International Journal of Human-Computer Studies 139 (2020) 102414

Contents lists available at ScienceDirect

International Journal of Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Comparing selection mechanisms for gaze input techniques in head-mounted displays

Augusto Esteves[a], Yonghwan Shin[b], Ian Oakley*[,b]

[a] ITI / LARSyS, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal
[b] Human Factors Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea

ARTICLE INFO

Keywords:
Hands-free input
Head pointing
Head-mounted display
Virtual-reality
Augmented-reality
Gaze input
Motion matching

ABSTRACT

Head movements are a common input modality on VR/AR headsets. However, although they enable users to control a cursor, they lack an integrated method to trigger actions. Many approaches exist to fill this gap: dedicated "clickers", on-device buttons, mid-air gestures, dwell, speech and new input techniques based on matching head motions to those of visually presented targets. These proposals are diverse and there is a current lack of empirical data on the performance of, experience of, and preference for these different techniques. This hampers the ability of designers to select appropriate input techniques to deploy. We conduct two studies that address this problem. A Fitts' Law study compares five traditional selection techniques and concludes that clicker (hands-on) and dwell (hands-free) provide optimal combinations of precision, speed and physical load. A follow-up study compares clicker and dwell to a motion matching implementation. While clicker remains fastest and dwell most accurate, motion matching may provide a valuable compromise between these two poles.

1. Introduction

Virtual- and Augmented-Reality (VR, AR) headsets are an emerging form factor and device platform suitable for applications from medicine (Sousa et al., 2017) to industry (Aromaa et al., 2016). While current devices (Controls - Interactive patterns - Designing for Google Cardboard, 2014; HoloLens, 2018b) aim to provide rich, high resolution graphics, there is less consensus on how users should interact with them. Common proposals include on-headset taps (Samsung, 2018c), swipes (Yu et al., 2016) and mid-air gestures captured using optical (Ha et al., 2014) or wearable sensors (Hsieh et al., 2016). However, while potentially powerful, these techniques occupy the hands, an undesirable situation in many practical wearable contexts (Baird and Barfield, 1999; Caudell and Mizell, 1992; Lukowicz et al., 2007; Ockerman and Pritchett, 1998; Tang et al., 2003; Zheng et al., 2015). To address this limitation, techniques based on tracking the eyes (Piumsomboon et al., 2017; Sidorakis et al., 2015; Tanriverdi and Jacob, 2000) and the head (Atienza et al., 2016; Clifford et al., 2017; Jr et al., 2017) (or a combination of both (Jalaliniya et al., 2015; Kytö et al., 2018; Piumsomboon et al., 2017; Qian and Teather, 2017)) have been widely proposed. Indeed, as high-fidelity head tracking is mature, cost-effective and integral to good quality VR and AR experiences, input based on head movements is already integrated into many device platforms (Controls - Interactive patterns - Designing for Google Cardboard, 2014; HoloLens, 2018b). Authors have also demonstrated it is accurate, comfortable, and convenient (Jalaliniya et al., 2014; Kytö et al., 2018), and it is often used as proxy for gaze (Nancel et al., 2013; Serrano et al., 2015).

However, while head-tracking input has been established an effective way to control a cursor, it lacks an intrinsic mechanism to confirm a highlighted selection. A large number of techniques have been proposed to address this limitation. A common solution is dwell or dwell time, a technique that triggers selection after a short pause (150-1500ms) over a target (Ramos et al., 2004). It suffers from a trade-off between the length of this pause (and overall target acquisition times) and the occurrence of 'Midas touches', or unintentional target activations during exploration or natural pauses in motion. Voice activation also provides a hands-free selection mechanism (HoloLens, 2016), but at the cost of lowering social acceptability (Rico and Brewster, 2010) and reduced effectiveness in noisy environments (Caudell and Mizell, 1992; Day et al., 2005). Other approaches co-opt the hands for triggering selection using a physical button, either on a dedicated device (Use the HoloLens clicker, 2017) or on the headset itself, or via making in-air gestures (HoloLens, 2018b) – the combination of head cursor with hand trigger seeks to minimize rather than completely remove reliance on the hands. Finally, a range of authors have explored selection via

* Corresponding author.
E-mail addresses: augusto.esteves@tecnico.pt (A. Esteves), yonghwanshin@unist.ac.kr (Y. Shin), ian.r.oakley@gmail.com (I. Oakley).
URLs: http://web.tecnico.ulisboa.pt/augusto.esteves/ (A. Esteves), http://interactions.unist.ac.kr/ (I. Oakley).

https://doi.org/10.1016/j.ijhcs.2020.102414
Received 26 September 2019; Received in revised form 14 February 2020; Accepted 17 February 2020
Available online 19 February 2020
1071-5819/

motion matching (Esteves et al., 2017; Khamis et al., 2018) – targets move in regular patterns and head motions are correlated against these changes. A selection is triggered when a similarity threshold is exceeded. While this technique requires more complex head motions, authors (Esteves et al., 2017) suggest it is less susceptible to producing Midas touches, while still keeping hands free and the input movements discrete, and not requiring command memorization.

This paper argues that while head based input is a good match for interaction with VR and AR headsets, there is a lack of clarity on the pros and cons of the different selection methods that can be combined with it. We contribute data from two studies to elucidate this issue. The first is a Fitts law (Fitts, 1954) study comparing five selection mechanisms; two hands-free (dwell, speech) and three hands-on (clicker, on-device, mid-air gesture). The second compares a motion matching approach to the peak performing hands-free (dwell) and hands-on (clicker) input techniques from the initial study. In addition to Fitts' and performance metrics, we report on participants perceived exertion and preference. The key contribution of this paper are these results – we argue establishing performance, exertion and preference baselines for head based target selection techniques will help researchers, designers and developers producing VR and AR systems make informed choices about the most appropriate input mechanisms to integrate into their systems. We close with a series of practical recommendations based on our study results.

2. Related work

In addition to reducing head/eye movement and attention switching, and supporting spatial cognition and mental transformations (Tang et al., 2003), allowing users to interact with head-mounted displays (HMDs) while minimizing the use of their hands will be important for AR and VR headsets to become workplace tools. This potential is exemplified in previous works exploring the use of HMDs to snap photos during industrial inspection tasks (Aromaa et al., 2016); as procedural information systems supporting maintenance tasks (Caudell and Mizell, 1992; Zheng et al., 2015), pre-flight inspections (Ockerman and Pritchett, 1998), and assembly tasks (Baird and Barfield, 1999; Day et al., 2005; Tang et al., 2003); and to provide support during medical analyses (Sousa et al., 2017) or emergency responses (Lukowicz et al., 2007). Several of these early examples illustrate proof-of-concept prototypes with very little input capabilities, ranging from single-state AR systems (Baird and Barfield, 1999), systems that rely on the HMD orientation (Caudell and Mizell, 1992), to systems that use a single speech command to navigate between content (Ockerman and Pritchett, 1998; Zheng et al., 2015). But while these provide a look at the potential of AR and VR headsets in the workplace, the need for richer input has been highlighted by expert users for, e.g., personalizing content or interface elements (Ockerman and Pritchett, 1998).

An approach that can provide seamless interaction with specific or highly structured work tasks is activity recognition: a system can provide appropriate customized content by identifying the tools a user is operating (such as surgical bayonets (Birkfellner et al., 2002)); by anchoring relevant content to specific locations (such as a patient's body during a laparoscopy (Fuchs et al., 1998)); by automatically identifying the task step or state during, e.g., assembly tasks (Henderson and Feiner, 2011); or by providing contextual information based on the user's location (such as in emergency rescue operations (Lukowicz et al., 2007)). However, supporting more open ended activities requires yielding control to a user by enabling techniques such eye-tracking (Piumsomboon et al., 2017; Sidorakis et al., 2015), head-tracking (Clifford et al., 2017; Jr et al., 2017) or their combination (Piumsomboon et al., 2017). While literature on these techniques is diverse, head-tracking offers practical advantages: the required sensing technology (inertial motion units) is mature, cheap (<10USD), small (<5mm square) and integrated into most current headsets. Further, existing comparisons between eye and head-tracking based input suggest head input increases comfort and accuracy (Jalaliniya et al., 2014; Kytö et al., 2018) while reducing workload and learning time (Bates and Istance, 2003).

Although head-tracking is a promising approach for hands-free HMD input capable of, for example, rapid and accurate control of a cursor (Controls - Interactive patterns - Designing for Google Cardboard, 2014), it lacks an integrated technique to trigger selections. Techniques such as dwell (Kjeldsen, 2001; Park et al., 2008), speech (HoloLens, 2016), gesture (HoloLens, 2018b) and the use of physical controls in the hand (Use the HoloLens clicker, 2017) or on the headset (Samsung, 2018c) typically fill this role. While these techniques inherently target and emphasize different qualities of interaction, we argue there is a lack of consistent and systematic comparisons that empirically contrast between them. This hampers the ability of designers and developers to make effective choices about what techniques to deploy. In the same way that prior work has provided an actionable characterization of the differences between head- and eye-pointing HMD input with dwell (Bernardos et al., 2016; Blattgerste et al., 2018) and 'clicker' selection confirmation (Hansen et al., 2018), we argue for the importance of data comparing a broad range of selection mechanisms for head-based input in generic selection tasks (unlike, e.g., text-entry comparisons (Yu et al., 2017)).

Beyond these traditional selection mechanisms, recent work on motion matching (Fekete et al., 2009; Vidal et al., 2013; Williamson and Murray-Smith, 2004) has focused on VR (Khamis et al., 2018; Piumsomboon et al., 2017) and AR (Esteves et al., 2017; Kangas et al., 2016) scenarios. In motion matching, interface elements move across distinct trajectories, and users interact with these by tracking their movement for a short period of time. While the technique is commonly applied to eye-tracking, its has also been shown to be effective when applied to head tracked motions (Esteves et al., 2017). We identify motion-matching as a final candidate selection technique for head motion based input on HMDs and seek to provide empirical data contrasting performance with this full range of viable selection techniques.

3. Head pointing study — Fitts' law

The first comparative study employed a Fitts' Law (ISO 9241-9) design across five of the most popular selection mechanisms for head-based input in VR/AR headsets. These are: dwell (e.g., Google Cardboard (Controls - Interactive patterns - Designing for Google Cardboard, 2014)); speech; clickers (dedicated device, typically held in the hand, featuring a button operated by a finger); mid-air gestures (e.g., HoloLens HoloLens, 2016; HoloLens, 2018b; Use the HoloLens clicker, 2017) and; on-device input (e.g., a button, as on the Samsung Gear VR Samsung, 2018c) – see Fig. 1. The goal was to compare these five common techniques across various quantitative and qualitative measures. We used a Fitts' law design as it is a representative and widely studied targeting task that is commonly used to study selection mechanisms (Hansen et al., 2018). However, we note that Fitts' law models movement, not selection time; model fit decreases if selection methods add to movement times (Vertegaal, 2008). As such our quantitative analysis focuses on simple time and accuracy metrics more than Fitts' law model fit and throughput.

3.1. Participants

We recruited 20 participants (7F), aged between 21 and 57 (M = 37.45, SD = 10.33). They were postgraduate students, researchers, and staff at a local institution. Using a 7-point Likert scale (higher is better), they rated their experience with head-mounted displays at 2.95 (SD = 1.54). One participant recorded no correct trials and was removed from all analyses.

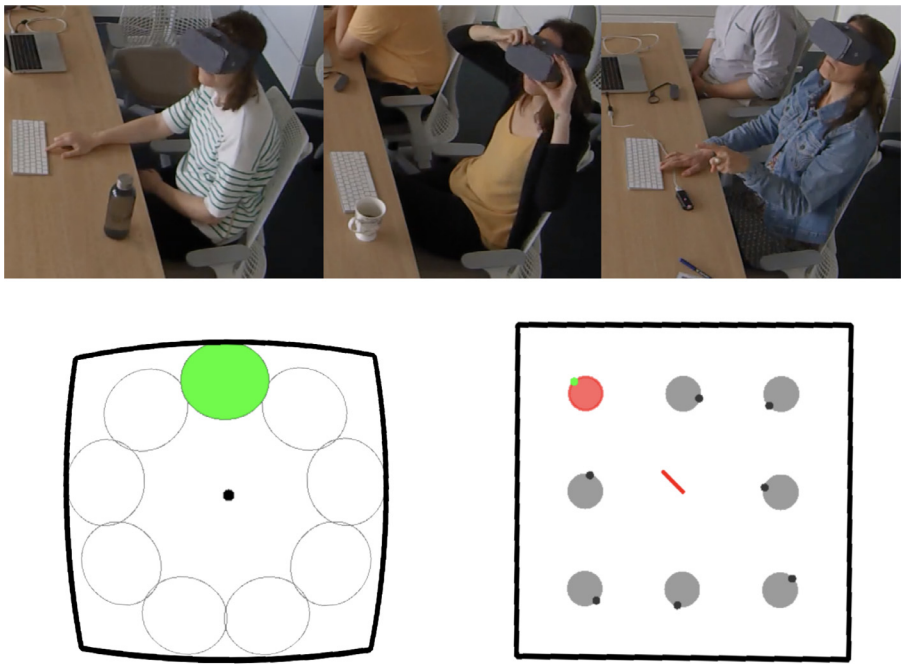

Fig. 1. Top: participants in the head pointing study, interacting in the clicker (left), on-device (middle), and mid-air gesture (right) conditions. Bottom, from left to right: study interface for the head pointing (VR view of ID 1.94 blocks with current target in green and head cursor in black) and motion matching (2D rendering of study graphics in the motion matching condition) studies. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Design

The experiment followed a within-subjects design (balanced using a Latin Square) and used a standard Fitts' reciprocal tapping task – in each block of trials, users selected each one of a ring of nine circular targets, crossing the center of the ring between each selection (Soukoreff and MacKenzie, 2004). These targets were drawn in a square 2D canvas (400px) that was positioned at approximately 200px from the user. Following recent recommendations (Guiard, 2009) regarding how to vary the index of difficulty (ID) of different blocks, we kept the diameter of the ring, and therefore the distance between targets, constant (at 73.37° of visual angle) and adjusted the diameter of the targets among 8 levels (ID levels: 1.94, 2.39, 2.74, 3.25, 3.54, 3.92, 4.48 and 5.44, ranging from 25.9° to 1.73° of visual angle). For each of the five selection conditions participants completed three sets of eight blocks of trials, one block per ID. The first set of eight blocks was discarded as practice. The order of the five device conditions was balanced in a Latin square design and the order of blocks within sets was randomized. In total, we retained 13,680 trials for analysis: five selection mechanisms × eight ID levels × two blocks × nine targets × 19 participants.

3.3. Materials

The study was implemented on a Google Pixel phone (1080p, 441ppi) for a Google Daydream VR headset and using the Processing programming environment. Selections were triggered by the following mechanisms:

Clicker. For simplicity, we use a Bluetooth keyboard as a physical controller. Participants were able to rest their arms on the table and position this keyboard as they saw fit. The keyboard was connected directly to the Pixel and selections were triggered by a tap on the space bar. In the seated pose used in this study, we note the difference between a hand-held clicker and a keyboard key sitting comfortably under participants' fingers is minimal.

On-device input. Users pressed either of the Pixel's inbuilt volume keys, located at the top left of the headset. Participants were encouraged to keep their hands on the buttons at all times.

Gesture. Gestures were detected using the Leap Motion, a dedicated finger gesture sensor noted for its low latency. This was attached to a PC and streamed events to the Pixel via UDP. The sensor was positioned on the desk, under a user's hand (see Fig. 1). The app responded to hand out/in events with a distinctive background color change. We logged hand gesture start and end times and triggered selections on gesture end events using a Java library (Morawiec, 2019)). Participants made a finger circle gesture to issue a selection. We selected the circle gesture as subjective testing indicated it had optimal reliability.

Dwell. Selections occurred when participants' gaze was stationary (<7px) for 400ms. It enables rapid selection without the perception of a noticeable wait. Literature on dwell times reports thresholds from 150 to 1500ms (Bernardos et al., 2016; Jacob, 1990; Miniotas et al., 2006; Sibert and Jacob, 2000; Velichkovsky et al., 1997; vrview, 2018d), so this figure represents the lower end of the range. The use of a short dwell threshold better matches our Fitts task as this models movement, not selection time.

Speech. Selection was triggered on the first spoken word; participants were asked (but not enforced) to say "select". We used Android's stock *SpeechRecognizer* to enable this (processing delay of ∼ 132ms) and logged all uttered words.

3.4. Procedure and metrics

The study took place in a quiet room with participants seated in front of a desk. In each trial we measured the selection time (from the previous selection), the location at which the selection was made (not necessarily over the target), the correctness of the selection and *re-entries*, the number of additional times (i.e., beyond the first) the user entered the correct target prior to making a selection (Clarke and Gellersen, 2017). After each of the five selection conditions participants completed the Borg CR10 (Borg, 1998) scale of perceived exertion. This captured participants' perceived effort (head, shoulder, arm) on a scale of 0 to 10 (0.5 increments, higher is harder). Finally, upon completing all tasks, participants rated each selection mechanism from 5 (least favourite) to 1 (favourite), provided written comments on their favourite and least favourite mechanisms and, rated their previous experience with these in the context of HMDs.

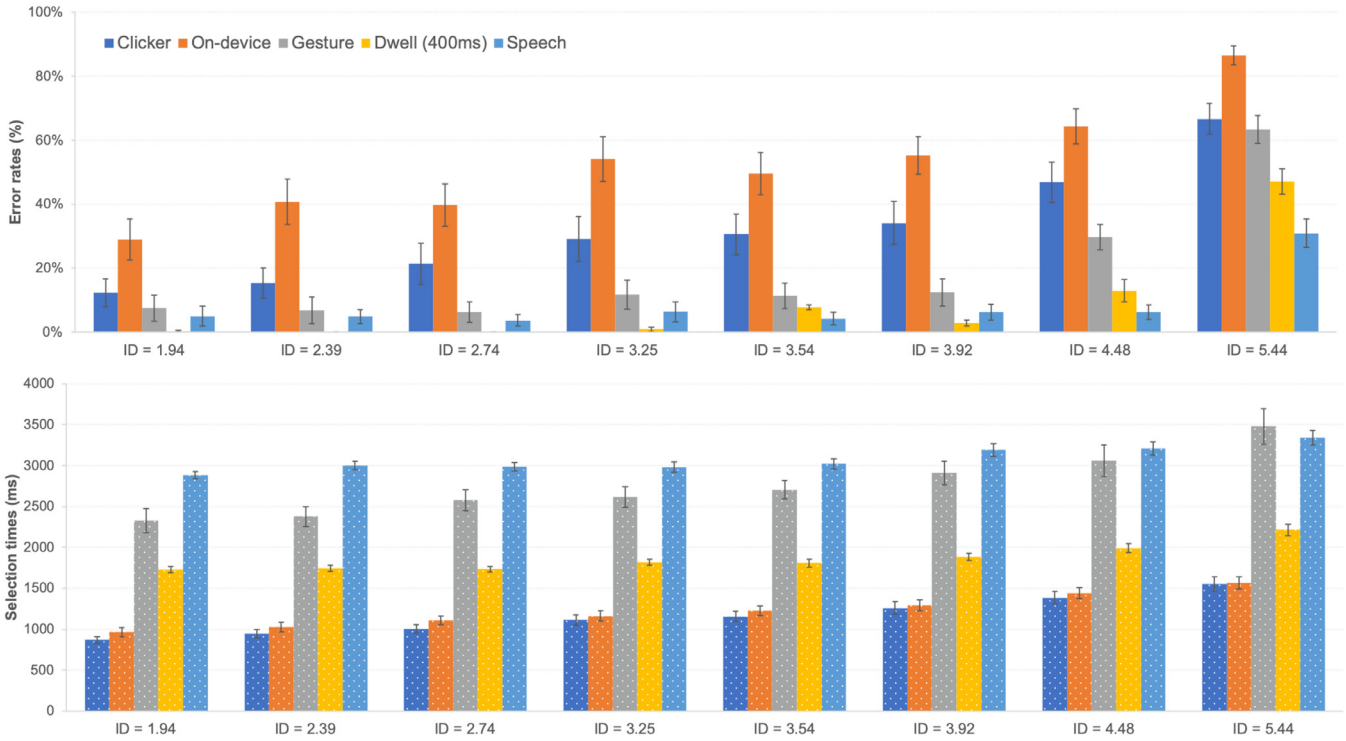

Fig. 2. Error rates (bars, top) and selection times (dotted bars, bottom) per index of difficulty (ID) for the Head Pointing study. Error bars show standard error.

3.5. Results

3.5.1. Objective measures

We first removed outliers more than three standard deviations from the mean in selection time or selection distance from the intended target – a total of 1.89% of trials. Mean error rates, reentries and selection times (from all trials) are shown in Figs. 2 (per ID) and 3 (aggregate). As expected in a Fitts' task, performance decreased substantially on targets with larger IDs – Fig. 2 shows trials involving the highest ID targets were quite challenging. In addition, mean selection times from only correct trials did not differ substantially from those in all trials (they were 7ms to 56ms faster, representing differences of between 0.23% to 4.58%). As such, we opted to analyze the full set of target completion times. We tested all data for normality with Shapiro-Wilk tests. Metrics in which data were normal were analyzed with one-way repeated measures ANOVAs, while metrics in which data failed normality checks were analyzed with Friedman tests. For ANOVAs, we report Greenhouse-Geisser corrected degrees of freedom in cases where sphericity was violated (Mauchly, 1940). We use general eta squared ($\hat{\eta}_G^2$) to express effect size for ANOVAs and Kendall's W for Friedman tests. Post-hoc testing was conducted with t-tests for parametric data and Friedman's Aligned Ranks tests for non-parametric data. All post-hoc pairwise comparisons include Bonferroni confidence interval adjustments. Main effects on these measures were all significant with large effect sizes: errors ($\chi^2(4) = 55.21$, $p < 0.001$, $W = 0.73$); time ($F(2.23, 40.08) = 202.04$, $p < 0.001$, $\hat{\eta}_G^2 = 0.851$) and; reentries ($\chi^2(4) = 58.67$, $p < 0.001$, $W = 0.78$). *Post-hoc* testing showed on-device led to significantly more errors than gesture and speech and clicker led to significantly more errors than speech and dwell (all $p <= 0.001$). Time differed between all conditions ($p <= 0.013$) except for between clicker and on-device ($p = 0.117$). Finally, reentries showed significant differences for all conditions ($p <= 0.015$) except for between clicker and on-device ($p = 0.74$), gesture and speech ($p = 1$), dwell and clicker ($p = 0.5$) and dwell and on-device ($p = 1$).

We also fit Fitts' law models to the data (Soukoreff and MacKenzie, 2004). As expected, r-squared varied substantially reflecting the difference in activation time for the different selection mechanisms. The mean r-squared of individual models calculated for each participant were: clicker (0.75), on-device (0.81), gesture (0.29), dwell (0.37), and speech (0.38). Mean throughput was: clicker (3.96),

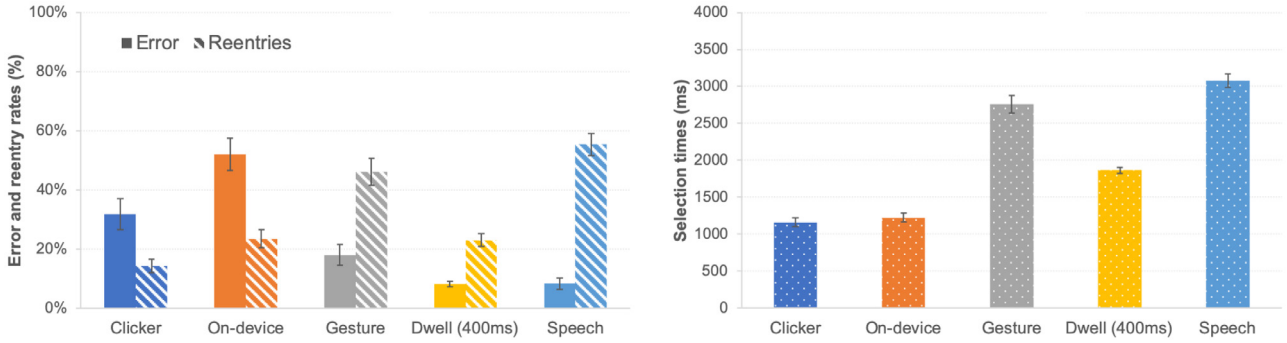

Fig. 3. Performance results for the Head Pointing study: error rates (solid bars) and reentry rates out of 144 trials (dashed bars) – left; selection times (dotted bars, right). Error bars show standard error.

on-device (3.24), gesture (1.65), dwell (2.83), and speech (1.69). A final RM ANOVA revealed these differences to be significant ($F(2.02, 36.35) = 99.89$, $p < 0.001$, $\hat{\eta}_G^2 = 0.756$) and all *post-hoc* t-tests were also significant ($p <= 0.007$) except between gesture and speech ($p = 0.650$).

3.5.2. Perceived exertion

Perceived exertion across conditions varied between *very weak*: clicker (1.13, $SD = 1.21$); to *weak*: dwell (1.63, $SD = 1.54$) and speech (1.79, $SD = 1.50$); to *moderate*: on-device input (2.97, $SD = 2.54$) and mid-air gestures (2.50, $SD = 1.98$). A related-samples Friedman test reveals these to be significantly different: $\chi^2(4) = 15.65$, $p = 0.004$. Pairwise comparisons reveal the clicker and dwell to be less tiring than on-device input ($p = 0.002$ and $p = 0.024$, respectively), and the clicker to be less tiring than mid-air gestures ($p = 0.021$). Only significant differences are presented for brevity.

3.5.3. Preference

We analyzed preference data with a Friedman test, showing significant variations in these rankings ($\chi^2(4) = 52.32$, $p < 0.001$). Post-hoc aligned rank tests indicated that *clicker* and *dwell* led to significantly better preference rankings than all other conditions (all $p < .005$). The 12 participants who picked the *clicker* as their favourite described it as being the fastest (9), easiest (4), most familiar (3), most comfortable (2), and most accurate (1) selection mechanism: "It was easy to coordinate (head) movement and press the space bar. I found that I could hit the targets at much quicker response speeds" (P16). Similarly, the six participants who picked *dwell* as their favourite described it as being the easiest (4), the most comfortable (3), fastest (1), most accurate (1), and most satisfying (1) selection mechanism: "(With) dwell (it) is much easier to select the object by just looking at (it); while other technique (s) require some trigger" (P3). Finally, one participant rated the *on-device input* as their favourite, describing it as the fastest – see Fig. 5.

The results for the least favourite selection mechanism are more varied. The 11 participants who picked the *mid-air gesture* as their least favourite described it as being the hardest and most straining to perform (10), the slowest (1), and highlighted a lack of user feedback (2) and control (2). The five participants who picked *speech* as their least favourite described it as the most frustrating (2) and slowest (1) selection mechanism. Of these participants, two highlighted experiential qualities of speech: P6 did not like the sound of her voice; and P11 did not appreciate the contrast between the "original acoustics" of his input, and the simple digital output in the study system. P12 was concerned that repeated speech selections could trigger a repetitive strain injury (RSI). Lastly, three participants described the *on-device input* as their least favourite due to discomfort (3): "I thought to quit the task" (P20). When asked to rate their previous experience with the selection mechanisms used in the study on a 7-point Likert scale (higher is better), participants rated their average experience with dwell at 2.00 ($SD = 1.56$); speech at 1.32 ($SD = 0.75$); clickers at 2.26 ($SD = 1.45$); mid-air gestures at 1.42 ($SD = 0.77$); and on-device input at 2.05 ($SD = 1.65$).

To sum up: based on a combination of good performance over the full spectrum of measures (task times, error rates, exertion scores and preferences), we highlight clicker as providing the best performance during hands-on input and dwell as providing the best performance during hands-free input. We opt to take these techniques forward for further study.

4. Motion matching study

Motion matching is an emergent hands-free interaction paradigm for VR (Khamis et al., 2018; Piumsomboon et al., 2017) and AR (Esteves et al., 2017; Kangas et al., 2016). In motion matching, interface elements move across distinct trajectories, and users interact with these not by pointing, but by tracking their movement with their eyes (Khamis et al., 2018) or head (Esteves et al., 2017). Because of this departure from pointing (and thus Fitts' Law), we compare a head-based motion matching implementation to the peak performing hands-free (dwell) and hands-on (clicker) input techniques from the first study using standard performance metrics: error rates and acquisition times.

4.1. Participants

We recruited 18 participants (8F), aged between 21 and 29 ($M = 23.06$, $SD = 2.10$). These were undergraduate and postgraduate students at a local institution. Using a 5-point Likert scale (higher is better), participants consistently rated their experience with head-mounted displays at 1.

4.2. Design

The experiment followed a within-subjects design with selection mechanism as the only independent variable: *clicker, dwell*, and *motion matching*. The order in which participants interacted with the selection conditions was balanced using a Latin Square, and participants completed 20 blocks per selection mechanism. The first and eleventh blocks (the block after a small break – see Procedure and Metrics) were regarded as practice and not retained for analysis. Each block consisted of eight trials, one randomly ordered selection on each of eight circular targets. The circular targets were displayed in a square arrangement: four at the corners of the canvas, four at each side's midpoint. Each was 40px in diameter, corresponding to 10% of the canvas width, 9.86°of visual angle and indexes of difficulty of either 1.98 (center) or 2.37 (corner) – see Fig. 1. These target locations and sizes were selected to yield IDs that were as similar as possible to the lowest two ID targets in the head pointing study (IDs of 1.94 and 2.39), in order to better compare and contrast data between the two studies. Participants were instructed to select the target highlighted in red (other targets were displayed in grey). In the motion matching condition, the 10 pixel wide moving dot travelling the contour of the intended target was displayed in green (as opposed to grey) – these moved at 180°/s and were equally spaced with a phase of 45°. To minimize visual search, a small line displayed at the center of the canvas pointed towards the intended target. Finally, trials timed out at five seconds and were logged as a wrong selection. 7776 trials were recorded: three selection mechanisms × eight trials × 18 blocks × 18 participants.

4.3. Materials

The VR setup (e.g., canvas size) and equipment was as described in the previous study, as well as the clicker implementation. Two changes were made to the earlier dwell implementation: we increased the selection threshold to 1000ms, and only allowed selections to take place when the gaze cursor was hovering over a target. These parameters are a better representation of commercial applications (e.g., Google Cardboard uses a 1200ms dwell time (Library, 2019)), and allow us to explore different possible implementations of the technique. This in turn expands our contribution by capturing additional performance and subjective data.

The motion matching selection followed a popular implementation (Clarke et al., 2017; Esteves et al., 2015; Khamis et al., 2018; Velloso et al., 2016) as first described by Vidal et al. (2013), where simple Pearson's correlations are computed for *xtarget-yaw* and *ytarget-pitch* relationships (yaw and pitch values calculated with HeadTransform (2018a)). If both exceed a *correlation threshold* of 0.8 for a given target, and no other displayed targets attain the same result (either individually or via an average of both results), the target is selected. The correlations operate in a *rolling window* of 1000ms, and start 500ms after a new set of targets is displayed (during which participants are engaged in open-loop orientating behavior marking the beginning

of a smooth pursuit movement). These parameters are derived from the closest state-of-the-art implementation described in Esteves et al. (2017). The system had a sample rate of 100Hz.

4.4. Procedure and metrics

The experiment was conducted in a quiet room with participants seated in front of a desk. Each session started with a brief explanation of the study, and by capturing participants' demographics and previous experience with HMDs. Each trial started after participants selected a single target displayed at the center of the canvas. This not only centered participants' head pointer prior to each trial, but also allowed them to pause the study between trials in order to take short discretionary breaks (e.g., to make adjustments to their pose). Furthermore, participants were required to take a break of at least one minute halfway through each condition (10 blocks). As before, and after completing a selection condition, participants were asked to fill in the Borg CR10 scale of perceived exertion. After all three conditions were completed, participants rated each selection mechanism from 3 (least fav.) to 1 (favourite) and provided written comments on their favourite and least favourite mechanisms. Finally, data on participants' error rates and selection times were recorded for further analysis – no target reentry data was captured as motion matching does not entail pointing.

4.5. Results

4.5.1. Objective measures

Shapiro-Wilk normality tests indicated our dwell results were not normally distributed: $W = 0.34$, $p < 0.001$. As such, we opted to conduct our analysis using related-samples Friedman tests. Effect sizes were calculated using Kendall's W. This revealed significant different across error rates ($\chi^2(2) = 27.507$, $p < 0.001$, $W = 0.764$) and selection times ($\chi^2(2) = 36.00$, $p < 0.001$, $W = 1$) – see Fig. 4. Pairwise comparisons (with Bonferroni corrections) revealed that dwell produced less errors than the clicker and motion matching selection mechanisms ($p = 0.005$ and $p < 0.001$, respectively), but was the slowest of the three ($p < 0.001$ and $p = 0.008$, respectively). While no statistically significant differences were found for the error rates between the clicker and motion matching conditions ($p = 0.166$), participants were significantly faster using clicker ($p = 0.008$).

4.5.2. Perceived exertion

Perceived exertion across conditions varied between *moderate*: clicker (2.86, $SD = 1.55$) and dwell (2.83, $SD = 1.46$); to *somewhat heavy*: motion matching (3.73, $SD = 1.66$). A related-samples Friedman test revealed these to be significantly different: $\chi^2(2) = 6.03$, $p = 0.049$. Pairwise comparisons showed dwell to be less tiring than motion matching ($p = 0.037$). Only significant differences are presented for brevity.

4.5.3. Preference

We analyzed preference data with a Friedman test. It showed significant variations ($\chi^2(2) = 9.33$, $p = 0.009$) and post-hoc aligned rank tests indicated that clicker led to significantly better preference rankings than motion matching ($p = .002$). Indeed, the majority of participants picked the clicker (10) and, to a lesser extent, dwell (7) as their favourite selection mechanisms, for mostly the same reasons as in the head pointing study (see Fig. 5). Interestingly, P13 favored the motion matching approach as it allowed for coarser and broader head motions. The same participant listed dwell as his least favourite mechanism for the opposite reason: it required precise pointing and dwelling. Four other participants also described dwell as their least favourite mechanism, mostly due to the high dwell threshold (3): "it was inconvenient to wait until it (triggered). I wanted to move quickly to the next [trial]" (P2). Eleven participants picked motion matching as their least favourite approach due to its low accuracy (5), being physically tiring (4), and because it was hard to sync their head movement with the moving targets (4): "sometimes I can't get the timing (right)" (P16). Finally, two participants reported the hand-clicker as their least favourite mechanism as its fast nature facilitated input errors (2).

5. Discussion

5.1. Head pointing study

The Head Pointing study demonstrated the benefits of the *clicker* and *dwell* as selection mechanisms for gaze-based HMDs. These picked by most participants as their favourites (12 and six participants, respectively – see Fig. 5) and also showed objective benefits compared to their peers (hands-on and hands-free mechanisms). While it was not faster, clicker did lead to greater throughput than on-device input – suggesting it may have been more accurate and stable – and was reported to be less taxing. Dwell's superiority was also clear: while it did not offer improvements in exertion ratings or errors, it was faster, led to greater throughput and was more stable than the gesture and speech methods.

Its worth speculating about the causes of some of these variations. In the on-device condition, error rates spike as IDs increase (see Fig. 2). This likely reflects the increasing impact of the disturbance to Head Pointing precision caused by the physical act of pressing and releasing a button on the headset – triggering the selection caused the head to wobble. The acts of gesture and speech input may have added similar physical disturbances (note the high number of re-entries for these conditions in Fig. 3), but the protracted nature of input in these modalities gave users time to re-target the cursor. While increasing accuracy, these long, involved targeting operations likely contributed to the low preference ratings for gesture and speech (respectively, 11 and five users rated these techniques are their least favourite). Finally, while hands-on methods (clicker, on-device) provide more rapid performance than hands-free methods, their accuracy is lower. We suggest this is

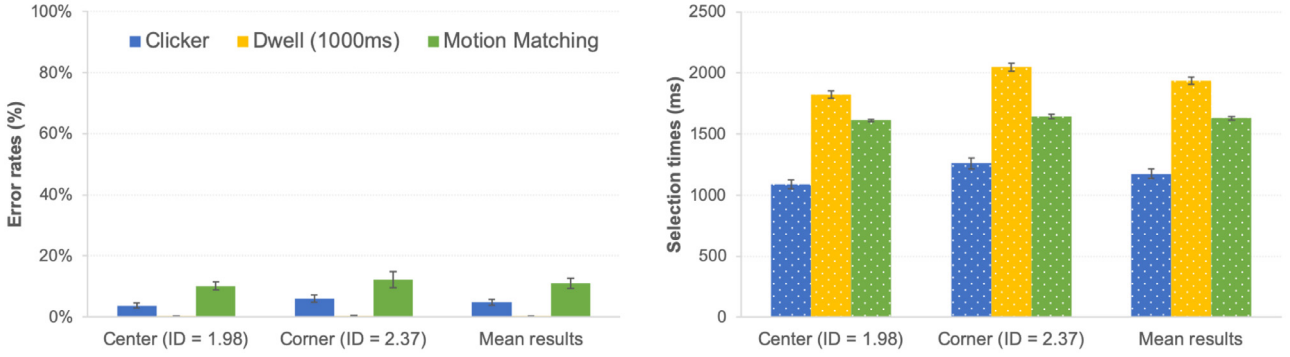

Fig. 4. Performance results for the Motion Matching study: error rates (bars, left) and selection times (dotted bars, right). Error bars show standard error.

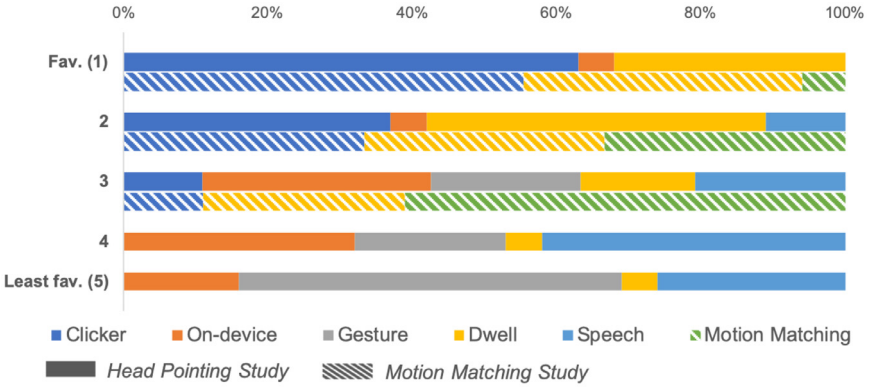

Fig. 5. Preference results for the head pointing (solid) and motion matching studies (dashed). Dwell implementations vary between these, using a 400ms selection threshold in the former and 1000ms in the latter.

partly due to a speed accuracy trade-off – the use of an atomic and highly familiar input trigger (simply pressing a button) encouraged participants to optimize for speed over accuracy. With the more constrained hands-free techniques, they were unable to make this choice (e.g., there is no way to "dwell faster"), resulting in longer but more accurate input.

5.2. Motion matching study

The Motion Matching study compared the peak performing hands-on (clicker) and hands-free selection (dwell) mechanisms to *motion matching*, a recent interaction technique that tracks users' head motions in response to moving widgets. Despite variations in the dwell implementation (longer threshold, simplified target selection), the results resemble those from the Head Pointing study: clicker was more rapid while dwell was more accurate – indeed, dwell errors approach zero. There were some notable differences between the two studies. Clicker was slower in the Motion Matching study (1174 ms) than equivalent ID level targets in the Head Pointing study (907 ms), most likely due to participants more strongly emphasizing accuracy – in the Motion Matching study erroneous trials were repeated, a process that penalizes errors. This observation is supported by the reduced clicker error rate in the Motion Matching study (4.9% vs 13.8%) for targets with equivalent IDs.

Dwell also showed only a modest increase in task times in the Motion Matching study – from 1735ms to 1935ms at equivalent IDs. That is less than the expected 600ms increase due to the longer dwell time (400 ms vs 1000 ms). We attribute this to the changes in the dwell implementation and task – in the Head Pointing study, travel distances were edge-to-edge, roughly twice the center-to-edge distances in the Motion Matching study. Furthermore, in the Head Pointing study dwell was triggered with a strict threshold for defining a stationary cursor (less than 7 pixels of movement, 1.73° of visual angle), while in the Motion Matching study, we used an implementation that simply required participants to remain over an on-screen target of interest (9.86° of visual angle). These variations likely made the task easier, partly offsetting the inevitable increase in selection times due to the longer dwell time.

Motion matching struck the middle ground between clicker and dwell. Errors were equivalent to those with clicker, while times were significantly reduced compared to dwell. Obviously, this result is dependent on features of the dwell and motion matching implementations (such as the 1000ms dwell/matching time) and may not generalize to all possible versions of these techniques. We also note that the study suggests some interesting differences between motion matching and conventional targeting. Specifically, both clicker and dwell show expected increases in selection time on the corner (ID = 2.37) over the center (ID = 1.98) targets – these took between 170ms (clicker) and 224ms (dwell) more time to select. This effect is much reduced for motion matching – differences between the corner and center targets are just 32ms. This independence of selection time to selection distance is a potentially beneficial property of the motion matching technique that deserves further attention in the future.

The subjective data from the Motion Matching study reinforced the notion that users favour the clicker and dwell techniques. Indeed, these preferences were maintained despite an uptick in the perceived exertion for both clicker and dwell, likely caused by the requirement to accurately select targets and longer dwell threshold. We also note dwell was still rated as requiring less exertion than motion matching. This suggests that motion matching may be most appropriate for sporadic or occasional tasks rather than protracted or continual ones.

6. Recommendations

The primary goal of this work is to elucidate the differences between the wide range of possible selection triggering techniques that can be applied to head gaze based pointing input. We close the paper with concrete recommendations distilled from the studies and data we present. We split these into recommendations for hands-on and hands-free tasks.

6.1. Hands-on selection

Despite the increasing enterprise use-cases for VR and AR, there are many situations in which hands-on interaction is still appropriate, including applications that can be enjoyed at home (e.g., entertainment) or in desktop scenarios (Sousa et al., 2017). In these situations, and despite the obvious limitations of having to hold/carry an additional peripheral for interaction, the *clicker* is the obvious choice for prolonged use. Compared to mid-air gestures, clicker was faster and preferred – this despite a higher error rate (albeit in a task where these were not penalized) and use of a dedicated, low latency and state of the art sensor to track the gestures. Compared to on-device, task times were similar – the buttons in both cases could be pressed with similar speed. However, the physical disturbance induced by on-device input led to higher error rates and frustrated users. Given the ready availability of this technique, it may be suitable for occasional use to select large, central (or well-separated) targets, such as those that might appear on confirmation or dialog boxes (in AR user interfaces) or facilitate teleportation in 360° VR scenes.

6.2. Hands-free selection

Dwell performs well during the hands-free input tasks we studied: at 400ms it is faster than and preferred to speech, due to it being inconspicuous and immune to issues such as environmental noise. At 1000ms

and compared to motion matching, it produces fewer input errors and is less physically taxing. Depending on the configuration of the techniques (e.g., the timing thresholds) dwell may be slower than motion matching – this held true for the versions examined in the Motion Matching study reported here (i.e., 1000ms). We also note our study data suggest that motion matching may be relatively immune to the impact of target distance. It may be just as easy to select a distant target as it is to select a more proximate one. We believe these results indicate that motion matching has promise for specific types of targeting task and believe that future work should explore the technique in more detail to improve its design and characterize how it can best be deployed. We detail some ideas for this in the next section.

7. Limitations and future work

A number of limitations impact our results and conclusions and suggest avenues for future work. Perhaps most critically, neither of our study tasks reproduce the "Midas touch" inadvertent selection problem. In both studies, participants did not need to engage in exploration or search sub-tasks, targets to select were always clearly highlighted and paths towards these targets were always free of distractors. In these simple scenarios, the performance of techniques such as dwell are likely over-stated – it is very difficult to inadvertently select an undesired target, meaning that problems related to this issue are simply never surfaced. In contrast, in more realistic systems, Midas touch selections can be sufficiently disruptive so as to require modal solutions in which users must explicitly active and deactivate dwell mediated selection (Istance et al., 2008).

Extending the work in this paper with more realistic tasks featuring distractors is a clear next step for this work. We believe the costs of techniques like dwell and benefits of techniques like motion matching may be more completely and clearly delineated in such tasks. We also note the motion matching technique studied in this paper could also be improved. This could take place both under the hood – different matching algorithms may offer improved performance (Drewes et al., 2018) – and also through improved design. For example, while all head pointing techniques studied in this paper feature interactive feedback (a cursor, color highlights when over a target), the motion matching implementation lacked this type of aid. We identify designing interactive feedback to support motion matching input as a topic for future work.

We suggest that combining dwell and motion matching (as they only require head-motion data) is a fruitful area for future study. One approach could look at dwell for selection of front and center controls while motion matching could access sporadic content in the periphery of the users' field-of-view (e.g., settings or volume controls). This would be particularly useful in devices with larger field-of-views than the Google Daydream (90°), where the large scale head movements and sustained holds inherent with dwell could become straining. Another approach could look to minimize the "Midas touch" problem of dwell in everyday AR scenarios by employing a short dwell time to trigger more explicit motion matching controls. This would also benefit motion matching implementations in AR by reducing the number of moving targets in the interface at any given time, and potentially reducing false activations during locomotion. Further, we would argue these challenges and opportunities are not restricted to AR, and could be expanded to other domains where gaze input is beneficial – such as interaction with smart environments and devices (Velloso et al., 2016) or public displays (Pfeuffer et al., 2013; Vidal et al., 2013).

There are also more general limitations. We used a VR headset, and it would be useful to verify the results and conclusions of this work with an AR system. Based on the low error rates observed with input techniques such in-air gestures (which feature highly visually salient cues in the form of hand motions) in the Head Pointing study, we believe our results will generalize well to AR settings. Additionally, while no participants reported motion sickness during our studies, most likely due to the minimal nature of the visual cues presented, future work should explore whether input tasks in richer graphical environments are more susceptible to this type of problem. Future work should also explore the effect of these selection mechanism when used with more refined gaze pointing implementations (Kytö et al., 2018).

Finally, we have identified two minor procedural limitations that are worth highlighting. The first is that, despite its simplicity, the mid-air gesture used might have required a longer learning phase than the other, more familiar techniques, in the first study. The second is that, despite no reports indicating difficulties were experienced, we did not explicitly check whether any participants had reduced color perception. This could have affected how able participants were to perceive the interface cues in the second study as they were displayed in red, green, and grey. Future studies in this area should take care to control for these issues.

8. Conclusion

This paper presented two comparative studies examining three hands-on (clicker, on-device input, mid-air gestures) and three hands-free (dwell, speech, motion matching) selection mechanisms for gaze input techniques in HMDs. We report on these techniques' performance, perceived exertion, and participant preference, with the goal of providing clarity and a baseline on the pros and cons of these different selection methods. Further, we present several takeaways from these results and highlight areas for future research. These include blended approaches for gaze-based input involving dwell and motion matching and studying novel implementations and feedback mechanisms for the latter. In sum, this paper contributes empirically grounded insights about the performance of different selection methods that can be combined with head pointing input. Researchers, designers and developers aiming to produce usable and mature VR and AR systems can use this data and advice to make better choices and design more usable systems.

CRediT authorship contribution statement

Augusto Esteves: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Writing - original draft, Writing - review & editing, Visualization, Project administration. **Yonghwan Shin:** Software, Investigation, Data curation. **Ian Oakley:** Conceptualization, Methodology, Software, Formal analysis, Resources, Data curation, Writing - original draft, Writing - review & editing, Project administration, Funding acquisition.

Declaration of Competing Interest

None.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (2017R1D1A1B 03031364).

References

Aromaa, S., Aaltonen, I., Kaasinen, E., Elo, J., Parkkinen, I., 2016. Use of wearable and augmented reality technologies in industrial maintenance work. Proceedings of the 20th International Academic Mindtrek Conference. ACM, New York, NY, USA, pp. 235–242. https://doi.org/10.1145/2994310.2994321.

Atienza, R., Blonna, R., Saludares, M.I., Casimiro, J., Fuentes, V., 2016. Interaction techniques using head gaze for virtual reality. 2016 IEEE Region 10 Symposium (TENSYMP). pp. 110–114. https://doi.org/10.1109/TENCONSpring.2016.7519387.

Baird, K.M., Barfield, W., 1999. Evaluating the effectiveness of augmented reality displays for a manual assembly task. Virtual Real. 4 (4), 250–259. https://doi.org/10.1007/BF01421808.

Bates, R., Istance, H.O., 2003. Why are eye mice unpopular? a detailed comparison of

head and eye controlled assistive technology pointing devices. Univ. Access Inf. Soc. 2 (3), 280–290. https://doi.org/10.1007/s10209-003-0053-y.
Bernardos, A.M., Gómez, D., Casar, J.R., 2016. A comparison of head pose and deictic pointing interaction methods for smart environments. Int. J. Hum. Comput.Interact. 32 (4), 325–351. https://doi.org/10.1080/10447318.2016.1142054.
Birkfellner, W., Figl, M., Huber, K., Watzinger, F., Wanschitz, F., Hummel, J., Hanel, R., Greimel, W., Homolka, P., Ewers, R., Bergmann, H., 2002. A head-mounted operating binocular for augmented reality visualization in medicine - design and initial evaluation. IEEE Trans. Med. Imaging 21 (8), 991–997. https://doi.org/10.1109/TMI.2002.803099.
Blattgerste, J., Renner, P., Pfeiffer, T., 2018. Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. Proceedings of the Workshop on Communication by Gaze Interaction. ACM, New York, NY, USA, pp. 1:1–1:9. https://doi.org/10.1145/3206343.3206349.
Borg, G., 1998. Borg'S perceived exertion and pain scales. Human kinetics.
Caudell, T.P., Mizell, D.W., 1992. Augmented reality: an application of heads-up display technology to manual manufacturing processes. Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences. ii. pp. 659–669 vol.2. https://doi.org/10.1109/HICSS.1992.183317.
Clarke, C., Bellino, A., Esteves, A., Gellersen, H., 2017. Remote control by body movement in synchrony with orbiting widgets: an evaluation of tracematch. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1 (3), 45:1–45:22. https://doi.org/10.1145/3130910.
Clarke, C., Gellersen, H., 2017. MatchPoint: spontaneous spatial coupling of body movement for touchless pointing. Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology. ACM, New York, NY, USA, pp. 179–192. https://doi.org/10.1145/3126594.3126626.
Clifford, R.M.S., Tuanquin, N.M.B., Lindeman, R.W., 2017. Jedi ForceExtension: Telekinesis as a Virtual Reality interaction metaphor. 2017 IEEE Symposium on 3D User Interfaces (3DUI). pp. 239–240. https://doi.org/10.1109/3DUI.2017.7893360.
Controls -, 2014. Interactive patterns - Designing for Google Cardboard. URL https://designguidelines.withgoogle.com/cardboard/interactive-patterns/controls.html.
Day, P.N., Ferguson, G., Holt, P.O., Hogg, S., Gibson, D., 2005. Wearable augmented virtual reality for enhancing information delivery in high precision defence assembly: an engineering case study. Virtual Real. 8 (3), 177–184. https://doi.org/10.1007/s10055-004-0147-8.
Drewes, H., Khamis, M., Alt, F., 2018. DialPlate: Enhancing the Detection of Smooth Pursuits Eye Movements Using Linear Regression. arXiv:1807.03713 [cs]. ArXiv: 1807.03713
Esteves, A., Velloso, E., Bulling, A., Gellersen, H., 2015. Orbits: Gaze Interaction for Smart Watches using Smooth Pursuit Eye Movements. Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology. ACM Press, pp. 457–466. https://doi.org/10.1145/2807442.2807499.
Esteves, A., Verweij, D., Suraiya, L., Islam, R., Lee, Y., Oakley, I., 2017. SmoothMoves: Smooth Pursuits Head Movements for Augmented Reality. Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology. ACM, New York, NY, USA, pp. 167–178. https://doi.org/10.1145/3126594.3126616.
Fekete, J.-D., Elmqvist, N., Guiard, Y., 2009. Motion-pointing: Target Selection Using Elliptical Motions. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 289–298. https://doi.org/10.1145/1518701.1518748.
Fitts, P.M., 1954. The information capacity of the human motor system in controlling the amplitude of movement. J. Exp. Psychol. 47 (6), 381–391. https://doi.org/10.1037/h0055392.
Fuchs, H., Livingston, M.A., Raskar, R., Colucci, D., Keller, K., State, A., Crawford, J.R., Rademacher, P., Drake, S.H., Meyer, A.A., 1998. Augmented reality visualization for laparoscopic surgery. Medical Image Computing and Computer-Assisted Intervention MICCAI 98. Springer, Berlin, Heidelberg, pp. 934–943. https://doi.org/10.1007/BFb0056282.
Guiard, Y., 2009. The problem of consistency in the design of fitts' law experiments: consider either target distance and width or movement form and scale. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 1809–1818. https://doi.org/10.1145/1518701.1518980.
Ha, T., Feiner, S., Woo, W., 2014. WeARHand: Head-worn, RGB-D camera-based, bare-hand user interface with visually enhanced depth perception. 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). pp. 219–228. https://doi.org/10.1109/ISMAR.2014.6948431.
Hansen, J.P., Rajanna, V., MacKenzie, I.S., Bækgaard, P., 2018. A Fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display. Proceedings of the Workshop on Communication by Gaze Interaction. ACM, New York, NY, USA, pp. 7:1–7:5. https://doi.org/10.1145/3206343.3206344.
Henderson, S.J., Feiner, S.K., 2011. Augmented reality in the psychomotor phase of a procedural task. 2011 10th IEEE International Symposium on Mixed and Augmented Reality. pp. 191–200. https://doi.org/10.1109/ISMAR.2011.6092386.
HoloLens, 2016. Interaction Model. URL https://blogs.windows.com/buildingapps/2016/01/21/hololens-interaction-model/.
HoloLens, 2018b. Use gestures. URL https://support.microsoft.com/en-us/help/12644/hololens-use-gestures.
Hsieh, Y.-T., Jylhä, A., Orso, V., Gamberini, L., Jacucci, G., 2016. Designing a willing-to-use-in-public hand gestural interaction technique for smart glasses. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 4203–4215. https://doi.org/10.1145/2858036.2858436.
Istance, H., Bates, R., Hyrskykari, A., Vickers, S., 2008. Snap clutch, a moded approach to solving the Midas touch problem. Proceedings of the 2008 symposium on Eye tracking research & applications. Association for Computing Machinery, Savannah, Georgia, pp. 221–228. https://doi.org/10.1145/1344471.1344523.
Jacob, R.J.K., 1990. What you look at is what you get: eye movement-based interaction techniques. ACM Press, pp. 11–18. https://doi.org/10.1145/97243.97246.
Jalaliniya, S., Mardanbegi, D., Pederson, T., 2015. MAGIC Pointing for Eyewear Computers. Proceedings of the 2015 ACM International Symposium on Wearable Computers. ACM, New York, NY, USA, pp. 155–158. https://doi.org/10.1145/2802083.2802094.
Jalaliniya, S., Mardanbeigi, D., Pederson, T., Hansen, D.W., 2014. Head and eye movement as pointing modalities for eyewear computers. 2014 11th International Conference on Wearable and Implantable Body Sensor Networks Workshops. pp. 50–53. https://doi.org/10.1109/BSN.Workshops.2014.14.
Jr, J.J.L., Kruijff, E., McMahan, R.P., Bowman, D., Poupyrev, I.P., 2017. 3D User Interfaces: Theory and Practice. Addison-Wesley Professional. Google-Books-ID: fxWSDgAAQBAJ
Kangas, J., Åpakov, O., Isokoski, P., Akkil, D., Rantala, J., Raisamo, R., 2016. Feedback for smooth pursuit gaze tracking based control. Proceedings of the 7th Augmented Human International Conference 2016. ACM, New York, NY, USA, pp. 6:1–6:8. https://doi.org/10.1145/2875194.2875209.
Khamis, M., Oechsner, C., Alt, F., Bulling, A., 2018. VRpursuits: interaction in virtual reality using smooth pursuit eye movements. Proceedings of the 2018 International Conference on Advanced Visual Interfaces. ACM, New York, NY, USA, pp. 18:1–18:8. https://doi.org/10.1145/3206505.3206522.
Kjeldsen, R., 2001. Head gestures for computer control. Proceedings IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems. pp. 61–67. https://doi.org/10.1109/RATFG.2001.938911.
Kytö, M., Ens, B., Piumsomboon, T., Lee, G.A., Billinghurst, M., 2018. Pinpointing: precise head- and eye-based target selection for augmented reality. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 81:1–81:14. https://doi.org/10.1145/3173574.3173655.
Library, 2019. for embedding immersive media into traditional websites.: googlearchive/vrview. https://github.com/googlearchive/vrview Original-date: 2015-10-13T01:45:58Z.
Lukowicz, P., Timm-Giel, A., Lawo, M., Herzog, O., 2007. Wearit@work: toward real-world industrial wearable computing. IEEE Pervasive Comput. 6 (4), 8–13. https://doi.org/10.1109/MPRV.2007.89.
Mauchly, J.W., 1940. Significance test for sphericity of a normal n-variate distribution. Ann. Math. Stat. 11 (2), 204–209.
Miniotas, D., Å pakov, O., Tugoy, I., MacKenzie, I.S., 2006. Speech-augmented eye gaze interaction with small closely spaced targets. Proceedings of the 2006 Symposium on Eye Tracking Research & Applications. ACM, New York, NY, USA, pp. 67–72. https://doi.org/10.1145/1117309.1117345.
Morawiec, D., 2019. Contributed library to use the leap motion in processing.: nok/leap-motion-processing. Original-date: 2013-02-03T17:15:34Z.
Nancel, M., Chapuis, O., Pietriga, E., Yang, X.-D., Irani, P.P., Beaudouin-Lafon, M., 2013. High-precision pointing on large wall displays using small handheld devices. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 831–840. https://doi.org/10.1145/2470654.2470773.
Ockerman, J.J., Pritchett, A.R., 1998. Preliminary investigation of wearable computers for task guidance in aircraft inspection. Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215). pp. 33–40. https://doi.org/10.1109/ISWC.1998.729527.
Park, H.M., Seok Han Lee, Jong Soo Choi, 2008. Wearable augmented reality system using gaze interaction. IEEE, pp. 175–176. https://doi.org/10.1109/ISMAR.2008.4637353.
Pfeuffer, K., Vidal, M., Turner, J., Bulling, A., Gellersen, H., 2013. Pursuit calibration: making gaze calibration less tedious and more flexible. Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology. ACM, New York, NY, USA, pp. 261–270. https://doi.org/10.1145/2501988.2501998.
Piumsomboon, T., Lee, G., Lindeman, R.W., Billinghurst, M., 2017. Exploring natural eye-gaze-based interaction for immersive virtual reality. 2017 IEEE Symposium on 3D User Interfaces (3DUI). pp. 36–39. https://doi.org/10.1109/3DUI.2017.7893315.
Qian, Y.Y., Teather, R.J., 2017. The eyes don't have it: an empirical comparison of head-based and eye-based selection in virtual reality. Proceedings of the 5th Symposium on Spatial User Interaction. ACM, New York, NY, USA, pp. 91–98. https://doi.org/10.1145/3131277.3132182.
Ramos, G., Boulos, M., Balakrishnan, R., 2004. Pressure widgets. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, Vienna, Austria, pp. 487–494. https://doi.org/10.1145/985692.985754.
Rico, J., Brewster, S., 2010. Gesture and voice prototyping for early evaluations of social acceptability in multimodal interfaces. International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction. ACM, New York, NY, USA, pp. 16:1–16:9. https://doi.org/10.1145/1891903.1891925.
Samsung, 2018c. Gear VR: How do I use the touch pad? | Samsung Support Saudi Arabia. URL https://github.com/googlevr/vrview Original-date: 2015-10-13T01:45:58Z.
Serrano, M., Ens, B., Yang, X.-D., Irani, P., 2015. Gluey: Developing a Head-Worn Display Interface to Unify the Interaction Experience in Distributed Display Environments. Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services. ACM, New York, NY, USA, pp. 161–171. https://doi.org/10.1145/2785830.2785838.
Sibert, L.E., Jacob, R.J.K., 2000. Evaluation of Eye Gaze Interaction. ACM Press, pp. 281–288. https://doi.org/10.1145/332040.332445.
Sidorakis, N., Koulieris, G.A., Mania, K., 2015. Binocular eye-tracking for the control of a 3d immersive multimedia user interface. 2015 IEEE 1st Workshop on Everyday Virtual Reality (WEVR). pp. 15–18. https://doi.org/10.1109/WEVR.2015.7151689.
Soukoreff, R.W., MacKenzie, I.S., 2004. Towards a standard for pointing device

evaluation, perspectives on 27 years of fitts law research in HCI. Int J Hum Comput Stud 61 (6), 751–789. https://doi.org/10.1016/j.ijhcs.2004.09.001.

Sousa, M., Mendes, D., Paulo, S., Matela, N., Jorge, J., Lopes, D.S., 2017. VRRRRoom: Virtual Reality for Radiologists in the Reading Room. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 4057–4062. https://doi.org/10.1145/3025453.3025566.

Tang, A., Owen, C., Biocca, F., Mou, W., 2003. Comparative effectiveness of augmented reality in object assembly. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 73–80. https://doi.org/10.1145/642611.642626.

Tanriverdi, V., Jacob, R.J.K., 2000. Interacting with eye movements in virtual environments. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 265–272. https://doi.org/10.1145/332040.332443.

Use, 2017. the HoloLens clicker. URL https://support.microsoft.com/en-us/help/12646/hololens-use-the-hololens-clicker.

HeadTransform, 2018a. | Google VR. URL https://developers.google.com/vr/reference/android/com/google/vr/sdk/base/HeadTransform.

Velichkovsky, B., Sprenger, A., Unema, P., 1997. Towards Gaze-mediated Interaction: Collecting Solutions of the “Midas Touch Problem”. Human-Computer Interaction INTERACT ’97. Springer, Boston, MA, pp. 509–516. https://doi.org/10.1007/978-0-387-35175-9_77.

Velloso, E., Wirth, M., Weichel, C., Esteves, A., Gellersen, H., 2016. AmbiGaze: direct control of ambient devices by gaze. Proceedings of the 2016 ACM Conference on Designing Interactive Systems. ACM Press, pp. 812–817. https://doi.org/10.1145/2901790.2901867.

Vertegaal, R., 2008. A Fitts law comparison of eye tracking and manual input in the selection of visual targets. Proceedings of the 10th International Conference on Multimodal Interfaces. ACM, New York, NY, USA, pp. 241–248. https://doi.org/10.1145/1452392.1452443.

Vidal, M., Bulling, A., Gellersen, H., 2013. Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, New York, NY, USA, pp. 439–448. https://doi.org/10.1145/2493432.2493477.

vrview:, 2018d. Library for embedding immersive media into traditional websites. https://github.com/googlevr/vrview Original-date: 2015-10-13T01:45:58Z.

Williamson, J., Murray-Smith, R., 2004. Pointing without a pointer. CHI ’04 Extended Abstracts on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 1407–1410. https://doi.org/10.1145/985921.986076.

Yu, C., Gu, Y., Yang, Z., Yi, X., Luo, H., Shi, Y., 2017. Tap, dwell or gesture?: exploring head-based text entry techniques for HMDs. Proceedings of the 2017CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 4479–4488. https://doi.org/10.1145/3025453.3025964. Event-place: Denver, Colorado, USA

Yu, C., Sun, K., Zhong, M., Li, X., Zhao, P., Shi, Y., 2016. One-dimensional handwriting: inputting letters and words on smart glasses. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 71–82. https://doi.org/10.1145/2858036.2858542.

Zheng, X.S., Foucault, C., Matos da Silva, P., Dasari, S., Yang, T., Goose, S., 2015. Eye-wearable technology for machine maintenance: effects of display position and hands-free operation. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 2125–2134. https://doi.org/10.1145/2702123.2702305.