Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality

Mikko Kytö[1,2], **Barrett Ens**[2], **Thammathip Piumsomboon**[2], **Gun A. Lee**[2], and **Mark Billinghurst**[2]
[1]Aalto University, Espoo, Finland
mikko.kyto@aalto.fi
[2]University of South Australia, Adelaide, Australia
{barrett.ens, thammathip.piumsomboon, gun.lee, mark.billinghurst} @unisa.edu.au

ABSTRACT
Head and eye movement can be leveraged to improve the user's interaction repertoire for wearable displays. Head movements are deliberate and accurate, and provide the current state-of-the-art pointing technique. Eye gaze can potentially be faster and more ergonomic, but suffers from low accuracy due to calibration errors and drift of wearable eye-tracking sensors. This work investigates precise, multimodal selection techniques using head motion and eye gaze. A comparison of speed and pointing accuracy reveals the relative merits of each method, including the achievable target size for robust selection. We demonstrate and discuss example applications for augmented reality, including compact menus with deep structure, and a proof-of-concept method for on-line correction of calibration drift.

Author Keywords
Eye tracking; gaze interaction; refinement techniques; target selection; augmented reality; head-worn display

ACM Classification Keywords
H.5.2 Information interfaces and presentation: User Interfaces: Input devices and strategies

INTRODUCTION
Recently available head-worn Augmented Reality (AR) devices will become useful for mobile workers in many practical applications, such as controlling networks of smart objects [15], situated analytics of sensor data [14], or in-situ editing of CAD or architectural models [36]. For users to be mobile and productive, it is important to design interaction techniques that allow precise selection and manipulation of virtual objects, without bulky input devices.

Eye gaze is a potentially useful input mode for AR applications, since it uses an innate human ability and doesn't require extra hardware to be carried. However, eye gaze is well known to be inaccurate, due to both human physiology and tracking system limitations. Head-pointing has been used as a proxy for gaze [42,53], and is fairly precise, but requires unnatural, fatiguing head movements [3,4,31]. Alternatively, researchers have developed multimodal techniques that use a secondary input mode to refine eye gaze selection. Researchers have investigated such techniques in several domains, including desktop displays [56], handheld devices [49] and virtual reality [52], however they have been little explored for wearable AR.

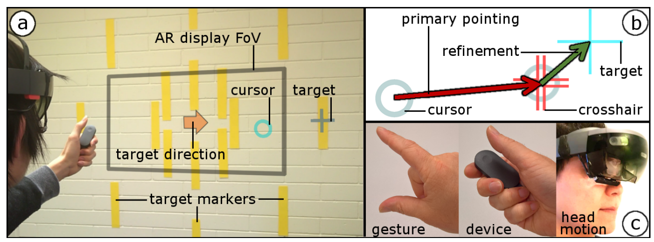

Figure 1. Pinpointing explores multimodal head and eye gaze selection for wearable AR a) Study layout of target markers, with feedback cues and HoloLens viewing field shown. b) Pinpointing techniques consist of a primary pointing motion plus secondary refinement. c) Refinement techniques: air-tap gesture, HoloLens clicker device, and head motion.

This paper explores Pinpointing: multimodal head and eye gaze pointing techniques for wearable AR (Figure 1). We build on prior work by adapting multimodal pointing refinement techniques for wearable AR, by combining gaze with hand gestures, handheld devices and head movement. Our exploration also includes head pointing, the current state-of-the-art pointing technique [30,35]. We further discuss the implications of these results for interface designers, and potential applications of Pinpointing techniques. We demonstrate two example implementations for precise menu selection and online improvement of gaze calibration.

KEY CONTRIBUTIONS
The contributions of the paper are:

- A broad comparison of target selection accuracy and speed for eye gaze, head pointing, and several multimodal techniques for improved accuracy. Results help clarify previous contradictory results for similar techniques, predict attainable target sizes for a wide range of techniques, and demonstrate previously unattained precision (< 0.2º) for head-based pointing.

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

ACM ISBN 978-1-4503-5620-6/18/04…$15.00
https://doi.org/10.1145/3173574.3173655

- Adaption of multimodal techniques for wearable AR, resulting in several previously unexplored implementations that refine coarse eye gaze and head pointing with fine hand gesture, device gyro and head motion input.
- Two example applications that demonstrate the potential of Pinpointing for improving wearable AR interaction: GazeBrowser uses gaze interaction with high precision pointing to navigate compact smart object menus. SmartPupil shows a novel online method for mitigating calibration drift of wearable eye trackers.

RELATED WORK ON GAZE BASED INTERACTION

Our user study investigates head- and eye gaze-based interaction techniques coupled with different refinement techniques. We review the related work in the following.

Head- and Eye-Based Target Selection

Our study explores eye gaze as an input method, as well as head pointing, which can provide a proxy for gaze, but has become a separate method in its own right.

Head-pointing

Together with hand-based interaction techniques, head-based interaction has been actively investigated in the field of 3D user interface, virtual reality (VR) [6,11], desktop GUIs [5,29], assistive interfaces [37], and wearable computing [7]. One of the earliest works in interaction techniques for virtual environments [40] included head directed navigation and object selection. Recently head-direction-based pointing has been widely adopted as a standard way of pointing at virtual objects without using hands or hand-held pointing devices (e.g., Oculus Rift [44] and Microsoft HoloLens [39]). Atienza et al. [1] further explored head-based interaction techniques in a VR environment. With wearable eye-tracking devices becoming affordable to use in combination with head-worn displays (e.g. Pupil Labs [32,51], FOVE [19]), researchers are increasingly exploring wearable eye gaze input [50,55].

Eye gaze

While initially used for measuring and understanding users' focus and attention [41], eye gaze has been actively investigated as an input method [27]. Gaze pointing uses eye tracking technology to identify which object a person is looking at. In one of the earliest investigations, Jacob [28] proposed basic interaction techniques using eye gaze on a desktop computer. Eye movement reflects not only conscious (explicit) but also unconscious (implicit) intent. As a result, eye-based input suffers from the well-known 'Midas Touch' problem [28] of involuntarily selection. Researchers have investigated solutions to this problem, mostly based on dwell time (e.g. [28,45,54,62]), smooth pursuits, where eye gaze follows continuously a target (e.g. [17,33,63,64]) and gaze gestures (e.g. [2,13,25,26]), but also by using a second modality for confirming selections (e.g. a button press or hand gesture as in HoloLens).

Inaccuracy of eye tracking causes challenges in designing gaze-based interactions. Feit et al. [18] showed that achieving a success-rate of 90% percent of target fixations requires targets as large as 5.9 cm in width and 6.2 cm, at 65 cm from the screen, although filtering eye-movements can decrease target size by 35% (3.9 cm width, 4.2 cm height). Such inaccuracy is more challenging with gaze-based interaction in limited field-of-view (FoV) head-worn displays, such as HoloLens (FoV approx. 30×17º), which we use in our work.

Beyond calibration issues, eye gaze interaction is limited by sensor noise in pupil detection and drift due to shifting of the eye tracking hardware [8,34,50]. We address this issue with an application that uses refinement input to improve the calibration as the system is used (See SmartPupil: Online Calibration Improvement, below).

Comparative Studies on Pointing Techniques

Head-pointing is well known for its benefit of providing hands-free interaction, yet its performance and usability has been considered inferior compared to hand-based input methods. Early investigation by Jagacinski and Monk [29] reported a joystick being faster than head-pointing. Lin et al. [35] compared head- and hand-directed pointing methods on a large stereoscopic projection display. The results suggest hand directed pointing has better overall performance, lower muscle fatigue, and better usability, yet head-pointing provides better accuracy. Bernardos et al. [4] compared pointing with an index finger and head- pointing on a wall-size projection screen in terms of speed and accuracy. They did not find a significant difference in terms of task performance, yet hand-based pointing showed better perceived usability.

In comparison to eye gaze interaction, head-pointing is more voluntary and stable. Bates and Istance [3] compared head- and gaze-based pointing techniques on a desktop computer, and found that eye gaze had worse performance, a steeper learning curve, was more uncomfortable to use, and required higher workload. Similarly, Jalaliniya et al. [31] compared eye and head pointing with mouse pointing, and found that eye gaze was faster than head or mouse, but head motion was more accurate and convenient.

Comparing eye gaze with hand pointing in VR, Tanriverdi and Jacob [61] found eye gaze performed faster, especially for distant objects, while participants' ability to recall spatial information was weakened. However, Cournia et al. [12] later found eye gaze performed worse for distant objects. They postulated this contradiction in results might have been resulted from a difference in interaction styles.

These works show that despite disadvantages, eye gaze interaction has many potential benefits for wearable AR. Because tracking limitations make gaze a poor selection tool, especially for very small objects, we explore how to improve accuracy by coupling gaze with other techniques, such as hand gestures or handheld devices. We furthermore investigate whether it is possible to use similar methods to substantially improve the precision of head pointing to fare better against standard pointing methods such as the mouse.

Combining Pointing Techniques

Next, we introduce prior works that have explored multi-modal interaction methods using gaze- or head-based input.

Combination of Eye Gaze and Head-Pointing

Closest to our work in focus on pointing accuracy, Spakov et al. [56] proposed using head movement to complement the low accuracy of gaze-based pointing. From a series of user experiments, they found head-assistance significantly improved accuracy without sacrificing efficiency.

The works discussed thus far, along with several others [38,43,57], have investigated eye gaze and head pointing primarily with desktop monitors. A handful of recent works have used wearable eye trackers with head-worn AR and VR displays. Jalaliniya et al. [30] used a Google Glass [65] combined with a custom eye tracker to investigate combination of head- and eye-based pointing. They found that their proposed refinement of quick eye-based pointing with subsequent head-motion, was faster than head pointing alone, without sacrificing accuracy. Conversely, Qian and Teather [52] recently found that head-pointing was faster than combined eye and head input in wearable VR. They also found, contrary to other previous studies [31,61], that head input was faster than eye gaze only. As such, it is still unclear how eye gaze and head-pointing should be combined in order to allow fast and accurate pointing.

Combination of Eye gaze and Manual Input

One of the earliest efforts to refine eye gaze input was MAGIC, proposed by Zhai et al. [66], which used a mouse to improve gaze accuracy. When the user looks at a target object, the mouse pointer appears at the gaze point, allowing users to refine its position. Through a pilot study they showed MAGIC pointing could reduce physical effort and fatigue compared to manual input alone, while providing greater accuracy than eye gaze alone.

Chuan and Sivaji [10] compared a combination of eye gaze and finger pointing against mouse and finger pointing alone on a desktop interface. They found the proposed method had lower error and faster performance on larger targets compared to finger pointing, while mouse outperformed both overall. Later, Chatterjee et. al. [9] investigated various desktop interaction methods combining eye gaze with hand gesture input. A Fitts' Law study showed a proposed method having a higher index of performance compared to eye gaze or hand gesture input alone.

Pfeuffer et al. investigated using eye gaze coupled with touch input in various setups including touch screen [46] multi-screen [48], touch pen [47], and tablet computer [49] setups. The most relevant work to our study is CursorShift, a technique that combined eye-gaze and touch for tablet interaction, using eye-gaze for low fidelity cursor position and touch for fine tuning the cursor position [49].Using a similar touch refinement approach, Stellmach and Dachselt developed various eye- and head-based refinement techniques on a distant screen using a handheld touch surface [58,59]. While their eye-based touch refinement technique was faster, a head-controlled zoom approach provided users with more feeling of control [59].

In summary, much prior research on improving gaze input has used hand input for refinement, while some works have used head motion, with mixed results. While head-pointing is shown to be less accurate than mouse input [31], there have not been any efforts, to our knowledge, to refine head movements as we explore in this work. Furthermore, most prior work has focused on desktop environments, whereas few studies have explored gaze refinement for wearable displays. Whereas most prior works focus on a single technique, we provide a broad comparison of both eye gaze and head pointing with several refinement methods (scaled head motion, hand gesture input, and handheld device input) for improved accuracy on a head-worn AR display.

PINPOINTING: SELECTION TECHNIQUES FOR AR

In this section, we discuss the primary design factors relevant to Pinpointing. We define Pinpointing as the use of multimodal pointing techniques for wearable (AR) interfaces. Specifically, these techniques use coarse pointing selection, followed by a secondary, local refinement motion (Figure 1) to provide pinpoint accuracy. In this exploratory work, we limit our investigation to Pinpointing on 2D surfaces, as might be used in menu selection, interactive visualization or in-situ CAD applications.

Target Applications

Our exploration of these techniques is aimed at wearable AR applications that require precise accuracy for selecting virtual or real objects. A wide variety of applications can make use of precise selection in menus, for instance to select system parameters, or to control various functions of a smart object. Because many head-worn AR displays have a limited FoV and because virtual menus or annotations can obstruct important real-world objects, menu item size should be minimized as much as possible. Also, many applications such as interactive data visualizations or in-situ CAD may require the selection of tiny visual features.

Design Requirements

We summarize the design requirements for Pinpointing interaction as follows:

R1) Pinpointing must balance the needs of selecting large objects with minimal speed and effort, with the ability to select very small targets when desired.

R2) Because wearable AR platforms overlay content directly on the real world, interaction should leverage the context provided by a user's visual focus.

R3) To afford mobility in interactive environments and provide a natural experience with virtual content, use of wieldy, external devices should be minimized.

R4) Interaction in AR applications should be as 'invisible' as possible, so that users are primarily focused on real and virtual objects, and not mechanics of the interface.

R5) Interactions that trigger noticeable object behaviours should be deliberate, so that users are not distracted by unintended consequences of actions.

Next, we discuss the primary elements of Pinpointing techniques – primary pointing mode, selection method, and refinement technique – and describe the options used in the current work. A summary of techniques used in our following study is outlined in Figure 2.

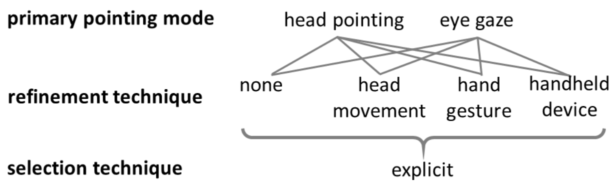

Figure 2. Pinpointing refinement methods. Each primary selection mode can be paired with any refinement method.

Primary Pointing Mode

Our Pinpointing exploration includes two primary input modes, eye gaze and head pointing. These modes have strong potential for head-worn AR devices, since they require low effort and the required sensors can be embedded into devices worn by the users. These methods observe requirements R2 and R4, since a user's focus of attention is a strong indicator of their intended actions.

Selection Method

Selection methods can be defined as implicit or explicit. Implicit selections are made without any conscious effort by the user, whereas explicit selections require a deliberate user action. Eye gaze and head pose can both be used to trigger selections implicitly, with coarse accuracy, by predicting when a user is focused on a particular object.

Currently, we explore only explicit selections. While mechanisms that require no additional input mode have been studied (R3), such as head tilt [57] and dwell [28], we instead use simple yet reliable methods that provide fast and deliberate interaction (R5). Our implementations use two simple triggers, a button click on a small device and a finger gesture, that both integrate cleanly with the refinement input modes described next.

Refinement Technique

Secondary refinement techniques provide precision beyond the limits of eye or head input alone, to allow the selection of very small targets; after the initial selection with the eye or head pointing mode, a second mode is used to provide additional input. In practical applications, this refinement phase can be made optional, however for study purposes we strictly delineate techniques with or without refinement.

In our study we explore three options that add minimal bulk for mobile users (R2): head motion, hand gesture input, and input from a small handheld device (Figure 1, Figure 2). These methods all provide a clear distinction from the primary mode (R5), and can be scaled for fine control (R1). We implemented and tested these methods in the user study described next.

PINPOINTING STUDY

We conducted a user study to compare various Pinpointing techniques to evaluate their efficacy as pointing techniques for AR. We investigate both eye gaze and head pointing as primary pointing modes, each combined with head motion, hand gesture and device refinement techniques (Figure 2). For the current study, we use only explicit selection techniques, triggered by a handheld device or hand gesture. This study explores only 2D selection, with all targets and feedback superimposed on a wall of the study environment.

Technique Implementation

These techniques were implemented on a Microsoft HoloLens device [39] running Windows 10. Mounted on the HoloLens is a Pupil Labs' eye tracker [51] (Figure 3), which captures the user's right eye at 120 Hz. The eye tracker is tethered to a desktop PC (Intel Core i7-770K with NVIDIA GeForce GTX 1070), which sends eye tracking data to the HoloLens via Wi-Fi. The eye tracker is calibrated using a separate desktop program linked to Pulil Labs' Pupil Capture software, remotely connected to the HoloLens. The selection techniques are implemented as follows and summarized in Table 1.

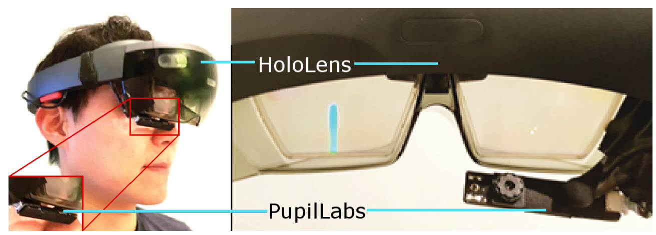

Figure 3. Pinpointing techniques were implemented on a Microsoft HoloLens with a Pupil Labs eye tracking device mounted below the user's right eye.

The baseline head-pointing technique (*Head Only*), is controlled by a 1:1 mapping of the user's head position. Selections are triggered explicitly using either the handheld HoloLens 'clicker' device, or by an 'air-tap' gesture.

Technique	Pre-refine feedback	Refinement triggered	Refinement feedback	Selection triggered
Head Only		none	none	click up
Head+ Gesture		air-tap down		air-tap up
Head+ Device		click down		click up
Head+ Head		click down		click up
Eye Only	none	none	none	click up
Eye+ Gesture	none	air-tap down		air-tap up
Eye+ Device	none	click down		click up
Eye+Head	none	click down		click up

Table 1. Refinement technique mechanisms and feedback.

The baseline eye gaze technique (*Eye Only*) works by simply directing the eye to the target centre and triggering an explicit selection. To make the technique as fast and natural as possible, we opted to not provide feedback on the detected gaze position. While a gaze-controlled cursor may help improve selection accuracy, it may also slow the selection speed and cause unwanted distraction [20,66].

The six refinement techniques consist of two phases, *pre-refinement* and *refinement*. In the pre-refinement phase, the user makes an initial target selection with the given baseline technique. For head conditions, a cursor is shown (without crosshairs). For eye conditions, the user looks at the target.

The refinement phase provides an opportunity for users to improve their initial selection. Refinement begins when the user holds down either the device button or air-tap gesture, and ends when the button or air-tap is released. As the start of the refinement phase, a crosshair cursor appears at the position of the head-pointer or detected gaze. The crosshair is controlled according to the corresponding technique:

The *Head+Device* and *Eye+Device* techniques use rotation of the HoloLens clicker device (Figure 1c) to control the crosshair. Control is similar to a gyro-mouse [21], where the device yaw and pitch control the x and y crosshair movement, respectively. The start and end of the refinement phase are controlled by the clicker button, which can be easily held down while rotating the device.

The *Head+Gesture* and *Eye+Gesture* techniques use horizontal and vertical movement of the hand to control the x and y crosshair movement, respectively. The start and end of the refinement phase are controlled by the air-tap gesture, which consists of moving an extended finger down toward the thumb to tap and back up to release (Figure 1c).

With the *Head+Head* and *Eye+Head* techniques, the crosshair is controlled by head motion, as with head pointing. Whereas the control-display (CD) ratio of head-pointing is 1:1, head refinement uses a higher CD ratio (set to 2:1 with pilot trials) to increase targeting precision. The CD ratio of device and gesture refinement techniques were tuned so that all three were similar. The refinement duration is controlled by holding and releasing the clicker button.

Task

Participants are required to select a point target, marked by a cross, using a crosshair cursor (Figure 1). From a central starting position, targets are found in one of 8 compass directions at 2 target distances. The direction factor is included to account for variation of eye calibration in different regions of the user's view, and for drift as the head is moved. As current AR displays typically have relatively a small display size, we test two distances to include targets that lie either inside or outside the device FoV. *Near* distance targets are initially visible within the device FoV, so that participants can ideally shift their eye gaze among all items using eye movement alone. Targets at the *Far* distance require participants to move their head before being able to see them.

To prevent the need to search for targets, all locations are anchored to visible, real-world markers (Figure 1). Markers are placed on a wall 2m in front of the participant, which coincides with the focal distance and virtual image plane distance of the Microsoft HoloLens.

To begin a trial, the participant must align a visible head pointer with a central starting target (1.2° diameter). After a random interval of 750-1250 ms, a virtual cross-shaped target appears at one of the 16 marker positions, chosen in random order. Since the target may not be initially visible within the FoV, a directional arrow appears in place of the central target to indicate the target's direction (Figure 1).

Participants are asked to align a crosshair cursor (see Design and Procedure, below) with the target as precisely as possible, and to do so as quickly as possible. To discourage excessive effort on accuracy at the expense of time, a limit of 5 s (based on pilot studies) is placed on each trial. Similarly, distance limit (5° at start of refinement and 2° for selection) is enforced to eliminate trials affected by mishaps such as sensor noise.

Design and Procedure

We used a within-participants design with 3 factors:

- selection **technique** (*Head Only*, *Eye Only*, *{Head/Eye}+{Gesture/Device/Head}*)
- target **angle** (*0*, *45*, *90*, *135*, *180*, *225*, *270*, *315*° measured anti-clockwise from the right axis)
- target **distance** (*Near* and *Far* – 7, 21° from centre, respectively)

For each technique, participants completed 3 blocks (in addition to 1 block for training) of all target combinations, for a total of 8 techniques × 8 directions × 2 distances × 3 blocks = 192 trials each.

To mitigate fatigue, the study was broken into two sessions lasting 40-90 minutes each, for head-based and eye-based techniques. To reduce learning effects, half of participants were randomly assigned to start with either session. Targets were presented in a random order within each block, and refinement techniques within each session were fully counter-balanced using a Latin square design.

For analysis, we collected the position of each selection on the target plane, including the pointing position both before and after the refinement phase. We also collected the time of each trial and each phase. Participants provided demographic information prior to the study and completed a questionnaire on completion, containing questions and comments on their general preference. After each technique, the perceived task load of each technique was measured using a raw-TLX questionnaire [23].

Participants

We recruited 12 participants (2 female, mean age 32 years, SD = 6.8) from our university. Due to the variety of technologies used for this study, we sought participants with at least some previous experience with one or more of them. All but 2 participants had at least intermediate experience using handheld AR and see-through AR displays such as the HoloLens, and 8 had intermediate or expert experience with eye-tracking equipment.

Analysis

We conducted a repeated-measures ANOVA (α = .05) for accuracy and time by having interaction technique, target angle and distance as independent variables. When the assumption of sphericity was violated (tested with Mauchly's test), we used Greenhouse-Geisser corrected values in the analysis. The post hoc tests were conducted using pairwise t-tests with Bonferroni corrections. Effect sizes are reported as partial eta squared (η_p^2). Time and error analyses included only successful target selections (6275 of 7263 total trials, or 86.4%). Accuracy results are reported in angular distance (degrees). For reference, a 1 degree angle is approximately 3.5 cm in width at a distance of 2 m.

RESULTS

The primary results are summarized in Figure 4 and Figure 5. For head-based input, all three refinement techniques exceed the accuracy of the state-of-the-art *Head Only* technique by two to three times. We also found out that *Head+Device* and *Head+Head* refinements did not increase the task load compared to *Head only*. *Head+Device* and *Head+Head* were subjectively preferred over *Head only*.

For the eye gaze techniques, refinement improved selection accuracy over the baseline *Eye Only* substantially, by five to six times, but remained less precise than *Head Only*. Our introduced device-gyro refinement technique (*Eye+Device*) performed well, providing slightly better accuracy than *Eye+Head* and slightly faster than *Eye+Gesture*. It was also subjectively preferred over *Eye+Head* and *Eye+Gesture* and required lower perceived task load. We provide more detailed analyses in the following.

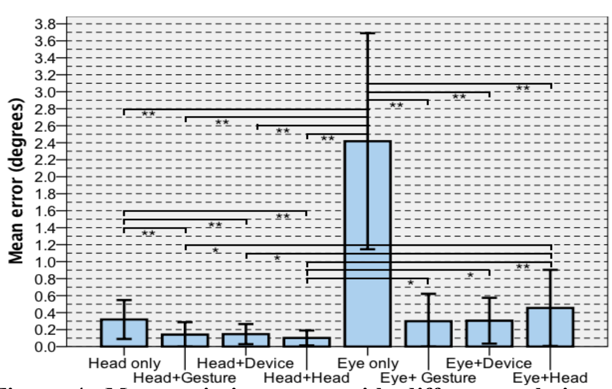

Figure 4. Mean pointing error with different techniques. Statistical significances are marked with stars (=p<.01 and *=p<.05). Error bars represent standard deviations.**

Accuracy

We found a significant main effect on accuracy only for interaction technique, $F(1.18, 13.01) = 63.69$, $p < .001$, $\eta_p^2 = .85$. No interaction effects were found. As expected, *Head Only* (mean error 0.32º, SD = 0.22) was more accurate than *Eye only* (2.42º, SD = 1.27, Figure 4).

Generally, the refinement techniques drastically improved pointing accuracy. All improvements in accuracy compared to the corresponding baseline mode were statistically significant (Figure 4). The accuracy of eye pointing in particular improved from 2.42º to be below 0.5º. Our results support findings from Spakov et al. [56], who showed that head movements can be used to improve accuracy of gaze on a desktop display. They found a nearly threefold improvement in accuracy, while we found nearly fivefold, from 2.42º to 0.49º. With *Gesture* refinement, improvement was even greater and the mean error decreased to 0.30º. As such, we extend the results of Chatterjee et al. [9], who used hand gestures to improve gaze accuracy in desktop setup, to wearable AR. In addition, our novel refinement technique *Eye+Device*, using a device's gyro sensor, was as accurate as gesture refinement.

With head pointing, accuracy was improved from 0.32º to be below 0.2º, a substantial improvement over the commonly used baseline technique. Interestingly, the eye-based refinement techniques remained slightly less accurate than *Head Only*. A similar result by Jalaliniya et al. [30] found that eye gaze with head refinement resulted in the same accuracy as head only. However, they used the same CD ratio for the baseline and refinement techniques, whereas we used higher CD-ratio (2:1) in the head-based refinement phase.

Overall, *Head+Head* appeared the most accurate; it was statistically more accurate than any of the eye based techniques, although not significantly better than *Head+Device* or *Head+Gesture*. *Head+Gesture* and *Head+Device* were more accurate than *Eye+Head*, which suffered noticeably from FoV limitations; head movements with *Eye+Head* were larger than with *Head+Head* because the refinement started an average of 1.6º further from the target (Figure 6a). These large movements often caused the target to move out of view with the *Eye+Head* technique. This problem could be mitigated by lowering the refinement phase CD ratios for the eye based techniques. However, for study purposes we kept CD ratios constant between eye- and head-based techniques.

Time

Main Effects

We found main effects on performance time of technique, $F(7, 77) = 40.03$, $p < .001$, $\eta_p^2 = .78$, distance $F(1,11) = 1000.00$, $p < .001$, $\eta_p^2 = .99$, and angle, $F(7,77) = 3.14$, $p = .006$, $\eta_p^2 = .22$. The difference for distance is easily explained by the required head movement for *Far* targets, but angle is more difficult to explain. Pairwise comparison

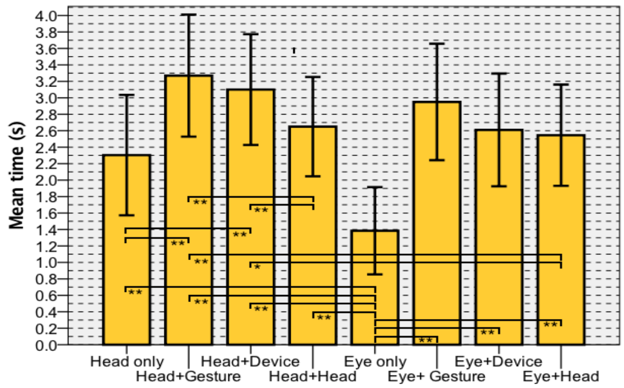

Figure 5. Mean times with different interaction techniques. The statistical significances are marked with stars (c=p<.01 and *=p<.05). Error bars represent standard deviations.**

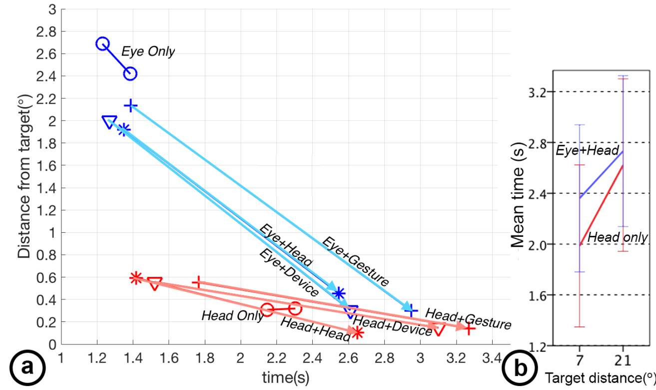

Figure 6. a) Time-error plot for different interaction techniques. Arrows represent differences at start and end of refinement[1] and b) interaction effect between distance and two techniques (*Head only* and *Eye+Head*) on time.

showed that *0°* (middle-right) was significantly faster than *225°* (bottom-left), and marginally faster than the other bottom targets (p = .084 for *270°* and p = .056 for *315°*). A similar study of target selection with a limited FoV by Ens et al. [16] also found slower selection for lower targets.

Whereas *Head Only* was more accurate than *Eye Only*, *Eye Only* was the faster technique (mean selection time 1.38s., SD = .73s versus 2.30s, SD = .73s). A similar conclusion was found by Jalaliniya et al. [31], however, Qian et al. [52] found the opposite, that head pointing was faster than eye pointing. The reason for this difference could be the presence of the cursor in Qian's study, although Jalaliniya [31] also visualized eye gaze using a the cursor. Thus, it is unclear what has caused the difference in results between these two studies, as they do not report radial distances to targets, which has been shown to have impact on results (see Interaction Effects, below).

Head movement as a refinement method was generally faster than gesture and device refinement. When paired with head-pointing, the difference was statistically significant (Figure 5). Moreover, *Eye+Head* was significantly faster than *Head+Device* and *Head+Gesture*. We expect the faster selection times with head refinement was partly because due to the lack of mode-switching during the selection task. In fact, refinement with the *Head+Head* condition started about 0.35s earlier than with *Head+Gesture* (Figure 6a). This difference was statistical significant (p =.026).

Interaction Effects

There was a significant interaction between technique and distance, F(3.77, 41.43) = 2.92, p = 0.035, η_p^2 = .21. From individual ANOVAs for each distance, we found that *Head Only* was significantly faster (p < .001) than *Eye+Head* for the *Near* distance but not for *Far* (p = 1.00, see Figure 6b). This result is similar to a study by Jalaliniya [30], which found that eye gaze with head refinement (CD ratio 1:1) was faster at large radial distances than head only. Possibly, all eye refinement techniques may improve in this way with a wider FoV, as the required head movements are reduced.

Task load

Results from the NASA TLX questionnaire showed differences for perceived task load between techniques, with significant effects for all, as well as the overall mean. The results from repeated-measures ANOVA were as follows: F_{Mental} (3.00) = 3.47, p = .027, η_p^2 = .24, $F_{Physical}$ (7) = 7.57, p < .001, η_p^2 = .41, $F_{Temporal}$ (3.04) = 3.57, p = .024, η_p^2 = .25, $F_{Performance}$ (3.38) = 5.84, p = .002, η_p^2 = .35, F_{Effort} (7) = 5.54, p < .001, η_p^2 = .34, $F_{Frustration}$ (7) = 5.09, p < .001, η_p^2 = .32 and F_{Mean} (3.50) = 7.81, p < .001, η_p^2 = .42. The results from pairwise comparisons between techniques can be seen in Figure 7 (next page). Eye-based interaction technique with gestural refinement (*Eye+Gesture*) revealed a much higher task load compared to other techniques.

For head-based techniques, *Head+Device* and *Head+Head* had the lowest task load for all attributes. Interestingly, the perceived *Physical load* and *Effort* were rated equally or even lower than *Head only*, which requires less head movement overall. For eye-based techniques, *Eye only* and *Eye+Device* had the lowest loads. *Eye+Head* scores suffered from the problem of the cursor disappearing beyond the FoV as explained earlier (see Accuracy, above).

When comparing corresponding head- and eye-based techniques together (i.e. *Eye+X* vs. *Head+X*), it can be noted that there were not significant differences between conditions, except with *Head* refinement in *Performance*. However, *Eye Only* generally had a lower task load than *Head Only* (though not significant). *Head* was lower than *Eye* for *Gesture* and *Head* refinements, where eye-based techniques resulted in a higher load, likely due to longer refinement motions.

[1] There is also a small change in mean error for *Eye Only*. As we used moving average to smooth eye movements, the fixations are filtered during the button click (≤ 200 ms), during which time the mean error is slightly decreased.

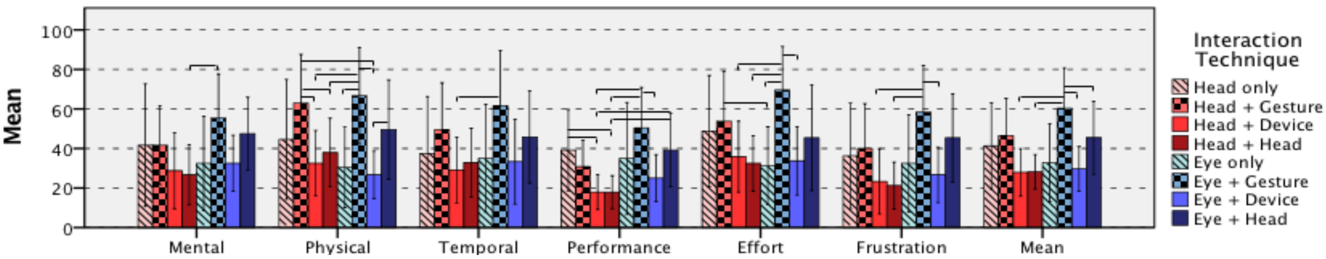

Figure 7. The mean responses for the attributes of NASA TLX questionnaire. The statistical significant differences are marked as connecting lines.

Preferences

Figure 8 shows the results of user preference. As well as being more accurate, head interaction was slightly preferred overall: 58.3% preferred *Head*, whereas 41.7% preferred *Eye*. Head pointing was often seen as the more familiar technique (P11: "*I usually speed through UI with a mouse and I think this reflects in this technique.*"). However, some participants found it difficult to stabilize head movements (P9: "*hard to keep the head steady for a longer time*") and this caused some issues to neck muscles (P5: "*I feel uncomfortable on my neck.*"). Eye pointing, conversely, was easy and fast (P15: "*Very easy and fast."),* but inaccurate and difficult to control (P4: "*It was frustrating that the eye tracking was not accurate enough. I was just hoping it will be close enough to the target. Cannot really control how accurately I can perform the task.*"). Interestingly, time for refinement techniques did not always match with preference. For example, *Head+Head* was faster than *Head+Device*, but was not preferred. *Device* was the most preferred refinement condition for both primary techniques.

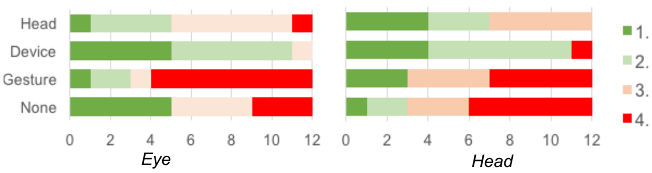

Figure 8. The Device refinement was the most preferred technique with both primary modes (1: best ~ 4: worst).

Defining target sizes

To design applications that use Pinpointing techniques, we computed the required target size for each technique using two methods. The first follows the approach presented by Feit et al. [18], who defined the width and height (S_{width} and S_{height}) for the rectangular targets according to Equation 1:

$$S_{width}/S_{height} = 2\,(O_{x/y} + 2\,\sigma_{x/y}) \qquad (1)$$

where $O_{x/y}$ is the mean x or y offset between target and values and $\sigma_{x/y}$ is the x or y component of the standard deviation. Then about 95% of values lie within two standard deviations of the mean for normally distributed data. Second, we computed ellipses using multivariate normal distribution with 95% confidence intervals. We used the maximum distance between the target position and ellipse edge to define the radii of circular targets. We expect the latter method to provide near-100% targeting accuracy in most conditions, although this has yet to be empirically tested. Sizes for rectangular (width and height) and circular (diameter) targets are shown in Table 2.

Our study resulted in a larger rectangular target size for *Eye Only* (7.0º × 12.7º) than reported by Feit et al. (5.2º × 5.5º). Both are extended along the *y*-dimension, however ours is much more so. This likely resulted from differences in the underlying eye tracking systems as well as drift introduced by our wearable configuration. We noticed the vertical alignment between the HoloLens display and eye tracker drifts more easily than the horizontal direction (Table 2).

Technique	Width (°)	Height (°)	Diameter (°)
Head Only	2.20	2.08	3.65
Head+Gesture	1.22	1.87	3.28
Head+Device	1.19	1.37	1.61
Head+Head	0.58	1.10	1.82
Eye Only	7.00	12.67	16.17
Eye+Gesture	2.06	2.70	4.70
Eye+Device	1.80	2.00	3.75
Eye+Head	1.98	3.95	6.54

Table 2. Target sizes for each technique.

Removing this drift is a potential application for refinement techniques (see SmartPupil: Online Calibration Improvement, below). The inaccuracy of eye tracking has been evident in other studies using head-worn displays. Qian et al. [52] found in their experiment with a commercial VR platform that the selection error rate with eye gaze for a target of 8º in diameter was approximately 15%, comparable to the 20.8% miss error rate for our 10º diameter distance threshold (not including time-out errors).

DISCUSSION

The accuracy of current state-of-the-art head pointing is not at the level where it should be to allow precise selections,

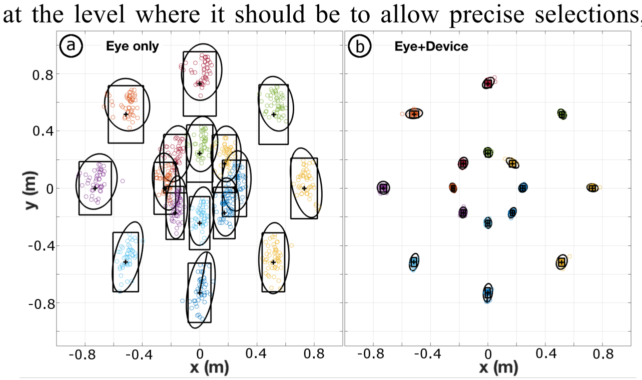

Figure 9. Examples of defined target sizes for a) *Eye only* and b) *Eye+Device* techniques. For both conditions, the diameter for circle targets was determined by the bottom-most target.

far worse than a mouse, for example. As such, there is clear motivation for Pinpointing refinement techniques. While previous studies have focused on refining eye gaze, according to best of our knowledge, no one has investigated refining head movements before. We showed in this study that refinement techniques can improve head pointing accuracy by a factor of three, in particular *Head+Head*, which provided the greatest precision. This technique allows interaction with very small objects such as small menu items (see Gaze Browser: Precise Menu Selection, below) or data points.

Previous studies investigating eye gaze refinement have implemented a single refinement technique, e.g., mouse [66], touch [49], head [56] or gesture [9], and have compared against eye pointing without refinement, but no wide-ranging comparative studies between refinement techniques have been conducted before this study. We studied two previously suggested refinement modalities (head and gesture) that are feasible for wearable AR scenarios and included device as a new addition. The latter technique was the most preferred and provided slightly faster selection compared to gesture and head refinement techniques. A drawback of these methods is the need for an external handheld device. This requirement could be eliminated by triggering refinement with a binary gesture (e.g. a tap gesture) or touch on a wearable item (e.g. a ring or clothing) that can be detected with higher reliability than continuous gestures.

Although we studied the refinement techniques with one particular AR platform, the interaction techniques can be ported to other wearable AR and VR systems. Time, error and preference results may generalize to other systems with similar display and eye tracking hardware, although results will generally improve with developments in display FoV, eye tracking reliability, and gesture sensing technologies.

APPLICATIONS FOR PINPOINTING

In this section, we show example AR applications where Pinpointing can be used to enhance interaction.

Gaze Browser: Precise Menu Selection

We developed a prototype application called GazeBrowser (Figure 10) for interacting with smart object menus. GazeBrowser demonstrates how both implicit and explicit selection techniques, with varying capabilities for targeting accuracy, can be combined in a single interface. Gaze [22,63] and AR [24] have both been previously explored in the context of controlling smart objects. Combining gaze and AR may be a very useful approach, since in future many such objects will become distributed in our physical environments. These may lack displays and input controls but can be controlled through network connections.

We designed a fractal radial menu that can be controlled by Pinpointing. The fractal menu is compact to minimize visual clutter of AR annotations; virtual annotations are initially hidden but users can "browse" objects with their gaze to implicitly reveal simple visual markers. When an object of interest is found, menu interaction begins with an explicit selection.

Once a menu is opened, users can make further selections to open recursive sub-menus. To keep information within an easily browsable space, each level becomes smaller, as shown in Figure 10. By controlling the refinement motion, users can navigate through several levels continuously, with a single selection. The fine obtainable precision of Pinpointing allows deep menu structures to be condensed within a small physical space. Refinement techniques allow such fractal radial menus that would be impractical using eye gaze alone, since menu items would need to be much larger, thus distributed over a wider area.

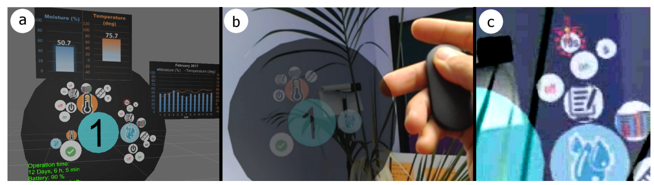

Figure 10. Compact fractal radial menus with deep structure in GazeBrowser. a) Example application controlling humidity and temperature sensors. b) HoloLens screenshot of menu used with device refinement. c) A close-up showing selection of very small radial menu items using Pinpointing. The selected item is the small yellow dot marked by the crosshair.

SmartPupil: Online Calibration Improvement

Our second prototype application, SmartPupil, is an online method for improving the gaze calibration. This method learns from data collected during the refinement stage of the Pinpointing techniques. Drift of gaze calibrations is a well-known issue especially on wearable eye trackers [50,60]. During our study, we observed that gaze calibration was typically less accurate for extreme upper and lower targets, and that overall accuracy tended to degrade over time, due to minute shifting of the HoloLens during head movement.

We implemented a basic method based on the *Eye+Device* technique, that learns from refinement motions to improve the accuracy of future interactions. Recall that at the start of the refinement phase, the cursor appears at the currently detected gaze position. With our technique, a correction vector, $\vec{v}$, is added to this position, and updated with each refinement motion. The correction vector is calculated as

$$\vec{v} = \vec{v}_{prev} + \alpha\vec{\epsilon} \qquad (2)$$

Where $\vec{v}_{prev}$ is the previous correction vector and α is a tuning parameter, set to 0.5 in our implementation. The update vector $\vec{\epsilon}$ is calculated as the distance from the initial cursor position p_{start} to the target centre t if the target is correctly selected, otherwise from p_{start} to the final cursor position p_{end} (Figure 11). If the cursor is not moved in the refinement phase, then no update is added for that selection.

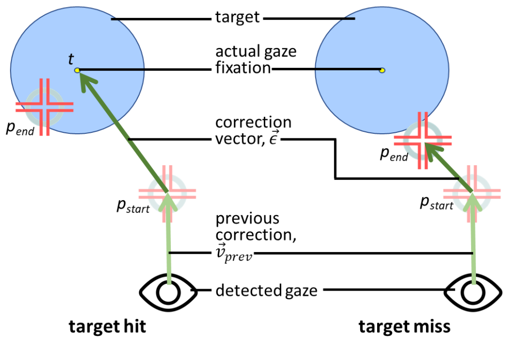

Figure 11. A basic online calibration improvement method, implemented for proof of concept.

We ran a small pilot study with four participants (all male, 1 left handed) to test our improvement method. With a design similar to the above user study, participants completed 6 blocks of trials using two variations of the *Eye+Device* technique, with and without SmartPupil updates (*improvement* and *no improvement*, respectively). Participants were asked to focus their gaze on a small marker at the centre of targets (diameter 2º) before clicking, so that the pre-refinement error could be measured.

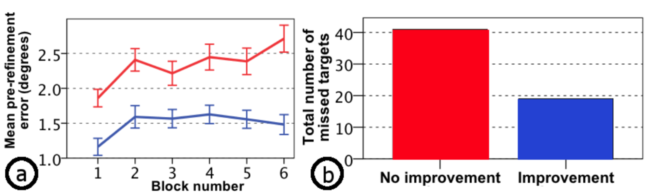

Figure 12. a) Gaze calibration error increases over time. With SmartPupil, degradation is less noticeable, and the mean pre-refinement cursor-placement error is reduced. b) SmartPupil led to a reduction the overall number of missed targets.

From the recorded successful trials (760 of 820 total), we examined this offset over time, and verified that the gaze tracking accuracy appeared to degrade. This degradation is visible in Figure 12a as an increase in the mean gaze-prediction error over the 6 blocks of trials. However, the degradation of the calibration was less noticeable when the improvement method was used. The tracking error was lower overall with the improvement method, with a mean pre-refinement offset of 1.50° (SD = 1.10) from the target centre when using SmartPupil versus 2.34° (SD = 1.38) without. Furthermore, the total number of target misses was lower when using SmartPupil (41/417 trials, 10%, vs. 19/403, 5%), as shown in Figure 12b.

Participants overall preferred SmartPupil over the baseline technqiue, and most noticed they were able to hit the targets more easily and more frequently. While this naïve method appears to be useful for improving gaze calibrations online, it could easily be improved. For instance, the current method stores a single vector, with the implicit assumption that calibration error is equal for all targets. Instead, a more sophisticated implementation could store multiple vectors, weighted by the user's current gaze and head orientation.

FUTURE WORK

In future work, we would like to first further study the proposed Pinpointing techniques on alternate platforms, particularly on AR devices with wider FoV. Such a study may reveal further benefits of gaze versus head pointing. We would further like to explore techniques for other types of interaction beyond target selection. For instance, applications such as CAD may benefit from precise techniques for scaling, rotating and placing objects. Additionally, interaction languages need to be developed around these precision techniques so that users can apply them in various ways or in combination. For instance, a user may want to select a face or vertex rather than a whole object, or apply multiple different operations in sequence. Additional precision-techniques for operations such as group selection and set subtraction will also be useful for applications with a large number of minute objects, such as analysis of 3D scatterplots or point clouds.

When considering the use of Pinpointing in real-world, 3D AR environments, it would be worthwhile to study precision techniques in the depth direction. For example, when interacting with smart objects or 3D visualizations, eye-based techniques can leverage focus and convergence cues, which is not possible with head-based techniques. In addition, the use of eye- and head-based selection techniques in real-world AR applications will raise questions about how the user's mobility (e.g., walking) affects the feasibility of these techniques. For example, walking may cause the visual attention to be directed towards navigating in the environment, making the use of Pinpointing techniques more difficult.

CONCLUSION

In summary, this work has taken a close look at a variety of multimodal techniques for precision target selection in AR. We investigated both eye gaze and head pointing combined with refinement provided by a handheld device, hand gesture input and scaled head motion. A user study showed trade-offs of different variations of Pinpointing. Confirming previous work, eye gaze input alone is faster than head pointing, but the head pointing allows greater targeting accuracy. A previously unexplored use of scaled head refinement proved to be the most accurate, although participants primarily preferred device input and found gestures required the most effort. We further demonstrated two applications for Pinpointing, compact menu selection and online correction of eye gaze calibration. Further work is required to reproduce these techniques on a wider variety of device hardware and to explore more sophisticated applications such as situated analytics and in-situ CAD.

ACKNOWLEDGMENTS

The first author was supported by a Jorma Ollila grant from the Nokia Foundation and a grant from the Academy of Finland (grant number 311090). We gratefully thank our study participants, reviewers for their valuable feedback, and Futurice Ltd. for loaning the research equipment.

REFERENCES

1. Rowel Atienza, Ryan Blonna, Maria Isabel Saludares, Joel Casimiro, and Vivencio Fuentes. 2016. Interaction techniques using head gaze for virtual reality. In *Proceedings - 2016 IEEE Region 10 Symposium, TENSYMP 2016*, 110–114. https://doi.org/10.1109/TENCONSpring.2016.7519387
2. Mihai Bace, Teemu Leppänen, David Gil De Gomez, and Argenis Ramirez Gomez. 2016. ubiGaze : Ubiquitous Augmented Reality Messaging Using Gaze Gestures. In *SIGGRAPH ASIA*, Article no. 11.
3. Richard Bates and Howell Istance. 2003. Why are eye mice unpopular? A detailed comparison of head and eye controlled assistive technology pointing devices. *Universal Access in the Information Society* 2, 3: 280–290. https://doi.org/10.1007/s10209-003-0053-y
4. Ana M. Bernardos, David Gómez, and José R. Casar. 2016. A Comparison of Head Pose and Deictic Pointing Interaction Methods for Smart Environments. *International Journal of Human-Computer Interaction* 32, 4: 325–351. https://doi.org/10.1080/10447318.2016.1142054
5. Martin Bichsel and Alex Pentland. 1993. *Automatic interpretation of human head movements*.
6. Doug Bowman, Ernst Kruijff, Joseph J. LaViola, and Ivan Poupyrev. 2004. *3D User Interfaces: Theory and Practice*. Addison Wesley Longman Publishing Co., Redwood City.
7. Stephen Brewster, Joanna Lumsden, Marek Bell, Malcolm Hall, and Stuart Tasker. 2003. Multimodal "eyes-free" interaction techniques for wearable devices. *Proceedings of the conference on Human factors in computing systems - CHI '03*, 5: 473. https://doi.org/10.1145/642611.642694
8. Benedetta. Cesqui, Rolf van de Langenberg, Francesco. Lacquaniti, and Andrea. D'Avella. 2013. A novel method for measuring gaze orientation in space in unrestrained head conditions. *Journal of Vision* 13, 8: 28:1-22. https://doi.org/10.1167/13.8.28
9. Ishan Chatterjee, Robert Xiao, and Chris Harrison. 2015. Gaze + Gesture : Expressive, Precise and Targeted Free-Space Interactions. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 131–138. https://doi.org/10.1145/2818346.2820752
10. Ngip Khean Chuan and Ashok Sivaji. 2012. Combining eye gaze and hand tracking for pointer control in HCI: Developing a more robust and accurate interaction system for pointer positioning and clicking. In *CHUSER 2012 - 2012 IEEE Colloquium on Humanities, Science and Engineering Research*, 172–176. https://doi.org/10.1109/CHUSER.2012.6504305
11. Rory M.S. Clifford, Nikita Mae B. Tuanquin, and Robert W. Lindeman. 2017. Jedi ForceExtension: Telekinesis as a Virtual Reality interaction metaphor. In *2017 IEEE Symposium on 3D User Interfaces, 3DUI 2017 - Proceedings*, 239–240. https://doi.org/10.1109/3DUI.2017.7893360
12. Nathan Cournia, John D. Smith, and Andrew T. Duchowski. 2003. Gaze- vs. Hand-based Pointing in Virtual Environments. In *Proc. CHI '03 Extended Abstracts on Human Factors in Computer Systems (CHI '03)*, 772–773. https://doi.org/10.1145/765978.765982
13. Heiko Drewes and Albrecht Schmidt. 2007. Interacting with the computer using gaze gestures. In *Proc. of the International Conference on Human-computer interaction'07*, 475–488. https://doi.org/10.1007/978-3-540-74800-7_43
14. Neven A M ElSayed, Bruce H. Thomas, Kim Marriott, Julia Piantadosi, and Ross T. Smith. 2016. Situated Analytics: Demonstrating immersive analytical tools with Augmented Reality. *Journal of Visual Languages and Computing* 36, C: 13–23. https://doi.org/10.1016/j.jvlc.2016.07.006
15. Barrett Ens, Fraser Anderson, Tovi Grossman, Michelle Annett, Pourang Irani, and George Fitzmaurice. 2017. Ivy: Exploring Spatially Situated Visual Programming for Authoring and Understanding Intelligent Environments. *Proceedings of the 43rd Graphics Interface Conference*: 156–162. https://doi.org/10.20380/gi2017.20
16. Barrett M. Ens, Rory Finnegan, and Pourang P. Irani. 2014. The personal cockpit: a spatial interface for effective task switching on head-worn displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3171–3180. https://doi.org/10.1145/2556288.2557058
17. Augusto Esteves, David Verweij, Liza Suraiya, Rasal Islam, Youryang Lee, and Ian Oakley. 2017. SmoothMoves : Smooth Pursuits Head Movements for Augmented Reality. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software & Technology*, 167–178. https://doi.org/10.1145/3126594.3126616
18. Anna Maria Feit, Shane Williams, Arturo Toledo, Ann Paradiso, Harish Kulkarni, Shaun Kane, and Meredith Ringel Morris. 2017. Toward Everyday Gaze Input: Accuracy and Precision of Eye Tracking and Implications for Design. *CHI '17 Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*: 1118–1130. https://doi.org/10.1145/3025453.3025599
19. FOVE. FOVE Eye Tracking Virtual Reality Headset. Retrieved September 19, 2017 from

https://www.getfove.com/

20. Sven-Thomas Graupner and Sebastian Pannasch. 2014. Continuous Gaze Cursor Feedback in Various Tasks: Influence on Eye Movement Behavior, Task Performance and Subjective Distraction. . Springer, Cham, 323–329. https://doi.org/10.1007/978-3-319-07857-1_57
21. Gyration. Gyration Air Mouse Input Devices. Retrieved September 18, 2017 from https://www.gyration.com/
22. Jeremy Hales, David Rozado, and Diako Mardanbegi. 2013. Interacting with Objects in the Environment by Gaze and Hand Gestures. In *Proceedings of the 3rd International Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction*, 1–9.
23. Sandra G. Hart and Lowell. E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology* 1: 139–183.
24. Valentin Heun, James Hobin, and Pattie Maes. 2013. Reality editor: programming smarter objects. *Proceedings of the 2013 ACM conference on Pervasive and Ubiquitous Computing adjunct publication (UbiComp '13 Adjunct)*: 307–310. https://doi.org/10.1145/2494091.2494185
25. Aulikki Hyrskykari, Howell Istance, and Stephen Vickers. 2012. Gaze gestures or dwell-based interaction? In *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '12*, 229–232. https://doi.org/10.1145/2168556.2168602
26. Howell Istance, Aulikki Hyrskykari, Lauri Immonen, Santtu Mansikkamaa, and Stephen Vickers. 2010. Designing gaze gestures for gaming: an Investigation of Performance. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA '10* 1, 212: 323. https://doi.org/10.1145/1743666.1743740
27. Rob Jacob and Sophie Stellmach. 2016. Interaction technologies: What you look at is what you get: Gaze-based user interfaces. *Interactions* 23, 5: 62–65. https://doi.org/10.1145/2978577
28. Robert J. K. Jacob. 1990. What you look at is what you get: eye movement-based interaction techniques. *Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI '90*: 11–18. https://doi.org/10.1145/97243.97246
29. Richard J. Jagacinski and Donald L. Monk. 1985. Fitts' Law in two dimensions with hand and head movements. *Journal of motor behavior* 17, 1: 77–95. https://doi.org/10.1080/00222895.1985.10735338
30. Shahram Jalaliniya, Diako Mardanbegi, and Thomas Pederson. 2015. MAGIC Pointing for Eyewear Computers. *Proceedings of the 2015 ACM International Symposium on Wearable Computers - ISWC '15*: 155–158. https://doi.org/10.1145/2802083.2802094
31. Shahram Jalaliniya, Diako Mardanbeigi, Thomas Pederson, and Dan Witzner Hansen. 2014. Head and eye movement as pointing modalities for eyewear computers. *Proceedings - 11th International Conference on Wearable and Implantable Body Sensor Networks Workshops, BSN Workshops 2014*: 50–53. https://doi.org/10.1109/BSN.Workshops.2014.14
32. Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication - UbiComp '14 Adjunct*, 1151–1160. https://doi.org/10.1145/2638728.2641695
33. Mohamed Khamis, Axel Hoesl, Alexander Klimczak, Martin Reiss, Florian Alt, and Andreas Bulling. 2017. EyeScout : Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays. In *UIST '17 Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, 155–166. https://doi.org/10.1145/3126594.3126630
34. Linnéa Larsson, Andrea Schwaller, Marcus Nyström, and Martin Stridh. 2016. Head movement compensation and multi-modal event detection in eye-tracking data for unconstrained head movements. *Journal of Neuroscience Methods* 274: 13–26.
35. Chiuhsiang Joe Lin, Sui Hua Ho, and Yan Jyun Chen. 2015. An investigation of pointing postures in a 3D stereoscopic environment. *Applied Ergonomics* 48: 154–163. https://doi.org/10.1016/j.apergo.2014.12.001
36. Alfredo Liverani, Giancarlo Amati, and Gianni Caligiana. 2016. A CAD-augmented Reality Integrated Environment for Assembly Sequence Check and Interactive Validation. *Concurrent Engineering* 12, 1: 67–77. https://doi.org/10.1177/1063293X04042469
37. Rainer Malkewitz. 1998. Head pointing and speech control as a hands-free interface to desktop computing. In *Proceedings of the third international ACM conference on Assistive technologies (Proceeding Assets '98)*, 182–188.
38. Diako Mardanbegi, Dan Witzner Hansen, and Thomas Pederson. 2012. Eye-based head gestures movements. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, 139–146.
39. Microsoft. The leader in Mixed Reality Technology | HoloLens. Retrieved September 14, 2017 from https://www.microsoft.com/en-us/hololens

40. Mark R Mine. 1995. *Virtual environment interaction techniques*. https://doi.org/10.1.1.38.1750
41. Richard A. Monty and John W. Senders. 1976. *Eye Movements and Psychological Processes*. Lawrence Erlbaum Associates, Hillsdale, NJ.
42. Mathieu Nancel, Olivier Chapuis, Emmanuel Pietriga, Xing-Dong Yang, Pourang P. Irani, and Michel Beaudouin-Lafon. 2013. High-precision pointing on large wall displays using small handheld devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 831. https://doi.org/10.1145/2470654.2470773
43. Tomi Nukarinen, Jari Kangas, Oleg Špakov, Poika Isokoski, Deepak Akkil, Jussi Rantala, and Roope Raisamo. 2016. Evaluation of HeadTurn - An Interaction Technique Using the Gaze and Head Turns. *Proceedings of the 9th Nordic Conference on Human-Computer Interaction - NordiCHI '16*: 1–8. https://doi.org/10.1145/2971485.2971490
44. Oculus. Oculus Rift. Retrieved September 19, 2017 from https://www.oculus.com/
45. Hyung Min Park, Seok Han Lee, and Jong Soo Choi. 2008. Wearable augmented reality system using gaze interaction. In *Proceedings - 7th IEEE International Symposium on Mixed and Augmented Reality 2008, ISMAR 2008*, 175–176. https://doi.org/10.1109/ISMAR.2008.4637353
46. Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. 2014. Gaze-touch: Combining Gaze with Multi-touch for Interaction on the Same Surface. In *Proceedings of the 27th annual ACM symposium on User interface software and technology - UIST '14*, 509–518. https://doi.org/10.1145/2642918.2647397
47. Ken Pfeuffer, Jason Alexander, Ming Ki Chong, Yanxia Zhang, and Hans Gellersen. 2015. Gaze-Shifting: Direct-Indirect Input with Pen and Touch Modulated by Gaze. *Proceedings of the 28th annual ACM symposium on User interface software and technology - UIST '15*: 373–383. https://doi.org/10.1145/2807442.2807460
48. Ken Pfeuffer, Jason Alexander, and Hans Gellersen. 2015. Gaze+ touch vs. touch: what's the trade-off when using gaze to extend touch to remote displays? In *Human-Computer Interaction – INTERACT 2015*, 349–367.
49. Ken Pfeuffer and Hans Gellersen. 2016. Gaze and Touch Interaction on Tablets. *Proceedings of the 29th Annual Symposium on User Interface Software and Technology - UIST '16*: 301–311. https://doi.org/10.1145/2984511.2984514
50. Thammathip Piumsomboon, Gun Lee, Robert W. Lindeman, and Mark Billinghurst. 2017. Exploring natural eye-gaze-based interaction for immersive virtual reality. *2017 IEEE Symposium on 3D User Interfaces, 3DUI 2017 - Proceedings*: 36–39. https://doi.org/10.1109/3DUI.2017.7893315
51. Pupil Labs. Pupil Labs. Retrieved September 19, 2017 from https://pupil-labs.com/
52. Yuanyuan 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 ACM Symposium on Spatial User Interaction*, 91–98.
53. Marcos Serrano, Barrett Ens, Xing-Dong Yang, and Pourang Irani. 2015. Gluey: Developing a Head-Worn Display Interface to Unify the Interaction Experience in Distributed Display Environments. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '15*, 161–171. https://doi.org/10.1145/2785830.2785838
54. Linda E. Sibert and Robert J. K. Jacob. 2000. Evaluation of eye gaze interaction. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '00*: 281–288. https://doi.org/10.1145/332040.332445
55. Nikolaos Sidorakis, George Alex Koulieris, and Katerina Mania. 2015. Binocular eye-tracking for the control of a 3D immersive multimedia user interface. In *2015 IEEE 1st Workshop on Everyday Virtual Reality, WEVR 2015*, 15–18. https://doi.org/10.1109/WEVR.2015.7151689
56. Oleg Špakov, Poika Isokoski, and Päivi Majaranta. 2014. Look and Lean : Accurate Head-Assisted Eye Pointing. *Proceedings of the ETRA conference* 1, 212: 35–42. https://doi.org/10.1145/2578153.2578157
57. Oleg Špakov and Päivi Majaranta. 2012. Enhanced Gaze Interaction Using Simple Head Gestures. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*, 705–710. https://doi.org/10.1145/2370216.2370369
58. Sophie Stellmach and Raimund Dachselt. 2012. Look & Touch : Gaze-supported Target Acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2981–2990. https://doi.org/10.1145/2207676.2208709
59. Sophie Stellmach and Raimund Dachselt. 2013. Still Looking: Investigating Seamless Gaze-supported Selection, Positioning, and Manipulation of Distant Targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 285–294. https://doi.org/10.1145/2470654.2470695
60. Yusuke Sugano and Andreas Bulling. 2015. Self-Calibrating Head-Mounted Eye Trackers Using

Egocentric Visual Saliency. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15).*, 363–372. https://doi.org/10.1145/2807442.2807445

61. Vildan Tanriverdi and Robert J. K. Jacob. 2000. Interacting with eye movements in virtual environments. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '00* 2, 1: 265–272. https://doi.org/10.1145/332040.332443

62. Boris Velichkovsky, Andreas Sprenger, and Pieter Unema. 1997. Towards Gaze-Mediated Interaction - Collecting Solutions of the "Midas touch problem." *Proceedings of the International Conference on Human-Computer Interaction (INTERACT'97)*: 509–516. https://doi.org/10.1007/978-0-387-35175-9_77

63. Eduardo Velloso, Markus Wirth, Christian Weichel, Augusto Esteves, and Hans Gellersen. 2016. AmbiGaze: Direct Control of Ambient Devices by Gaze. *Proceedings of the 2016 ACM Conference on Designing Interactive Systems - DIS '16*: 812–817. https://doi.org/10.1145/2901790.2901867

64. Melodie Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: Spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 439–448. https://doi.org/10.1145/2493432.2493477

65. X. Glass. Retrieved September 19, 2017 from https://x.company/glass/

66. Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and gaze input cascaded (MAGIC) pointing. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '99*, 246–253. https://doi.org/10.1145/302979.303053