Pointing by Gaze, Head, and Foot in a Head-Mounted Display

Katsumi Minakata
Technical University of Denmark, Kgs. Lyngby, Denmark
katmin@dtu.dk

John Paulin Hansen
Technical University of Denmark, Kgs. Lyngby, Denmark
jpha@dtu.dk

I. Scott MacKenzie
York University, Toronto, Ontario, Canada
mack@cse.yorku.ca

Per Bækgaard
Technical University of Denmark, Kgs. Lyngby, Denmark
pgba@dtu.dk

Vijay Rajanna
Sensel, Inc., Sunnyvale, California, USA
vijay.drajanna@gmail.com

Figure 1: (A) HTC VIVE head-mounted display (HMD) equipped with the Pupil Lab's binocular gaze tracking unit was used for gaze input. The Fitts' law multi-directional selection task that the participant sees inside the headset is shown on the monitor. The headset tracks orientation when used for head input. (B) A 3DRudder foot mouse with 360° movement control was used for foot input. Tilting the 3DRudder moves the cursor in the tilt direction.

ABSTRACT

This paper presents a Fitts' law experiment and a clinical case study performed with a head-mounted display (HMD). The experiment compared gaze, foot, and head pointing. With the equipment setup we used, gaze was slower than the other pointing methods, especially in the lower visual field. Throughputs for gaze and foot pointing were lower than mouse and head pointing and their effective target widths were also higher. A follow-up case study included seven participants with movement disorders. Only two of the participants were able to calibrate for gaze tracking but all seven could use head pointing, although with throughput less than one-third of the non-clinical participants.

CCS CONCEPTS

• **Human-centered computing → Pointing devices**.

KEYWORDS

Fitts' law, ISO 9241-9, foot interaction, gaze interaction, head interaction, dwell activation, head-mounted displays, virtual reality, hand controller, accessibility, disability

1 INTRODUCTION

Pointing in virtual reality (VR) is an important topic for interaction research [Sherman and Craig 2002]. In addition to hand-controllers,

head- and gaze-tracking are now available in VR headsets, thus presenting a number of questions for HCI researchers: How big should targets be if accessed using one of these input methods? How fast can a button be activated in VR?

In scenarios with situation-induced impairments, a user's hands may be unavailable for pointing or text entry [Freedman et al. 2006; Kane et al. 2008; Keates et al. 1998]. For example, a surgeon in need of medical image information during surgery cannot use his/her hands to interact with the display, since the hands are occupied and sterilized [Elepfandt and Grund 2012; Hatscher et al. 2017]. We expect HMDs to support such work tasks in the future, and therefore we need to understand interaction with UI elements and potential hands-free input modalities for HMDs.

According to a 2016 disability report, 7.1% of the United States population have an ambulatory disability that restricts the movements of the limbs [Capio et al. 2018], making it difficult or even impossible to work on a computer using a mouse or keyboard. Individuals without full hand or finger control rely on alternatives, for instance gaze, to accomplish simple tasks on a computer or to communicate with others [Edwards 2018; Majaranta et al. 2011]. Such individuals would benefit much from hands-free alternatives when interacting in HMDs. Accessibility in VR is particularly relevant, since this medium provides opportunities for people with mobility constraints to experience places and events not accessible in real life. Also, products, buildings, and surroundings can be modeled in VR to evaluate their accessibility [Di Gironimo et al. 2013; Hansen et al. 2019; Harrison et al. 2000].

These discussions bring forth questions on which hand-free input methods – gaze, head, or foot – work well when interacting with an HMD.

For half a century, Fitts' law has been used to quantitatively evaluate pointing methods and devices, such as the mouse, stylus, hand controllers, track pads, head pointing, and gaze pointing [Soukoreff and MacKenzie 2004]. The use of a standard procedure makes it possible to compare performance across pointing methods and across studies. However, only a few studies have applied this procedure to HMDs (e.g., [Qian and Teather 2017]).

A Fitts' law point-select task is easy to do. It can be performed by children and novice computer users, which is not the case for text entry. Therefore, such tasks hold potential for standardized assessments of input methods for individuals with motor and/or cognitive challenges. What should the target size be for accurate pointing? How many selections can occur per minute? The last issue also bears on communication speed while typing.

Virtual reality applications inherently display the world in 3D by positioning objects at different depths. Most VR applications also introduce motion in the user's field of view to enhance immersion. However, our research did not manipulate depth or introduce motion cues. Our goal was to establish a baseline of input characteristics without depth cues or background motion. Future studies will include depth cues and motion as independent variables to study how they interact with target size, movement amplitude, and pointing method. Additionally, we only examined one equipment setup for each pointing method and, as a practical consideration, we only included two target widths and two movement amplitudes. Finally, we only consider the input methods as uni-modal, leaving out their use in combination.

The main contributions of this paper are (i) comparing three hands-free input methods (gaze, head, foot) for HMDs and (ii) presenting a standardized test procedure for VR and accessible computing.

2 EVALUATION USING FITTS' LAW

Our evaluation used the Fitts' law procedure in the ISO 9241-9 standard for non-keyboard input devices [ISO 2000]. The most common task is two-dimensional with targets of width W arranged around a layout circle. Selections proceed in a sequence moving across and around the circle (see Figure 2). Each movement covers an amplitude A – the diameter of the layout circle (which is the distance between the centers of two opposing targets). The movement time (MT, in seconds) is recorded for each trial and averaged over the sequence.
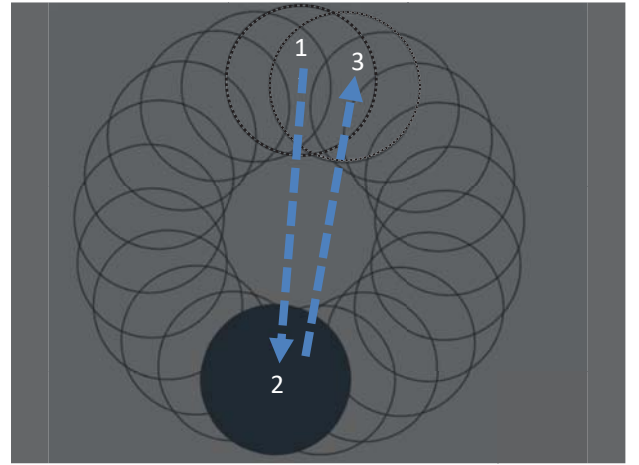
Figure 2: Two-dimensional target selection task in ISO 9241-9. After pointing at a target (1) for a prescribed time (e.g., 300 ms), selection occurs and the next target appears in solid (2). When this target is selected, the next target (3) is highlighted. Arrows, numbers, and target annotations were not shown in the actual task display.

The difficulty of each trial is quantified using an index of difficulty (ID, in bits) and is calculated from A and W as

$$ID = log_2\left(\frac{A}{W} + 1\right). \tag{1}$$

The main performance measure in ISO 9241-9 is throughput (TP, in bits/second) which is calculated over a sequence of trials as the ID:MT ratio:

$$TP = \frac{ID_e}{MT}. \tag{2}$$

The standard specifies calculating throughput using the effective index of difficulty (ID_e). The calculation includes an adjustment for accuracy to reflect the spatial variability in responses:

$$ID_e = log_2\left(\frac{A_e}{W_e} + 1\right) \tag{3}$$

where

$$W_e = 4.133 \times SD_x. \qquad (4)$$

The term SD_x is the standard deviation in the selection coordinates computed over a sequence of trials. Selections are projected onto the task axis, yielding a single normalized x-coordinate for each trial. The factor 4.133 adjusts the target width for a nominal error rate of 4% under the assumption that the selection coordinates are normally distributed. The effective amplitude (A_e) is the actual distance traveled along the task axis. (See [MacKenzie 2018] for additional details.)

Throughput is a potentially valuable measure of human performance because it embeds both the speed and accuracy of participant responses. Comparisons between studies are therefore possible, with the proviso that the studies use the same method in calculating throughput.

3 RELATED WORK

The ISO 9241-9 standard for pointing devices has been used to compare methods to control a mouse cursor using a numeric keypad [Felzer et al. 2016]. In a user study with non-disabled participants, throughput values were about 0.5 bits/s among the methods compared. A user from the target community (who has the neuromuscular disease Friedreich's Ataxia) achieved throughput values around 0.2 bits/s, reflecting a lower performance. In a second study using the same ISO 9241-9 standard, this individual participated in a comparison of click actuation methods for users of a motion-tracking mouse interface [Magee et al. 2015]. In this case, mean throughput values of about 0.5 bits/s were achieved. Keates et al. [2002] found a significant difference in throughput with the mouse between able-bodied users (4.9 bits/s) and motion-impaired users (1.8 bits/s).

The ISO-9241-9 standard has also been used for gaze and head tracking, but not involving individuals with motor challenges. Zhang and MacKenzie [2007] measured a throughput of 2.3 bits/s for long dwell-time selection (750 ms) with gaze and 3.1 bits/s for short dwell-time selection (500 ms). For head tracking systems, De Silva et al. [2003] reported a throughput of 2.0 bits/s, and Roig-Maimó et al. [2018] reported 1.3 bits/s when interacting with a tablet by head movements detected by the device's front-facing camera.

3.1 Eye-gaze and head pointing with HMDs

Pointing at objects in VR is mainly achieved through an external controller and the ray-casting method [Mine 1995]. HMDs such as the HTC VIVE, Oculus Rift, Samsung Gear VR, provide the user with a hand-held controller to interact with objects. This seriously limits the user's mobility where controllers are tracked within a fixed space. Also, the user's hands are occupied in operating the controllers.

Lubos et al. [2014] conducted a Fitts' Law experiment in VR with direct selection by hand positioning in the user's arm reach. They reported an average throughput of 1.98 bits/s.

Qian and Teather [2017] compared head-only input to eye+head and eye-only inputs in a click-selection task in a FOVE HMD, which was the first commercially available headset with built-in gaze tracking. They used a 3D background decoration and volume surface on targets. Head-only input yielded a significantly lower error rate

(8%) compared to eyes-only input (40%) and eyes+head inputs (30%). Consequently, head-only input achieved the highest throughput (2.4 bits/s) compared to eyes-only input (1.7 bits/s) and eyes+head input (1.7 bits/s). Task completion time and subjective ratings also favoured the head-only input method.

Hansen et al. [2018] compared pointing with the head to pointing with gaze and mouse inputs in a Fitts' law experiment with the same FOVE HMD, but within a neutral 2D (i.e., "flat") scene. Dwell (300 ms) and mouse click were used as selection methods. Overall, throughput was highest for the mouse (3.2 bits/s), followed by head pointing (2.5 bits/s) and gaze pointing (2.1 bits/s).

Blattgerste et al. [2018] investigated if using head or gaze pointing is an efficient pointing method in HMDs. The authors tested pointing and selection on two interfaces: a virtual keyboard and a selection menu typical in VR applications. Aiming with gaze significantly outperformed aiming with the head in terms of time-on-task. When interacting with the virtual keyboard, pointing with the eyes reduced the time-on-task by 32% compared to pointing with the head. Similarly, when working on the menu, gaze input reduced the aiming time by 12% compared to head input. However, Blattgerste et al. [2018] used point-and-click tasks on a keyboard and a menu, also the field-of-view was varied in the experiment. Because this was not a standard Fitts' law task, the constraints of Fitts' law experiment were not met.

Kytö et al. [2018] investigated gaze and head pointing in Augmented Reality (AR) and they also found gaze pointing faster than head pointing, while head pointing allowed the best target accuracy.

3.2 Foot pointing with HMDs

While foot input has been used for tasks such as moving or turning in HMDs for gaming [Matthies et al. 2014], point-and-select interactions in HMDs with foot input are yet not explored. There is prior research on foot input in desktop settings for point-and-select interactions, however [Velloso et al. 2015]. Pearson and Weiser [1986] first demonstrated an input device called "Moles" that functions similar to a mouse. Pakkanen and Raisamo [2004] presented a foot-operated trackball used for non-accurate pointing tasks. Dearman et al. [2010] demonstrated tapping on a foot pedal as a selection trigger for text entry on a mobile device.

Recently, foot input has been combined with gaze input in a multi-modal interaction setup. Göbel et al. [2013] first combined gaze input with foot input for secondary navigation tasks like pan and zoom in zoomable information spaces. Furthermore, gaze input has been combined with foot input, achieved through a wearable device, for precise point-and-click interactions [Rajanna and Hammond 2016] and for text entry [Rajanna 2016]. Lastly, Hatscher et al. [2017] demonstrated how a physician performing minimally-invasive interventions can use gaze and foot input to interact with medical image data presented on a display.

3.3 Summary

The ISO 9241-9 standard procedure has been applied in numerous studies of input devices, some involving users with motor disabilities. Foot-assisted point-and-click interaction has been examined in some studies, but none involved HMDs. Three studies compared head and gaze input in HMDs, two using a FOVE HMD found head

throughput values superior to gaze values, while the last study, applying a high-precision gaze tracker, found gaze time-on-task to outperform head input with rather large margins.

4 METHOD

Our methodology includes an experiment, and a case study. These are presented together in the following sections.

4.1 Participants

The experiment included 27 participants (M = 25 yrs; 17 male, 10 female). The mean inter-pupillary distance was 60.3 mm (SD = 3.15 mm). Most (45%) had tried HMDs several times before, and 28% only one time before. Some (45%) had previously tried gaze interaction. None had tried the foot-mouse. All participants had normal or corrected-to-normal vision.

The case study included 8 individuals from a care center for wheelchair users (M = 44 yrs; 6 male, 2 female). All had a movement disorder such as cerebral palsy.

4.2 Apparatus

An HTC VIVE HMD was used. The HMD has a resolution of 2160 × 1200 px, renders at a maximum of 90 fps, and has a field of view of 110° visual angle. A Pupil Labs binocular eye-tracking add-on system for the HTC VIVE was installed and collected data at 120 Hz. The vendor specifies a gaze-accuracy and precision of < 1° and 0.08° visual angle, respectively. The HMD weighs 520 grams and has IR-based position tracking plus IMU-based orientation tracking.

A Logitech corded M500 mouse was used for manual input. A foot-mouse was used as an input device for foot pointing; it is capable of controlling a cursor on a monitor via foot movements. The foot-mouse is made by 3DRudder and is a "foot powered VR and gaming motion controller" with the capability of 360° movement. See Figure 1.

Software to run the experiment was a Unity implementation of the FittsTaskTwo software (now GoFitts) developed by MacKenzie[1]. The Unity version[2] includes the same features and display as the original; that is, with circular targets presented on a flat 2D-plane. See Figure 2.

4.3 Procedure

The participants were greeted upon arrival and were asked to sign a consent form after they were given a short explanation of the experiment. Next, participants' inter-pupillary distance was measured and the HTC VIVE headset was adjusted accordingly. Then, participants were screened for color blindness with an online version of the Ishihara test [Ishihara 1972]. This was done to ensure that they could distinguish colored targets from a colored background.

For the experiment, participants completed a baseline mouse condition as their first pointing condition. Since the mouse is familiar this serves as an easy introduction to the ISO-9241-9 test procedure. The other three conditions were head-position, foot-mouse, and gaze-pointing, with order counterbalanced according to a Latin square. Selection was performed by positioning the cursor inside

---

[1]available at http://www.yorku.ca/mack/GoFitts/
[2]available at https://github.com/GazeIT-DTU/FittsLawUnity

the target for the prescribed dwell time. Based on the findings of Majaranta et al. [2011], a dwell time of 300 ms was chosen .

For each pointing method, four levels of index of difficulty (*ID*) were tested, composed of two target widths (50 pixels, 75 pixels) and two movement amplitudes (80 pixels, 120 pixels) resulting in layout circle centres approximately 2.5° and 3.8° from the centre. The target widths spanned visual angles of about 3.1 ° and 4.7°, respectively.

Spatial hysteresis was set to 2.0. When targets were entered, their size doubled, while visually remaining constant. For each of the four *ID*s, 21 targets were presented for selection. As per the ISO 9241-9 procedure, targets were highlighted one-by-one in the same order for all levels, starting with the top position (12 o'clock). When this target was selected, a target at the opposite side was highlighted (approximately 6 o'clock), then when activated a target at 1 o'clock was highlighted, and so on, moving clockwise. See Figure 2. The first target at 12 o'clock is not included in the data analysis in order to minimize the impact from initial reaction time.

The target layout was locked in world space and did not move with head motion. The color of the targets were blue and the color of the background was a dark brown. The pointer (i.e., cursor) was visible at all times and appeared as a violet-red dot. For the mouse pointing condition, the cursor was the mouse position on screen. For gaze pointing, this was the location the participants' looked at on the monitor, as defined by the intersection of the two gaze vectors from the centre of both eyes on the target plane. For head pointing, the cursor was the central point of the headset projected directly forward.

Failing to activate 20% of the targets in a 21-target sequence triggered a repeat of that sequence. Sequences were separated, allowing participants a short rest break as desired. Additionally, they had time to rest for a couple of minutes when preparing for the next pointing method.

Before testing the gaze pointing method, participants performed a gaze calibration procedure. This consisted of a bull's eye target that moved to one of six locations in the field of view. Participants were asked to fixate on the target each time it moved to one of the six locations. This sequence was repeated twice (i.e., once for calibration and once for verification of the calibration). The verification sequence was further conducted before the last pointing condition and was collected in order to determine the calibration stability and quality from the beginning to the end of the experiment. Completing the full experiment took approximately 30 minutes for each participant.

Upon the end of the experiment, participants completed a questionnaire soliciting demographic information and ratings of the different pointing methods for mental and physical workload and comfort. In addition, participants were encouraged to submit their impressions of each method and were asked to rank the pointing methods from most preferred to least preferred.

4.4 Design

The design and independent variables (with levels) for each phase of evaluation were as follows:

Experiment − 4 × 2 × 2 within-subjects:

- Pointing method (mouse, gaze, head-position, foot-mouse)

- Target amplitude (80 pixels, 120 pixels)
- Target width (50 pixels, 75 pixels)

Case study – 2 × 2 × 2 within-subjects:

- Pointing method (head-position, gaze)
- Target amplitude (80 pixels, 120 pixels)
- Target width (50 pixels, 75 pixels)

Pointing method was the primary independent variable. Target amplitude (A) and target width (W) were included to ensure the conditions covered a range of task difficulties (ID), which were suitable for an HMD, rather than a screen monitor. With the A-W values above, the IDs ranged from $\log_2(\frac{80}{75} + 1) = 1.05$ bits to $\log_2(\frac{120}{50} + 1) = 1.77$ bits.

The dependent variables were time to activate (ms), throughput (bits/s), and effective target width (pixels), calculated according to the procedures in ISO 9241-9. Errors were not possible because all targets were activated via dwell selection. In addition, we measured pupil size at 120 Hz throughout the experiment. Analysis of the pupil data are reported elsewhere [Bækgaard et al. 2019].

For each sequence, 21 trials were performed. There were four such sequences in a block, one sequence for each ID condition. Four such blocks were performed for each pointing method. There were 27 participants in total. In all, 27 Participants × 2 Pointing Methods × 2 Target Amplitudes × 2 Target Widths yielded 4536 trials in the experiment (27 Participants × 4 Pointing Methods × 2 Target Amplitudes × 2 Target Widths). Trials with an activation time greater than two SDs from the mean were deemed outliers and removed. Using this criterion, 128 out of 9072 trials (1%) in the experiment were removed.

5 RESULTS

Repeated-measures ANOVAs were used on the data for the two phases of evaluation (experiment and case study).

For the experiment, two additional analyses were executed on the time-to-activate dependent variable. That is, two median-split, dichotomous, independent variables were created for the accuracy and precision of the eye-tracking variables: level of accuracy (low vs. high) and level of precision (low vs. high). These measured independent variables were utilized to determine whether accuracy and precision interacted with any of the primary independent variables. For the accuracy analysis, neither the main effect of accuracy nor the interactions involving accuracy were statistically significant ($F(1, 25) = 0.34$, $p = .57$). The same pattern was found in the precision analysis ($F(1, 25) = 1.46$, $p = .24$). Thus, the accuracy and precision results were not further explored.

Moreover, subjective measures were collected at the end of testing. That is, participants rated the pointing methods for mental workload, physical workload, and comfort. Ratings ranged from 1 (lowest) to 10 (highest). These measures were analyzed with an omnibus Friedman test followed by post hoc Wilcoxon signed-rank tests. All post hoc tests included a Bonferroni correction to mitigate type-I errors.

5.1 Experiment

5.1.1 Time to Activate. The grand mean for time to activate per trial was 785 ms. From fastest to slowest, the means were 697 ms (mouse), 721 ms (head pointing), 823 ms (foot pointing), and finally

898 ms (gaze pointing). In terms of target amplitude, the 80-pixel condition ($M = 708$ ms) yielded lower mean times to activate relative to the 120-pixel condition ($M = 861$ ms). Regarding the target-width factor, the 75-pixel condition ($M = 719$ ms) yielded lower mean times to activate relative to the 50-pixel condition ($M = 851$ ms). See Figure 3.
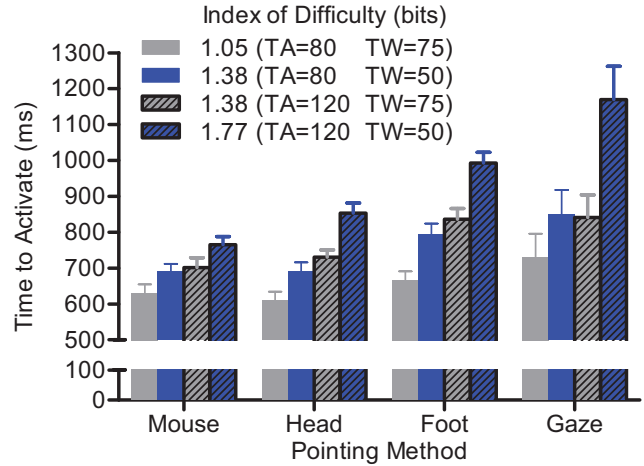
Figure 3: Time to Activate (ms) by pointing method and index of difficulty represented as target amplitude (TA) and target width (TW) in pixels. Error bars denote one standard error of the mean.

The main effect on time to activate was significant for pointing method ($F(3, 78) = 8.95$, $p = .00004$), target amplitude ($F(1, 26) = 123.96$, $p < .00001$), and target width ($F(1, 26) = 92.24$, $p = .00001$). As well, the interaction effects were significant for pointing method by target amplitude ($F(3, 78) = 4.57$, $p = .005$), pointing method by target width ($F(3, 78) = 7.98$, $p = .0001$), and target amplitude by target width ($F(3, 78) = 14.43$, $p = .0008$).

To check if the mean time to activate differed as a function of target orientation, a polar plot was created. The first selection at the 0° target orientation was excluded from the analysis. See Figure 4. The gaze-pointing conditions exhibited an upper hemifield bias with a higher time to activate at the lower-target orientations relative to the lateral-target orientations. The other pointing methods yielded no orientation dependence with time to activate.

5.1.2 Throughput. The grand mean for throughput was 3.10 bits/s. The means by pointing method were 3.87 bits/s (mouse), 3.40 bits/s (head-position), 2.58 bits/s (foot-mouse), and 2.55 bits/s (gaze). In terms of target amplitude, the 80-pixel condition ($M = 3.16$ bits/s) yielded a higher mean throughput relative to the 120-pixel condition ($M = 3.04$ bits/s). Regarding target width, the 75-pixel condition ($M = 3.22$ bits/s) yielded higher a mean throughput relative to the 50-pixel condition ($M = 2.98$ bits/s). See Figure 5.

As the target amplitude decreased from 120 to 80 pixels, throughput increased; however, this pattern significantly reversed only for the mouse pointing condition. Similarly, as the target width increased from 50 to 75 pixels, throughput increased; however,
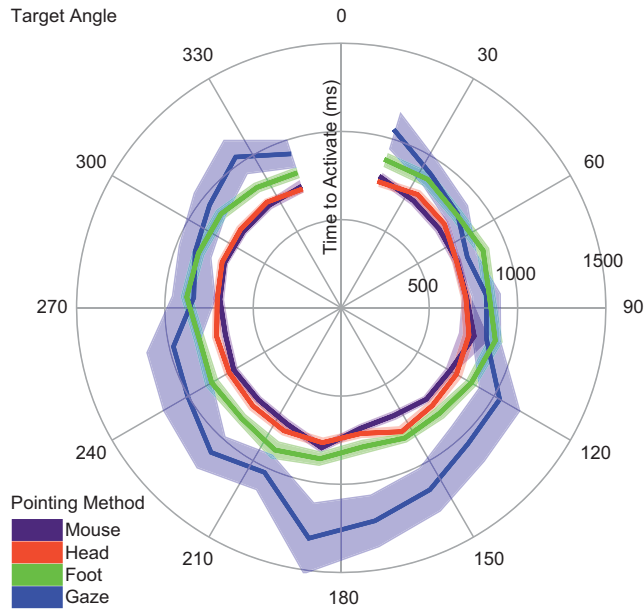
Figure 4: Polar plot showing the distribution of the mean time to activate according to target orientation. One standard error of the mean is depicted with a shaded area around the mean.
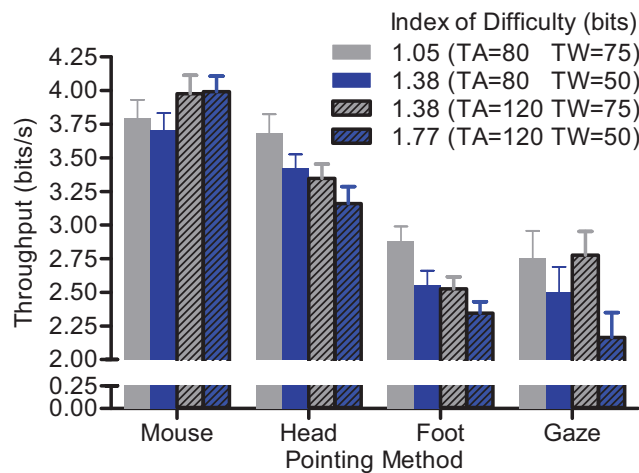

Figure 5: Throughput (bits/s) by pointing method and index of difficulty represented as target amplitude (TA) and target width (TW) in pixels. Error bars denote one standard error of the mean.

this pattern was not significantly different for the mouse pointing condition.

The main effects on throughput were significant for pointing method ($F(3, 78) = 57.913$, $p < .00001$), target amplitude ($F(1, 26) = 6.94$, $p < .014$), and target width ($F(1, 26) = 39.74$, $p < .00001$). These main effects were qualified by significant interaction effects between pointing method and target amplitude ($F(3, 78) = 3.98$, p

= .011), and between pointing method and target width ($F(3, 78) = 3.83$, $p = .013$).

5.1.3 Effective Target Width. The grand mean for effective target width was 50.9 pixels. The mean effective target width for mouse pointing was 35.4 pixels, followed by head-position pointing at 46.1 pixels, then gaze pointing at 60.9 pixels, and finally foot-mouse pointing at 61.4 pixels. In terms of target amplitude, the 80-pixel condition ($M = 48.2$ pixels) yielded smaller mean effective target width relative to the 120-pixel condition ($M = 53.7$ pixels). Regarding target width, the 75-pixel condition ($M = 55.5$ pixels) yielded larger mean effective target widths relative to the 50-pixel condition ($M = 46.4$ pixels; see Figure 6).
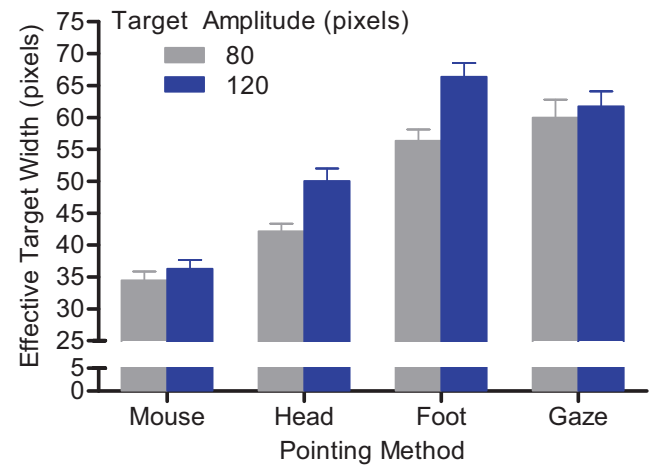

Figure 6: Effective target width (pixels) by pointing method and target amplitude. Error bars denote one standard error of the mean.

The main effects on effective target width were significant for pointing method ($F(3, 78) = 42.45$, $p < .00001$), target amplitude ($F(1, 26) = 29.21$, $p < .00001$), and target width ($F(1, 26) = 75.24$, $p < .00001$). However, these effects were qualified by a significant interaction effect between pointing method and target amplitude ($F(3, 78) = 4.76$, $p = .004$).

5.2 Case study

Eight individuals (M = 44 yrs; 6 male, 2 female) from a care center for individuals, which have a movement disorder such as cerebral palsy, were recruited and participated in the case study. The mean age was 44 years and 2 participants were female.

Only two of the participants were able to achieve a sufficiently good calibration such that they could control gaze pointing. Therefore, we only analyzed data from the head-pointing condition. However, data from the two gaze recordings are included in Figure 5.

5.2.1 Time to Activate. The grand mean for time to activate was 1812 ms. The mean time to activate differed depending on whether the target amplitude was 80 pixels (1637 ms) or 120 pixels (1987 ms), $p = .007$. The mean time to activate dependent variable also varied as a function of target width such that it was lower for the 75 pixel target width (1631 ms) relative to the 50 pixel target width
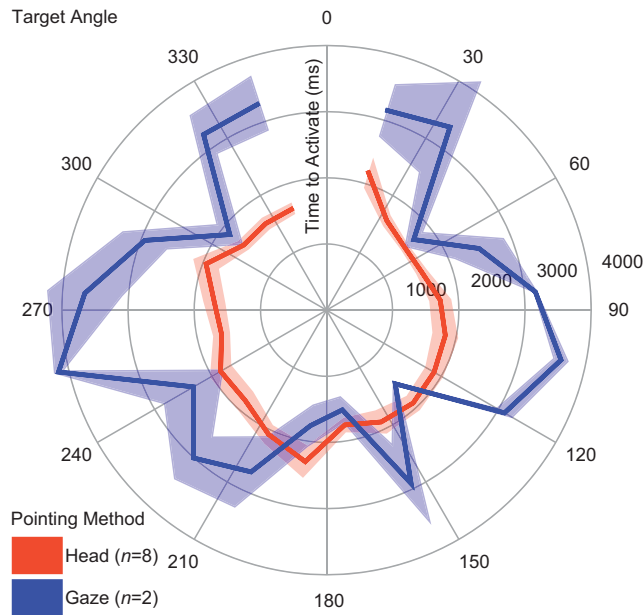
Figure 7: Polar plot showing the distribution of the mean time to activate by head (n = 8) and gaze (n = 2) pointing according to target orientation for the clinical data. One standard error of the mean is depicted with a shaded area around the mean.

(1994 ms), p = .012. The main effects of target amplitude and target width were statistically significant, $F(1, 6)$ = 15.70, p = .007 and $F(1, 6)$ = 12.61, p = .012, respectively. However, the target amplitude by target width interaction was non-significant, p > .05.

5.2.2 Throughput. The grand mean for throughput was 0.99 bits per second. Throughput did not differ as a function of target amplitude. Throughput varied as a function of target width, as it was lower for the 50-pixel target (.89 bits/sec) relative to the 75-pixel target (1.11 bits/sec), p = .047. The main effect of target width was statistically significant, $F(1, 6)$ = 6.22, p = .047. However, the target amplitude main effect and the target amplitude by target width interaction were non-significant, p > .05

5.2.3 Effective target width. The grand mean for effective target width was 47 pixels. Effective target width did not differ as a function of target amplitude and target width, ps > .05. Additionally, the target amplitude by target width interaction was non-significant, p > .05.

5.3 Subjective Ratings

The Friedman test on the pointing-method ranking in the experiment was significant, $\chi^2(3)$ = 19.52, p = .0002. Participants preferred mouse pointing (n = 11) over gaze (n = 9), head (n = 4), and foot pointing (n = 0). However, this difference was only significant between head and mouse pointing and foot and mouse pointing, Z = -1.39, p = .0049 and Z = -2.82, p = .00001, respectively. No other differences were significant regarding the pointing-method rankings (p > .10).

The Friedman test on the mental workload ratings was significant, $\chi^2(3)$ = 25.04, p = .00001. Participants rated gaze pointing as the most mentally demanding, followed by the foot-mouse, then head-position; finally, mouse pointing was rated the least mentally demanding. However, gaze pointing was only significantly different from the mouse and head-position methods, Z = -3.49, p = .0005 and Z = 3.19, p = .0014. The foot-mouse pointing method was significantly more mental demanding than the head-position and mouse pointing methods, Z = -3.25, p = .0012 and Z = -3.53, p = .0004. No other differences were significant regarding the pointing method ratings on mental workload (p > .10).

The Friedman test on the physical workload ratings was significant, $\chi^2(3)$ = 22.49, p = .00005. Participants rated foot pointing as the most physically demanding, followed by head, then gaze; finally, mouse pointing was rated the least physically demanding. Foot pointing was significantly different from mouse pointing, Z = -4.18, p = .00003. Head pointing was significantly different from mouse pointing, Z = -3.85, p = .0001. Finally, gaze pointing was rated more physically demanding than mouse pointing, Z = -2.71, p = .01. No other differences were significant regarding the pointing method on physical workload (p > .10).

The Friedman test on the comfort ratings was non-significant (p > .30). As a result, no post hoc analyses were executed on the level of comfort by pointing method.

5.4 User Comments

Participants' responses in the experiment regarding the different pointing methods were summarized and reviewed for qualitative patterns. Some users provided more than one comment regarding their rating, which is why the comments do not equal the rating tallies. Overall, when participants indicated their preferred pointing method, mouse pointing was in first place. Comments include, *"It was easy"* (n = 5) and *"I have lots of experience with it"* (n = 6). Participants that rated gaze pointing as the most preferred said, *"It was the easiest"* (n = 7), *"It was quick"* (n = 2), and *"It's the least demanding"* (n = 3). Participants that rated head-position pointing as their most preferred mentioned that, *"It was easy"* (n = 3), and *"It was natural"* (n = 1).

When participants rated the pointing methods in terms of their least preferred, many chose the foot-mouse, saying *"It was mentally demanding"* (n = 4) or *"It was too tiresome"* (n = 5). Participants that rated head-position pointing as their least preferred said *"It was uncomfortable"* (n = 5) or *"It was annoying"* (n = 2). Participants who rated gaze pointing as their least preferred said it had *"bad calibration"* (n = 2), that *"the accuracy was off"* (n = 8), or *"it was uncomfortable"* (n = 3).

6 DISCUSSION

Throughput values in the current experiment suggest head pointing is an efficient input method, with just 12% less throughput than the mouse. Gaze throughput was 34% lower - and so was foot input. In addition, effective target width and time to activate were better for head than gaze input and users rated gaze more mentally demanding than head input.

Our findings in the experiment confirm the findings of Qian and Teather [2017] and Hansen et al. [2018] who reported that

head input outperforms gaze. In the present study, with a different hardware setup, we also found the head better than gaze in terms of throughput, time to activate, and effective target width. Eye tracking has been reported to be inaccurate for many years [Ware and Mikaelian 1987], and still seems to be, even with an HMD, where noise from head motion is eliminated.

As a practical consideration we only examined two target widths and two movement amplitudes. Also, the *ID*s ranged from 1.05 to 1.77 bits, which is in the lower end of *ID*s normally applied in Fitts' law studies. Blattgerste et al. [2018] suggest their results might be different from Qian and Teather [2017] because they used a more precise gaze tracking system. Further research with more and higher *ID*s is needed to clarify if more precise gaze tracking in HMDs is significantly better when measured with a standard procedure like the ISO 9241-9.

The observed gaze throughput of the HTC VIVE HMD was 2.55 bits/s. The throughput of the FOVE HMD tested by Qian and Teather [2017] and Hansen et al. [2018] was 1.7 bits/s and 2.1 bits/s, respectively. There may be several reasons for the differences. One study [Qian and Teather 2017] used a 3D environment while the other [Hansen et al. 2018] used a standard 2D interface. However, the head throughputs for these two studies were almost identical, 2.4 bits/s and 2.5 bits/s, respectively. This suggests that 3D vs. 2D is important for gaze but not for head interaction. The difference between the gaze throughput reported herein and the throughput found previously [Hansen et al. 2018] is most likely due to the equipment used, because Hansen et al. had identical 2D set-ups. Future study is needed, for example, comparing 2D vs. 3D interfaces, moving vs. static backgrounds, and targets in head space vs. targets in world space. For instance, Rajanna and Hansen [2018] reported that motion in the background decreased gaze typing performance on an overlay keyboard in a FOVE HMD.

The system we tested was particularly slow in the lower hemisphere (see Figure 4), while Hansen et al. [2018] reported the FOVE HMD to have the same movement time in all directions. The gaze tracking system in the present experiment may solve this issue with improvements in software or hardware. With the current version of this system, our observations suggest placing UI elements on the horizontal plane or in the upper hemifield when pointing time is a priority.

Foot input is an option when seated. This input method performed similarly to gaze for most of the performance measures, and user ratings did not differ between the two. However, this input method requires an extra device, while both head and gaze tracking are internal to the HMD.

Gaze tracking may require a calibration process that is difficult for people with cognitive challenges [Majaranta et al. 2011]. In our case study, only two of seven participants could be properly calibrated. A reason for this was that when wearing an HMD, it is difficult to communicate with the participants, for instance to point at the monitor where they should look.

Head-mounted displays may be attractive for a first assessment of gaze interaction because the cost of quality gaze communication systems is higher than an HMD with built-in gaze tracking. Basic issues of whether an individual has the capability to perform a calibration or has the eye motor control needed to maintain a fixation may be clarified with an off-the-shelf HMD. This raises an

important question of how performance measures on commodity hardware generalize to real-life task situations, for example, using a computer or driving a wheelchair. Rajanna and Hansen [2018] found gaze typing in HMDs around 9.4 wpm (dwell 550 ms) for first-time users, which compares well to performance of novice users of remote gaze trackers for on-screen qwerty keyboards (5 to 10 wpm for dwell times 1000 to 450 ms) [Majaranta et al. 2009] but further research needs to clarify if this is also the case for a Fitts' law task and for individuals with disabilities.

When the ISO 9241-9 procedure was applied by Felzer et al. [2016], the user from the target group had a throughput less than half of the non-clinical participants. In our case study, the throughput for head interaction was less than one-third of the throughput for non-clinical participants. Hence, we might not expect performance levels within the average when assessing an individual but rather look for his or her best performance. In addition, it should of course be considered if the highest efficiency was achieved with a device that would otherwise be effective for the user. Additional questions remain: Can the HMD be used with glasses? Do caregivers know how to operate the device?

7 CONCLUSION

In conclusion, head input outperforms gaze input on throughput, effective target width, and time to activate, while gaze performed similar to foot input. Our case study suggests head pointing to be more effective than gaze pointing for the clinical group we observed.

REFERENCES

Per Bækgaard, John Paulin Hansen, Katsumi Minakata, and I. Scott MacKenzie. 2019. A Fitts' Law Study of Pupil Dilations in a Head-Mounted Display. In *2019 Symposium on Eye Tracking Research and Applications (ETRA '19), June 25–28, 2019, Denver, CO, USA*. ACM, New York, NY, USA. https://doi.org/10.1145/3314111.3319831

Jonas Blattgerste, Patrick Renner, and Thies Pfeiffer. 2018. Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. In *Proceedings of the Workshop on Communication by Gaze Interaction – COGAIN '18 (Article 1)*. ACM, New York, 1–9. https://doi.org/10.1145/3206343.3206349

Catherine M. Capio, Cindy H. P. Sit, Bruce Abernethy, W. Erickson, C. Lee, and S. von Schrader. 2018. 2016 Disability status report: United States. *Ithaca, NY: Cornell University Yang Tan Institute on Employment and Disability (YTI)* (2018).

Gamhewage C. De Silva, Michael J. Lyons, Shinjiro Kawato, and Nobuji Tetsutani. 2003. Human factors evaluation of a vision-based facial gesture interface. In *Computer Vision and Pattern Recognition Workshop – CVPRW'03*, Vol. 5. IEEE, New York, 52–52.

David Dearman, Amy Karlson, Brian Meyers, and Ben Bederson. 2010. Multi-modal text entry and selection on a mobile device. In *Proceedings of Graphics Interface 2010 – GI '10*. Canadian Information Processing Society (CIPS), Toronto, 19–26. http://dl.acm.org/citation.cfm?id=1839214.1839219

Giuseppe Di Gironimo, Giovanna Matrone, Andrea Tarallo, Michele Trotta, and Antonio Lanzotti. 2013. A virtual reality approach for usability assessment: case study on a wheelchair-mounted robot manipulator. *Engineering with Computers* 29, 3 (2013), 359–373.

Alistair Edwards. 2018. Accessibility. In *The Wiley Handbook of Human Computer Interaction*, Kent Norman and Jurek Kirakowski (Eds.). Vol. 2. Wiley, Hoboken, NJ, 681–695.

Monika Elepfandt and Martin Grund. 2012. Move it there, or not? The design of voice commands for gaze with speech. In *Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human-Machine Interaction (Article 12)*. ACM, New York, 1–12. https://doi.org/10.1145/2401836.2401848

Torsten Felzer, I. Scott MacKenzie, and John Magee. 2016. Comparison of two methods to control the mouse using a keypad. In *International Conference on Computers Helping People with Special Needs – ICCHP '16*. Springer, Berlin, 511–518. https://doi.org/10.1007/978-3-319-41267-2_72

Vicki A. Freedman, Emily M. Agree, Linda G. Martin, and Jennifer C. Cornman. 2006. Trends in the use of assistive technology and personal care for late-life disability, 1992–2001. *The Gerontologist* 46, 1 (2006), 124–127. https://doi.org/10.1093/geront/46.1.124

Fabian Göbel, Konstantin Klamka, Andreas Siegel, Stefan Vogt, Sophie Stellmach, and Raimund Dachselt. 2013. Gaze-supported foot interaction in zoomable information spaces. In *Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems – CHI EA '13*. ACM, New York, 3059–3062. https://doi.org/10.1145/2468356.2479610

John Paulin Hansen, Vijay Rajanna, I. Scott MacKenzie, and Per Bækgaard. 2018. A Fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display. In *Workshop on Communication by Gaze Interaction – COGAIN '18 (Article 7)*. ACM, New York. https://doi.org/10.1145/3206343.3206344

John Paulin Hansen, Astrid Kofoed Trudslev, Sara Amdi Harild, Alexandre Alapetite, and Katsumi Minakata. 2019. Providing access to VR through a wheelchair. In *Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems – CHI '19*. ACM, New York. https://doi.org/10.1145/3290607.3299048

C. S. Harrison, P. M. Dall, P. M. Grant, M. H. Granat, T. W. Maver, and B. A. Conway. 2000. Development of a wheelchair virtual reality platform for use in evaluating wheelchair access. In *3rd International Conference on Disability, VR and Associated Technologies, Sardinia, Edited by P. Sharkey*.

Benjamin Hatscher, Maria Luz, Lennart E. Nacke, Norbert Elkmann, Veit Müller, and Christian Hansen. 2017. GazeTap: Towards hands-free interaction in the operating room. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction – ICMI '17*. ACM, New York, 243–251. https://doi.org/10.1145/3136755.3136759

Shinobu Ishihara. 1972. *Tests for colour-blindness*. Kannehara Shuppan Co, Ltd., Tokyo, Japan.

ISO. 2000. *Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices (ISO 9241-9)*. Technical Report Report Number ISO/TC 159/SC4/WG3 N147. International Organisation for Standardisation.

Shaun K. Kane, Jeffrey P. Bigham, and Jacob O. Wobbrock. 2008. Slide rule: Making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility – ASSETS '08*. ACM, New York, 73–80. https://doi.org/10.1145/1414471.1414487

S. Keates, P. J. Clarkson, and P. Robinson. 1998. Developing a methodology for the design of accessible interfaces. In *Proceedings of the 4th ERCIM Workshop*. 1–15.

Simeon Keates, Faustina Hwang, Patrick Langdon, P. John Clarkson, and Peter Robinson. 2002. Cursor measures for motion-impaired computer users. In *Proceedings of the Fifth International ACM Conference on Assistive Technologies – ASSETS '02*. ACM, New York, 135–142. https://doi.org/10.1145/638249.638274

Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head-and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 81.

Paul Lubos, Gerd Bruder, and Frank Steinicke. 2014. Analysis of direct selection in head-mounted display environments. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 11–18.

I. Scott MacKenzie. 2018. Fitts' law. In *Handbook of Human-Computer Interaction*, K. L. Norman and J. Kirakowski (Eds.). Wiley, Hoboken, NJ, 349–370. https://doi.org/10.1002/9781118976005

John Magee, Torsten Felzer, and I. Scott MacKenzie. 2015. Camera Mouse + ClickerAID: Dwell vs. single-muscle click actuation in mouse-replacement interfaces. In *Proceedings of HCI International – HCII 2015 (LNCS 9175)*. Springer, Berlin, 74–84. https://doi.org/10.1007/978-3-319-20678-3_8

Päivi Majaranta, Ulla-Kaija Ahola, and Oleg Špakov. 2009. Fast gaze typing with an adjustable dwell time. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems – CHI '09*. ACM, New York, 357–360. https://doi.org/10.1145/1518701.1518758

Päivi Majaranta, Hirotaka Aoki, Mick Donegan, Dan Witzner Hansen, John Paulin Hansen, Aulikki Hyrskykari, and Kari-Jouko Räihä (Eds.). 2011. *Gaze interaction and applications of eye tracking: Advances in assistive technologies*. IGI Global, Hershey, PA.

Denys Matthies, Franz Müller, Christoph Anthes, and Dieter Kranzlmüller. 2014. ShoeSoleSense: Demonstrating a wearable foot interface for locomotion in virtual environments. In *Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems – CHI EA '14*. ACM, New York, 183–184. https://doi.org/10.1145/2559206.2579519

Mark R. Mine. 1995. Virtual environment interaction techniques. *UNC Chapel Hill CS Dept* (1995).

Toni Pakkanen and Roope Raisamo. 2004. Appropriateness of foot interaction for non-accurate spatial tasks. In *Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems – CHI EA '04*. ACM, New York, 1123–1126. https://doi.org/10.1145/985921.986004

Glenn Pearson and Mark Weiser. 1986. Of moles and men: The design of foot controls for workstations. In *ACM SIGCHI Bulletin*, Vol. 17. ACM, 333–339.

Yuan Yuan Qian and Robert J. Teather. 2017. The eyes don't have it: An empirical comparison of head-based and eye-based selection in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction – SUI '17*. ACM, New York, 91–98. https://doi.org/10.1145/3131277.3132182

Vijay Rajanna. 2016. Gaze typing through foot-operated wearable device. In *Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility – ASSETS '16*. ACM, New York, 345–346. https://doi.org/10.1145/2982142.2982145

Vijay Rajanna and Tracy Hammond. 2016. GAWSCHI: Gaze-augmented, wearable-supplemented computer-human interaction. In *Proceedings of the ACM Symposium on Eye Tracking Research and Applications – ETRA '16*. ACM, New York, 233–236. https://doi.org/10.1145/2857491.2857499

Vijay Rajanna and John Paulin Hansen. 2018. Gaze typing in virtual reality: Impact of keyboard design, selection method, and motion. In *Proceedings of the ACM Symposium on Eye Tracking Research and Applications – ETRA '18 (Article 15)*. ACM, New York. https://doi.org/10.1145/3204493.3204541

Maria Francesca Roig-Maimó, I. Scott MacKenzie, Cristina Manresa-Yee, and Javier Varona. 2018. Head-tracking interfaces on mobile devices: Evaluation using Fitts' law and a new multi-directional corner task for small displays. *International Journal of Human-Computer Studies* 112 (2018), 1 – 15. https://doi.org/10.1016/j.ijhcs.2017.12.003

William R. Sherman and Alan B. Craig. 2002. *Understanding virtual reality: Interface, application, and design*. Morgan Kaufmann, San Francisco, CA.

R. William Soukoreff and I. Scott MacKenzie. 2004. Towards a standard for pointing device evaluation: Perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies* 61, 6 (2004), 751–789. https://doi.org/10.1016/j.ijhcs.2004.09.001

Eduardo Velloso, Dominik Schmidt, Jason Alexander, Hans Gellersen, and Andreas Bulling. 2015. The feet in human–computer interaction: A survey of foot-based interaction. *ACM Comput Surveys* 48, 2, Article 21 (2015), 35 pages. https://doi.org/10.1145/2816455

Colin Ware and Harutune H. Mikaelian. 1987. An evaluation of an eye tracker as a device for computer input. In *Proceedings of the CHI+GI'87 Conference on Human Factors in Computing Systems*. ACM, New York, 183–188.

Xuan Zhang and I. Scott MacKenzie. 2007. Evaluating eye tracking with ISO 9241 - Part 9. In *Proceedings of HCI International - HCII '07 (LNCS 4552)*. Springer, Berlin, 779–788. https://doi.org/10.1007/978-3-540-73110-8_85