The More the Merrier? Investigating Multimanual Interaction in a 3D Fitts' Law Task

Finn Rietz, Bennet Leiss, Lukas Kleber, Lasse Hafke

Abstract

Modern input devices allow the user to interact with virtual worlds in a natural or sometimes even supernatural, magical way. However, when interacting in a 3D space, users have to move their arms in a wide range, which results in physical exertion – especially if used over longer periods of time. We designed a *Multimanual User Interface (MUI)* allowing the user to reach targets further away without having to move the full distance.

Therefore, we divided the interaction space into smaller sections, each with a virtual pair of hands. Using our MUI, the entire interaction space is reachable but interaction requires only reduced physical exertion. Users are able to activate the different hand pairs in these sections by means of different switch methods. For hands-free control of switching between the virtual hand pairs, interaction methods using foot, voice and gaze were evaluated. We conducted an experiment evaluating the different switch methods with multiple hand pairs. The measured data showed the preference of users to use MUIs over only using one regular hand pair.

Introduction

With the improvement of Head-mounted Displays (HMDs) and compatible input devices, more ways of interaction with virtual 3D-Environments became possible. One way is to track the hands and forearms. The Leap Motion maps hand movements to a virtual pair of hands, which allows a more natural and intuitive interaction in stereoscopic environments. In 3D selection tasks, virtual hand techniques are considered to be the most natural way of interaction (Lubos et al., 2014). Nevertheless, interaction in virtual stereoscopic environment is still fundamentally different from interaction in the real world (Argelaguet and Andujar, 2013). Bruder et al., 2013 dubbed the phenomenon "touching the void", since there is no tactile feedback, and usually no complete representation of the user's body, at best a virtual hand.

$$ID = \left\{ \frac{D}{W} + 1 \right\} \tag{1}$$

Figure 1: Fitt's Law: Index of Difficulty (Fitts, 1992)

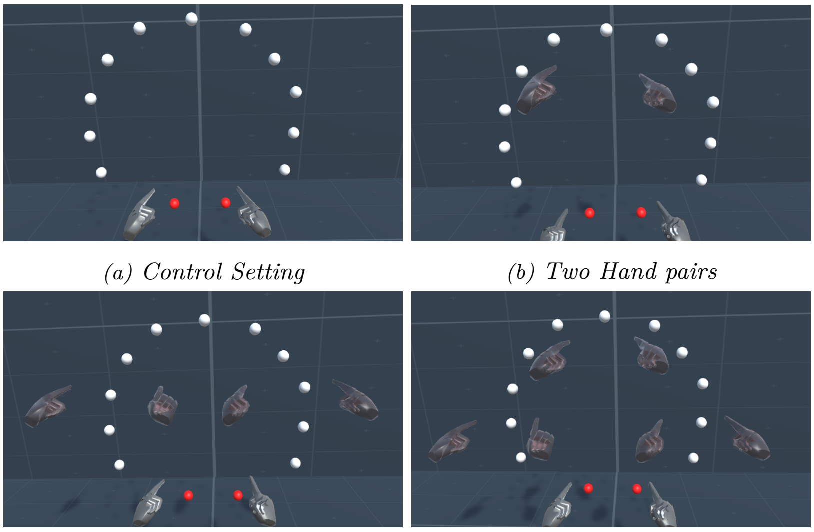
The result is a longer control phase of the user's movement, as there is no supporting feedback to adjust the movement — lengthening the overall movement (Liu et al., 2009). The control phase, where the user adjusts their movement to hit the target, is the second part of each goal-directed movement. The first part is a ballistic phase, which is a fast movement in the general goal-direction (Woodworth, 1899). This concept is valid in the real world and in VR (Nieuwenhuizen et al., 2009). Our approach to improve the comfort of virtual hand interaction is to remove most of the ballistic phase of the movement by placing a pair of hands close to the target and giving the user multiple methods to switch between those pairs. According to Fitt's Law (cf. Fig.1) reducing the distance D equals a less difficult task (Fitts, 1992). However, the question arises whether a lower ID also means less physical demand during the task. Compared to a real environment most movements are more demanding in VR (Bruder et al., 2013). It is unclear how general movement in VR compares to movement utilizing multiple hand pairs.

Related Work

Superimposing the virtual hands with the physical hands has shown to offer optimal performance (Wang and MacKenzie, 1999). However, optimal performance is not guaranteed over prolonged time spans, since fatigue has a negative influence on performance (Lubos et al., 2015), which means that a more comfortable system may offer higher performance for prolonged use. The movements differ in demand, due to the extra cognition users have to invest when operating multiple pairs of hands (Lubos et al., 2014). As users became accustomed to the additional pair of hands, the difference in time between bi- and quadmanual interfaces decreased. During their study, either one or two pairs of hands could be used to reach the target. The second pair was activated by looking in its direction.
We extended this idea to *multimanual interaction*, as our setup consisted of more than one or two pairs of hands. It featured conditions with 1, 2, 3 and 4 pairs of hands, which could be activated by one of four different methods. The methods we designed and tested were speech recognition, line of gaze, a button positioned at the user's foot, and a combination of the button and line of gaze method. Each method allowed for either a direct or a successive selection of hands.

MUI: multimanual user interfaces

Figure 2a shows the control setting, a regular bimanual user interface, where the users have to move their real hands for the full distance. Thus, this setting has the longest



(a) Control Setting *(b) Two Hand pairs*

(c) Three Hand pairs *(d) Four Hand pairs*

Figure 2: The different settings

ballistic phase. Multimanual user interfaces are shown in Figure 2b, Figure 2c and Figure 2d. In those, the user is able to operate multiple pairs of hands, however, only one pair of hands is active at any given time. The inactive hand pairs are displayed semi-transparently and can be activated by one of the four given switch methods. The active hand pair is then displayed fully opaque. This results in the mapping of the real-life hands to the selected pair of hands. Depending on the setting, the number of hands available varies. The general interaction space is divided into smaller interaction sections, each equipped with an inactive pair of hands. This method reduces in general the distance between targets and another pair of hands, shortening the ballistic phase (Lubos et al., 2014).

Switch Methods

We designed four different methods for switching between the present pairs of hands. None restricted the hand usage - they remain freely movable while switching.

line of gaze We implemented the selection method from (Lubos et al., 2014), which allows users to select hand pairs by looking at them. However, we extended the setup by adding neutral areas in-between the hands, which keep the current selection active, while the users are free to look around without losing control of the active hands, removing one of the shortcomings of the original implementation, in which hand pair switches happened too quickly.

Button This switch method is using a button placed on the ground, close to the user's feet. By pushing the button, the selection cycles clockwise one after another through all available hand pairs. This method is heavily dependent on the condition, as one cycle is always as long as the number of hand pairs available.

Line of gaze & button method A combination of aforementioned methods. The hand pair, which would be activated by line of gaze, first gets highlighted before getting selected, giving the user visual feedback. To activate the highlighted pair, the user has to press the button.

Speech recognition Using the Cortana API, we assigned keywords to the different pairs of hands. The pair could then be activated by saying the corresponding keyword. Keywords were chosen by the general direction in which the pair of hands were located. Note that the entire experiment was conducted in German and therefore the German words for "up, down, left, right" were used. There was a slight input delay ($< 1\,s$) to process and recognize the spoken words.

Experiment

Our experiment consisted of 13 tasks. Each task featured a different combination of hand pairs and switch methods. We had 4 switch methods and 3 settings (2, 3 and 4 pairs of hands), resulting in 12 tasks, and an additional task with just one hand pair without any switch method. After the completion of each task, participants answered four questionnaires, including the Simulator Sickness Questionnaire (Kennedy et al., 1993), the NASA-TLX (Hart and Staveland, 1988), the System Usability Scale (Brooke et al., 1996) and a variation of the Borg-Questionnaire for perceived exertion (Lindegård et al., 2012). At the end of our 2.5 hours experiment, we collected the participant's demographics, prior VR-experience and their overall favorite switch method, for short-time tasks, as well as for long-time tasks.
During one task participants had to complete 44 trials in VR. The VR environment consisted of a circle of targets in front of the participant, a dark ground plane, and a dark checkered background. The targets were small, white spheres. While these elements are in the background and do not improve the performance, research indicated that they are helpful for the user, e.g. orientation and depicting scale (Teather and Stuerzlinger, 2014). Each task the participants had three test trials in order to get familiar with the switch method. Each trial started with the two bottom spheres (the reset spheres) being highlighted in red. The participant then had to touch both spheres simultaneously. After two seconds of continuous contact, another sphere in the circle of targets turned red. The participant then had to reach the highlighted target. Upon completion, the bottom two spheres turned red again and a new trial started. The circle consists of 11 targets that are arranged symmetrical (excluding the two reset spheres). Over one task, each sphere, except the reset spheres, was highlighted randomly four times, resulting in 44 trials. Therefore we had a total of $44 * 13 = 572$ trials per participant.
At the start of the experiment, participants were instructed to be as fast and precise as possible.

We evaluated the following hypotheses:

H1: The participants prefer using multiple hands over only one hand pair.

H2: Having only one hand pair for completing the task is physically more demanding compared to having multiple pairs of hands available.

H3: Participants are slower using multiple pairs of hands compared to bimanual interaction.

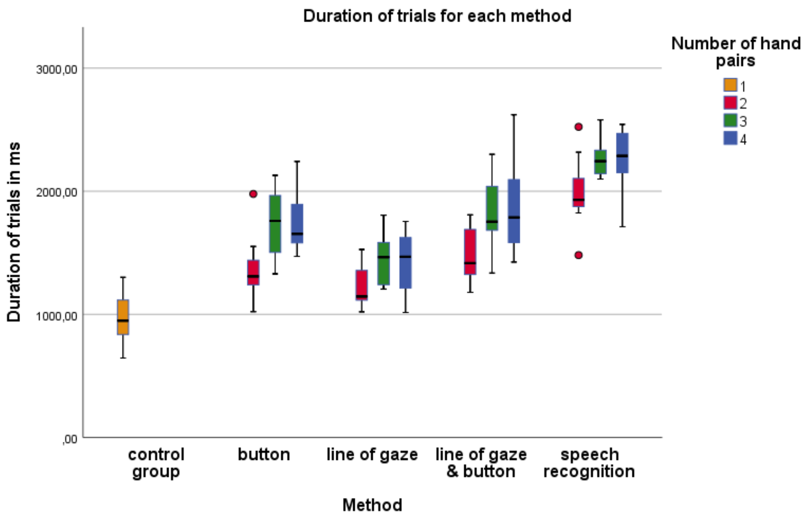


Figure 3: The median of the duration of each method

Participants

11 participants took part in our experiment. Ten of them were male, one was female. The age varied from 19 to 30 ($M = 22.1$). All the participants were right-handed. None of the participants had any medical conditions regarding their shoulders or arms. Nine of the participants had already made experiences with virtual reality, two participants had not made any previous experiences with virtual reality. Eight participants had never used a Leap Motion before, three had made experiences with the Leap Motion.

Materials

Users wore an Oculus Rift CV1 HMD, tracked by the Oculus Sensor. The distance measured between the Oculus HMD and Sensor was 1 meter. The Leap Motion Controller was attached to the table in the optimal calibrated position for each participant (cf. Figure 5). The 3D Environment (cf. Figure 2) was rendered with Unity3D on a computer with an Intel Core I7 3.4GHz CPU and Nvidia GeForce GTX 780Ti, at more

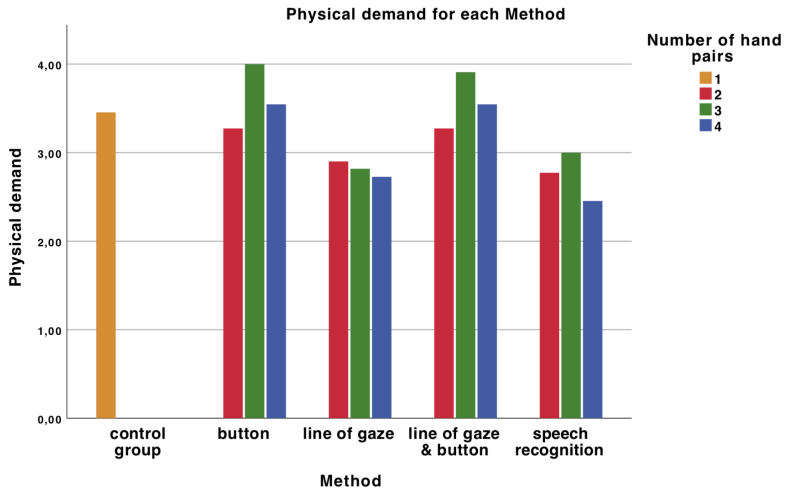


Figure 4: The median of the TLX physical demand of each method

than 90 frames per second the entire time. For the button condition, a Griffin PowerMate USB was used. Both the chair and the table, where the participants were seated at, were adjusted in height to account for different statures. The participants were seated according to the Guidebook of Human Factors (Sanders and McCormick, 1993). With a special focus on the 90 °-bend in their knees, an upright back, elbows on the table, and the Leap Motion positioned at the middle of the participant's palm, while their arms are flat on the table. This brought the participant's hands in the optimal tracking position above the Leap Motion, while maintaining a 90°-bend in their elbow during the trials.

Methods

For our experiment, we chose a $3 \times 4 \times 44 + 44$ within-subject design. The factors were the number of hand pairs available (setting: 1, 2, 3, 4) and the switch method used for selecting the hand pairs (button, line of gaze, line of gaze & button, speech recognition). The setting and control was ordered according to a Latin square. During each setting, the participant got to use all switch methods. We tested every switch method in every setting. Note that, only after the completion of one setting with all switch methods, the next setting would be tested. As an example participants first used all switch methods in a 4-pair setting, before using them again in a 3- and 2-pair setting. The participant could freely choose which pair of hands to activate, with the only restriction being that it was only possible to touch the reset targets with the bottom hand pair.

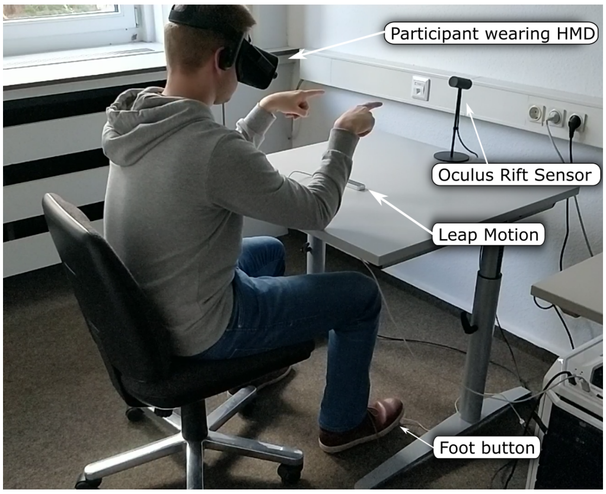


Figure 5: Illustration of the experiment setup

Results

The results were evaluated at the 5% significance level. Since the results were not normally distributed, they were evaluated using non-parametric Two-Way Friedman ANOVAs. The times needed to complete the task were significantly different, $\chi^2(12) = 108.563,\ p < .001$. We did three Wilcoxon signed-rank tests, which showed that the users were slower with two, three and four pairs of hands (independent of the switch method) compared to only one pair of hands ($Z = -2.934,\ p < .05$). The result of the measurement of the time required by the participants to complete the tasks is shown in Figure 3.

The results from the "Mental Demand" part of the NASA-TLX showed that there were significant differences ($\chi^2(12) = 19.234,\ p < .05$) regarding the mental demand for the combinations of settings and switch methods. A Wilcoxon signed-rank test showed that the switch method line of gaze & button had been rated significantly more demanding than the control setting for any number of available hand pairs. The switch method line of gaze had only been rated significantly more demanding in the setting with three available hand pairs ($Z = -2.232,\ p < .05$).

For the "Physical Demand" component we found no significant differences regarding the physical demand ($\chi^2(12) = 9.199,\ p = .686$). However, we found significant differences when comparing the Borg questionnaire data ($\chi^2(12) = 27.666,\ p < .05$). Post-Hoc tests showed that the line of gaze condition was significantly less demanding than other conditions. Compared to the control condition, the a Bayesian T-Test showed no significance, but anecdotal evidence for H2 ($T(10) = 2.014, p = .72. BF_{01} = .891$). After completing every task, the participants were asked to fill out a subjective questionnaire containing questions about the switch methods. When asked which switch method they liked most overall, 63.6% (7) of the participants said they would prefer the switch method line of gaze. 18.2% (2) preferred having only one hand pair and no switch

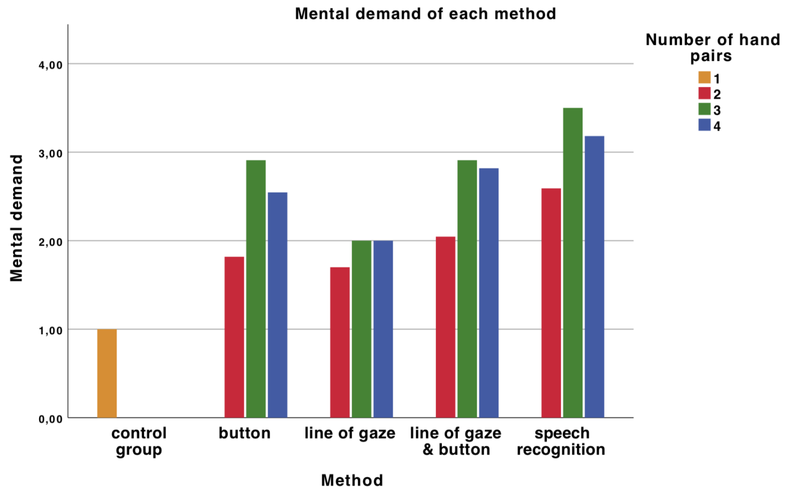


Figure 6: The median of the mental demand of each method

method. One participant favored the speech recognition method, one last the line of gaze & button method.
When the participants were asked specifically which method they would use for short tasks, 54.5% (6) participants said they would choose the method line of gaze, 27.3% (3) said they would choose one pair of hands without any switch methods, 9.1% (1) participant chose speech recognition and another one chose the button method.
When the participants were asked specifically which method they would use for long tasks, 63.6% participants said they would choose the method line of gaze, 18.2% said they would prefer line of gaze & button and another 18.2% said they would choose speech recognition.

Discussion

The results from the subjective questionnaire at the end of the experiments show that the participants generally preferred using multiple pairs of hands. Specifically for long tasks participants wanted to use multiple pairs of hands combined with the switch method line of gaze. For short tasks, some participants preferred having one hand pair. This might be, because short tasks are less physically demanding and therefore the increased mental demand of controlling multiple hands is not beneficial. This partly confirms our hypothesis H1.
Results from the NASA-TLX showed that there was no significant difference in the physical demand, results from the Borg-questionnaire showed only anecdotal evidence for a difference between the best switch method (line of gaze) and the control condition. Thus we cannot confirm our hypothesis H2. A further study with more participants

would have to evaluate H2.
As seen in Figure 6, the mental demand for most switch methods was higher for three hand pairs than for four. We argue, that this is because in the setting with three pairs of hands available, the participants had to decide which hand pair they use for targets located in the upper center. For the setting with two and four pairs of hands, this decision omitted, because there was a section especially assigned to the top center. (cf. Figure 2c, Figure 2d).
The time needed to complete the tasks was higher when participants were to use multiple pairs of hands. The time greatly varied depending on the combination of the setting and which switch method were used. Overall, there was no combination of settings and switch methods where the mean duration of trials was lower than the mean duration of trials in the control setting. This confirms our hypothesis H3 as well as corresponding with the findings of (Lubos et al., 2014). Whereas the lower speed indicates a lower performance, the user's preference for multiple hand pairs indicates that a further evaluation of MUIs should still be considered.

Conclusion

As an extension of quadmanual user interfaces (Lubos et al., 2014), we introduced multimanual user interfaces (MUIs). To evaluate our switch methods, we let participants test the methods in each setting. From our analysis, we conclude that the best switch method is line of gaze, as it was the fastest overall and was picked as the favored switch method by most of the participants. We evaluated physical and mental demand for each switch method and compared it with a bimanual user interface. This comparison indicated that controlling multiple pairs of hands is mentally more demanding.
Further research in this topic could take a deeper look at how to divide the interaction space into smaller sections, for example by dynamically creating more pairs of hands, so it is not necessary to have them in fixed positions in the scene. Our proposed switch methods could be extended by new and different methods, but also be enhanced by, i.e. using a more ergonomic foot button. For example a gas pedal, where the users can rest their foot on could reduce the physical exertion, especially for long tasks. Even though there were no significant differences in physical demand, participants preferred using multiple pairs of hands and having a higher mental demand than having only one pair of hands available and having a lower mental demand. Once more, we want to remark that when asked which switch method the participants would choose for long tasks, not a single participant wanted to use the regular hand pair. This leaves a very promising foundation for future research on the topic of MUIs.

References

Argelaguet, F. & Andujar, C. (2013). A survey of 3D object selection techniques for virtual environments. *Computers & Graphics*, *37*(3), 121–136.

Brooke, J. et al. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, *189*(194), 4–7.

Bruder, G., Steinicke, F., & Stuerzlinger, W. (2013). Touching the void revisited: Analyses of touch behavior on and above tabletop surfaces. In *IFIP Conference on Human-Computer Interaction* (pp. 278–296). Springer.

Fitts, P. M. (1992). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology: General*, *121*(3), 262.

Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139–183). Elsevier.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, *3*(3), 203–220.

Lindegård, A., Wahlström, J., Hagberg, M., Vilhelmsson, R., Toomingas, A., & Tornqvist, E. W. (2012). Perceived exertion, comfort and working technique in professional computer users and associations with the incidence of neck and upper extremity symptoms. *BMC musculoskeletal disorders*, *13*(1), 38.

Liu, L., van Liere, R., Nieuwenhuizen, C., & Martens, J.-B. (2009). Comparing aimed movements in the real world and in virtual reality. In *Virtual Reality Conference, 2009. VR 2009. IEEE* (pp. 219–222). IEEE.

Lubos, P., Bruder, G., & Steinicke, F. (2014). Are Four Hands better than Two? Bimanual Interaction for Quadmanual User Interfaces. In *Proceedings of the 2nd ACM Symposium on Spatial User Interaction (SUI)* (pp. 123–126). ACM Press.

Lubos, P., Bruder, G., & Steinicke, F. (2015). Influence of Comfort on 3D Selection Task Performance in Immersive Desktop Setups. *Journal of Virtual Reality and Broadcasting*, *12(2015)*(2).

Nieuwenhuizen, K., Liu, L., van Liere, R., & Martens, J.-B. (2009). Insights from dividing 3D goal-directed movements into meaningful phases. *IEEE computer graphics and applications*, *29*(6), 44–53.

Sanders, M. S. & McCormick, E. J. (1993). *Human factors in engineering and design.* McGRAW-HILL book company.

Teather, R. J. & Stuerzlinger, W. (2014). Visual aids in 3D point selection experiments. In *Proceedings of the 2nd ACM symposium on Spatial user interaction* (pp. 127–136). ACM.

Wang, Y. & MacKenzie, C. L. (1999). Effects of orientation disparity between haptic and graphic displays of objects in virtual environments. In *INTERACT* (Vol. 99, pp. 391–398).

Woodworth, R. S. (1899). Accuracy of voluntary movement. *The Psychological Review: Monograph Supplements*, *3*(3), i.