Three-Point Interaction: Combining Bi-manual Direct Touch with Gaze

Adalberto L. Simeone
University of Portsmouth, UK
adals@acm.org

Andreas Bulling
Max Planck Institute for Informatics, Germany
bulling@mpi-inf.mpg.de

Jason Alexander
Lancaster University, UK
j.alexander@lancaster.ac.uk

Hans Gellersen
Lancaster University, UK
h.gellersen@lancaster.ac.uk

ABSTRACT

The benefits of two-point interaction for tasks that require users to simultaneously manipulate multiple entities or dimensions are widely known. Two-point interaction has become common, e.g., when zooming or pinching using two fingers on a smartphone. We propose a novel interaction technique that implements three-point interaction by augmenting two-finger direct touch with gaze as a third input channel. We evaluate two key characteristics of our technique in two multi-participant user studies. In the first, we used the technique for object selection. In the second, we evaluate it in a 3D matching task that requires simultaneous continuous input from fingers and the eyes. Our results show that in both cases participants learned to interact with three input channels without cognitive or mental overload. Participants' performance tended towards fast selection times in the first study and exhibited parallel interaction in the second. These results are promising and show that there is scope for additional input channels beyond two-point interaction.

CCS Concepts

•Human-centered computing → Empirical studies in HCI; Interaction techniques;

1. INTRODUCTION

Bi-manual interaction has expanded interaction towards new paradigms. By using the non-dominant hand in an active way, user interfaces can double the number of input channels at their disposal. The key advantage of this approach is the increased bandwidth that directly improves user performance [17]. The design of bi-manual interaction techniques (ITs) and the suitability of the task to which they are applied play a key role [12]. Tasks that can be simultaneously performed with both hands and can still be thought of as a single activity (i.e. resizing and moving a rectangle) will benefit most from two-handed ITs [19]. These parallelisable tasks allow the user to divide the workload between their two hands so that both can operate independently, with neither hand having to wait for the other.

AVI '16, June 07 - 10, 2016, Bari, Italy

ISBN 978-1-4503-4131-8/16/06…$15.00

DOI: http://dx.doi.org/10.1145/2909132.2909251

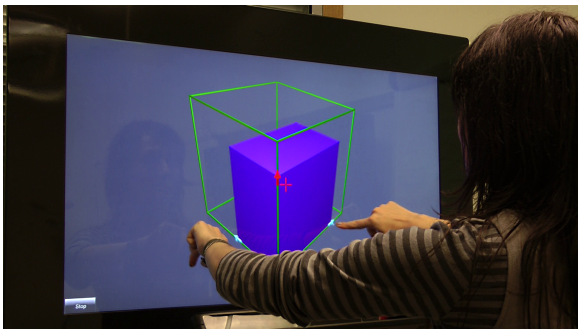

Figure 1: A participant using Three-Point Interaction to match the extents of the blue box to those of the green frame. The X and Z axes are controlled with the fingers and the Y axis with the eyes.

In this work we investigate the fundamental question of whether bi-manual input is a human limit or whether "three-point" input is feasible. As we expect a third channel to increase users' cognitive burden, we use a combination of channels that appear to have the smallest combined cognitive footprint, namely direct touch and gaze. We believe that three-point interaction can be useful in situations in which the combined manipulation of three different entities contributes toward a common goal. Such scenarios are commonplace in everyday and professional application, e.g., tasks that require concurrent manipulation of two attributes of an interface element or tasks that consist of multiple parallelisable steps, such as the specification of 3D primitives.

The primary goal of this work is to better understand human abilities in using ITs. Specifically, we aim to understand (1) how well people perform three actions in parallel and (2) how we can adapt this technique to become a meaningful alternative to uni-manual and bi-manual input. To answer both questions we conducted two user studies. In the first we analyse which factors potentially impact users' performance when using three input channels in a task requiring *discrete* selection of three targets. In the second study we

analyse how well users can perform three *continuous* actions simultaneously in a task requiring the matching of two three-dimensional objects. Results from the first study show that users are able to perform three near-simultaneous actions as long as they happen in A region of 24 cm of radius. In the second study, users performed in complete parallelism for 13% of the overall time spent in the task. These results allow us to identify the optimal conditions to use three-point ITs. We then discuss the design implications and the applicability of this metaphor to real-world scenarios.

2. RELATED WORK

Several studies have explored interaction using two modalities, e.g., by comparing hand-to-device mappings such as using fingers or two mice, styluses, tablets, etc. [3, 7, 21, 15]. We chose direct touch for hand-based input as touch input does not require an intermediary between the user and the screen. We believe the disadvantages of direct-touch (e.g. occlusion and finger-related inaccuracies) do not outweigh its benefits in our context. Gaze is interesting as a third input channel given that the eyes are always available whenever we interact with our hands and given their long history for human-computer interaction purposes. In contrast to touch, gaze is not spatially constrained and we can use our eyes and our hands independently thanks to proprioception, the knowledge of the inherent positions of our limbs.

In the following we describe other existing approaches to using two or more modalities to provide user input.

2.1 Bi-manual Continuous Interaction

The literature has documented the benefits that two-handed user interfaces (2H UIs) have over single-handed (1H) UIs. The primary motivation for bimanual interfaces has emerged from the way we normally interact with the physical world [31]. In everyday life we use both the dominant hand and the non-dominant hand to perform a multitude of tasks. Buxton and Myers [4] demonstrated how 2H UIs allowed users to parallelise tasks and to perform better than 1H approaches. Evidence suggests that symmetrical mappings of the two channels (e.g., they both alter the state of the system in the same way) can increase performance and parallelism over asymmetric designs [16].

Bi-manual interaction has been explored in 2D design environments and 3D applications where tasks involving continuous manipulation of 2D/3D points are commonplace. Zeleznik et al. [31] applied 2H interaction to 3D manipulation techniques. They noted that special attention should be given to the mapping between the degrees of freedom (DOF) provided by input devices and their effect in the actual scene. Users reported difficulties when interacting with mappings that had little relevance to real-world analogues. Schultheis et al. [23] compared a 2H UI to a wand and a traditional mouse finding that after sufficient training the 2H UI is able to outperform the alternatives by an order of magnitude. Training and tasks which inherently require parallelisable steps to be performed were reported to benefit from a 2H UI. Multi-touch UIs have also found several applications. Simeone et al. compared indirect bi-manual 3D ITs to state-of-the-art direct 3D ITs, finding that net manipulation times are comparable between the two paradigms [26] and that indirect touch results in 30% less errors in a collision avoidance task than direct touch [25]. Song et al. used a Kinect to allow bi-manual 3D interaction through a handlebar metaphor [27]

2.2 Multipoint Interaction

In the past, researchers have given different meanings to the concept of using 'three hands' for interaction. In an early work, Cutler et al. pondered if there could be such thing as 'three-handed' interaction. Using the example of two-handed interaction with tabletops they conceptualized this idea as an interaction *"where the user provides two hands and the computer a third to assist"* [5]. A successive work by Aguerreche et al. [1] explored the notion of 'three-handed' manipulation of virtual objects. In their work the 'three hands' were provided by two or three different users and not by a single person. Using the head orientation, instead of eye-tracking, to control the location of a 2D cursor is another possibility, as demonstrated by Virtual Reality applications. Lubos et al. [18] explored "quadmanual user interfaces", where a single user controls two sets of hands.

Hancock et al. [9] investigated one-, two- and three-finger techniques for interacting with 3D objects at 'shallow-depth' on a tabletop system. Their three-touch technique allowed the user to perform 6 DOF manipulation with three contact points, two belonging to the dominant hand and the third to the non-dominant hand. Their results showed that the three-touch technique performed better than the two-touch alternative with both outperforming the single-touch technique. Multipoint selection techniques on multi-touch surfaces were further studied by Kin et al. [14] who compared the performance of direct touch, bi-manual and multi-finger input. They found that there was no significant advantage once the number of fingers used for interaction exceeded two.

2.3 Multimodal and Gaze-supported User Interfaces

A large body of work in multimodal user interfaces explored the possibility of using two or more input modalities, such as gestures, voice or eye gaze. Multimodal UIs combine different sources of input in order to complement and disambiguate each other. In Bolt's work on multimodal UIs, pointing gestures were used to acquire an object or a location in space and specify an action to be performed [2]. Gaze estimation by a head sensor was also used to define manipulation targets that, together with gesture and speech recognizers, are then integrated into a single action [13]. Multimodal UIs can decompose a complex interaction task into multiple smaller sub-tasks. In contrast, the 2H UIs previously described allow users to express two distinct inputs at the same time and thus to perform two different tasks simultaneously.

Stellmach et al. have explored the concept of using eye tracking not as the sole input channel but rather, complemented by another other input modality so as to compensate the inherent inaccuracies associated to eye tracking. The authors present two solutions for gaze-supported interaction: one exploring selection using gaze and a mobile device [29] and another which uses the same setup [28] for panning and zoom. Successive work by Pfeuffer et al. [20] explored the combination of indirect touch input and gaze, with the latter modality allowing indirect manipulation of objects in the area focused by gaze.

Whole-body interaction is also another emerging research direction. Daiber et al. [6] present a system where multi-touch interaction is combined with the Wii Balance board which senses feet input. By shifting one's weight in specific ways users are able to perform additional interactions while maintaining the use of the hands. However, only movement

in 8 distinct directions was able to be sensed. Simeone et al. [24] investigated the combination of mouse input with feet input. They found that feet and mouse can work in parallel, but feet are better suited for secondary tasks that do not require high precision.

2.4 Eye-hand coordination

At the core of three-point interaction using gaze and touch is a proper spatial and temporal coordination of eye and hand movements. Eye-hand coordination is a complex process and has been studied a lot in experimental psychology and human movement science. For example, Gowen and Miall [8] investigated the interactions between eye and hand during tracing and drawing of different simple shapes. Their results suggested a bidirectional relationship between the eye and hand. Sailer et al. [22] investigated how gaze behaviour and eye-hand coordination changed when subjects learned a challenging visuomotor task. They found that learners first established basic mapping rules between manual actions and eye-movement commands that were then implemented and refined during skill acquisition and refinement. Johansson et al. [11] analysed the coordination between gaze behaviour, fingertip movements, and movements of the object in an object manipulation task. They concluded that gaze supports hand movement planning by marking key positions to which the fingertips or grasped object are subsequently directed. Finally, Hayhoe et al. [10] investigated the temporal dependencies of natural vision by measuring eye and hand movements while subjects made a sandwich. They found that these dependencies are limited and that much natural vision could be accomplished with just-in-time representations.

3. THREE-POINT INTERACTION SYSTEM

To gain a better understanding of three-point interaction, we developed a system capable of interpreting both hand and eye gaze input. The system comprises a Microsoft Pixelsense 1080p screen mounted vertically on a trolley and a Tobii X300 eye tracker. The screen has a diagonal of 101.6 cm (40 in) and measures 89 cm × 49.5 cm of visible screen with its base placed at a height of 120 cm from the ground. The tracker is mounted on a custom-built shelf attached to a glass panel bolted to the trolley (see Figure 2).

To correctly follow the user's eyes, the tracker is placed in line with the bottom of the visible area of the screen. Users were instructed to stand at a distance of 65 cm from the screen to maximise the coverage extent of the tracker and to still allow touch-based interaction. The standing distance from the screen is a trade-off between allowing participants to reach the screen and increasing the eye-tracking area. With this setup, pilot testing identified two areas of the screen where eye-tracking was unreliable: the top part of the screen (as participants would be required to stand too far away for full tracker coverage) and a lower area of the screen where the participants' arms occlude the tracker. We designed the experimental trials so that users did not need to interact in these areas (the top $^1/_6$ and the bottom $^1/_6$ of the screen). The tracker provides readings at a frequency of 60 Hz, which are then processed through a moving average algorithm before being used by the application.

4. STUDY 1: DISCRETE POINT SELECTION

The goal of the first user study is three-fold: (1) validate if users are able to perform discrete three-point interaction, (2) understand how point size and separation influences selection time; and (3) evaluate the existence of any learning effect. We asked users to select three on-screen targets, one with each hand and one with their eyes (see Figure 3).

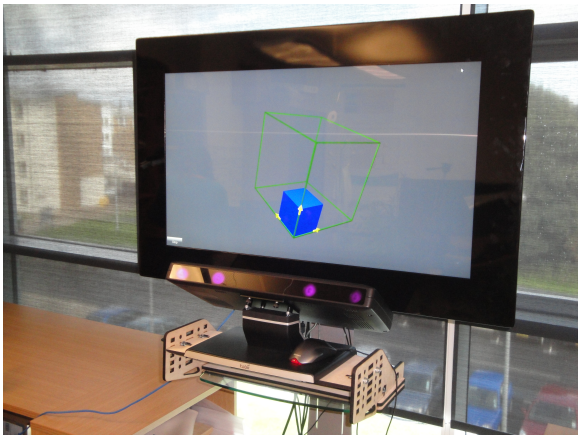

Figure 2: The Microsoft Pixelsense mounted vertically and the Tobii X300 eye tracker placed on a custom laser cut wooden shelf.

4.1 Participants

Fifteen participants (14 male, one female) aged between 21 and 41 years old ($M = 26.3\,years, SD = 4.81$) participated in the study. Data collected from our post-hoc questionnaires (measured on a scale ranging from 1: very low to 7: very high) shows that they all had considerable experience with touchscreen devices ($M = 6.33, SD = 0.98$) while experience with eye trackers was mixed ($M = 3.64, SD = 1.80$). This study lasted for approximately 30 minutes. Participants were paid £8 for their participation.

4.2 Experimental Task

The task required users to select three discrete on-screen points: one with each hand and one with their eyes. To avoid any target-to-input decision time overhead, targets to be touched were square (one in the left half of the screen, one in the right), and the target to be selected with the eyes was circular and appeared in the central vertical third of the screen. Participants were asked to select all three targets as fast as possible by touching the left and right targets and looking at the third target.

4.3 Design and Procedure

The experiment was a 3×3 factor within-subject, repeated measures design, with the independent variables of target size (in pixels: 24, 44, and 64; in cm circa: 1, 2, and 3, respectively) and distance (in pixels: 320, 512, and 768; in cm circa: 15, 24, and 35, respectively). The eye-acquired target was randomly placed in the middle vertical third of the screen. The distance *(d)* parameter was then used to determine the other targets assigned to the left and right hand by placing them at random locations on a circle centred at the eye target and of radius *d*. All locations appear in areas of the screen where the tracking is optimal. By varying the *size* (used as the side for the square targets; the target

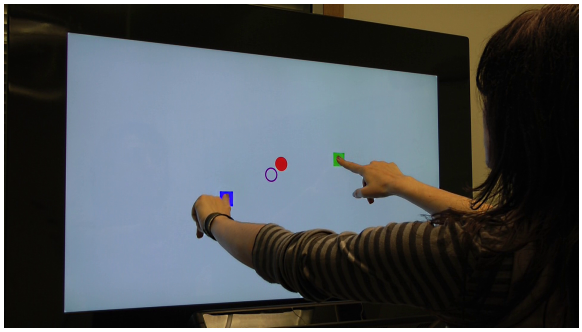

Figure 3: The Selection task. After selecting the leftmost and rightmost targets, the participants' gaze (indicated by the purple circle) is moving towards the middle target.

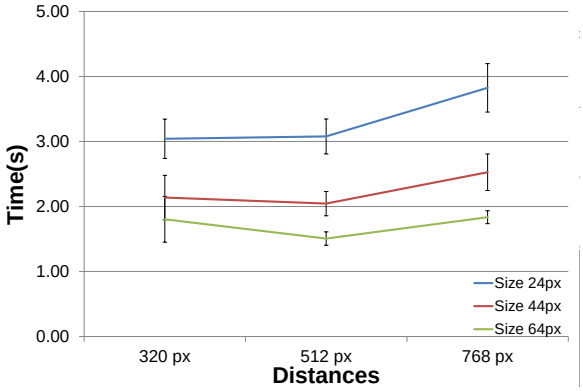

Figure 4: Mean completion times grouped by size and distance with standard error bars.

associated to the eyes remains constant in size and has a radius of 32 pixels) and *distance* between the targets we seek to evaluate their impact on the selection performance.

At the beginning of the study, participants were briefly introduced to a training scenario where the system showed feedback of the locations on the screen they touched or looked at. The experimenter described how they were only allowed to use one finger per hand and that the crosshair represented the tracked location of their eyes as interpreted by the system. Each trial then followed the same procedure as described here. Before starting, participants are presented with a blank screen. After they stated that they were ready, the experimenter would start a new trial. This prompted the user to place their left hand finger and right hand finger on their respective starting positions (indicated by rectangles placed at the middle of the leftmost and rightmost vertical thirds of the screen) and look inside a circle (placed in the middle of the central third of the screen).

In order to trigger the start of the trial itself, the system had to sense an intersection across all three areas. We introduced this in order not to bias the results by having participants find their hands close to the position in which the targets would appear. When these conditions were met (and held) a countdown appeared in the center of the screen. After three seconds have passed, the three targets appear and participants simultaneously touched and gazed at the required positions. A task was considered successful as soon as simultaneity of touch and gaze on all three positions was recorded, stopping time measurement. However, if a touch did not intersect the associated target, that trial was marked as an error but participants were still allowed to complete it.

Each participant performed 45 trials, with five repetitions of each distance/size condition, and were presented in counterbalanced order. The system prompted the participants to have a break after every 5 trials. The experimenter then waited for the participant to be ready to resume the task. For each task we measured the task completion time and logged all touch and gaze input events for subsequent analysis. At the end of all trials we collected the participants' subjective feedback through a questionnaire.

4.4 Results

In this study we found that participants were able to complete more than 90% of the tasks. They quickly understood how to use the extra input channel without experiencing any significant cognitive or physical overload. Furthermore we observed that when interface elements are grouped together in the same region of the screen, users can continue to interact with all three channels independently. On our setup, we found this optimal value consisting of an area having a radius of 512 px (24 cm). Increasing target size also contributes to improve results.

Feasibility of Discrete Three-Point Input — We collected a total of 15 participants × 45 target acquisitions, resulting in a total of 675 data points. Of these, 624 trials trials were completed (92%). However, since 13 of these 51 failed trials were due to a single participant (25%), we removed this subject from the subsequent analysis, resulting in a new total of 630 trials with 592 trials considered successfully completed (94%) and 38 considered as failed. We believe these results shows that participants were able to control an additional input channel. Figure 4 shows the mean task completion time for each size/distance condition.

We asked participants to rate how easily they learned to use the system on a 7-point scale. The resulting answers ($M = 6.47, SD = 0.99$) indicate that the concept of three-point selection was not difficult to learn. The other qualitative results show that participants did not find the task to be neither mentally ($M = 2.53, SD = 1.46$) nor physically demanding ($M = 3.87, SD = 1.55$). They also rated themselves as being able to select the targets very quickly ($M = 5.00, SD = 1.31$) without feeling frustrated or discouraged at any point ($M = 2.00, SD = 1.25$).

Impact of Size and Distance on Performance We performed an ANOVA on the collected data and found a significant effect of *size* ($F_{2,\ 26} = 15.81, p < 0.01$). Pairwise comparisons among different size levels showed significant differences for the pairs: 24 px and 44 px ($p < 0.01$); 44 px and 64 px ($p = 0.01$); 24 px and 64 px ($p < 0.01$). This was reflected by the participants' behaviour during the trials. Participants struggled with the smallest size condition, which required more accurate pointing. A total of 18 selection errors (47%) were related to this size condition. This result confirms that bigger target sizes contribute to faster performances, even when interacting with a three-point IT.

The *distance* variable also had a significant impact on task completion times ($F_{2,26} = 3.458, p < .05$). The discernible implications of altering this factor also impacted the users'

behaviour in selecting targets. We observed that while targets were close together users did not have to use their peripheral vision as all targets appeared approximately in the foveal area. When increasing the distance to 768 pixels (36 cm), we noticed a difference from 512 px (24 cm) ($p < 0.05$) as participants had to focus their attention onto the two targets to be selected by their hands. This finding hints at the optimal spacing of targets selectable with a three-point IT. When targets are placed in an area of 24 cm of radius, users were able to see them all in the region of the screen they were currently focusing. When this radius increased, we observed that they often had to divert their attention as the leftmost and rightmost targets were now outside their peripheral vision.

This study required users to express three independent interaction intents. While we asked participants to try to do so as simultaneously as possible, in absolute terms the events were still performed sequentially. The next study investigates the implications of performing three continuous actions concurrently.

5. STUDY 2: CONTINUOUS INPUT

This study applies the technique to a task requiring continuous three-point input. The goals of this study are: (1) to verify that users can interact simultaneously across each input channel; (2) determine the impact of non-symmetrical input and (3) understand if their performance improves with practise. Results show that participants attained a final average performance tending towards parallelism. We further observed that, contrary to our initial assumptions, three-point performance does not seem to be affected by non-symmetric or oblique movements.

5.1 Participants and Apparatus

The same participants who completed the first user study (with the exception of the second participant who was later excluded) also participated to the second, after taking a break. The study lasted approximately 30 minutes. We used the same apparatus as in the first study.

5.2 Experimental Task

The task requires the participants to match the extents of a user-manipulable 3D cube to those of a target box in a 3D environment (see Figure 1). The user manipulates the dimensions of a cube by dragging arrows superimposed on three of the edges (the closest vertical edge, and the two closest bottom edges). The leftmost and rightmost arrows are sensitive to touch input while the central arrow is sensitive to eye-gaze input. Participants must first "lock" onto the central arrow through an eye fixation. The system recognises this dwelling by changing the color of the arrow from yellow to orange. After an interval of 500 ms the arrow is considered "locked" and can be controlled with the eyes.

Users manipulate the arrows uni-dimensionally: the 3D location of the touch point is used to find a ray intersecting the plane on which the arrow is located. Only the displacement along the relevant axis to the intersected arrow (if one is found) is used to manipulate it. We are primarily interested in the parallelism of task completion (not user accuracy) and so we included a snapping mechanism to aid completion. Snapping is automatically engaged when the user moves an arrow within 30 px of the target. When this happens, the length of the side associated to the arrow is set to the frame's length and the arrow itself turns green to indicate completion.

To understand whether non-symmetric continuous input had any impact, we alternated the direction of input along each of the three axes. Users had to either increase or decrease the length of side. Each side was initialised to a minimum length if users had to increase that side, and vice versa to its maximum length. Thus, users were subjected to all eight direction combinations for the three manipulable axes. Some configurations have completely synchronised movements (i.e. all axes set to increase or decrease), while the remainder provide various combinations of movements.

Finally, to investigate whether vertical gaze movements were easier to perform, we also tilted the cube 15° both left and right of the vertical axis. We wanted to estimate the performance of eye movements along an oblique path.

5.3 Study Design and Procedure

We used a $3 \times 8 \times 2$ within subjects, repeated measures study design, with three independent variables: *axis tilt* (left, none, right), *input direction* and *repetition*. Each participant was calibrated using the standard procedure in Tobii Studio.

At the beginning of the study participants were asked to familiarise themselves with the 3D arrows attached to the blue box. In the training mode, they could freely interact with the three arrows by using touch and gaze without the presence of any target frame. Whenever they felt comfortable enough, the trial could be started by pressing a button in the top left corner. After a countdown phase elapsed, the blue box and the green frame appeared with the dimensions and axis tilt associated to the particular trial. The blue box always appeared aligned to the edge closest to the screen of the target frame. In the tilted condition, the ensemble of box, widget and target frame is rotated.

As instructed, participants were told to simultaneously manipulate the three arrows so that they all contributed to the task of matching the dimensions of the blue box to those of the green frame. When all three dimensions were "snapped" to the target frame, the trial automatically ended. All axis tilt/input direction pairs were repeated twice, for a total of 720 trials. At the end of the study, participants completed a subjective feedback questionnaire.

5.4 Metrics

In order to estimate the coordination between the two hands and gaze we adopted Zhai et al.'s measure of *Translation Coordination* (T_c) [32]. It is defined as the ratio between the shortest distance to the goal and the length of the actual distance travelled. It results in a value between 0 and 1, with 1 representing perfect coordination (the user did not deviate from the shortest path possible) and intermediate values representing superfluous movement. Participants' distance to the goal was sampled every 25 ms. To analyse the results, we resampled all progression data in 100 evenly spaced points, so that results from all the different trials could be compared directly. By plotting the mean T_c values from the three independent channels we were able to observe whether participants leaned towards serial operation of the three channels or towards parallel coordination.

Additionally, we also used the *NDC* metric (*Number of Degrees of freedom Combined* [30]), which calculates a value between 1 and 3 describing how many DOF were used simultaneously in any given instant. From these values, it is

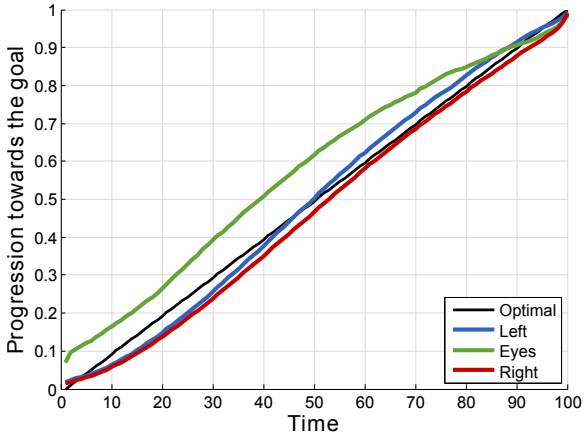

Figure 5: The three curves represent each channel's mean progression data across all participants and trials, resampled in 100 evenly spaced points. The curves begin at $t = 1$ (trial started, $Progression = 0$) and end at $t = 100$ (goal reached, $Progression = 1$).

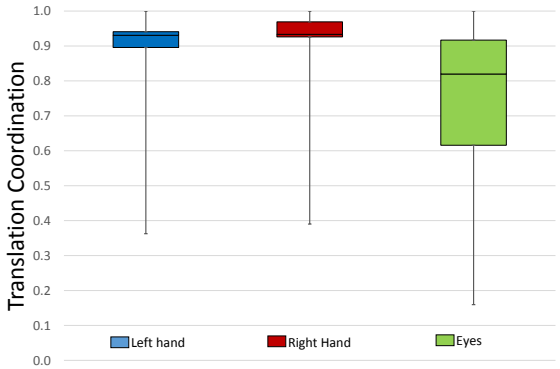

Figure 6: Mean Translation Coordination values by input channel.

possible to calculate the percentage of time spent manipulating 1, 2, or 3 DOF simultaneously. These percentages give us an insight on the parallelism rate achieved by users.

6. RESULTS

Simultaneous Interaction — The first goal was to understand whether or not participants are able to perform three continuous movements simultaneously. Mean T_c values for the left and right hand are respectively 0.91 and 0.94, values very close to the optimum. T_c values for the eyes resulted in a value of 0.74 (see Fig. 6). Recall that each channel operated uni-dimensionally. The lower value for the eyes reflects the greater difficulty in using gaze as continuous input and that more movement than what was strictly necessary was performed. An analysis of variance on the study conditions highlighted the fact that the T_c measure is not affected by neither inclination nor the direction of motion.

Figure 5 shows the mean progression towards the goal averaged at each of the 100 time-points for each of the three input channels across all participants. The left and right hands overlap to a substantial degree and are very close to the optimal progression (the black line in Figure 5). Regarding eye gaze, the curve starts at higher initial progression than the manual channels. We believe this was caused by the dwelling mechanism, which only allowed movement after unlocking the arrow, thus causing the arrow to jump ahead towards the location users were fixating.

The mean NDC values during the task was 1.66 ($SD = 0.21$). For each trial, we then calculated the percentage of time users manipulated 1, 2, or 3 DOF: 40% (1 DOF), 47% (2 DOF), and 13% (complete parallelism for all 3 DOF). The NDC metric does not indicate which of the three DOF were being manipulated. Since this work is the only such example focusing on parallelism across three input channels, we cannot say whether this is value is high or low. However, it shows that it is possible to perform three actions simultaneously for a substantial period of time.

Non-Symmetric Interaction and Axial Tilt — We performed an analysis of variance on the data which did not provide any evidence to support that either the *tilt* or *input direction* had any significant impact on users' performance. However, one participant stated that *"moving the arrows in to a cuboid was much easier than a rectangular shaped [parallelepiped] challenge."* On a cognitive level it might be less demanding to perform three symmetric movements, but we did not find any evidence.

Learning Effects — We had the participants perform the trials in groups of eight, after which they rested. We analysed whether there was any learning effect on their overall performance measured at each of the six blocks. A regression analysis on these overall means showed that there their performance improved from the beginning to the end of the task (R^2=0.83).

Qualitative Results — The three-point technique was rated as having an average cognitive load ($M = 4.47, SD = 1.51$, on a scale ranging from 1 – very low, to 7 – very high) and physical demand ($M = 3.67, SD = 1.72$). It caused low frustration ($M = 3.33, SD = 1.95$) and was easy to learn ($M = 5.27, SD = 1.49$).

7. DISCUSSION

Our main research question was to investigate whether or not users are able to use three-point interaction and how well users are able to form meaningful interactions with this technique. Results have shown that three-point interaction is a viable strategy and it can also be improved in a relatively short time frame. One participant's comment exemplifies this: *"[I] feel like I would have improved further with additional practice."* Indeed, no participant had prior experience in such an interaction technique, as opposed to a lifetime of bi-manual interaction. To this end, a participant commented that *"I was frustrated with myself for 'forgetting' to move my eyes sometimes"* (referring to the second task). In the following, we discuss the effectiveness of three-point interaction in the tasks we studied.

7.1 Discrete Input

In the first scenario, results show that users are able to manage the selection of the three targets relatively quickly. Using gaze to provide another interaction point did not pose any significant extra cognitive burden on the users. These two results lead us to believe that in this setting, the technique can be used effectively in all situations which require three simultaneous entities to be selected. As we have initially described, a three-point IT can be applied to interaction tasks

which require three actions to be performed or to group more interactions into one (e.g., applying two status changes to the same object).

Based on our findings we can say that such a three-point design should ideally avoid laying interface elements in area of radius wider than 24 cm as we have found that by increasing this distance too much we introduce a 'switching cost' as users need to switch their attention in order to focus on elements that are outside their foveal area. On the topic of peripheral vision, one participant commented that *"[I] assumed it would be ample to select points to the left/right. [I] found it harder than expected and had to concentrate more for smaller points."* Overall, the fast completion times and the good feedback received should make this approach a prime candidate for further investigation in a real application context.

7.2 Continuous Input

The ability to control three points simultaneously can be learnt if given enough training. However, we noticed that the necessity of dividing one's attention between the various elements of the task frustrated some users in the second study. Indeed, for as long as gaze tracking was enabled, the system continued to use it to affect its state. This introduced some frustration where unintended eye saccades altered the shape of the manipulable cube. To overcome this problem, we suggest "discretising" gaze input so that only specific areas of the screen would be sensitive to gaze. For example, the system could display a graduated ruler (representing potential heights for the cube), providing users something to fixate. Based on the results of the first study this might increase performance at the expense of a loss of precision.

7.3 Parallelism

We sought to define parallelism through the two metrics we calculated: the *Translation Coordination* [32] and the *Number of Degrees of freedom Combine*d [30]. These two analyses show that three actions can be parallelised to a certain extent. We believe the result that 13% of overall time is spent in complete parallelism will be useful as a baseline for future three-point ITs.

Our research also identified a new open question: how much parallelism is required for a three-point IT? As we primarily focused in the discovery of the limits of this novel interaction technique, we cannot derive from the outcome what is the optimal trade-off in terms of parallelism and effectiveness. Indeed, even if a three-point IT is used in a more serialised way, it still retains the theoretical advantages of a greater bandwidth and decreased switching costs, as opposed to 2H or 1H alternatives. We leave the characterisation of the design space of three-point ITs to future research.

7.4 Application Scenarios

We believe that three-point ITs are best suited in those tasks where high expressive power is needed. In order to evolve the three-point ITs from a concept into real application scenarios, we briefly discuss how it could be applied in real world applications.

7.4.1 3-Way Status Change

There are various situations in which it would be helpful to simultaneously visualize how two parameters influence the appearance of an interface element. For example, in photo-editing applications, gaze could be used to focus on a specific part of the picture while our two hands could be used to modify the level of shadows, midtones or highlights. Their combined interaction is difficult to estimate and this specific interaction is usually performed by choosing a value and experimenting how it interacts with different values for the other parameters. A three-point IT would allow us to visualize how two parameters interact on a specific part of the picture (controlled by our eyes).

7.4.2 Graphical Editing

The task in our second study is representative of a very common occurrence in 3D modelling, where specifying the extents of a geometric primitive is an activity requiring several substeps (i.e. specifying a base, then the height, for a 3D parallelepiped). Three-point IT can be designed around the behaviour commonly found in graphic applications. For example, touch input can be used to specify the two corners of the base of a paralleliped while our eyes control its height. In the visualization context, pointing at three different objects could determine the instantiation of a volume of space containing those objects and all others in between.

7.5 Design Implications

Based on our observations and results of our studies we propose guidelines for the design of effective three-point ITs.

- Define a clear way to activate and deactivate three-point interaction, e.g., enable it while both fingers are on the screen and disable it when they are removed.
- Provide feedback to inform the user when an input channel is activated.
- Provide visual assists for users to focus on with their eyes, for example a grid of points to help in choosing the extents of a 3D solid.
- Avoid gaze input interfering in the regions interested by manual input.
- Layout the interface elements so that they are well within the foveal area (within a radius of 24 cm from the focal point).
- Design interface elements so that they have at least a radius of 2 cm, to avoid selection inaccuracies.

8. CONCLUSIONS

In this paper we have presented a novel interaction technique which, by combining eye-tracking and direct touch input, can provide three simultaneous points of interaction. This work aimed to answer some fundamental questions concerning the technique, namely which factors can influence its performance; how effective can users become after practise and whether or not it is feasible to parallelise three independent actions across different input channels. To this end we have designed two user studies, each focusing on a different perspective of the technique. The first study investigated three simultaneous discrete input points. The results show that the participants were able to learn the technique quickly and easily and perform better when targets are at least 2 cm in radius and no more than 24 cm apart.

The second study sought to answer a more fundamental question regarding whether or not users are able to learn how to perform three actions concurrently. The results show that

after a brief training, the participants were able to interact in complete parallelism for 13% of the time spent in the task.

This work has shown that bi-manual interaction is not a hard limit and that we are able to learn how to use more than two input channels. Our research gave us the opportunity to understand which factors affect the performance of three-point interaction. It also highlighted the need for future research efforts in this design space and on the adaptation of this metaphor to real application scenarios.

9. REFERENCES

[1] L. Aguerreche, T. Duval, and A. Lécuyer. 3-hand manipulation of virtual objects. In *Proc. JVRC 2009*.

[2] R. A. Bolt. Put-that-there: Voice and gesture at the graphics interface. In *Proc. SIGGRAPH 1980*, pages 262–270. ACM.

[3] P. Brandl, C. Forlines, D. Wigdor, M. Haller, and C. Shen. Combining and measuring the benefits of bimanual pen and direct-touch interaction on horizontal interfaces. In *Proc. AVI 2008*, pages 154–161. ACM.

[4] W. Buxton and B. Myers. A study in two-handed input. *ACM SIGCHI Bulletin*, 17(4):321–326, Apr. 1986.

[5] L. D. Cutler, B. Frölich, and P. Hanrahan. Two-handed direct manipulation on the responsive workbench. In *Proc. I3D 1997*, pages 107–114. ACM.

[6] F. Daiber, J. Schöning, and A. Krüger. Whole body interaction with geospatial data. In *Smart Graphics*, volume 5531 of *LNCS*, pages 81–92. Springer Berlin / Heidelberg, 2009.

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

[8] E. Gowen and R. C. Miall. Eye-hand interactions in tracing and drawing tasks. *Human Movement Science*, 25(4-5):568 – 585, 2006. Advances in Graphonomics: Studies on Fine Motor Control, Its Development and Disorders.

[9] M. Hancock, S. Carpendale, and A. Cockburn. Shallow-depth 3D interaction. In *Proc. CHI 2007*, pages 1147–1156. ACM.

[10] M. M. Hayhoe, A. Shrivastava, R. Mruczek, and J. B. Pelz. Visual memory and motor planning in a natural task. *Journal of Vision*, 3(1):49–63, 2 2003.

[11] R. S. Johansson, G. Westling, A. Backstrom, and J. R. Flanagan. Eye-hand coordination in object manipulation. *Journal of Neuroscience*, 21(17):6917–6932, 2001.

[12] P. Kabbash, W. Buxton, and A. Sellen. Two-handed input in a compound task. In *Proc. CHI 1994*, pages 417–423. ACM.

[13] E. Kaiser, A. Olwal, D. McGee, H. Benko, A. Corradini, X. Li, P. Cohen, and S. Feiner. Mutual disambiguation of 3D multimodal interaction in augmented and virtual reality. In *Proc. ICMI 2003*, pages 12–19. ACM.

[14] K. Kin, M. Agrawala, and T. DeRose. Determining the benefits of direct-touch, bimanual, and multifinger input on a multitouch workstation. In *Proc. GI 2009*, pages 119–124. CIPS.

[15] G. Kurtenbach, G. Fitzmaurice, T. Baudel, and B. Buxton. The design of a gui paradigm based on tablets, two-hands, and transparency. In *Proc. CHI 1997*, pages 35–42. ACM.

[16] C. Latulipe, C. S. Kaplan, and C. L. A. Clarke. Bimanual and unimanual image alignment: an evaluation of mouse-based techniques. In *Proc. UIST 2005*, pages 123–131. ACM.

[17] A. Leganchuk, S. Zhai, and W. Buxton. Manual and cognitive benefits of two-handed input: an experimental study. *ACM Trans. Comput.-Hum. Interact.*, 5(4):326–359, Dec. 1998.

[18] P. Lubos, G. Bruder, and F. Steinicke. Are 4 hands better than 2?: bimanual interaction for quadmanual user interfaces. In *Proc. SUI 2014*, pages 123–126.

[19] R. Owen, G. Kurtenbach, G. Fitzmaurice, T. Baudel, and B. Buxton. When it gets more difficult, use both hands: exploring bimanual curve manipulation. In *Proc. GI 2005*, pages 17–24. CIPS.

[20] K. Pfeuffer, J. Alexander, M. K. Chong, Y. Zhang, and H. Gellersen. Gaze-shifting: Direct-indirect input with pen and touch modulated by gaze. In *Proc. UIST 2015*, pages 373–383. ACM.

[21] J. Rekimoto. Smartskin: an infrastructure for freehand manipulation on interactive surfaces. In *Proc. CHI 2002*, pages 113–120. ACM.

[22] U. Sailer, J. R. Flanagan, and R. S. Johansson. Eye-hand coordination during learning of a novel visuomotor task. *Journal of Neuroscience*, 25(39):8833–8842, 2005.

[23] U. Schultheis, J. Jerald, F. Toledo, A. Yoganandan, and P. Mlyniec. Comparison of a two-handed interface to a wand interface and a mouse interface for fundamental 3D tasks. In *Proc. 3DUI 2012*, pages 117–124. IEEE.

[24] A. L. Simeone, E. Velloso, J. Alexander, and H. Gellersen. Feet movement in desktop 3D interaction. In *Proc. 3DUI 2014*, pages 71–74. IEEE.

[25] A. L. Simeone and H. Gellersen. Comparing direct and indirect touch in a stereoscopic interaction task. In *Proc. 3DUI 2015*, pages 105–108.

[26] A. L. Simeone. Indirect touch manipulation for interaction with stereoscopic displays. In *Proc. 3DUI 2016*. IEEE.

[27] P. Song, W. B. Goh, W. Hutama, C.-W. Fu, and X. Liu. A handle bar metaphor for virtual object manipulation with mid-air interaction. In *Proc. CHI 2012*, pages 1297–1306. ACM.

[28] S. Stellmach and R. Dachselt. Investigating gaze-supported multimodal pan and zoom. In *Proc. ETRA 2012*, pages 357–360. ACM.

[29] S. Stellmach and R. Dachselt. Look & touch: gaze-supported target acquisition. In *Proc. CHI 2012*, pages 2981–2990. ACM.

[30] M. Veit, A. Capobianco, and D. Bechmann. An experimental analysis of the impact of touch screen interaction techniques for 3-d positioning tasks. In *Proc. VR 2011*, pages 75–82. IEEE.

[31] R. C. Zeleznik, A. S. Forsberg, and P. S. Strauss. Two pointer input for 3D interaction. In *Proc. SI3D 1997*, pages 115–120. ACM.

[32] S. Zhai and P. Milgram. Quantifying coordination in multiple dof movement and its application to evaluating 6 dof input devices. In *Proc. CHI 1998*, pages 320–327. ACM.