Eye&Head: Synergetic Eye and Head Movement for Gaze Pointing and Selection

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

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

ABSTRACT

Eye gaze involves the coordination of eye and head movement to acquire gaze targets, but existing approaches to gaze pointing are based on eye-tracking in abstraction from head motion. We propose to leverage the synergetic movement of eye and head, and identify design principles for *Eye&Head* gaze interaction. We introduce three novel techniques that build on the distinction of head-supported versus eyes-only gaze, to enable dynamic coupling of gaze and pointer, hover interaction, visual exploration around pre-selections, and iterative and fast confirmation of targets. We demonstrate Eye&Head interaction on applications in virtual reality, and evaluate our techniques against baselines in pointing and confirmation studies. Our results show that Eye&Head techniques enable novel gaze behaviours that provide users with more control and flexibility in fast gaze pointing and selection.

Author Keywords

Gaze interaction; Eye-head coordination; Eye tracking; Target selection; Virtual Reality; 3D Interaction

CCS Concepts

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

INTRODUCTION

Eye gaze naturally involves a coordination of head and eye movement [4]. As we interact with the world, we shift our gaze from object to object. Where gaze shifts are small, they may be achieved by eye movement alone but generally they involve a contribution of head movement [6]. Even though our eyes have a physical range of 50°, they rarely rotate beyond 30° relative to the head [24]. A gaze shift will typically start with eye movement but be supported by head movement, not only to reach further, but also to stabilise the eyes in a comfortable position after reaching a target [55]. However, in spite of the prevalence of head motion in gaze, eye-head coordination has not been reflected in the design of gaze interfaces.

Prior work on gaze for pointing and selection has treated head movement as a problem that interferes with eye tracking, to the extreme of using chin rests to suppress them [3, 33]. State of the art eye trackers are less restrictive but compensate for head motion in gaze estimation [11, 62]. In contrast, we propose to use eye and head movement in tandem, for multimodal *Eye&Head Gaze* interaction. By understanding eye-head coordination we can design novel interactions that leverage concurrent input from eye and head tracking. This has particular relevance for gaze interaction beyond the computer screen, for example with larger display surfaces, head-mounted displays (HMD), and virtual, augmented or real 3D environments, as these expose wider fields of view (FOV).

Gaze is attractive for interaction as we naturally look at objects that we consider for manipulation. Also, we are able to move our gaze faster to a target than our hands or a manually controlled cursor. However, users rely on gaze primarily for visual information seeking and overlaying this with gaze input has well-known problems. Coupling gaze with continuous feedback supports target selection but can be distracting when it follows every eye movement [17, 43]. As gaze is "always on" there is the Midas touch problem of deciding when to select input, necessitating a separate confirmation mechanism [16]. In gaze-only interfaces, this is addressed with dwell methods but these require users to fixate their gaze unnaturally long on targets while having to avoid incidental dwell on other objects [14]. As we show in this work, these problems can be tackled in new ways by combining eye and head movement.

In this work, we first identify design principles for Eye&Head gaze interaction, grounded in eye-head coordination literature, and then apply these in three novel gaze interaction techniques. The first one, *Eye&Head Pointing*, lets users point with their gaze, but the pointer is only updated to a new gaze position when the gaze shift is supported by head movement. The effect is that head movement acts like a clutch, for dynamic coupling of the user's gaze and the pointer. *Eye&Head Dwell* is a complementary confirmation technique, where a dwell timer is only triggered by a head-supported gaze shift but can be paused and resumed with eyes-only gaze. *Eye&Head Convergence* is an alternative to dwell for fast target confirmation by aligning both the eye pointer and the head pointer over a target, which we show to be distinctive as signal of intent. All three techniques have been implemented in a head-mounted virtual reality (VR) environment, with application examples that demonstrate their advantages. We also evaluated them against gaze pointing and dwell selection baselines.

UIST '19, October 20-23, 2019, New Orleans, LA, USA.

ACM ISBN 978-1-4503-6816-2/19/10 ...$15.00.
http://dx.doi.org/10.1145/3332165.3347921

The contributions of this work are: (1) principles of eye-head coordination for gaze interaction; (2) three novel techniques that leverage head-supported versus eyes-only gaze to tackle pointing and selection problems; (3) demonstration of advantages of Eye&Head interaction in VR applications; (4) validation of the principles and user benefits in empirical evaluation.

RELATED WORK

Prior work on eye and head movement for interaction has treated the two modalities as separate rather than integral. Head as well as gaze pointing were developed in the eighties, as alternative to mouse input for users with limited motor control [8, 13, 45, 57]. In comparison, eye movement is faster and requires less energy, while head motion is less jittery and more controlled [3, 23, 44]. We focus on the combination of the two modalities, to take advantage of their relative strengths and synergetic relationship.

Gaze Tracking based on Eye versus Head Movement

Computer displays are normally set up to be comfortably viewable without need for head movement. The visual angle of the display width usually does not exceed 40°, and gaze shifts of up to 20° from a central position are comfortably achieved with eye-in-head rotation [6]. Desktop gaze tracking has consequently focussed on eye movement and treated any head movement as accidental [11, 33, 62]. Gaze tracking with multiple screens can be facilitated with head-mounted eye tracking but this can lead to problems as gaze shifts from one screen to another are performed with a combination of head and eye movement. For example, when a user looks down from a screen in-front to a screen in-hand, they use a different eye-in-head range for viewing [54].

Conversely, gaze attention over wider visual fields is often approximated by head pointing and ignores eye-in-head movement. Various works have used face pose tracking for gaze pointing on large displays [34, 35]. Early work on VR explored gaze directed input but solely based on head orientation [30, 59]. However, also recent products such as Microsoft's HoloLens 1 include gaze abstractions that are based on head movement without any eye-tracking. These works on head-only gaze reflect the major role head movement plays in larger gaze shifts. However, as we will show in this work, it is a fallacy to assume that users would be looking straight ahead when they have completed a gaze shift, even when the shift in attention is supported by head movement.

Gaze-based Pointing and Selection

Selection by pointing generally involves two phases [30]. In a first phase, the user identifies an intended target object by pointing at it. In a second phase, they confirm the target via a further action. Eye movement is highly effective for the pointing phase as we can direct our gaze more quickly toward a target than our hands or any other pointing device. Where gaze shifts involve head movement, it is the eyes that reach a target first while the head follows more slowly [4]. However, eye movement is jittery and a variety of techniques employ gaze for coarse-grained positioning in combination with mouse, pen or touch for refinement [40, 51, 53, 60]. While conventional pointing uses a cursor metaphor, it is less clear how best to provide feedback with gaze as eye movement is primarily engaged in information seeking rather than target identification [17]. We introduce a novel technique addressing this with feedback that follows the eyes only when the head also moves, freeing the eyes to explore around a potential target without distraction by cursor motion.

For the confirmation phase, the most common eyes-only technique is dwell selection by prolonged fixation of a target [9, 16, 57], while it is also possible to use eye gestures [32], smooth pursuit if targets are in motion [56], or additional confirmation buttons [25]. Eyes-only techniques need to be based on gaze behaviours that are distinct from natural viewing, and consequently can be experienced as awkward and tiring. Alternatively, eye gaze can be combined with a separate modality for confirmation, such as key, mouse or button click which also enable higher throughput [61]. This work introduces two novel techniques where confirmation instead is based on eye-head coordinated movement, enabling hands-free confirmation while addressing usability limitations of dwell time methods.

Gaze Interaction in 3D Environments

Eye trackers have become more prevalent in 3D environments such as VR with several commercial products available such as the HoloLens 2, HTC Vive Pro Eye and the FOVE 0. This development has sparked an increase in eye tracking research in 3D environments, and eye tracking has been used for measuring and leveraging gaze behaviour [1, 2, 47, 48], foveated rendering [38], as well as redirected walking [58].

3D environments expose additional challenges for pointing and selection, as targets can appear at different depths, and in larger fields of regard around the user. The user's FOV is naturally controlled by head movement but there is no universally preferred selection method. The prevalent pointing metaphor for targets beyond manual reach is ray-casting [12]. Gaze has been found to be faster than hand pointing, especially for distant objects [52]. A range of works have compared eye and head pointing showing that eye gaze is faster and less strenuous, while head pointing is often preferred as more stable, controlled and accurate [5, 10, 18, 23, 44]. As in 2D contexts, eye pointing can be combined with fast manual confirmation by click or hand gesture [41, 46], or with dwell time or other specific eye movement for hands-free selection [20, 31, 42]. In contrast to the 2D desktop setting, gaze in VR inherently involves eye-head coordination due to the wider FOV. This work is first to reflect how the naturally synergetic movement of head and eyes can be leveraged in design of gaze interactions.

Combination of Eye and Head Movement

Head movement has been used to support gaze pointing in a variety of ways. Head gestures such as nodding have been proposed for confirmation of targets users look at [27, 50]. These methods exploit eye-head coordination implicitly as they track the compensatory eye movement during a head gesture, without need for separate head tracking. In extension, head turning has been proposed for scalar input to controls fixated by gaze [36] and 3D target disambiguation [28]. In *EyeSeeThrough*, head movement controls a toolglass that can be moved over gaze-fixated targets [29]. Other work has supplemented eye pointing with subsequent refinement of the selection by head movement [18, 19, 22, 49]. Recently,*Pinpointing* compared head versus eyes as primary pointing modes, and a variety of techniques for subsequent selection refinement [23].

The existing body of work has in common that eye and head movement are treated in separation, for use of one after the other. In contrast, this work proposes pointing and selection techniques that build on the *integral* relationship of eye and head movement in gaze interaction.

EYE&HEAD DESIGN PRINCIPLES

There are no prior studies of eye-head coordination in HCI but we can build on fundamental insights from neuroscience:

- Larger gaze shifts require head movement. The eyes have a physical range of 50° but rarely rotate beyond 30° relative to a central position in the head [24]. The head is therefore needed to explore further.
- Not all gaze shifts involve head movement. Gaze shifts up to ~20° can be performed with only eye movement [6, 24].
- The decision and timing to support a gaze shift with head movement is influenced by multiple factors, such as expected duration of maintaining gaze in the vicinity of the new direction, position of the next target, and initial eye-in-head position [7, 37].
- The head will start or continue to move after a target is first reached by the eyes, and the eyes perform compensatory movement based on the vestibulo-ocular reflex (VOR) to maintain focus on the target [4, 6].
- The head will not typically move fully toward a target, as head movement requires more energy while a comfortable eye-in-head position is reached sooner [6].

A key design implication is that we can distinguish two types of gaze shift: *head-supported* gaze shifts where eye movement is accompanied by head movement, and *eyes-only* gaze shifts that are performed without contribution by the head. We identify three design principles that build on this dichotomy:

Head-supported gaze is more stable than eyes-only gaze. The head does not contribute to every gaze shift; it only supports the eyes when the attention shift is more substantial such that it requires or warrants a recentering of the area the eyes can comfortably explore. Consequently, gaze points selected with head support are less volatile than gaze points that change with every eye fixation.

Eyes-only gaze explores around objects selected by head-supported gaze. Objects acquired with head support are significant in that they constitute a base from which other objects in the vicinity are explored eyes-only, with less effort.

Head-eye alignment can signal intent. The head does not normally rotate all the way with the eyes to acquire an object, and an offset remains between head and eye at the end of a gaze shift. Though it may seem counterintuitive, it follows that we practically never look exactly straight ahead. Alignment of head and eye can therefore be available to signal intent.

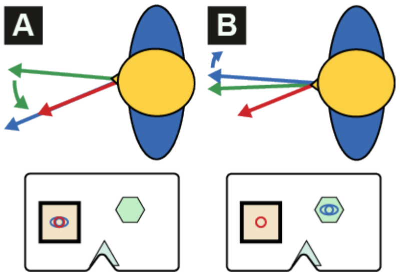


Figure 1. Eye&Head Pointing. The pointer moves to a new gaze position only when the gaze shift is head-supported. A: The pointer (red) follows the user's gaze (blue) to the square as the user is also moving their head (green). B: The user shifts their gaze to the hexagon, but as the shift is eyes-only without head movement the pointer remains on the square.

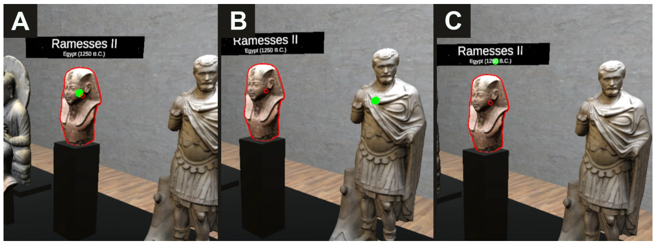


Figure 2. Eye&Head Pointing in a virtual museum. The green dot indicates the gaze position. A: An artefact is selected with head-supported gaze, triggering contextual information. B: The selection is maintained when the user looks at nearby artefacts, using eyes-only gaze. C: The context display can be viewed without needing to carefully maintain gaze on the artefact.

Note the fundamental limitation of head pointing. As the head does not normally move the full distance to the target, head pointing does not accurately identify objects of interest, even when the gaze shift is supported by head motion.

EYE&HEAD GAZE INTERACTION TECHNIQUES

We designed three novel pointing and selection techniques based on the identified principles for Eye&Head gaze. The techniques are gaze-only, based on where the user looks, but combine information from both eye and head in the process.

Eye&Head Pointing

The first technique is for gaze pointing modulated by head movement. As shown in Figure 1, the *Eye&Head Pointer* moves to a new gaze position when the gaze shift is accompanied by head motion. When a user moves their gaze without head movement, the pointer remains at the last position acquired with head support. Note that head movement only modulates the pointer. The points selected are gaze points, and it is not taken into account where the head points relative to a target. The technique can be implemented with a cursor or implicit feedback highlighting objects selected by the pointer. It has the effect that a cursor or object selection is more stable and less distracting than a conventional gaze cursor that follows eye movement continuously.

The Eye&Head Pointer enables users to fluidly couple gazing and pointing. They can move their eyes to look at objects without the pointer following. However as soon as they also move their head, the pointer will jump to where they are looking. The coupling can be implicit and entirely based on naturally eye-head coordination, but users can also choose to move their head to have the pointer follow them to a gaze target they might otherwise have attained eyes-only.

We developed a virtual museum application to illustrate Eye&Head pointing for exploration of artefacts by gaze (Fig. 2). Head-supported gaze shifts trigger a contextual display over the artefact, in the same way as a mouse hover might in a desktop interface. Eyes-only gaze can be used to view other artefacts while the selection is maintained. The hover selection only changes once the user shifts their gaze with head support, turning to explore another artefact. The hover selection can also be extended by using a manual trigger (or other "click" method) to expand the contextual display for more detail. This demonstrates several advantages of the Eye&Head technique:

- Hover interaction can be driven by gaze while avoiding that the display changes with every gaze shift.
- There is no need for users to carefully maintain their gaze on an object in order to maintain the selection.
- Users are free to visually explore the interface while a gaze selection is maintained.
- A gaze selection can be confirmed with a click method even when the user is no longer looking at the selected object.

Eye&Head Dwell

This technique complements Eye&Head pointing with a novel dwell method for confirmation of selected targets. As shown in Figure 3, a dwell timer is triggered only for targets that have been selected with head-supported gaze. If the user looks away from the target with eyes-only gaze, the timer is paused, and it resumes when they return their gaze to the target. If the user performs a head-supported gaze shift before dwell time has completed, selection is aborted and the timer reset.

Figure 4 shows Eye&Head Dwell with *Euler's Constellations*, a puzzle game we developed for illustration. User are tasked to draw a star constellation by successive gaze selection of stars, with the challenge to draw in one line without traversing any path more than once. As such, users have to plan their selections ahead and revisit past selections to solve the puzzle. The Eye&Head Pointer combined with Eye&Head Dwell allows users to gaze on past and future selections without any time pressure and without risking that a selection is committed accidentally. This demonstrates key advantages the Eye&Head technique has over conventional gaze point-and-dwell:

- Users are free to dwell on potential targets without risk of unintended selection. This is useful for cognitively demanding tasks where thorough consideration of a choice can induce prolonged fixation.
- Selection can be paused while other options are inspected, and users save selection time when they decide to return to their first choice.

Eye&Head Convergence

Eye&Head Convergence is an alternative to dwell for confirmation, and can be combined with Eye&Head or conventional gaze pointing. The technique applies the principle of using

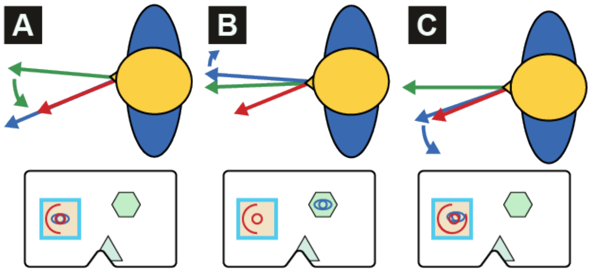


Figure 3. Eye&Head Dwell. A: Eye&Head pointing at an object triggers a dwell timer (red). B: The timer is paused when the user moves their gaze (blue) away from the object without also moving their head (green). C: The dwell timer resumes when the gaze returns to the selected object.

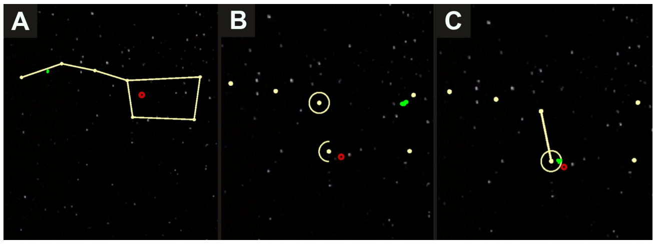


Figure 4. Eye&Head Dwell in a puzzle game (green dot: gaze; red dot: Eye&Head pointer). A: Display of a constellation to be reproduced by gaze. B: The user started selecting a star (near red dot) but moves their gaze to other stars to re-evaluate the selection, causing the dwell timer to pause. C: The user returns their gaze and completes the selection.

head-eye alignment for confirmation. The underlying assumption is that the head does not fully align with the gaze vector when a new target is reached, for which we provide empirical support below. The additional head movement to "close the gap" can then be used to confirm the target selection. Figure 5 illustrates our implementation of the technique based on a cursor metaphor. When the user's gaze reaches a target, the gaze cursor expands to display a convergence area and additionally the head cursor is shown. The user can then confirm the selection by moving the head cursor to within the convergence area. If the head cursor is already within the convergence area, a timer is started during which the eyes and head have to remain within the threshold to confirm the selection.

The size of the convergence area is defined by an angular threshold around the gaze point, and impacts the behaviour of the technique. With larger angular thresholds, less head movement is needed and selections are faster. However a larger threshold also increases the likelihood that a head cursor is already within the convergence area, and the risk of accidental selection. A lower threshold reduces this risk, but requires more head movement. The required time to confirm a target (Equation 1), t, is defined as the angular distance from the gaze cursor to the head cursor, d, subtracting the angular threshold, a, divided by the angular head velocity, v. If the gaze and head distance are equal to or smaller than the chosen radius, the confirmation time is equal to a chosen dwell time, t_d.

$$t = \begin{cases} (d-a)/v & \text{if } d > a \\ t_d & \text{else} \end{cases} \tag{1}$$

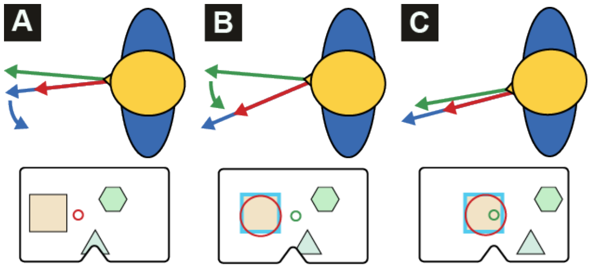
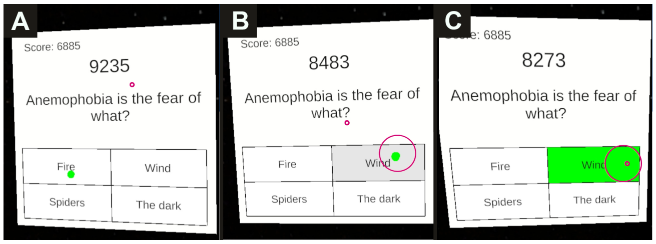


Figure 5. Eye&Head Convergence with conventional gaze. A: The pointer (red) follows the gaze (blue) toward the square. B: As gaze reaches the target, the cursor expands to define a convergence area and a head pointer (green) appears. C: The target selection is confirmed by moving the head pointer into the convergence area.



Score: 6885
9235
Anemophobia is the fear of what?
Fire
Wind
Spiders
The dark

Score: 6885
8483
Anemophobia is the fear of what?
Fire
Wind
Spiders
The dark

e: 6885
8273
Anemophobia is the fear of what?
Fire
Wind
Spiders
The dark

Figure 6. Eye&Head Convergence with Eye&Head Pointing in a quiz application. The green dot indicates gaze, and the red dot the head point. A: The user inspects answers without triggering interaction. B: When the user move their head toward an answer at which they look, the cursor jumps to their gaze point and a head pointer appears as the cursor expands to the convergence area. C: The user commits the selection by moving the head pointer to within the convergence area.

Figure 6 illustrates the convergence technique in a quiz game, combined with Eye&Head Pointing. Users are shown quiz questions and tasked to select the correct answer as quickly as possible. An erroneous answer awards no points and as such, users have to select both quickly and accurately. The application highlights a number of advantages of both Eye&Head Pointing and Eye&Head Convergence:

- Users are free to visually inspect choices without any distraction by cursor movement. This is useful when both speed and accuracy are important.
- Potential targets can be inspected for as long as users need to, without risking unintended selection.
- Users can traverse their gaze or head across other options when reaching for a target, without deselecting a currently highlighted option. This affords more freedom in the layout of choices on the interface.
- Selection by convergence can be faster than a conservative dwell time, as the required head motion can be performed in shorter time.
- Convergence is less error-prone for selection than short dwell times, as users have better control over their head movement than over the duration of gaze fixations.

PERFORMANCE EVALUATION

The applications we developed demonstrate qualitative advantages of Eye&Head over conventional eye pointing and selection. In addition, we conducted two user studies for evaluation, one on pointing and the other on confirmation. The objective of these studies was to compare user performance with Eye&Head against eye gaze pointing and conventional dwell selection as baselines. The pointing study had the additional objective to quantify the offset between head and eye during naturalistic gaze, to test our assumption about head-eye alignment and inform the choice of a threshold for Eye&Head Convergence.

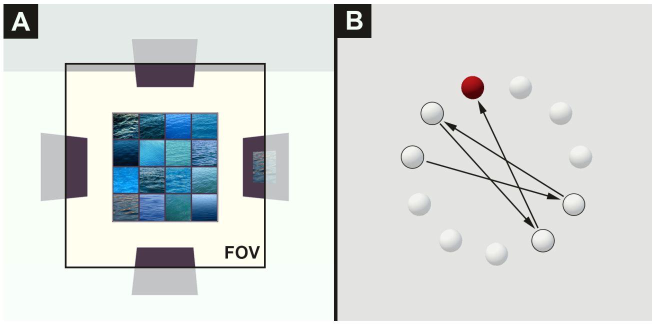


Figure 7. User study tasks. A: Participants were shown an image on one of the side panels and had to locate and select the matching image on the grid. Note, only one panel at a time was visible to the user. B: Participants were tasked to select the highlighted targets (red) in succession.

We designed two tasks for the purposes of our study (Fig. 7). The first one is a search task, designed to require consideration of potential targets (and thus more gaze shifts) prior to selection. The second task, in contrast, highlighted individual targets in a predictable sequence, so that it could be performed with single gaze shifts. We used these tasks instead of our applications for evaluation, as they are more general and better suited for fair comparison against baselines.

Search Task. Participants had to find and select a matching picture within a 4x4 grid (Fig. 7A). The original picture was shown on one of four panels surrounding the grid. 8 of the 16 pictures were shown consecutively at each panel, each picture being shown twice in total (32 trials per participant and condition). Each panel had a corresponding grid shuffle. The panel order and their corresponding stimuli order were randomised. All trials were performed sequentially without pause or head realignment to mimic a typical scenario where users perform multiple selections. The grid was placed at 4 metres distance from the participants and had a width and height of 50°. The centre of the adjacent panels was placed at 50° eccentricity from the grid centre to encourage movement outside the typical eye rotation range. We measured completion time of the whole task, number of errors (incorrect selections), amount of head and gaze movement, and offset between head and eye.

Circular Task. Here, participants had to select targets across a circular layout in a predictable sequence based on the ISO 9241-9 standard [15] (Fig. 7B). The interface displayed eleven targets at 4 metre distance from the viewer. When all eleven targets had been selected, a new circle would appear. The target size was 4° in diameter, chosen to be large enough to minimise the effect of eye tracker error while avoiding target overlap. The diameter of the circle of targets was varied in four conditions with different pointing range (10°, 20°, 30°, 40°). Diameter sizes were chosen to have a mix of eccentricities where head movement would be unlikely (<20°) and likely (>20°), while avoiding that targets would move out of view and confound pointing with search. Participants performed five circles per condition (55 trials per condition), in a randomised order. We measured selection time, error rate, and the amount of head and gaze movement. Note this task was not included in analysis of head-eye alignment, as back-and-forth pointing discourages head following and would bias results.

Apparatus

We developed both tasks using Unity version 2017.4.3d1 and used an HTC Vive with the Tobii Pro VR Integration eye tracker (120Hz) for both studies.

POINTING USER STUDY

This study compared two pointing techniques, Eye&Head and Gaze, for pointing. In the baseline technique (Gaze), the pointer followed gaze continuously. For both technqiues, a cursor indicated the participant's pointing position and the HTC Vive hand-controller trigger was used to confirm selection. Eye&Head Pointing had a head rotational threshold of 15°/s and a translational threshold of 0.1m/s to activate the cursor, originating from informal testing.

Procedure

12 participants (5 female, 26.25 ±3.65 years) recruited from the local university participated in the study. Eleven participants had occasional VR experience, and one used VR daily. Ten participants had occasional or no eye tracking experience and two participants had daily experience. Participants first signed a consent form and answered a demographic questionnaire. Participants were then seated and put on the HMD. Participants started with the search task using both techniques before performing the circular task. Participants performed a five-point eye tracking calibration before each session. After calibration, participants had a training session before the test session. The pointing technique order was counterbalanced with a Latin square. After completing a task with a technique, participants removed the HMD and filled out a questionnaire consisting of eight 7-point Likert items based on common usability factors adopted from previous work [42]. A semi-structured interview was conducted after each completed task to extract preferences. The study took 30 minutes to complete.

Results

Search Task

Paired samples t-tests showed no significant time or error difference (Table 1). However, Eye&Head Pointing showed significantly higher head movement. Fig. 8 shows questionnaire responses. Wilcoxon signed-rank tests showed no significant differences between the pointers.

Table 1. Search task performance and paired-samples t-test results.

	Trial time (s)	Error (#)	Head motion (deg)
Gaze	196.79 (±52.37)	5.80 (±5.18)	1508 (±566)
Eye&Head	188.99 (±27.37)	5.30 (±4.05)	2651 (±678)
t-test	$t(11) = .601$, $p = .563$	$t(11) = .447$, $p = .668$	$t(11) = 4.401$, $p = .002$

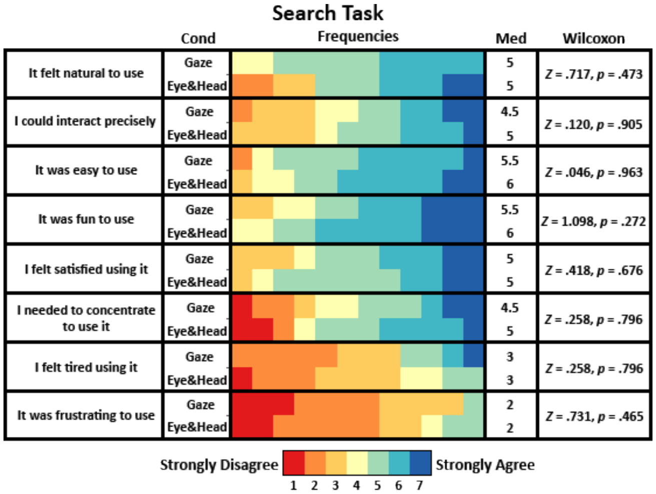


Figure 8. Search task questionnaire and Wilcoxon signed-rank test results.

Interview results showed that participants' opinions were split. Eight expressed favourable opinions towards Eye&Head Pointing. P5 claimed *"Eye&Head Pointing was much easier to use, and you did not have to focus as much since the pointer did not constantly follow your gaze"*. P2 commented *"Eye&Head Pointing is a nice way to get rid of all extra distractions and movements in the background"*. P2 also commented on the pointer's naturalness *"The head movement was already there, so for the majority of the selection I did not realise I was using the technique and that I needed an extra head movement"*. Finally, P9 noted *"Gaze pointing is more dependent on the eyes, and I felt like it was too responsive. I felt like I needed to concentrate on controlling it and it was distracting me"*. However, four participants preferred regular gaze pointing. Their main reason was the annoyance of being forced to use their head. P6 commented *"The Eye&Head Pointer was more of a challenge. It did not feel as natural as gaze pointing". The participants of this group were among the five participants with the lowest head movement.*

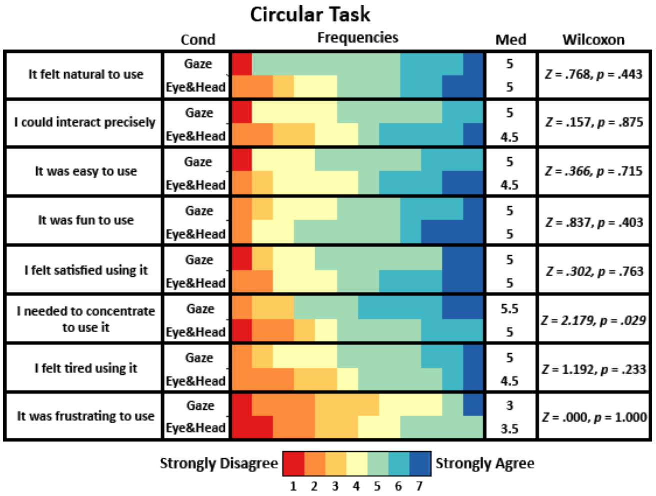
Circular Task

Analysis showed no significant differences in time, error rate or throughput. See Table 2 for comprehensive results. Eye&Head Pointing had significantly higher head movement at all distances but there was no significant difference in pointing performance. This was surprising for distances of 10° and 20° where targets are reachable eyes-only, and where we expected head motion to slow down pointing. However, efficiency of Eye&Head Pointing is explained by enabling users to still confirm a target while their gaze is already moving on.

Fig. 9 shows the questionnaire results. The Eye&Head Pointer required significantly less perceived concentration. As in the search task, eight participants preferred Eye&Head Pointing while four preferred gaze pointing. Participants often performed mistakes with gaze pointing where the gaze would move faster between targets than they could press the button, whereas the Eye&Head Pointer provided more control over the

Table 2. Circular task performance and paired-samples t-test results.

		Time (s)	Error (%)	Throughput (bit/s)	Head motion (deg)
10°	Gaze	.43 (±.07)	7.4 (±4.9)	3.55 (±.55)	1.28 (±1.07)
	Eye&Head	.48 (±.14)	8.7 (±7.5)	3.41 (±1.33)	10.35 (±1.89)
	t-test	$t(11) = 1.509$, $p = .165$	$t(11) = .417$, $p = .685$	$t(11) = .403$, $p = .695$	$t(11) = 11.961$, $p < .001$
20°	Gaze	.50 (±.12)	10.0 (±5.8)	4.70 (±1.79)	2.39 (±2.31)
	Eye&Head	.49 (±.12)	13.4 (±6.7)	4.30 (±1.38)	12.88 (±2.46)
	t-test	$t(11) = .197$, $p = .848$	$t(11) = 1.277$, $p = .230$	$t(11) = .981$, $p = .348$	$t(11) = 9.861$, $p < .001$
30°	Gaze	.55 (±.09)	17.3 (±7.2)	4.34 (±1.15)	5.55 (±4.67)
	Eye&Head	.53 (±.12)	13.3 (±7.4)	4.72 (±1.60)	16.50 (±3.11)
	t-test	$t(11) = .758$, $p = .468$	$t(11) = 1.312$, $p = .348$	$t(11) = .993$, $p = .342$	$t(11) = 4.988$, $p = .002$
40°	Gaze	.62 (±.10)	19.4 (±7.1)	4.24 (± 1.21)	10.85 (± 2.54)
	Eye&Head	.59 (±.12)	16.4 (± 7.3)	4.51 (± 1.35)	21.85 (±3.88)
	t-test	$t(11) = .451$, $p = .451$	$t(11) = 1.264$, $p = .342$	$t(11) = .701$, $p = .498$	$t(11) = 3.546$, $p = .009$



Circular Task

	Cond	Frequencies	Med	Wilcoxon
It felt natural to use	Gaze		5	$Z = .768, p = .443$
	Eye&Head		5	
I could interact precisely	Gaze		5	$Z = .157, p = .875$
	Eye&Head		4.5	
It was easy to use	Gaze		5	$Z = .366, p = .715$
	Eye&Head		4.5	
It was fun to use	Gaze		5	$Z = .837, p = .403$
	Eye&Head		5	
I felt satisfied using it	Gaze		5	$Z = .302, p = .763$
	Eye&Head		5	
I needed to concentrate to use it	Gaze		5.5	$Z = 2.179, p = .029$
	Eye&Head		5	
I felt tired using it	Gaze		5	$Z = 1.192, p = .233$
	Eye&Head		4.5	
It was frustrating to use	Gaze		3	$Z = .000, p = 1.000$
	Eye&Head		3.5	

Figure 9. Circular task questionnaire and Wilcoxon signed-rank tests results.

selection. P1 expressed *"It was good that the pointer moved more discretely, it made it clearer to me what I was currently selecting"*. Participants also complained that the cursor was continuously following their gaze. P10 commented *"The cursor was distracting, and I felt I needed to concentrate more to select a target so that my gaze would not go somewhere else"*. However, four participants disliked the Eye&Head Pointer's required head motion, especially for shorter distances. P4 added *"I preferred gaze pointing for the shorter distances because the required head movement was annoying when I did not need it. However, it did not matter for the longer distances as I moved my head anyway"*.

Summary across Tasks

Performance results were consistent across tasks. Head modulation of the gaze pointer increases effort in terms of required head movement but this did not affect pointing speed and accuracy. Additionally, we found no significant differences between VR or eye tracking experiences. A participant majority (8 of 12) expressed favourable opinions towards Eye&Head Pointing as it gave them more control and fewer distractions. Results also indicated that Eye&Head Pointing required less concentration from participants. However, a participant subgroup that showed tendencies to rely less on the head did not favour Eye&Head Pointing due to annoyance or effort caused by the extra head motion.

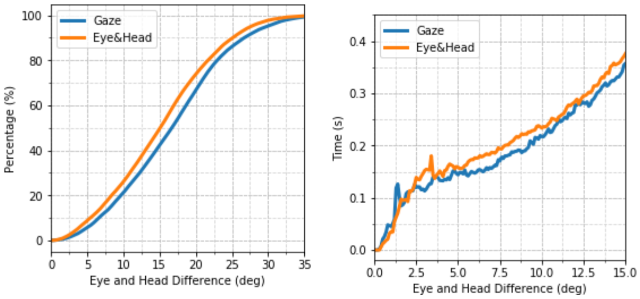


Figure 10. Left: Average percentage of search task spent within the angular difference between the eyes and head. Right: Average time spent within the angular difference between the eyes and head.

Head-Eye Alignment

Head-eye alignment was analysed based on the search task (Fig. 10). Unlike the circular task, the search task required users to inspect and compare images. This induced gaze shifts over different ranges which we deemed representative of natural gaze behaviour. We found that the offset between the head pointer and the gaze pointer was considerable for most of the time. On average participants would only spend 7.5% of the whole trial within 5° eye and head difference (Fig. 10, left). Closer inspection showed that a closer alignment within this angle generally only occurred when the eyes would move across the head. Instances where head and eye were within 3° angular proximity were of short duration, with average time at 0.11-0.15s, which is significantly shorter than the time required for a gaze fixation.

We found no significant difference between regular gaze and Eye&Head Pointing. Eye&Head Pointing led to 75% more head motion compared with unmodulated gaze, but this did not have any significant effect on head-eye alignment. The results support the proposed utility of head-eye alignment for signalling intent. This confirms the conceptual basis for the Eye&Head Convergence technique, and suggests a practical choice of angular threshold at 3° between head and eye.

CONFIRMATION USER STUDY

Our second study had the objective to evaluate the Eye&Head Dwell and Convergence techniques in comparison to regular gaze dwell. We used the same equipment, tasks and conditions as in the pointing study (Fig. 7). The participants performed the tasks with four techniques: Gaze + Dwell (G + D), Eye&Head Pointing + Eye&Head Dwell (EH + D), Gaze + Eye&Head Convergence (G + C), and Eye&Head Pointing + Eye&Head Convergence (EH + C).

We chose a dwell time of 700ms for both dwell techniques, designed to be proficiently usable by novice users and comparable to dwell times in prior similar work [5, 42, 46]. Other work has used dwell times as short as 300ms [10, 26] but such dwell times are for highly practised users and specific tasks [26]. Eye&Head Dwell had an angular threshold of 2° between the gaze point and cursor chosen via informal testing.

Table 3. Search task performance and repeated measures ANOVA.

	Trial time (s)	Error (#)	Head motion (deg)
G + D	173.77 (± 36.76)	9.25 (± 8.51)	1559 (± 705)
EH + D	182.61 (± 39.55)	5.92 (± 2.84)	2391 (± 846)
G + C	181.38 (± 29.93)	5.42 (± 2.27)	3213 (± 711)
EH + C	176.30 (± 32.37)	5.83 (± 3.90)	3382 (± 678)
ANOVA	$F(3, 33) = .454$, $p = .716$	$F(3, 33) = 1.863$, $p = .155$	$F(3, 33) = 28.300$, $p < .001$

Eye&Head Convergence parameters were decided via data collected from the pointing study (Fig. 10). We set the angular threshold to 3° as participants had spent less than 2-3% of the search task time within this close range of head and gaze alignment. Occurences of alignment within this range had only lasted 110-150ms on average, well below the dwell time.

Procedure

12 participants (3 female, 28.08 ±3.55 years) participated in the study. Eleven reported occasional previous experience with VR, and one reported daily to weekly VR use. Ten had occasional experience with eye tracking and two participants reported daily to weekly experience. Six had participated in the pointing study. The same procedure was used as in the first study. The study took 45 minutes to complete.

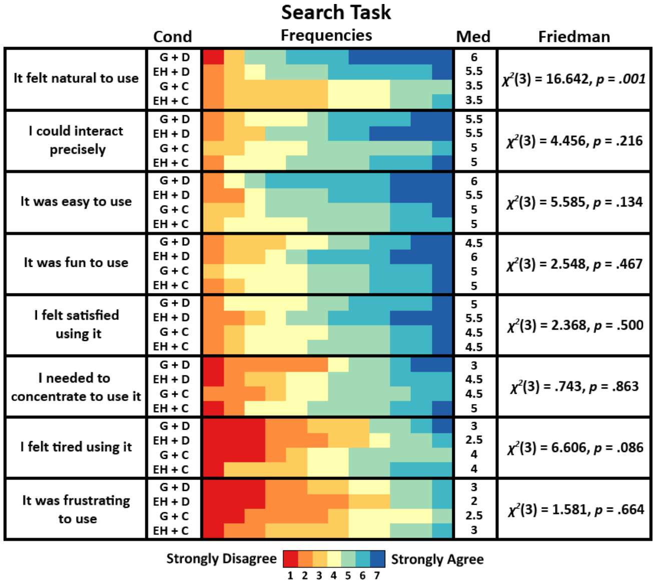
Results

Search Task

Repeated measures ANOVA showed no significant time or error rate differences (Table 3). However, G + D had a larger error count variance compared to the rest. Significant differences were found in head motion. Further Bonferroni corrected post-analysis showed that G + D also had significantly lower head motion compared EH + D ($p = .030$) and both Convergence combinations (both $p < .001$). Additionally, EH + D had a significantly lower head motion compared to both G + C ($p < .001$) and EH + C ($p = .008$). The head motion differences are not surprising as Eye&Head Convergence to some extent require head pointing, and Eye&Head Pointing also requires head movement to update the cursor position.

Fig. 11 shows the usability ratings from the search task. Friedman tests showed significant differences in naturalness. Further Bonferroni adjusted Wilcoxon analysis showed that G + D was significantly more natural than G + C ($z = 2.434$, $p = .015$) and EH + C ($z = 2.297$, $p = .022$). EH + D was also significantly more natural than G + C ($z = 2.383$, $p = .017$) and EH + C ($z = 2.683$, $p = .007$).

Participants' opinions about the techniques were again varied. Six participants preferred EH + D. P9 expressed *"EH + D was good because it did not feel as stressful as G + D and not as tiring as G + C and EH + C"*. P10 added *"EH + D was very useful as it allowed me to move more freely without making a selection compared to the other techniques"*. P11 also stated *EH + D gave me more control over my selections and it suited very well with the search task*. Three participants preferred EH + C. P4 stated *"EH + C felt more natural, and I think I am inclined to be more precise when also using my head. The dwell techniques were tiring because you had to stare at a target which felt unnatural"*. Three participants mentioned that the expanded cursor used in Eye&Head Convergence was distracting and added that the EH + C was less distracting than G + C due to its discrete nature. However, the remaining participants expressed no major difference between them. Just as in the first study, three participants who tended to use less head movement preferred G + D which required the least head movement. P12 stated *"Moving my gaze felt more natural and effortless compared to moving my head"*.



	Cond	Med	Friedman
It felt natural to use	G + D	6	$\chi^2(3) = 16.642$, $p = .001$
	EH + D	5.5	
	G + C	3.5	
	EH + C	3.5	
I could interact precisely	G + D	5.5	$\chi^2(3) = 4.456$, $p = .216$
	EH + D	5.5	
	G + C	5	
	EH + C	5	
It was easy to use	G + D	6	$\chi^2(3) = 5.585$, $p = .134$
	EH + D	5.5	
	G + C	5	
	EH + C	5	
It was fun to use	G + D	4.5	$\chi^2(3) = 2.548$, $p = .467$
	EH + D	6	
	G + C	5	
	EH + C	5	
I felt satisfied using it	G + D	5	$\chi^2(3) = 2.368$, $p = .500$
	EH + D	5.5	
	G + C	4.5	
	EH + C	4.5	
I needed to concentrate to use it	G + D	3	$\chi^2(3) = .743$, $p = .863$
	EH + D	4.5	
	G + C	4.5	
	EH + C	5	
I felt tired using it	G + D	3	$\chi^2(3) = 6.606$, $p = .086$
	EH + D	2.5	
	G + C	4	
	EH + C	4	
It was frustrating to use	G + D	3	$\chi^2(3) = 1.581$, $p = .664$
	EH + D	2	
	G + C	2.5	
	EH + C	3	

Figure 11. Questionnaire and Friedman test results for the search task.

Circular Task

We found significant circular task performance differences between the techniques (Table 4). Bonferroni corrected pairwise comparisons showed that G + C and EH + C compared to G + D and EH + D at all distances had significantly faster selection times (all $p < .001$, except EH + D and G + C at 40° ($p = .037$)), confirm times (all $p < .001$), and higher throughput (all $p < .001$ except EH + D and G + C at 30° ($p = .058$) and 40° ($p = .241$)). Head motion was significantly lower for G + D compared to all other techniques at all distances (all $p < .001$ except G + D and EH + D at 30° ($p = .026$), and 40° ($p = .345$)). Additionally, EH + D had a significantly lower head motion compared to both Convergence techniques at all distances (all $p < .001$).

In a post hoc analysis, we simulated shorter dwell times to investigate whether significant differences in selection and confirmation times were only due to the more conservative choice of dwell time. Results showed that both EH + C and G + C were significantly faster for all four distances also with a lower dwell time of 500ms; with dwell time chosen as low as 300ms, EH + C and G + C were still significantly faster for 10° distance, but not the larger distances. Note, that a larger Eye&Head Convergence angular threshold would lead to shorter confirm times and thus shorter times and higher throughput as less head motion would be required.

Table 4. Circular task performance and repeated measures ANOVA.

		Time (s)	Confirm time (s)	Throughput (bit/s)	Head motion (deg)
10°	G + D	.94 (±.05)	.70 (±.03)	2.75 (±.14)	2.06 (±2.05)
	EH + D	1.08 (±.03)	.73 (±.03)	2.34 (±.11)	10.80 (±1.38)
	G + C	.54 (±.10)	.24 (±.08)	4.95 (±.89)	13.96 (±1.57)
	EH + C	.52 (±.12)	.20 (±.06)	5.12 (±.97)	14.40 (±1.91)
	ANOVA	$F(3, 33) = 162.6$, $p < .001$	$F(3, 33) = 504.9$, $p < .001$	$F(3, 33) = 68.6$, $p < .001$	$F(3, 33) = 155.7$, $p < .001$
20°	G + D	1.00 (±.06)	.70 (±.02)	3.52 (±.18)	4.35 (±4.03)
	EH + D	1.09 (±.07)	.73 (±.06)	3.20 (±.22)	12.29 (±2.07)
	G + C	.74 (±.17)	.35 (±.11)	5.00 (±.80)	23.73 (±1.89)
	EH + C	.70 (±.15)	.33 (±.08)	5.28 (±.78)	24.87 (±2.03)
	ANOVA	$F(3, 33) = 26.6$, $p < .001$	$F(11) = 105.1$, $p < .001$	$F(3, 33) = 40.8$, $p < .001$	$F(3, 33) = 176.4$, $p < .001$
30°	G + D	1.14 (±.13)	.76 (±.07)	3.46 (±.54)	7.85 (±6.91)
	EH + D	1.05 (±.09)	.75 (±.08)	3.69 (±.55)	15.03 (±2.82)
	G + C	.89 (±.20)	.42 (±.13)	4.75 (±1.00)	32.65 (±1.46)
	EH + C	.84 (±.17)	.41 (±.16)	4.96 (±.88)	34.08 (±2.39)
	ANOVA	$F(3, 33) = 15.4$, $p < .001$	$F(3, 33) = 33.5$, $p < .001$	$F(3, 33) = 19.0$, $p < .001$	$F(3, 33) = 128.7$, $p < .001$
40°	G + D	1.33 (±.26)	.79 (±.10)	3.29 (± .52)	12.40 (± 8.39)
	EH + D	1.20 (±.28)	.77 (± .08)	3.84 (± .65)	19.36 (±5.05)
	G + C	1.02 (±.19)	.47 (± .13)	4.49 (± .80)	42.19 (±2.12)
	EH + C	.93 (±.18)	.42 (± .13)	4.93 (± .91)	43.46 (±2.47)
	ANOVA	$F(3, 33) = 21.0$, $p < .001$	$F(3, 33) = 41.5$, $p < .001$	$F(3, 33) = 23.8$, $p < .001$	$F(3, 33) = 101.7$, $p < .001$

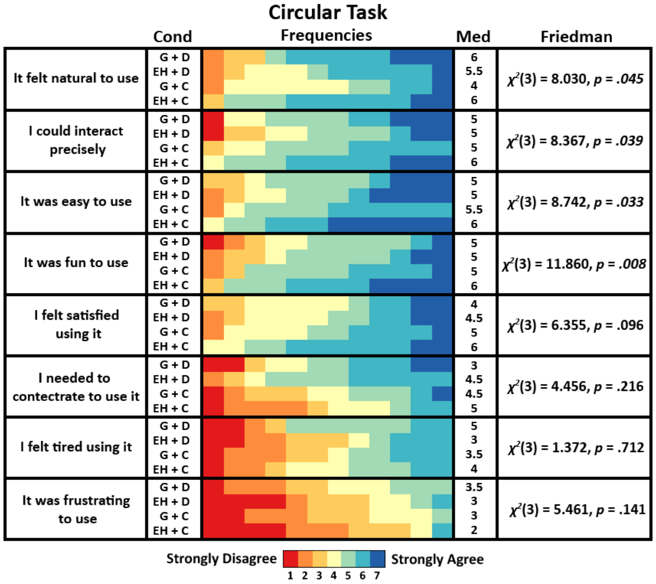


Figure 12. Questionnaire and Friedman test results for the circular task.

Friedman tests on usability ratings showed significant differences in naturalness, precision, easiness and enjoyment (Fig. 12). Bonferroni corrected Wilcoxon analysis showed that participants considered EH + C to be significantly more natural ($z = 2.919$, $p = .004$) and easier to use ($z = 2.972$, $p = .003$) than G + C . Participants also considered EH + C to be significantly more precise than G + C ($z = 2.714$, $p = .007$) and EH + D ($z = 2.200$, $p = .028$). Finally, EH + C was considered more fun than G + D ($z = 2.139$, $p = .032$).

Eight participants favoured the EH + C technique. P9 stated *"Eye&Head Convergence is easy and quick to select something when you know its position"*. P5 added *"I preferred EH + C for the circular task. It was more responsive than EH + D and G + D and less distracting than G + C"*. The chosen dwell time had a clear effect on the participants' responses. However, participants that expressed favourable opinions on Eye&Head Convergence thought that the chosen dwell time did not matter as the Eye&Head Convergence selection was instantaneous when reaching the angular threshold. One participant preferred EH + D for the circular task. Similarly to the search task, three participants preferred G + D. P12 stated *"It was hard and annoying to use my head all the time. I preferred the techniques where I could rely more on the eyes"*. Finally, P6 commented *"Convergence was really easy for close targets. But not for big movements, then I preferred selection by gaze"*.

Summary across Tasks

The type of task affected both performance and preference. No significant performance differences were observed when participants had to search for targets to select. However, when participants knew the target in advance they were significantly faster using Eye&Head Convergence for confirmation than with a dwell technique. In addition, we found no significant differences between VR or eye tracking experiences, nor new participants or participants who took part in both studies. Participants had differing opinions regarding their preferred technique, but mainly expressed favourable opinions for the combination of EH + D for the search task and the combination of EH + C for the circular task. Participants expressed preference for the Eye&Head Pointer as it provided more control and was found less distracting. As in the first study, a subgroup of the participants favoured regular gaze techniques due to annoyance or effort caused by the extra head movement needed with Eye&Head techniques.

DISCUSSION

At the core of Eye&Head interaction is the distinction between head-supported gaze and eyes-only gaze. Head movement requires more effort and energy than eye movement, and an attention shift supported by the head can be considered to represent a higher level of investment and interest. Based on the distinction, different behaviours can be attached to objects, depending on whether they are turned to by both head and eyes, looked at without head turn, or not gazed at. In our application examples, we have attached automated gaze behaviour only to the higher level of interest, to allow for exploratory attention to objects without side effect. However, other mappings are possible. In visual search, for example, all objects looked at could be marked as viewed, and head-supported attention could additionally trigger selection.

The three principles we proposed for Eye&Head interaction are validated by the application examples and study results. The first principle refers to stability of head-support gaze and is directly reflected in the design of the Eye&Head Pointer. The museum and puzzle applications show how the pointer facilitates stable gaze selection and feedback decoupled from individual fixations. Also, in both studies the majority of users found the Eye&Head pointer to provide more control and less distraction. The second principle is that eyes-only gaze affords exploration around objects selected by head-supported gaze. All our applications illustrate this, for example, with free exploration around an artefact of interest in the virtual museum, and examination of alternative choices after initial selection in the puzzle. The third principle is that head-eye alignment can be used as explicit input. In our first study, we showed that head and eye do not normally become completely aligned, a premise for using alignment as deliberate signal. We applied the principle in the Eye&Head Convergence technique, and the results of our second study show that the technique is robust and effective for fast confirmation of gaze targets.

A principal advantage of Eye&Head pointing is that objects can be pre-selected by gaze but that gaze is free to wander before the selection is finalised. This can be useful for many scenarios, for instance double-checking other conditions before finalising selection, or completing selection in sync with other events. It also avoids that the selection focus is lost prior to completion, for example caused by jitter in the eye movement, eye-tracking inaccuracy, or a visual distraction.

Questionnaire results from the pointing study showed no significance difference between pointers (Fig. 8 and Fig. 9). A possible explanation for these results could be that our study tasks were too simple for the participants to notice the interaction benefits from Eye&Head Pointing. The results could also mean that head modulation does not affect usability factors in a significant manner. However, a key insight from our studies is that gaze target acquisition is not slowed down by head modulation. This is significant as the demonstrated advantages of moving and updating the pointer only with head-supported gaze are gained without comprising performance. Eye&Head Pointing was designed with exploratory gaze applications in mind but our results show that it is also efficient for pointing at known targets in fast succession.

A specific benefit for fast pointing is that targets can still be confirmed when the eyes are already moving to the next object. This is compelling for applications where gaze is combined with a separate “click” modality such as a button or other manual trigger, as it requires less concentration from the user for timing of gaze shifts. The user’s eyes can move on when they are ready and do not need to wait until their hands have caught up. This matches natural eye-hand coordination, where the eyes lead manual action [60].

Eye&Head interaction can also be completely hands-free for which we introduced two novel “click” alternatives. Eye&Head Dwell extends the advantages of Eye&Head Pointing to dwell selection. A dwell-timer is only triggered when gaze is accompanied by head movement, providing more control over selection, and leaving the user free to pause selection to look at other objects. User can gaze at alternatives choices and consider them without time pressure, as the dwell-timer is only activated when both eyes and head move. This also addresses problems of gaze interaction with large objects where users require more time for visual inspection, a problem with regular dwell as it can result in unintended selection.

Eye&Head Convergence presents an alternative to dwell and employs alignment of the head pointer with the line of sight as the “click” alternative. The two techniques bring different strength to different applications. The convergence technique is faster and preferred when speed matters or targets are known, for example when selecting a tool from a menu and quickly returning to where the action is. Eye&Head Dwell, in contrast, is perceived as more natural when the tasks involves search and consideration of targets for gaze selection. However both techniques have in common that they support more stable selection, for example in crowded environments with selectable objects overlapping or in close proximity.

Eye&Head techniques can support users limited to eye and head movement for interface operation, as well as users leveraging gaze in conjunction with other modalities. We observed that a majority but not all users preferred Eye&Head over unmodulated gaze. Eye-head coordination literature suggests that there are “head-movers” versus “non-head-movers” [7], which our studies appear to confirm. Users preferring regular gaze disliked that our techniques required head movement, and on average moved their heads less than other participants. These results indicate the possibility of the Eye&Head techniques becoming burdensome over time due to the additional head movements. The techniques may also be less appropriate for users with physical disabilities or injuries that restrict head movement. Eye&Head pointing accommodates a reluctance to move the head as it requires only little head movement and does not depend on the extent to which the head supports a gaze shift. The sensitivity of our techniques could also be adapted for non-head-movers, for example lowering the rotational threshold at which head motion is detected, and increasing the selection radius in Eye&Head Convergence. Further longitudinal studies of the Eye&Head techniques’ and dynamic thresholds based on head movement tendency would thus be of interest.

All our results were obtained in VR. Studies have observed users moving their head more in VR than in comparable real-world tasks, caused by peripheral FOV limitations of HMDs [21, 39]. However, we do not expect this to limit applicability of our techniques, as they build on basic eye-head coordination behaviours that are consistent with observations in real-world tasks [24]. Synergetic eye and head movement is more prevalent when interactions span a wider FOV, for instance on large displays or across devices in smart rooms, but our techniques are also applicable with narrower FOV displays and deliberate head movement.

CONCLUSION

This work introduced Eye&Head gaze interaction with design principles and techniques that we validated in application prototypes and user studies. Our main conclusions for gaze interaction design are: (1) It proves useful to distinguish between head-supported and eyes-only gaze; (2) Modulation of a gaze pointer and/or dwell timer by head motion provides users with more stable feedback, better control, and freedom to roam with eyes-only gaze, without compromising pointing efficiency; (3) Eye-head convergence is viable as signal of intent, and enables fast hands-free target confirmation.

REFERENCES

[1] Rawan Alghofaili, Yasuhito Sawahata, Haikun Huang, Hsueh-Cheng Wang, Takaaki Shiratori, and Lap-Fai Yu. 2019a. Lost in Style: Gaze-driven Adaptive Aid for VR Navigation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI ’19).*

ACM, New York, NY, USA, Article 348, 12 pages. DOI: http://dx.doi.org/10.1145/3290605.3300578

[2] Rawan Alghofaili, Michael Solah, Haikun Huang, Yasuhito Sawahata, Marc Pomplun, and Lap-Fai Yu. 2019b. Optimizing Visual Element Placement via Visual Attention Analysis. In *Proceedings of the 26th IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR 2019)*.

[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 (Oct. 2003), 280–290. DOI: http://dx.doi.org/10.1007/s10209-003-0053-y

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

[5] 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. DOI: http://dx.doi.org/10.1145/3206343.3206349

[6] Edward G. Freedman. 2008. Coordination of the Eyes and Head During Visual Orienting. *Experimental Brain Research* 190, 4 (Oct. 2008), 369–387. DOI: http://dx.doi.org/10.1007/s00221-008-1504-8

[7] James H. Fuller. 1992. Head Movement Propensity. *Experimental Brain Research* 92, 1 (Dec. 1992), 152–164. DOI:http://dx.doi.org/10.1007/BF00230391

[8] J. Gunderson and G. Vanderheiden. 1984. Direct optical head-pointing input to standard computers using a TV screen based keyboard. In *Discovery '83: Computers for the Disabled*. Stout Vocational Rehabilitation Institute, University of Wisconsin-Stout, Menomonie, WI, USA, 85–89.

[9] John Paulin Hansen, Anders Sewerin Johansen, Dan Witzner Hansen, Kenji Itoh, and Satoru Mashino. 2003. Command Without a Click: Dwell Time Typing by Mouse and Gaze Selections. In *Human-Computer Interaction - INTERACT 2003*. IOS Press, Amsterdam, The Netherlands, 121–128.

[10] John Paulin Hansen, Vijay Rajanna, I. Scott MacKenzie, and Per Bækgaard. 2018. A Fitts' Law Study of Click and Dwell Interaction by Gaze, Head and Mouse with a Head-mounted Display. In *Proceedings of the Workshop on Communication by Gaze Interaction (COGAIN '18)*. ACM, New York, NY, USA, Article 7, 5 pages. DOI: http://dx.doi.org/10.1145/3206343.3206344

[11] Craig Hennessey, Borna Noureddin, and Peter Lawrence. 2006. A Single Camera Eye-gaze Tracking System with Free Head Motion. In *Proceedings of the 2006 Symposium on Eye Tracking Research & Applications (ETRA '06)*. ACM, New York, NY, USA, 87–94. DOI: http://dx.doi.org/10.1145/1117309.1117349

[12] Ken Hinckley, Randy Pausch, John C. Goble, and Neal F. Kassell. 1994. A Survey of Design Issues in Spatial Input. In *Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology (UIST '94)*. ACM, New York, NY, USA, 213–222. DOI: http://dx.doi.org/10.1145/192426.192501

[13] Thomas E. Hutchinson, K. Preston White, Worthy N. Martin, Kelly C. Reichert, and Lisa A. Frey. 1989. Human-Computer Interaction Using Eye-Gaze Input. *IEEE Transactions on Systems, Man, and Cybernetics* 19, 6 (Nov. 1989), 1527–1534. DOI: http://dx.doi.org/10.1109/21.44068

[14] 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)*. ACM, New York, NY, USA, 229–232. DOI: http://dx.doi.org/10.1145/2168556.2168602

[15] ISO 9241-9 2000. *Ergonomic requirements for office work with visual display terminals (VDTs) – Part 9: Requirements for non-keyboard input devices*. Standard. International Organization for Standardization.

[16] Robert J. K. Jacob. 1990. What You Look at is What You Get: Eye Movement-based Interaction Techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '90)*. ACM, New York, NY, USA, 11–18. DOI: http://dx.doi.org/10.1145/97243.97246

[17] Robert J. K. Jacob. 1993. Eye movement-based human-computer interaction techniques: Toward non-command interfaces. *Advances in human-computer interaction* 4 (1993), 151–190.

[18] Shahram Jalaliniya, Mardanbegi, Thomas Pederson, and Dan Witzner Hansen. 2014. Head and Eye Movement as Pointing Modalities for Eyewear Computers. In *2014 11th International Conference on Wearable and Implantable Body Sensor Networks Workshops*. IEEE, 50–53. DOI: http://dx.doi.org/10.1109/BSN.Workshops.2014.14

[19] 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. DOI: http://dx.doi.org/10.1145/2802083.2802094

[20] Mohamed Khamis, Carl Oechsner, Florian Alt, and Andreas Bulling. 2018. VRpursuits: Interaction in Virtual Reality Using Smooth Pursuit Eye Movements. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces (AVI '18)*. ACM, New York, NY, USA, Article 18, 8 pages. DOI: http://dx.doi.org/10.1145/3206505.3206522

[21] 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. DOI: http://dx.doi.org/10.1145/1743666.1743696

[22] 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. DOI:http://dx.doi.org/10.1145/2769493.2769550

[23] 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. DOI:http://dx.doi.org/10.1145/3173574.3173655

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

[25] 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. DOI: http://dx.doi.org/10.1145/2807442.2807461

[26] Päivi Majaranta, Ulla-Kaija Ahola, and Oleg Špakov. 2009. Fast Gaze Typing with an Adjustable Dwell Time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 357–360. DOI: http://dx.doi.org/10.1145/1518701.1518758

[27] Diako Mardanbegi, Dan Witzner Hansen, and Thomas Pederson. 2012. Eye-based Head Gestures. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12)*. ACM, New York, NY, USA, 139–146. DOI: http://dx.doi.org/10.1145/2168556.2168578

[28] 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. DOI: http://dx.doi.org/10.1145/3290605.3300842

[29] Diako Mardanbegi, Ken Pfeuffer, Alexander Perzl, Benedikt Mayer, Shahram Jalaliniya, and Hans Gellersen. 2019b. EyeSeeThrough: Unifying Tool Selection and Application in Virtual Environments. In *Proceedings of the 26th IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR 2019)*.

[30] Mark R Mine. 1995. *Virtual environment interaction techniques*. Technical Report. UNC Chapel Hill CS Dept.

[31] Pallavi Mohan, Wooi Boon Goh, Chi-Wing Fu, and Sai-Kit Yeung. 2018. DualGaze: Addressing the Midas Touch Problem in Gaze Mediated VR Interaction. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 79–84. DOI: http://dx.doi.org/10.1109/ISMAR-Adjunct.2018.00039

[32] Emilie Møllenbach, Martin Lillholm, Alastair Gail, and John Paulin Hansen. 2010. Single Gaze Gestures. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (ETRA '10)*. ACM, New York, NY, USA, 177–180. DOI: http://dx.doi.org/10.1145/1743666.1743710

[33] Carlos H. Morimoto and Marcio R. M. Mimica. 2005. Eye Gaze Tracking Techniques for Interactive Applications. *Computer Vision and Image Understanding* 98, 1 (April 2005), 4–24. DOI: http://dx.doi.org/10.1016/j.cviu.2004.07.010

[34] Omar Mubin, Tatiana Lashina, and Evert van Loenen. 2009. How Not to Become a Buffoon in Front of a Shop Window: A Solution Allowing Natural Head Movement for Interaction with a Public Display. In *Human-Computer Interaction – INTERACT 2009*. Springer Berlin Heidelberg, Berlin, Heidelberg, 250–263. DOI: http://dx.doi.org/10.1007/978-3-642-03658-3_32

[35] Yasuto Nakanishi, Takashi Fujii, Kotaro Kiatjima, Yoichi Sato, and Hideki Koike. 2002. Vision-Based Face Tracking System for Large Displays. In *UbiComp 2002: Ubiquitous Computing*. Springer Berlin Heidelberg, Berlin, Heidelberg, 152–159. DOI: http://dx.doi.org/10.1007/3-540-45809-3_11

[36] 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. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. ACM, New York, NY, USA, Article 43, 8 pages. DOI: http://dx.doi.org/10.1145/2971485.2971490

[37] Brian S. Oommen, Ryan M. Smith, and John S. Stahl. 2004. The Influence of Future Gaze Orientation upon Eye-head Coupling during Saccades. *Experimental Brain Research* 155, 1 (March 2004), 9–18. DOI: http://dx.doi.org/10.1007/s00221-003-1694-z

[38] Anjul Patney, Marco Salvi, Joohwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. 2016. Towards Foveated Rendering for Gaze-tracked Virtual Reality. *ACM Trans. Graph.* 35, 6, Article 179 (Nov. 2016), 12 pages. `DOI:` **`http://dx.doi.org/10.1145/2980179.2980246`**

[39] 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. `DOI:` **`http://dx.doi.org/10.1145/3225153.3225157`**

[40] 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. `DOI:` **`http://dx.doi.org/10.1145/2807442.2807460`**

[41] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze + Pinch Interaction in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction (SUI '17)*. ACM, New York, NY, USA, 99–108. `DOI:` **`http://dx.doi.org/10.1145/3131277.3132180`**

[42] Thammathip Piumsomboon, Gun Lee, Robert W. Lindeman, and Mark Billinghurst. 2017. Exploring Natural Eye-gaze-based Interaction for Immersive Virtual Reality. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 36–39. `DOI:` **`http://dx.doi.org/10.1109/3DUI.2017.7893315`**

[43] 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. `DOI:` **`http://dx.doi.org/10.1145/1743666.1743741`**

[44] 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. `DOI:`**`http://dx.doi.org/10.1145/3131277.3132182`**

[45] Robert G. Radwin, Gregg C. Vanderheiden, and Mei-Li Lin. 1990. A Method for Evaluating Head-Controlled Computer Input Devices Using Fitts' Law. *Human Factors* 32, 4 (Aug. 1990), 423–438. `DOI:` **`http://dx.doi.org/10.1177/001872089003200405`**

[46] Vijay Rajanna and John Paulin Hansen. 2018. Gaze Typing in Virtual Reality: Impact of Keyboard Design, Selection Method, and Motion. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications (ETRA '18)*. ACM, New York, NY, USA, Article 15, 10 pages. `DOI:` **`http://dx.doi.org/10.1145/3204493.3204541`**

[47] Ludwig Sidenmark and Anders Lundström. 2019. Gaze Behaviour on Interacted Objects During Hand Interaction in Virtual Reality for Eye Tracking Calibration. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications (ETRA '19)*. ACM, New York, NY, USA, Article 6, 9 pages. `DOI:` **`http://dx.doi.org/10.1145/3314111.3319815`**

[48] Vincent Sitzmann, Ana Serrano, Amy Pavel, Maneesh Agrawala, Diego Gutierrez, Belen Masia, and Gordon Wetzstein. 2018. Saliency in VR: How Do People Explore Virtual Environments? *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (April 2018), 1633–1642. `DOI:` **`http://dx.doi.org/10.1109/TVCG.2018.2793599`**

[49] 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. `DOI:` **`http://dx.doi.org/10.1145/2578153.2578157`**

[50] 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)*. ACM, New York, NY, USA, 705–710. `DOI:`**`http://dx.doi.org/10.1145/2370216.2370369`**

[51] 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. `DOI:` **`http://dx.doi.org/10.1145/2207676.2208709`**

[52] Vildan Tanriverdi and Robert J. K. Jacob. 2000. Interacting with Eye Movements in Virtual Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '00)*. ACM, New York, NY, USA, 265–272. `DOI:` **`http://dx.doi.org/10.1145/332040.332443`**

[53] Jayson Turner, Andreas Bulling, Jason Alexander, and Hans Gellersen. 2014. Cross-device Gaze-supported Point-to-point Content Transfer. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '14)*. ACM, New York, NY, USA, 19–26. `DOI:` **`http://dx.doi.org/10.1145/2578153.2578155`**

[54] Jayson Turner, Andreas Bulling, and Hans Gellersen. 2012. Extending the Visual Field of a Head-mounted Eye Tracker for Pervasive Eye-based Interaction. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12)*. ACM, New York, NY, USA, 269–272. `DOI:` **`http://dx.doi.org/10.1145/2168556.2168613`**

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

[56] 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 (UbiComp '13)*. ACM, New York, NY, USA, 439–448. DOI: http://dx.doi.org/10.1145/2493432.2493477

[57] Colin Ware and Harutune H. Mikaelian. 1986. An Evaluation of an Eye Tracker As a Device for Computer Input2. *SIGCHI Bulletin* 17, SI (May 1986), 183–188. DOI:http://dx.doi.org/10.1145/30851.275627

[58] Markus Zank and Andreas Kunz. 2016. Eye tracking for locomotion prediction in redirected walking. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*. 49–58. DOI:http://dx.doi.org/10.1109/3DUI.2016.7460030

[59] Robert C. Zeleznik, Andrew S. Forsberg, and Jürgen P. Schulze. 2005. *Look-That-There: Exploiting Gaze in Virtual Reality Interactions*. Technical Report. Brown University.

[60] 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. DOI: http://dx.doi.org/10.1145/302979.303053

[61] Xuan Zhang and I. Scott MacKenzie. 2007. Evaluating Eye Tracking with ISO 9241 - Part 9. In *Proceedings of the 12th International Conference on Human-computer Interaction: Intelligent Multimodal Interaction Environments (HCI'07)*. Springer-Verlag, Berlin, Heidelberg, 779–788. DOI: http://dx.doi.org/10.1007/978-3-540-73110-8_85

[62] Zhiwei Zhu and Qiang Ji. 2007. Novel Eye Gaze Tracking Techniques Under Natural Head Movement. *IEEE Transactions on Biomedical Engineering* 54, 12 (Dec. 2007), 2246–2260. DOI: http://dx.doi.org/10.1109/TBME.2007.895750