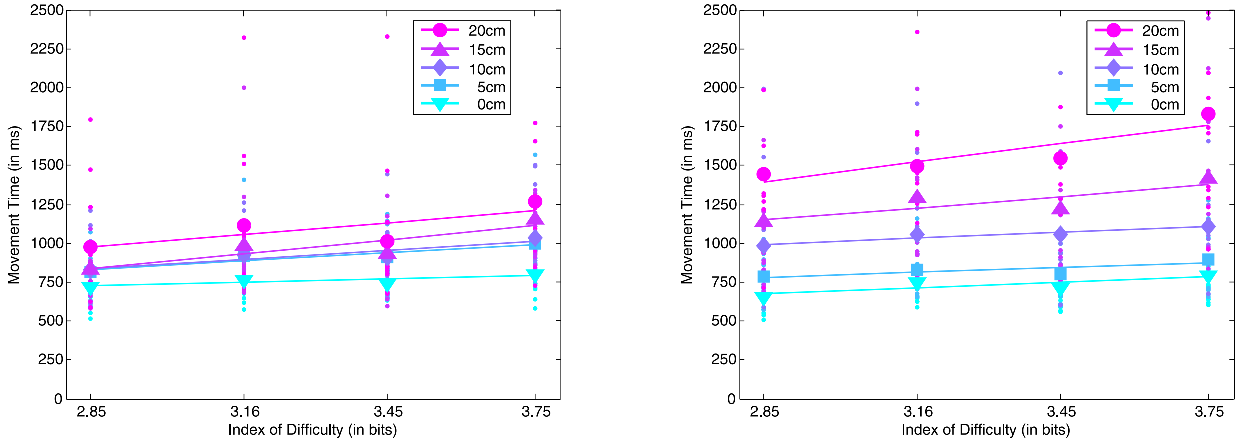
At the end of the experiment, we collected additional subjective preferences in an informal debriefing session. One subject remarked here notably:

"Selecting low objects was much easier on the surface – though it seemed counterintuitive at first!"

This comment was representative for many responses regarding the 2D touch technique. All but one subject preferred touching through 3D objects for objects close to the display surface.

5. DISCUSSION

The results from the Fitts' Law experiment reveal distinct characteristics of the 2D touch and 3D mid-air selection techniques, which impact their performance and applicability for interaction with objects displayed stereoscopically at different parallaxes. For



(a) 3D Mid-Air Selection (b) 2D Touch Selection

Figure 7: Models for (a) 3D mid-air selection and (b) 2D touch selection: solid lines are regressions of the measured movement time for the five target heights.

3D objects displayed up to 10cm above the display surface, touching objects in 2D on the surface by touching "through" the stereoscopic projection outperforms 3D mid-air selection in all considered metrics. Since much research has shown that 3D mid-air selection of virtual objects suffers from low accuracy and precision [8], e. g., due to visual conflicts, including vergence-accommodation mismatch, diplopia, and distance misperception [11], it is a promising finding that the reduction of 3D selection tasks to 2D input with the 2D touch technique can improve performance for tabletops with stereoscopically displayed objects. However, while interactions with both techniques are equal for objects at 0cm height, the results also show that the performance for the 2D touch technique decreases drastically for large negative parallax in comparison to 3D mid-air selection. At 20cm height, 2D touch performance is less than half in terms of throughput compared to performance at the screen. 3D mid-air selection performance drops much more slowly, decreasing only by about 30% at 20cm height.

For scenarios with stereoscopic visualization on (multi-)touch surfaces, the findings are still encouraging. They suggest that interactive 3D objects do not have to be constrained at the zero-parallax level, but may deviate up to 10cm before performance with the 2D touch technique is significantly degraded. For such distances, touch input is a good choice. Overall, our results show that it is indeed possible to leverage stereoscopic distance and interposition cues over a considerable range in touch-sensitive tabletop setups for improved spatial understanding of virtual data sets.

In our experiment, we compensated for different viewer heights by raising all subjects to a consistent head level. We did this to compensate for the potential confound that a lower viewpoint has a smaller 3D view volume due to (relatively) earlier clipping by the far and near sides of the display. In future commercial systems, we expect that stereoscopic touch tables could be height adjusted to accommodate for the height of each user.

In summary we suggest the following guidelines for the realization of touch interaction in 3D stereoscopic tabletop setups: For tabletop setups using the 2D touch technique, interactive virtual objects (e. g., buttons or other elements of graphical user interfaces) should not be displayed more than 10cm above the interactive display surface. Above that, the disadvantages outperform the benefits and 3D interaction techniques should be used.

6. CONCLUSION AND FUTURE WORK

In this paper we compared interaction techniques for tabletop setups with stereoscopic display. We analyzed the differences between 3D mid-air selection and a technique based on reducing the 3D selection problem to two dimensions by touching "through" the stereoscopic impression of 3D objects, i. e., a 2D touch on the display. The experimental results show a strong interaction effect between input technique and the stereoscopic parallax of virtual objects for all performance metrics, including movement time, errors, and effective throughput. Our main findings are:

- The 2D touch technique outperforms 3D mid-air selection for objects up to ca. 10cm height above the display surface.
- 3D mid-air selection is a better alternative for higher targets.
- Performance decreases faster for the 2D touch technique than for 3D selection with increasing height of virtual objects.

The results are encouraging for stereoscopic visualization in future touch-sensitive tabletop setups, since no additional tracking technology is needed for objects with small negative parallax. Recent sensing technologies for finger poses above display surfaces (e. g., Leap Motion [3]) will thus realize their benefits mostly only for objects at least about 10cm above the surface.

As a direction for future work, we cannot yet tell if these results hold for portable setups, where the orientation of the touch sensitive surface can change during interaction. We will pursue this topic to design more compelling user experiences as well as effective user interfaces for touch-sensitive stereoscopic display surfaces.

7. REFERENCES

[1] iMUTS - Interscopic Multi-Touch Surfaces. http://imuts.uni-muenster.de/, 2013.

[2] InSTInCT - Touch-based interfaces for Interaction with 3D Content. http://anr-instinct.cap-sciences.net/, 2013.

[3] Leap Motion. http://www.leapmotion.com/, 2013.

[4] Nintendo 3DS. http://www.nintendo.com/, 2013.

[5] H. Benko and S. Feiner. Balloon selection: A multi-finger technique for accurate low-fatigue 3D selection. In *Proc. of IEEE 3DUI*, pages 79–86, 2007.

[6] H. Benko, A. D. Wilson, and R. Balakrishnan. Sphere: multi-touch interactions on a spherical display. In *Proc. of ACM UIST*, pages 77–86, 2008.

[7] H. Benko, A. D. Wilson, and P. Baudisch. Precise selection techniques for multi-touch screens. In *Proc. of ACM CHI*, pages 1263–1272, 2006.

[8] F. Berard, J. Ip, M. Benovoy, D. El-Shimy, J. R. Blum, and J. R. Cooperstock. Did "minority report" get it wrong? Superiority of the mouse over 3D input devices in a 3D placement task. In *Proc. of INTERACT*, pages 400–414, 2011.

[9] G. Bruder, F. Steinicke, and W. Stuerzlinger. Touching the void revisited: Analyses of touch behavior on and above tabletop surfaces. In *Proc. of INTERACT*, 17 pages, 2013.

[10] G. Bruder, F. Steinicke, and W. Stuerzlinger. Effects of visual conflicts on 3D selection task performance in stereoscopic display environments. In *Proc. of ACM 3DUI*, pages 115–118, 2013.

[11] L.-W. Chan, H.-S. Kao, M. Y. Chen, M.-S. Lee, J. Hsu, and Y.-P. Hung. Touching the void: Direct-touch interaction for intangible displays. In *Proc. of ACM CHI*, pages 2625–2634, 2010.

[12] B. R. De Araújo, G. Casiez, J. A. Jorge, and M. Hachet. Mockup builder: 3D modeling on and above the surface. *Computers & Graphics*, 37:165–178, 2013.

[13] J.-B. de la Rivière, C. Kervégant, E. Orvain, and N. Dittlo. Cubtile: a multi-touch cubic interface. In *Proc. of ACM VRST*, pages 69–72, 2008.

[14] A. Y. Dvorkin, R. V. Kenyon, and E. A. Keshner. Reaching within a dynamic virtual environment. *J. NeuroEng. Rehabil.*, 4(23):182–186, 2007.

[15] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psych.*, 47:381–391, 1954.

[16] L. Geniva, R. Chua, and J. T. Enns. Attention for perception and action: task interference for action planning, but not for online control. *Exp. Brain Res.*, 185(4):709–717, 2008.

[17] M. Hachet, B. Bossavit, A. Cohe, and J.-B. de la Rivière. Toucheo: multitouch and stereo combined in a seamless workspace. In *Proc. of ACM UIST*, pages 587–592, 2011.

[18] M. Hancock, S. Carpendale, and A. Cockburn. Shallow-depth 3D interaction: design and evaluation of one-, two- and three-touch techniques. In *Proc. of ACM CHI*, pages 1147–1156, 2007.

[19] O. Hilliges, S. Izadi, A. D. Wilson, S. Hodges, A. Garcia-Mendoza, and A. Butz. Interactions in the air: Adding further depth to interactive tabletops. In *Proc. of ACM UIST*, pages 139–148, 2009.

[20] International Organization for Standardization. *ISO/DIS 9241-9 Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices*, 2000.

[21] G. Liu, R. Chua, and J. T. Enns. Attention for perception and action: task interference for action planning, but not for online control. *Exp. Brain Res.*, 185:709–717, 2008.

[22] J. M. Loomis and J. M. Knapp. Visual perception of egocentric distance in real and virtual environments. In *Virtual and adaptive environments*, pages 21–46. 2003.

[23] C. L. MacKenzie, R. G. Marteniuka, C. Dugasa, D. Liskea, and B. Eickmeiera. Three-dimensional movement trajectories in Fitts' task: Implications for control. *Q.J. Exp. Psychology-A*, 39(4):629–647, 1987.

[24] I. S. MacKenzie and P. Isokoski. Fitts' throughput and the speed-accuracy tradeoff. In *Proc. of ACM CHI*, pages 1633–1636, 2008.

[25] A. Martinet, G. Casiez, and G. Grisoni. The design and evaluation of 3D positioning techniques for multi-touch displays. In *Proc. of IEEE 3DUI*, pages 115–118, 2010.

[26] J. P. McIntire, P. R. Havig, and E. E. Geiselman. What is 3D good for? A review of human performance on stereoscopic 3D displays. *Proc. of the SPIE, Head- and Helmet-Mounted Displays XVII*, 8383:1–13, 2012.

[27] J. Pierce, A. Forsberg, M. Conway, S. Hong, R. Zeleznik, and M. Mine. Image plane interaction techniques in 3D immersive environments. In *Proc. of ACM I3D*, pages 39–44, 1997.

[28] D. Pyryeskin, M. Hancock, and J. Hoey. Comparing elicited gestures to designer-created gestures for selection above a multitouch surface. In *Proc. of ACM ITS*, pages 1–10, 2012.

[29] J. L. Reisman, P. L. Davidson, and J. Y. Han. A screen-space formulation for 2D and 3D direct manipulation. In *Proc. of ACM UIST*, pages 69–78, 2009.

[30] J. Schöning, F. Steinicke, D. Valkov, A. Krüger, and K. H. Hinrichs. Bimanual interaction with interscopic multi-touch surfaces. In *Proc. of INTERACT*, pages 40–53, 2009.

[31] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Phil. Trans. R. Soc. B*, 364:3549–3557, 2009.

[32] S. Strothoff, D. Valkov, and K. H. Hinrichs. Triangle cursor: Interactions with objects above the tabletop. In *Proc. of ACM ITS*, pages 111–119, 2011.

[33] R. J. Teather and W. Stuerzlinger. Pointing at 3D target projections with one-eyed and stereo cursors. In *Proc. of ACM CHI*, 10 pages, 2013.

[34] R. J. Teather and W. Stuerzlinger. Pointing at 3D targets in a stereo head-tracked virtual environment. In *Proc. of IEEE 3DUI*, pages 87–94, 2011.

[35] D. Valkov, A. Giesler, and K. H. Hinrichs. Evaluation of depth perception for touch interaction with stereoscopic rendered objects. In *Proc. of ACM ITS*, pages 21–30, 2012.

[36] D. Valkov, F. Steinicke, G. Bruder, and K. H. Hinrichs. 2D touching of 3D stereoscopic objects. In *Proc. of ACM CHI*, pages 1353–1362, 2011.

[37] D. Valkov, F. Steinicke, G. Bruder, K. H. Hinrichs, J. Schöning, F. Daiber, and A. Krüger. Touching floating objects in projection-based virtual reality environments. In *Proc. of JVRC*, pages 17–24, 2010.

[38] A. Viau, A. G. Feldman, B. J. McFadyen, and M. F. Levin. Reaching in reality and virtual reality: A comparison of movement kinematics in healthy subjects and in adults with hemiparesis. *J. NeuroEng. Rehabil.*, 1(11), 2004.

[39] D. Whitney, D. A. Westwood, and M. A. Goodale. The influence of visualmotion on fast reaching movements to a stationary object. *Letters to Nature*, 423:869–873, 2003.

[40] P. Willemsen, A. A. Gooch, W. B. Thompson, and S. H. Creem-Regehr. Effects of stereo viewing conditions on distance perception in virtual environments. *Presence-Teleop. Virt.*, 17(1):91–101, 2008.

[41] A. D. Wilson, S. Izadi, O. Hilliges, A. Garcia-Mendoza, and D. Kirk. Bringing physics to the surface. In *Proc. of ACM UIST*, pages 67–76, 2008.