BimodalGaze: Seamlessly Refined Pointing with Gaze and Filtered Gestural Head Movement

Ludwig Sidenmark
Lancaster University
Lancaster, United Kingdom
l.sidenmark@lancaster.ac.uk

Diako Mardanbegi
Adhawk Microsystems
Kitchener, Ontario, Canada
diako@adhawkmicrosystems.com

Argenis Ramirez Gomez
Lancaster University
Lancaster, United Kingdom
a.ramirezgomez@lancaster.ac.uk

Christopher Clarke
Lancaster University
Lancaster, United Kingdom
c.clarke1@lancaster.ac.uk

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

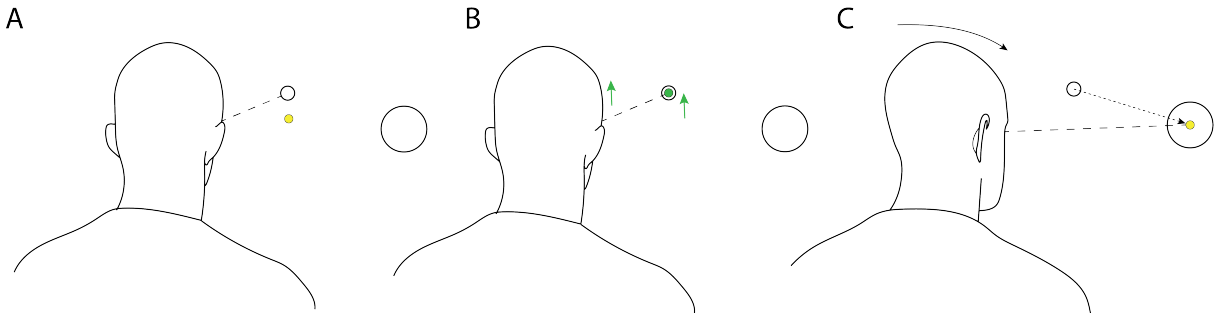

Figure 1: BimodalGaze enables users to point by gaze and to seamlessly refine the cursor position with head movement. A: In *Gaze Mode*, the cursor (yellow) follows where the user looks but may not be sufficiently accurate. B: The pointer automatically switches into *Head Mode* (green) when *gestural head movement* is detected. C: The pointer automatically switches back into Gaze Mode when the user redirects their attention. Note that the Head Mode is only invoked when needed for adjustment of the cursor. Any *natural head movement* associated with a gaze shift is filtered and does not cause a mode switch.

ABSTRACT

Eye gaze is a fast and ergonomic modality for pointing but limited in precision and accuracy. In this work, we introduce BimodalGaze, a novel technique for seamless head-based refinement of a gaze cursor. The technique leverages eye-head coordination insights to separate natural from gestural head movement. This allows users to quickly shift their gaze to targets over larger fields of view with naturally combined eye-head movement, and to refine the cursor position with gestural head movement. In contrast to an existing baseline, head refinement is invoked automatically, and only if a target is not already acquired by the initial gaze shift. Study results show that users reliably achieve fine-grained target selection, but we observed a higher rate of initial selection errors affecting overall performance. An in-depth analysis of user performance provides insight into the classification of natural versus gestural head movement, for improvement of BimodalGaze and other potential applications.

ETRA '20, ,

ACM ISBN 978-1-4503-1234-5/17/07.
https://doi.org/10.1145/8888888.7777777

CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques**; **Virtual reality**; *Mixed / augmented reality*;

KEYWORDS

Eye tracking, Gaze interaction, Refinement, Eye-head coordination, Virtual reality

ACM Reference format:
Ludwig Sidenmark, Diako Mardanbegi, Argenis Ramirez Gomez, Christopher Clarke, and Hans Gellersen. 2020. BimodalGaze: Seamlessly Refined Pointing with Gaze and Filtered Gestural Head Movement. In *Proceedings of 2020 Symposium on Eye Tracking Research and Applications, , (ETRA '20),* 9 pages.
https://doi.org/10.1145/8888888.7777777

1 INTRODUCTION

Gaze is natural and fast for pointing at objects across our visual field, but poor for fine-grained cursor control and selection of detail. Gaze shifts are highly efficient as they use saccadic eye movement to quickly align objects of interest over the fovea, in natural coordination with head movement [Land 2004]. However, alignment over a target is not precise and induces uncertainty into gaze estimation, further exacerbated by calibration and measurement limitations of

eye tracking. As such, a solely gaze-based pointer is problematic to use with a conventional cursor metaphor for selection. A cursor will indicate the estimated gaze position but if it is off-target due to inaccuracy, there is no direct way for the user to nudge the cursor to the actual gaze position for correct selection [Porta et al. 2010].

A variety of work has addressed gaze inaccuracy by resorting to another modality for refinement of gaze input. Head pointing, in particular, is interesting for complementing gaze, as head movement affords stable and precise input while retaining the advantage of hands-free pointing [Bates and Istance 2003]. Head correction of gaze has been demonstrated on displays with narrow-field-of-view (FOV), where gaze shifts were assumed to be performed by the eyes alone, thus allowing head movement to be treated as independent input for relative cursor displacement [Jalaliniya et al. 2015; Kurauchi et al. 2015; Kytö et al. 2018; Špakov et al. 2014]. However, eye movement research has shown that only small gaze shifts are performed solely with eye movement, whereas more significant shifts naturally feature head movement to reach targets and maintain a comfortable eye-in-head position [Freedman and Sparks 2000; Land 2004]. It is therefore not straightforward to use head movement for gaze correction, in particular when gaze is considered for pointing across a larger FOV, such as on large displays, across devices, or in virtual and augmented reality (VR/AR).

In this work, we introduce *BimodalGaze* as a novel technique for gaze pointing with seamless head refinement. At the heart of the technique is a distinction between head movement that is *natural* in terms of eye-head coordination of gaze shifts, and head movement that we adopt as *gestural* as it is independent from gaze. As illustrated in Figure 1, users point primarily with gaze (A) but can seamlessly transition to refine the cursor position (B). The *Head Mode* for refinement is automatically invoked when gestural head movement is detected, while users are free to use natural head movement in *Gaze Mode* (C).

BimodalGaze is a generic gaze pointing technique. However, a key motivation was to support precise pointing over larger fields of view, and we therefore implemented and studied the technique in VR. We compared our technique with *Eye+Head Pinpointing* as a recent baseline for head-assisted gaze pointing, with the primary difference that the pinpointing technique relies on manual toggling between gaze pointing and head refinement modes. BimodalGaze overcomes the need for a manual switch but depends on effective detection of gestural head movement. While users were able to reliably achieve fine-grained selection we observed a comparatively higher rate of initial input error that affected selection time. We provide a detailed analysis of participant behaviour with our technique that gives insight into different causes for error and how they relate to design choices.

The contributions over our work thus comprise the BimodalGaze technique for precise hands-free pointing as well as insights into natural versus gestural head movement of wider relevance to interaction design with eye and head movement.

2 RELATED WORK

Gaze is fast in comparison with other pointing modalities but inherently limited in precision and accuracy. This has spurred design of techniques where gaze is combined with a complementary modality. *MAGIC* pointing is an early example, where gaze moves a cursor close to a target upon which the selection is completed with mouse input [Zhai et al. 1999], and *Gaze-shifting* demonstrated the same principle for direct touch and pen input [Pfeuffer et al. 2015]. However, the underlying conceptual model is to use gaze to support the manual input, and the design of the techniques enforces that gaze stops short of directly selecting a target. In contrast, techniques such as *Look&Touch* [Stellmach and Dachselt 2012] and *Cursor-shift* [Pfeuffer and Gellersen 2016] are based on a gaze-centric model, where gaze makes the initial selection which is then refined with touch input. Our design of BimodalGaze follows a gaze-centric model with gaze as primary pointing mode, and head pointing as complementary modality that is only used when the gaze input requires refinement.

A range of works have compared and integrated eye and head movement. In comparison, eye movement is faster and requires less energy, while head motion is more stable and affords better control [Bates and Istance 2003; Blattgerste et al. 2018; Kytö et al. 2018; Qian and Teather 2017]. *Look&Lean* was first to demonstrate combined use for gaze-centric precise pointing, but limited to use of lateral head motion observed by an eye tracker as corrective input [Špakov et al. 2014]. Other work, such as HMAGIC, has been based on models of gaze-assisted head-pointing [Jalaliniya et al. 2015; Kurauchi et al. 2015]. *Pinpointing* compared head versus eyes as primary pointing modes, and a variety of techniques for subsequent selection refinement [Kytö et al. 2018]. An assumption underlying these works is that eye and head are independent as inputs. This is problematic as gaze involves natural eye-head coordination, where the movement of the eyes and head are coupled in performing gaze shifts and stabilising gaze on targets [Bizzi 1974; Guitton and Volle 1987]. BimodalGaze, in contrast, allows for natural head support in the gaze mode while separate gestural head movement is detected for refinement.

The recent work on Pinpointing is particularly relevant to ours, as it explored combined eye and head pointing in a head-mounted display [Kytö et al. 2018]. Their work showed that head correction of gaze can be preferable even if manual input is available, as it is as effective and less effort compared to manual raycasting. Among the specific techniques proposed, Eye+Head Pinpointing is similar to BimodalGaze in providing distinct modes for gaze versus head control of the cursor, and was therefore adopted as baseline for comparison. However, the technique differs from ours in requiring manual toggling between the pointing modes, whereas BimodalGaze is designed to make the switch implicit and seamless.

There is numerous other work addressing limited gaze accuracy, with zooming [Lankford 2000], incremental disambiguation [Lutteroth et al. 2015] or specialist cursors that can be nudged via gaze buttons [Porta et al. 2010]. These techniques can be implemented with gaze alone but are slow as they require additional interaction steps. BimodalGaze has in common with these techniques that it is hands-free but it complements gaze with small head movement for more efficient cursor refinement.

For the design of BimodalGaze, we are leveraging insight from the eye-head coordination literature [Bizzi 1974; Freedman 2008; Guitton and Volle 1987; Land 2004]. Small gaze shifts may be achieved by eye movement alone but generally the head contributes to gaze. The eyes have a physical range of about 50 degrees to left

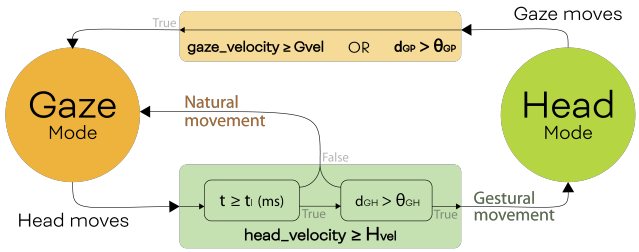

Figure 2: BimodalGaze state diagram.

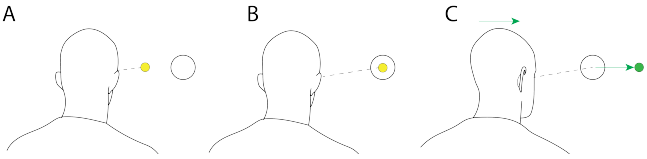

Figure 3: Typical situation without head movement filtering. A: The user use a gaze cursor (yellow) for pointing. B: The user moves their gaze and pointer onto the target. C: The user's gaze stays on target while performing a head movement as a natural part of the gaze shift. The pointer thus switches to head pointing (green) which drags the cursor away from the target during the natural head movement.

and right but rarely rotate beyond 30 degrees [Stahl 1999]. Head movement supports gaze to reach further and to maintain a comfortable eye-in-head position [Land and Tatler 2012]. The temporal relationship between eye and supporting head movement is complex. The head is slower to start and follow the eyes toward target. During a gaze shift, head movement augments the saccadic movement of the eye, such that the movements are additive toward reaching the target [Guitton and Volle 1987]. When a gaze target has been reached by the eyes, the head will typically continue to move while the eyes fixate the target by performing compensatory eye movement in the opposite direction, mediated by the vestibulo-ocular reflex (VOR) [Tweed et al. 1995]. It is therefore not straightforward to switch from gaze to refinement mode once the eyes have reached the target.

There are also a number of studies of eye-head movement in virtual reality that inform this work. In contrast to natural viewing, head-mounted displays limit the user's view. This has been observed to lead to less eye rotation and a comparatively larger contribution of head movement to gaze shifts [Kollenberg et al. 2010; Pfeil et al. 2018]. For this work, we specifically built on quantitative insight into eye, head and torso coordination from a recent study of gaze shifts in VR [Sidenmark and Gellersen 2019a]. That work informs the criteria we use to filter gestural head movement from head movement that is naturally coupled with gaze.

Even though the head naturally contributes to gaze, there has been only little work in the field that builds on insight into eye-head coordination. One application area is in gaze models for virtual characters, with the aim to generate realistic gaze [Itti et al. 2006] or aiding animators in creating specific communicative effects (e.g., glances out of the corner of eye) [Pejsa et al. 2016]. Here, gaze is typically rendered as eyes-only when gaze shifts are below a threshold of 10-15°, and coupled with head movement otherwise [Ruhland et al. 2015]. Recent work also introduced a distinction of eyes-only versus head-supported gaze for point and dwell input [Sidenmark and Gellersen 2019b]. Other work has build on coordinated eye and head movement for estimation of gaze depth and target disambiguation in 3D interfaces [Mardanbegi et al. 2019a,b].

3 BIMODALGAZE

BimodalGaze uses gaze as the primary modality for quick, effortless and hands-free pointer control. When gaze is not accurate or stable enough for selection, BimodalGaze allows automatic switching to head pointing when users perform deliberate head movements for further pointer refinement. See *Figure 2* for a system overview.

BimodalGaze leverages eye-head coordination insights to determine if a head movement is *gestural* for refinement; or a *natural* movement during a gaze shift. When a head movement is classified as natural, it is ignored and the user's gaze controls the pointer. When the movement is classified as gestural, BimodalGaze switches to head pointing for further pointer refinement. We define two criteria to differentiate natural from gestural head movements:

(1) A natural head movement starts t_l ms after an eye movement.
(2) A natural head movement will move in a similar direction to the prior eye movement (θ_{GH}).

The first criterion relates to the timing of head movements. Natural head movements are used to further the range of the eyes or to move the eyes into a comfortable position [Tweed et al. 1995]. Research has shown natural head movement occurs at the same time as, or a shortly after, the initial eye movement to maintain a comfortable eye-in-head position [Sidenmark and Gellersen 2019a]. If the head does not move within a certain time (t_l) after an eye movement we can assume that the head movement is gestural. Secondly, as the purpose of natural head movements is to increase the eyes' reach, or to move the eyes closer to their central position, it is reasonable to believe that a natural head movement will move in a similar direction as the eyes. As such, if the angular difference between the trajectory of the eyes and head (d_{GH}) are within a certain range (θ_{GH}), we assume the head movement to be natural.

Switching to Head Mode is suppressed during a deliberate gaze movement, which we define as any gaze movement over a velocity of G_{vel}. Alternatively, the system will switch back to Gaze Mode if the distance between the gaze and pointer position, d_{GP}, is greater than θ_{GP} (measured in visual angle). The latter condition is to prevent the cursor becoming too detached from the gaze due to misclassification of head movement (e.g. a natural movement detected as gestural fig. 3), or due to eye movements which are not detected as a deliberate gaze movement.

3.1 Implementation

We implemented BimodalGaze in VR. The selection of thresholds has a large impact on BimodalGaze's behaviour and requires careful consideration. For saccade detection, G_{vel} was set to $160°/sec$ which is a relatively high value to avoid unintentional switches

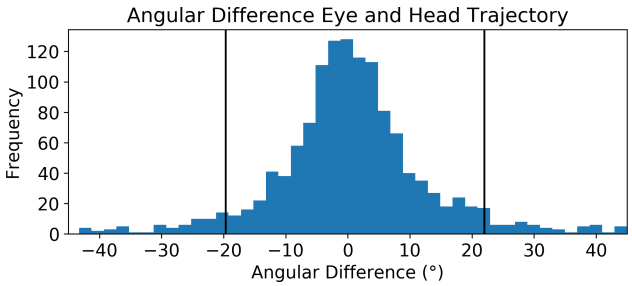

Figure 4: Angular difference between eye and head trajectory during a gaze shift. The zone within the lines represent 90% of all gaze shifts with accompanied head movement.

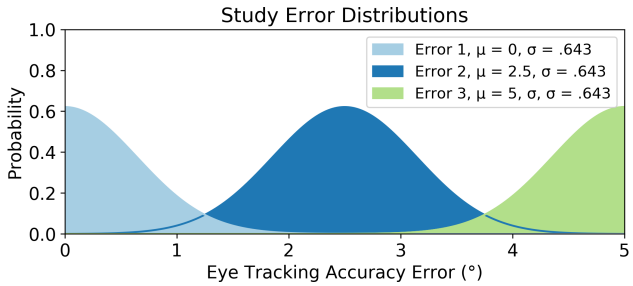

Figure 5: The different accuracy error distributions used for each accuracy condition. Note that the minimum value of Acc. 1 and maximum value of Acc. 3 was restricted to 0° and 5° respectively.

to Gaze Mode during refinement caused by corrective saccades, vestibulo-ocular reflex, or smooth pursuit eye movements. The sensitivity of head movement detection, H_{vel}, was set to $15°/sec$, to be low enough to detect smaller head movements but high enough to ignore minor unintentional head shifts caused by the user. We based t_l and θ_{GH} on prior work on eye-head coordination in VR. Sidenmark and Gellersen found that head movements generally start moving 150ms after the eye movement [Sidenmark and Gellersen 2019a], therefore we set t_l to 150 ms. We used the data from Sidenmark and Gellersen's work to calculate the angular difference between the eye and head trajectory (fig. 4). We found that for 90% of all gaze shifts with accompanying head movement, the eye and head trajectory were within 20° of each other. As such, we set θ_{GH} to 20°. Finally, we set θ_{GP} to 10°, so that users can freely adjust the pointer position, while not being able to move the pointer too far out in the periphery.

4 EVALUATION

We conducted a user study in VR to evaluate BimodalGaze and gather insights regarding its performance and user feedback. The flexibility of BimodalGaze raises the question of how often users transition to Head Mode, and how often selections can be made with gaze only under different conditions. We also want to assess how effective the automatic switching of BimodalGaze is, by comparing it with the Eye+Head Pinpointing technique, where the switch from gaze pointing to refinement via head movement is done manually by the user via a button click [Kytö et al. 2018].

4.1 Task

Participants were required to select spherical targets with a diameter of 3° at 2m distance. From a central starting position, targets were found in one of eight directions (cardinal and intercardinal) at three target distances (10, 25, 40°). Smaller target distances are within reach of gaze and do not necessitate head movement, whereas we would expect users to move their heads naturally towards the larger targets. The accuracy of the eye tracker could dictate how much the techniques rely on Head Mode. We investigated this effect by artificially inducing an offset into the eye tracker's gaze estimation to simulate different levels of eye tracker accuracy. For each trial we select a random gaze accuracy from one of three normal distributions. The accuracy distributions were varied to cover a range that included both minor and significant accuracy errors, see Figure 5. The direction of the induced eye tracking accuracy error was randomly selected for each trial.

To begin a trial, the participant had to align a visible cross-shaped head pointer and their gaze with a central starting target (2° diameter). After 500ms, a spherical target appeared at one of the 24 predefined positions, chosen in random order. Participants were instructed to select the target as precisely and quickly as possible.

The user study employed a within-subjects design, with independent variables and levels as follows:

- *Technique:* BimodalGaze, Eye+Head Pinpointing
- *Gaze estimation accuracy:* Acc. 1, Acc. 2, Acc. 3
- *Target direction:* Up, Down, Right, Left, Up-right, Up-left, Down-right, Down-left
- *Target distance:* 10, 25, 40°

For each technique, each participant completed 3 blocks (one for each gaze estimation accuracy) of 72 trials (8 directions x 3 distances x 3 repetitions). Half of participants performed Pinpointing first followed by BimodalGaze, and the other half performed the reverse. As such, the total number of trials per participant was 2 Techniques x 3 Blocks x 72 Trials = 432.

4.2 Apparatus

The techniques and task were developed in Unity version 2017.4.3. An HTC Vive with an integrated Tobii Pro Eye Tracker and data output frequency of 120Hz was used to record eye and head movement. We used the directional vectors of the eyes and head to calculate movements. We were able to record data at a mean gaze accuracy of $0.981 \pm 0.232°$ and a mean gaze precision of $0.427 \pm 0.152°$. The standard hand-held controller of the HTC Vive was used for manual input. In BimodalGaze, a cursor was always visible to show the current pointing position. The cursor colour was used to show if the pointer was currently in Gaze Mode (yellow) or Head Mode (green). A selection was made by pressing the hand-controller trackpad. As described by Kytö et al., the Head Mode of the Pinpointing technique was triggered by pressing and holding down the hand-controller trackpad and a selection was made by releasing the trackpad. The cursor was only visible during the refinement stage for Pinpointing [Kytö et al. 2018].

4.3 Procedure

We recruited 12 participants (8 male, 4 female, age: 25.42 ± 2.91) for our user study. Ten participants reported none or occasional VR experience, while two participants reported weekly VR experience. Eleven participants reported none or occasional eye tracking experience, while one reported daily eye tracking experience. Participants first signed a consent form and answered a demographic questionnaire. The participants were then seated and put on the HMD and handed the controller. The user study consisted of six test sessions, where the participants performed the task with one technique and one accuracy condition per session. The participants performed a five-point eye tracking calibration before each test session. The order of technique and accuracy conditions was counterbalanced with a Latin square. After each test session, participants removed the HMD and filled out a post-task questionnaire consisting of seven 5-point Likert items based on common usability factors (Precision, Ease, Learnability, Concentration, Physical effort, Frustration, Accurate switching), and were offered the opportunity to rest. A semi-structured interview was conducted after each completed task. The study took 45 minutes to complete.

4.4 Analysis

For each trial we measured the completion time, incorrect selections, time spent in Head Mode and total head movement. We conducted a four-way repeated-measures ANOVA ($\alpha = .05$) for performance metrics with interaction technique, gaze estimation accuracy, direction 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. Usability Likert-scale data was analysed with Friedman tests with Bonferroni-corrected Wilcoxon tests for post-hoc analysis.

5 RESULTS

In this section we analyse performance metrics for BimodalGaze and Pinpointing. For BimodalGaze we investigate how often refinement is required for selection, and reflect on this in our analysis of error rates and selection time.

5.1 Refinement

BimodalGaze provides the flexibility to use gaze pointing when it is sufficient, or to enter Head Mode when necessary. Out of all 2594 selections made with BimodalGaze, 454 (17.5%) were made in Gaze Mode and 2138 (82.5%) were made in Head Mode. Further insights (Table 1) showed that both eye tracker accuracy and target distance affects the prevalence of Gaze Mode selections for BimodalGaze. Lower eye tracking accuracy and further target distances led to more selections made in Head Mode. Direction had no effect on the prevalence of selections made in Gaze or Head Mode.

Table 1: Prevalence of BimodalGaze selections made in Gaze Mode based on accuracy error and target distance.

	10°	25°	40°	Total
Acc. 1	53.1%	33.7%	17.5%	34.7%
Acc. 2	19.1%	14.9%	11.8%	15.3%
Acc. 3	1.7%	2.4%	3.5%	2.5%
Total	24.7%	17.0%	10.9%	17.5%

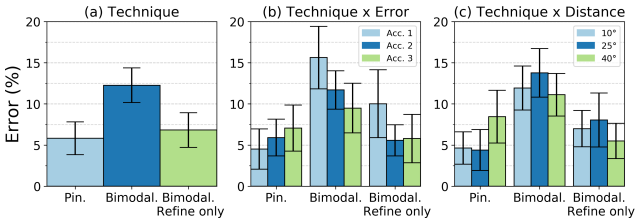

Figure 6: Mean error rate. Error bars represents mean 95% confidence interval.

5.2 Error Rate

All participants completed all trials with both Pinpointing and BimodalGaze. As such, we define an error as when the participant missed the target prior to a correct selection. We report the number of errors as the error rate, i.e. the number of trials resulting in an error divided by the total number of trials (fig. 6).

We found no significant four-way or three way-interactions. However, we found a significant Technique × Accuracy two-way interaction ($F_{1.33,14.66}$=7.17, p=.012). Investigation of simple main effects revealed the error rate for Pinpointing was unaffected by eye tracking accuracy ($F_{2,22}$=2.26, p=.128). However, the error rate significantly decreased for BimodalGaze as eye tracking accuracy decreased ($F_{2,22}$=5.94, p=.009). Looking across techniques at each level of accuracy showed that Pinpointing had fewer errors than BimodalGaze at Acc. 1 ($F_{1,11}$=52.15, p<.001), and Acc. 2 ($F_{1,11}$=28.06, p<.001), but not at Acc. 3 ($F_{1,11}$=1.47, p=.250).

We also found a significant Technique × Distance interaction ($F_{2,22}$=8.46, p=.002). Distance had a significant simple main effect on Pinpointing ($F_{2,22}$=5.89, p=.009), revealing that participants made more errors as the distance increased. This could have been caused by the increase in head motion at larger distances, and thus the increased risk of timing issues between clicking and head movement, see fig. 3. For BimodalGaze, distance had no effect on error rate ($F_{2,22}$=1.94, p=.168). Looking at individual distances, we found Pinpointing resulted in significantly fewer error than BimodalGaze at 10° ($F_{1,11}$=44.23, p<.001) and 25° distance ($F_{1,11}$=59.99, p<.001), but not at 40° distance ($F_{1,11}$=3.01, p=.110).

Further investigation of errors made with BimodalGaze revealed over half were made in Gaze Mode, see Table 2. Based on this finding, we investigated the number of errors made in Head Mode for BimodalGaze by calculating the error rates based on trials in which Head Mode was used, and in which an error was made during Head Mode. A three-way repeated measures ANOVA of Technique × Accuracy × Distance, revealed no significant difference between techniques for error rate (Pinpointing: 5.8% ± 0.9%, BimodalGaze: 7.2% ± 0.9%). We also found significant two-way interactions for Technique × Distance ($F_{2,22}$=10.21, p=.001) and Technique × Accuracy ($F_{1.4,15.1}$=4.13, p=.05). Further investigation of the simple main effects revealed that Pinpointing had significantly lower error

Table 2: Prevalence of BimodalGaze errors made in Gaze Mode based on accuracy error and target distance

	10°	25°	40°	Total
Acc. 1	65.2%	58.5%	50.0%	58.5%
Acc. 2	61.8%	51.4%	68.8%	60.4%
Acc. 3	26.1%	41.9%	50.0%	40.2%
Total	55.3%	52.1%	56.3%	54.4%

rate for the highest accuracy condition compared with BimodalGaze (Pinpointing: 4.5%, BimodalGaze: 10%, p=.018).

5.3 Selection Time

We were interested in how automating the switch between Gaze and Head Mode affects selection time. We define selection time as the time between the start of a trial to a successful selection, irrespective to the amount of prior incorrect selections.

We found no significant interactions for selection time. Technique had a significant main effect ($F_{1,11}$=5.17, p=.044) where Pinpointing was significantly faster overall than BimodalGaze (Fig. 7a). However, the mean overall difference was only 140ms. Accuracy also had a significant main effect ($F_{2,22}$=46.84, p<.001) where a lower accuracy, and therefore a higher reliance on Head Mode, lead to higher selection times. Post hoc-tests showed significant differences at all levels (all p<=0.38). Finally, Distance had a main effect ($F_{2,22}$=61.52, p<.001). Unsurprisingly, higher distance lead to significantly higher selection time. Post-hoc tests showed significant differences between all levels (all p<=.001).

Similarly to error rate, we analysed the average selection time for all trials in which Head Mode was used to select the target with BimodalGaze (and thus discard trials when selection was made only with gaze) using a three-way repeated measures ANOVA of Technique × Accuracy × Distance. We found BimodalGaze's selection time was significantly slower than Pinpointing by 220ms when we discounted selections made using only gaze (Pinpointing: 1.44±.56s, BimodalGaze: 1.66±.95s, $F_{1,11}$=11.57, p=.006).

5.4 Refinement Time

We excluded the trials where no refinement was used to investigate the average time spent in Head Mode for each technique (fig. 8). We found a main effect of Technique which showed users spend significantly more time in Head Mode with Pinpointing (0.75s) compared to BimodalGaze (0.65s) ($F_{1,11}$=15.78, p=.002). In Pinpointing

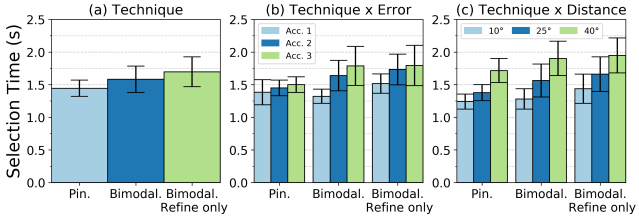

Figure 7: Mean selection time. Error bars represents mean 95% confidence interval.

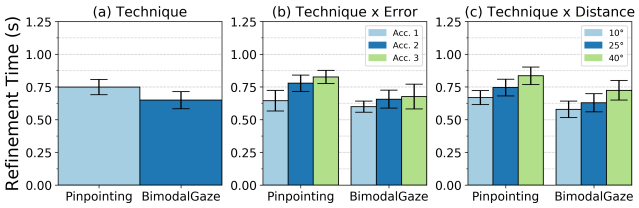

Figure 8: Mean refinement time for trials where refinement was used. Error bars represents 95% confidence interval.

the pointer is not visible before the user clicks the button to enter Head Mode, which could lead to participants spending more time finding and processing the pointer position at the start of refinement stage. BimodalGaze's lower refinement time combined with its higher selection time compared to Pinpointing, suggests that the switching could be further optimised. Eye tracking accuracy ($F_{2,22}$=44.25, p<.001) and Distance ($F_{2,22}$=71.20, p<.001) also had significant main effects. Post-hoc tests showed that decreasing accuracy or increasing distance leads to more refinement time for both techniques (all p<.001).

5.5 Head Movement

Investigation into head movement during trials showed showed no significant interactions (fig. 9). Significant main effects of Accuracy ($F_{2,22}$=31.55, p<.001) and Distance ($F_{1.26,13.87}$=101.39, p<.001), indicate that lower eye tracking accuracy, or larger distances led to more head movement for both techniques. Direction seemed to have no effect on the prevalence of selections in Gaze Mode. This demonstrates that users did not need to perform significantly larger head movements for BimodalGaze compared with Pinpointing.

5.6 Qualitative results

Friedman tests on usability ratings showed significant differences on all metrics except learnability, however Bonferroni corrected Wilcoxon post-hoc tests showed no significant differences between conditions. In general, participant preferences were split between the two techniques (BimodalGaze: 5, Pinpointing: 7), with each offering unique advantages.

Participants praised the seamless switching of BimodalGaze which was "effortless" (P1), and because "it worked all the time and I did not have to do anything" (P11). The automatic switching between modes did not appear to work as well for some participants, which appeared "a bit random" (P4). Participants also noted that

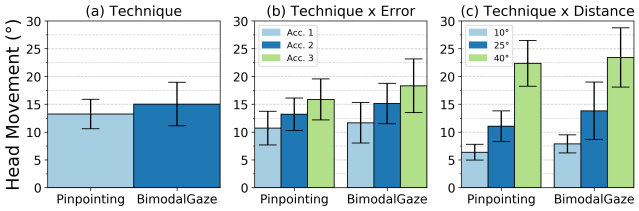

Figure 9: Mean head movement. Error bars represents mean 95% confidence interval.

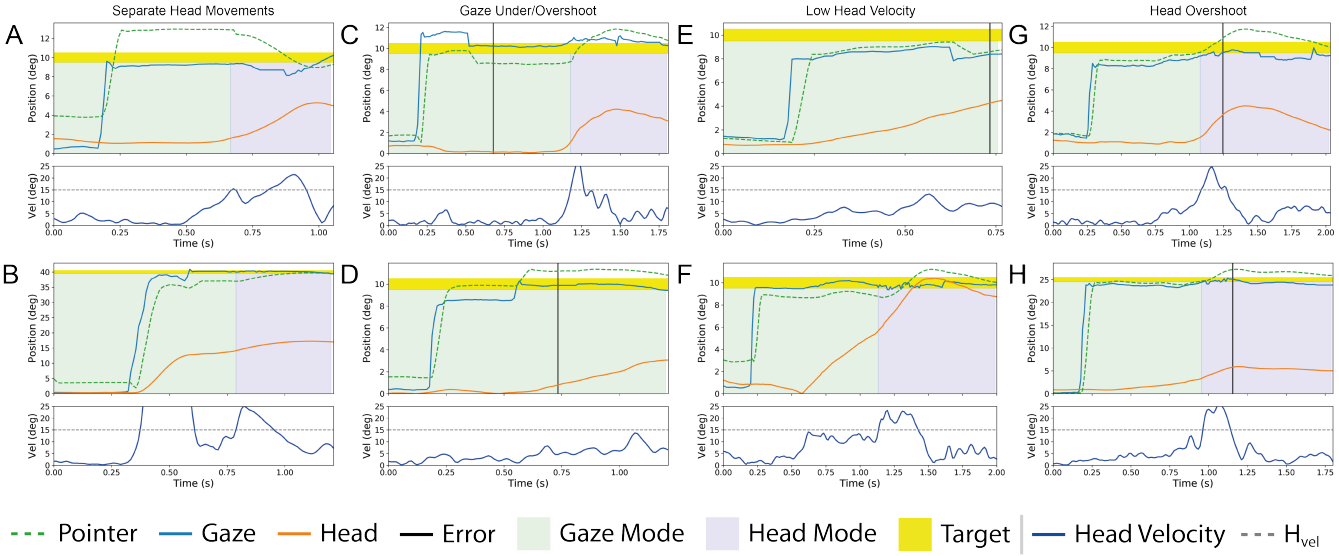

Figure 10: Trial examples of selections made with BimodalGaze.

the technique worked better for conditions with low eye tracker accuracy where "head movement was more pronounced which made the clutch control feel more natural" (P11), because for conditions with high eye tracking accuracy "I only had to do a small head movement, which was not always identified by the system" (P11). Participants also positively mentioned BimodalGaze's continuous feedback which provided opportunities to make informed decisions of whether to enter Head Mode - "I knew immediately if and how I should adjust the pointer" (P5). However, participants also mentioned that the continuous feedback could be distracting "especially when the eyes were controlling the cursor" (P2) or "when the eye tracker was inaccurate" (P12).

For Pinpointing, the main advantage was that participants felt "more in control [because] I decided when to show the feedback" (P2) because of the manual mode switching, which in turn meant for some "it was easier to select the target" (P10). However, some participants expressed difficulties with timing their button presses. P5 stated "I would move my gaze towards the object and then start aligning my head. During this time I would press the button which resulted in me overshooting the target and then having to readjust. Sometimes, I would also press the button preemptively and then realise that my gaze is still at the centre", a point reiterated by P11 "I found that I pressed the button too early. I would press the button too early so that the cursor was in the wrong place or the head would drag the cursor away from the target."

5.7 Participant Behaviour

Based on our analysis and feedback, we further investigated individual trials to unveil participants' selection and gaze behaviour during pointing and selection using BimodalGaze (fig. 10). From this in-depth investigation, we found interesting characteristics of how eye tracker accuracy and system thresholds affected selection.

We observed users pausing to assess whether or not to enter Head Mode before making a gestural head movement. If a natural head shift was not needed to reach the target comfortably, participants would perform a gaze shift, assess the pointer position and then perform a gestural head shift (e.g. fig. 10a). Likewise, in the event of a natural head movement, users would stop or significantly slow down their natural head movement before performing a gestural head movement – they would very rarely transition from natural to gestural during a single head movement (e.g. fig. 10b).

Participants performed more selection errors in conditions where Gaze Mode was initially used for selection, see Table 2. A common situation that leads to errors in Gaze Mode was the natural overshooting (e.g. fig. 10c) or undershooting (e.g. fig. 10d) of the primary saccade which may last up to 500 ms [Bahill et al. 1975]. This, combined with the eye tracking accuracy error, may cause the pointer to appear on the target when the participant's gaze was not. This situation may cause the user to press the button for selection - which according to the Keystroke-Level Model proposed by Card, Moran, and Newell [Card et al. 1980] could take anywhere between 80-280ms. However, by the time the button press was registered by the system, the users' gaze had moved onto the target, which in turn moves the pointer outside the target. This situation is more common for conditions with higher eye tracking accuracy, and not as prevalent for low accuracy conditions because the eye tracking accuracy error is larger than the error induced by over/undershooting, and as such the pointer rarely appears on the target.

We also found that our choice of a higher value for H_{vel} led to an inability of BimodalGaze to identify small gestural head movements for small refinements. As such users would rely entirely on Gaze Mode, which either caused incorrect selections (e.g. fig. 10e) or resulted in exaggerated head movements (e.g. fig. 10f). The latter led to overshooting the target when in Head Mode, which in turn led to unsuccessful selections as participants attempted to select the target whilst in the process of overshooting (fig. 10g, h). This phenomenon is more common when the eye tracking accuracy was higher, as only very small head movements would be needed for

accurate refinement, and may explain the significant difference in error rate we see between the techniques for the highest eye tracker accuracy condition.

6 DISCUSSION

The study results validate the principal approach of distinguishing between natural and gestural head movement for head-refinement of a gaze cursor. The comparative evaluation against a manually switched technique points to performance limitations that we discuss below. However, the headline results are:

- Users are effective with BimodalGaze. All participants successfully completed all selection tasks. When initial selections were off-target (i.e. counted as error in the study), users had no problem correcting their input.
- BimodalGaze demonstrates that gestural head movement can be reliably differentiated from natural head movements that are implicit with gaze, and harnessed as explicit input. Our work shows that the switching between gaze and head modes can be automated using the underlying knowledge of the way in which our head movement supports gaze.
- BimodalGaze enables selection of targets that users are not able to successfully acquire with gaze alone. Despite only accounting for 17.5% of selections, gaze-only errors accounted for over 50% of errors. This highlights that with gaze alone, users may not be able to select a target at all if the eye tracking accuracy is too poor, while our results show that participants are able to select all targets successfully with BimodalGaze, irrespective of scale of eye tracking error.

Compared to Pinpointing, BimodalGaze automates the mode switch between gaze and head input. The advantage is that our technique does not require explicit input for mode-switching, resulting in a more seamless transition. This also avoids the need for an additional input modality such as manual input in Pinpointing. Note that in our study, BimodalGaze was combined with a button click for selection confirmation. However, it could also be combined with alternate confirmation techniques such as dwelling to make the whole selection hands-free. This can be useful in situations where the hands are busy or unavailable.

A second feature by which BimodalGaze is different from Pinpointing is that head-refinement is optional rather than enforced. As our results show, over half of selections were made by gaze alone when eye tracker error was lower and target distance shorter. This validates the design choice in principle. However, 54.4% of selections made in gaze mode were off-target and required correction, which compromised the advantage and led to users spending longer time in gaze mode. Users made errors in gaze mode as a result of premature selection whilst under/overshooting the target. Clearer pointing feedback, such as target highlighting, or target acquisition techniques (e.g. BubbleCursor [Grossman and Balakrishnan 2005]) could be used to improve this aspect. In the latter case, Head Mode would be necessary in cases where the size of the bubble exceeds the density of the targets.

In performance comparison, BimodalGaze was on average slower for selection than Pinpointing, though not substantially (≈ 140 ms). This was due to longer time spent in gaze mode, while BimodalGaze was faster in Head Mode (≈100 ms). Overall, there was no significant difference in error rate in Head Mode. However, the error rate with BimodalGaze dropped when eye tracking error increased, and users made fewer errors than with Pinpointing when eye tracking error was highest. This indicates that our technique is particularly beneficial when eye tracking accuracy is poor.

The performance results suggest the mode switching could be further optimised. The time spend in gaze mode can be reduced by techniques that address premature selection, as discussed above. Another area of improvement are the criteria for entering Head Mode. The low H_{vel} resulted in difficulties entering Head Mode when only small movements were required, and in turn causes the head to overshoot as a result of exaggerated head movement. We selected a value of H_{vel} heuristically to minimise consistent mode switching. Optimisation of H_{vel}, or use of more sophisticated techniques (e.g. accuracy-dependent thresholds), could alleviate this problem. Instead of a rule-based approach as used in this work, machine learning could be adopted for classifying head movements, or to optimise the system's parameters.

The distinction of natural and gestural head movements makes it possible to attach different behaviours to objects that take gestural head movement as input while avoiding unwanted behaviours caused by natural head movements. In BimodalGaze, we used gestural head movements to refine a gaze cursor but other mappings are possible. For example, assuming that gaze pointing is accurate enough, gestural head movements could be used to manipulate (scaling, rotating, etc.) gazed on objects.

All our results were obtained in VR. However, we do not expect this to limit the applicability of BimodalGaze, as the technique builds on eye-head coordination behaviours that are consistent with observations in real-world tasks [Land and Tatler 2012]. Head movement is more common when interactions span a wider FOV, for instance on large screens or across devices, but BimodalGaze is also applicable with narrower FOV displays where natural head movements is less prevalent.

7 CONCLUSION

In this work we introduced BimodalGaze, a novel technique for fine-grained control of a gaze cursor. The technique enables users to refine a gaze-cursor with head movement and is entirely hands-free, which is useful for situations where the hands are busy or unavailable. The transition from a gaze mode for initial cursor placement to refinement by head movement is implicit, based on detection of gestural head movement, which is significant as it removes the need for any manual or other explicit input thus making the process more seamless. Evaluation of the technique highlights advantages of the technique in particular when eye tracking accuracy is poor but also points to performance limitations in the present implementation, which this work addressed with in-depth analysis of user performance and errors observed. This not only provides insight for improvement of BimodalGaze, but also generally into the classification of gestural versus natural head movement. The notion of classifying natural and gestural head movements extends beyond refinement of a pointer, opening up new opportunities for mapping gestural head movement without affecting the head's natural ability to support gaze.

REFERENCES

A. Terry Bahill, Michael R. Clark, and Lawrence Stark. 1975. Glissades-eye movements generated by mismatched components of the saccadic motoneuronal control signal. *Mathematical Biosciences* 26, 3-4 (1975), 303–318.

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 (Oct. 2003), 280–290. https://doi.org/10.1007/s10209-003-0053-y

Emilio Bizzi. 1974. The coordination of eye-head movements. *Scientific American* 231, 4 (Oct. 1974), 100–109.

Jonas Blattgerste, Patrick Renner, and Thies Pfeiffer. 2018. Advantages of Eye-gaze over Head-gaze-based Selection in Virtual and Augmented Reality Under Varying Field of Views. In *Proceedings of the Workshop on Communication by Gaze Interaction (COGAIN '18)*. ACM, New York, NY, USA, Article 1, 9 pages. https://doi.org/10.1145/3206343.3206349

Stuart K. Card, Thomas P. Moran, and Allen Newell. 1980. The Keystroke-level Model for User Performance Time with Interactive Systems. *Commun. ACM* 23, 7 (July 1980), 396–410. https://doi.org/10.1145/358886.358895

Edward G. Freedman. 2008. Coordination of the eyes and head during visual orienting. *Experimental Brain Research* 190, 4 (oct 2008), 369–387. https://doi.org/10.1007/s00221-008-1504-8

Edward G. Freedman and David L. Sparks. 2000. Coordination of the eyes and head: movement kinematics. *Experimental Brain Research* 131, 1 (mar 2000), 22–32. https://doi.org/10.1007/s002219900296

Tovi Grossman and Ravin Balakrishnan. 2005. The Bubble Cursor: Enhancing Target Acquisition by Dynamic Resizing of the Cursor's Activation Area. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. ACM, New York, NY, USA, 281–290. https://doi.org/10.1145/1054972.1055012

Daniel Guitton and Michel Volle. 1987. Gaze control in humans: eye-head coordination during orienting movements to targets within and beyond the oculomotor range. *Journal of neurophysiology* 58, 3 (1987), 427–459.

Laurent Itti, Nitin Dhavale, and Frédéric Pighin. 2006. Photorealistic attention-based gaze animation. In *2006 IEEE International Conference on Multimedia and Expo*. IEEE, 521–524. https://doi.org/10.1109/ICME.2006.262440

Shahram Jalaliniya, Diako Mardanbegi, and Thomas Pederson. 2015. MAGIC Pointing for Eyewear Computers. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers (ISWC '15)*. ACM, New York, NY, USA, 155–158. https://doi.org/10.1145/2802083.2802094

Tobit Kollenberg, Alexander Neumann, Dorothe Schneider, Tessa-Karina Tews, Thomas Hermann, Helge Ritter, Angelika Dierker, and Hendrik Koesling. 2010. Visual search in the (un)real world: how head-mounted displays affect eye movements, head movements and target detection. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (ETRA '10)*. ACM, New York, NY, USA, 121–124. https://doi.org/10.1145/1743666.1743696

Andrew Kurauchi, Wenxin Feng, Carlos Morimoto, and Margrit Betke. 2015. HMAGIC: Head Movement and Gaze Input Cascaded Pointing. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '15)*. ACM, New York, NY, USA, Article 47, 4 pages. https://doi.org/10.1145/2769493.2769550

Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 81, 14 pages. https://doi.org/10.1145/3173574.3173655

Michael Land and Benjamin Tatler. 2012. *Looking and Acting: Vision and Eye Movements in Natural Behaviour*. Oxford University Press, United Kingdom. https://doi.org/10.1093/acprof:oso/9780198570943.001.0001

Michael F. Land. 2004. The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations. *Experimental Brain Research* 159, 2 (01 Nov 2004), 151–160. https://doi.org/10.1007/s00221-004-1951-9

Chris Lankford. 2000. Effective Eye-gaze Input into Windows. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (ETRA '00)*. ACM, New York, NY, USA, 23–27. https://doi.org/10.1145/355017.355021

Christof Lutteroth, Moiz Penkar, and Gerald Weber. 2015. Gaze vs. Mouse: A Fast and Accurate Gaze-Only Click Alternative. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 385–394. https://doi.org/10.1145/2807442.2807461

Diako Mardanbegi, Tobias Langlotz, and Hans Gellersen. 2019a. Resolving Target Ambiguity in 3D Gaze Interaction Through VOR Depth Estimation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 612, 12 pages. https://doi.org/10.1145/3290605.3300842

Diako Mardanbegi, Ken Pfeuffer, Alexander Perzl, Benedikt Mayer, Shahram Jalaliniya, and Hans Gellersen. 2019b. EyeSeeThrough: Unifying Tool Selection and Application in Virtual Environments. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 474–483. https://doi.org/10.1109/VR.2019.8797988

Tomislav Pejsa, Daniel Rakita, Bilge Mutlu, and Michael Gleicher. 2016. Authoring Directed Gaze for Full-body Motion Capture. *ACM Trans. Graph.* 35, 6, Article 161 (Nov. 2016), 11 pages. https://doi.org/10.1145/2980179.2982444

Kevin Pfeil, Eugene M. Taranta, II, Arun Kulshreshth, Pamela Wisniewski, and Joseph J. LaViola, Jr. 2018. A Comparison of Eye-head Coordination Between Virtual and Physical Realities. In *Proceedings of the 15th ACM Symposium on Applied Perception (SAP '18)*. ACM, New York, NY, USA, Article 18, 7 pages. https://doi.org/10.1145/3225153.3225157

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. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 373–383. https://doi.org/10.1145/2807442.2807460

Ken Pfeuffer and Hans Gellersen. 2016. Gaze and Touch Interaction on Tablets. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 301–311. https://doi.org/10.1145/2984511.2984514

Marco Porta, Alice Ravarelli, and Giovanni Spagnoli. 2010. ceCursor, a Contextual Eye Cursor for General Pointing in Windows Environments. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (ETRA '10)*. ACM, New York, NY, USA, 331–337. https://doi.org/10.1145/1743666.1743741

Yuan Yuan Qian and Robert J. Teather. 2017. The Eyes Don'T Have It: An Empirical Comparison of Head-based and Eye-based Selection in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction (SUI '17)*. ACM, New York, NY, USA, 91–98. https://doi.org/10.1145/3131277.3132182

K. Ruhland, C. E. Peters, S. Andrist, J. B. Badler, N. I. Badler, M. Gleicher, B. Mutlu, and R. McDonnell. 2015. A Review of Eye Gaze in Virtual Agents, Social Robotics and HCI: Behaviour Generation, User Interaction and Perception. *Computer Graphics Forum* 34, 6 (Sept. 2015), 299–326. https://doi.org/10.1111/cgf.12603

Ludwig Sidenmark and Hans Gellersen. 2019a. Eye, Head and Torso Coordination During Gaze Shifts in Virtual Reality. *ACM Trans. Comput.-Hum. Interact.* 27, 1, Article Article 4 (Dec. 2019), 40 pages. https://doi.org/10.1145/3361218

Ludwig Sidenmark and Hans Gellersen. 2019b. Eye&Head: Synergetic Eye and Head Movement for Gaze Pointing and Selection. In *Proceedings of the 32Nd Annual ACM Symposium on User Interface Software and Technology (UIST '19)*. ACM, New York, NY, USA, 1161–1174. https://doi.org/10.1145/3332165.3347921

Oleg Špakov, Poika Isokoski, and Päivi Majaranta. 2014. Look and Lean: Accurate Head-assisted Eye Pointing. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '14)*. ACM, New York, NY, USA, 35–42. https://doi.org/10.1145/2578153.2578157

John S. Stahl. 1999. Amplitude of human head movements associated with horizontal saccades. *Experimental Brain Research* 126, 1 (apr 1999), 41–54. https://doi.org/10.1007/s002210050715

Sophie Stellmach and Raimund Dachselt. 2012. Look & Touch: Gaze-supported Target Acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 2981–2990. https://doi.org/10.1145/2207676.2208709

D. Tweed, B. Glenn, and T. Vilis. 1995. Eye-head Coordination during Large Gaze Shifts. *Journal of Neurophysiology* 73, 2 (Feb. 1995), 766–779. https://doi.org/10.1152/jn.1995.73.2.766

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)*. ACM, New York, NY, USA, 246–253. https://doi.org/10.1145/302979.303053