Vergence Matching: Inferring Attention to Objects in 3D Environments for Gaze-Assisted Selection

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

Christopher Clarke
University of Bath
Bath, United Kingdom
cjc234@bath.ac.uk

Joshua Newn
Lancaster University
Lancaster, United Kingdom
j.newn@lancaster.ac.uk

Mathias N. Lystbæk
Aarhus University
Aarhus, Denmark
mathiasl@cs.au.dk

Ken Pfeuffer
Aarhus University
Aarhus, Denmark
ken@cs.au.dk

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

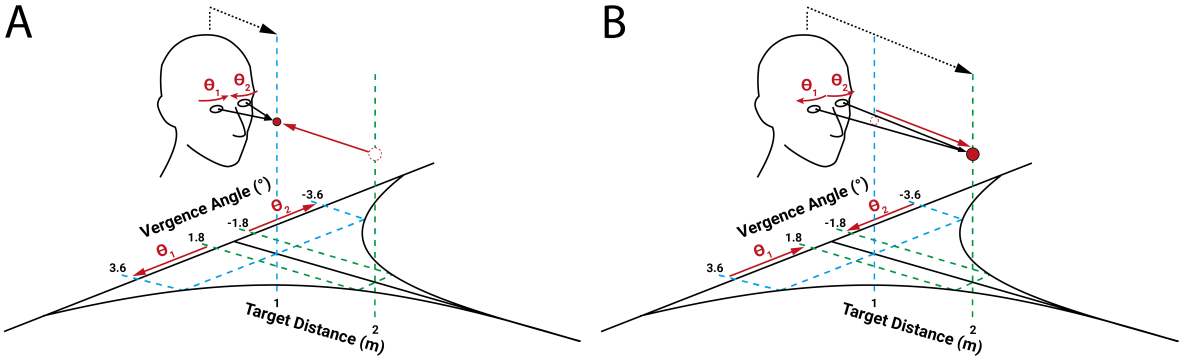


Figure 1: We present Vergence Matching, an interaction technique which uses the principle of motion correlation for selection of small targets in 3D environments. To select a target, smooth depth changes are induced perpendicular to the user: (a) when the target moves closer, the eyes move inwards increasing the vergence angle (convergence), (b) vice versa the vergence angle decreases (divergence) when the target moves away from the user. The relative vergence movement of the eyes are then correlated with the depth changes of the object to determine which target the user is attending to.

ABSTRACT

Gaze pointing is the de facto standard to infer attention and interact in 3D environments but is limited by motor and sensor limitations. To circumvent these limitations, we propose a vergence-based motion correlation method to detect visual attention toward very small targets. Smooth depth movements relative to the user are induced on 3D objects, which cause slow vergence eye movements when looked upon. Using the principle of motion correlation, the depth movements of the object and vergence eye movements are matched to determine which object the user is focussing on. In two user studies, we demonstrate how the technique can reliably infer gaze attention on very small targets, systematically explore how different stimulus motions affect attention detection, and show how the technique can be extended to multi-target selection. Finally, we provide example applications using the concept and design guidelines for small target and accuracy-independent attention detection in 3D environments.

CCS CONCEPTS

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

KEYWORDS

Selection; Attention Detection; Virtual Reality; Gaze, Vergence, Motion Correlation, Small Targets

ACM Reference Format:
Ludwig Sidenmark, Christopher Clarke, Joshua Newn, Mathias N. Lystbæk, Ken Pfeuffer, and Hans Gellersen. 2023. Vergence Matching: Inferring Attention to Objects in 3D Environments for Gaze-Assisted Selection. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3544548.3580685

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3580685

1 INTRODUCTION

When we know whether an object is in the focus of the user's attention then we can support the interaction in compelling ways, including implicit adaptation of the interface to the user's focus [9], object selection by gaze [21] and gaze-assisted manipulation and input [40, 41]. The conventional mechanism for detecting attention on objects in eye-tracked interfaces is to detect gaze fixations and match these spatially against the area an object covers in the visual field. However, this has inherent precision limitations as the eyes are not still during fixation and as fixations are detected as a spatial dispersion of gaze points [45]. The gaze points sampled in the process represent estimates prone to inaccuracy due to limitations in tracking, calibration and mapping of gaze to coordinates in display space. Therefore, inference of gaze attention to objects is usually limited to objects that are larger in size and spaced well apart.

We propose *Vergence Matching* as a novel technique for detecting attention to objects in 3D environments. Vergence Matching is independent of target size and we propose it to assist with selection of small objects that are difficult to attain by raycasting from hand, head, or eyes due to jitter or imprecise tracking. The core mechanism, inspired by Ahn et al.'s *Verge-It*, is to present stimuli that, when looked at, induce a vergence response of the eyes [2]. In our technique, we do this by moving any object of interest back and forth in the user's line of sight, and infer attention when we observe vergence movement of the eyes that corresponds with the motion of an object. When a user's focus of attention is on an object that is moving closer, the eyes will respond with converging movements to maintain binocular vision, and vice versa with diverging movements when the distance of the object from the eyes increases (see Figure 1). We scale objects in the process to remain constant in apparent size, to provide the user with a stable view of the relative position and spacing of objects. Users may notice that their eyes move but they do not need to pay any specific attention to the process as the eyes naturally adapt to changes in focal depth.

From the user's perspective, a system response is invoked when they maintain their gaze focus on an object until a vergence match is detected. This is not instantaneous, as the eyes need time to react and adapt to the object's motion, and the system needs to observe the eyes' response over time to evaluate its correspondence with object motion. The correspondence is determined by motion correlation [55], matching changes in vergence angle against changes in object distance from the eyes (Figure 1). Motion correlation is independent of object size as it is based on the correspondence of relative movements, as opposed to intersection of an input vector with an area in display space. The technique requires binocular tracking of eye movement but calibration is not required as the vergence angle is derived from movement of the eyes relative to each other. While an eye tracker might estimate gaze to be on an adjacent object, Vergence Matching will still match the object the user is gazing at, by its movement toward and away from the eyes.

In this work we present a study into the fundamentals of Vergence Matching, and the design and evaluation of two gaze-assisted selection techniques in which we use Vergence Matching to disambiguate pointing input for precise selection. Our first study was designed to test the feasibility of our concept and to establish that attention can be robustly inferred from vergence. As we use continuous motion to modulate the focal distance of objects, we specifically induce *slow vergence* for detection of attention, and we show that this is robust against false activation by *fast vergence* which is observed when users shift their attention between objects at different depths in the scene [10]. We further show that Vergence Matching can differentiate between up to four motions presented simultaneously with shifts in phase, which demonstrates feasibility of vergence-based selection from among multiple options. Our study also gives insight into design parameters, for example observing better performance of Vergence Matching with harmonic motion of objects back and forth than with linear motion.

We designed two techniques in which Vergence Matching is combined with raycasting to tackle the problem of selection ambiguity for small and closely spaced targets. In both techniques, raycasting is used for pre-selection of candidate targets, complemented by Vergence Matching to complete the selection. In *threshold-based* Vergence Matching, input is triggered as soon as a match of vergence with one of the candidates is detected, based on a pre-set motion correlation threshold. In *trigger-assisted* Vergence Matching users instead receive continuous feedback on the best-matching candidate and use a trigger to confirm the match. We evaluated the techniques using head cone casting as the modality for pre-selection of candidates, and comparing threshold- versus trigger-assisted selection from 2 or 4 candidates closest to the head ray. The results demonstrate the techniques' ability to select targets as small as 0.25° in width, and showcase Vergence Matching as a valid complement to conventional gaze-based interaction, with a trade-off in time for robust selection (3-4 seconds). In comparison, threshold-based Vergence Matching affords hands-free selection with only head and eye movement while trigger-assisted Vergence Matching affords the user with more control for faster and more accurate selection.

Vergence Matching provides a unique approach to infer visual attendance to objects of very small size in 3D environments, relying only on subtle manipulation of targets in the scene and the relative movement of the pupil positions. The technique provides new opportunities for gaze-assisted interaction that we illustrate with three applications in augmented and virtual reality (AR/VR). The applications, a wrist watch, context menu and notifications, highlight the benefit of being able to select small targets to reduce required display real estate and demonstrate the versatile combination of Vergence Matching with hand, head or gaze pointing to initiate selection. In sum, we provide the following contributions:

(1) Vergence Matching – a calibration- and size-independent technique for detecting visual attention on objects in XR environments. Vergence Matching leverages vergence eye movements and motion correlation by inducing motion on objects while maintaining their observed visual angle, thus minimising changes to the scene.
(2) Two selection techniques that use Vergence Matching for very small target selection, and three applications that show how Vergence Matching can provide unique advantages, allowing subtle and discreet gaze-based interfaces.
(3) The results of two user studies exploring the fundamentals of Vergence Matching and its use as a selection technique. The results show that Vergence Matching is viable for attention

detection without the risk of accidental detections of natural vergence eye movements, and Vergence Matching selection techniques can be used to select targets significantly smaller than the accuracy of the eye tracker.

2 RELATED WORK

For the design of Vergence Matching, we build on insights from pointing in 3D environments, stimuli-response eye movements, and prior work on vergence-based interaction.

2.1 Attention Detection and Selection in 3D Environments

Vergence Matching is designed to detect visual attention towards objects at any distance from the user in 3D environments. Typically, visual attention is derived from the user's gaze direction from which the system can react to [3, 4]. However, detecting attention on small targets or targets at a distance is limited by the natural jitter in eye fixations [66] and the accuracy and precision of eye trackers [15]. These limitations are present in fixation detection algorithms commonly used to infer visual attention on objects. For example, inherent noise in eye tracking leads to fixation detection algorithms with dispersion-based designs that have predefined large fixation areas (usually 1° in radius) [8, 45], or velocity-based algorithms where fixation areas can vary significantly in size [20, 37, 45].

In 3D interaction, the most commonly used metaphor for selection is *Ray-casting* where the user controls a ray via a controller [5, 35], or other body parts, such as gaze and head [1, 41, 43, 48, 51]. However, selecting small targets or targets at a large distance is also limited by users' motor skills [15, 36, 66]. HCI researchers have explored several approaches to improve interaction accuracy with multimodal techniques that enhance the pointing with a second user input [25, 49, 50, 52, 66], volume selection techniques where the pointing area is enlarged and multiple targets are disambiguated [7, 47, 65], or techniques that use heuristics or contextual information for selection [11, 17, 46, 53]. We use eye movements that respond to moving stimuli to select small and distant targets.

2.2 Stimuli-Response Eye Movements

Research has investigated interaction with eye movements that respond to external factors and can be performed as long as objects are visible. For example, smooth pursuit eye movements occur naturally when following moving stimuli. Contrasting saccades, smooth pursuits are characterised by the continuous and smooth motion of the eyes and has been exploited for interaction by inducing motion to targets and correlating the target and eye movement to identify which target the user is following [55, 59]. This technique enables the selection of small targets, for example, on a smartwatch [14], and has proven useful for general and occluded selection in VR [24, 47]. Similarly to Vergence Matching, smooth pursuit can be used for explicit interaction by a user [59] or implicitly by the system [42].

Furthermore, EyeGrip [23] uses optokinetic nystagmus eye movement, i.e., the sudden shift of attention when a moving object leaves the user's field of view (FOV). In their study, a sequence of scrolling images is shown, and EyeGrip detected the image of interest accurately through reactive movements. Additionally, vestibulo-ocular reflex (VOR) that occurs when the eyes fixate on an object but the head moves, leading to a reflex of eye movements to ensure stable vision, has been used for input [32] and has proven viable for target disambiguation with monocular eye trackers [31, 33]. The common main advantage of these methods is that they are not dependent on gaze calibration, can detect attention implicitly, and thus improve gaze interaction quality [42]. However, the necessity to make targets or the user move to induce these movements can be a challenge as most digital content is stationary. In VR and AR, proximity of multiple targets and small targets are key challenges for 3D interaction due to the varying relative positions of users and objects [13]. Using pursuit-based motion correlation (e.g. VRPursuits by Khamis et al. [24]) is difficult when multiple targets are near due to the risk of occlusion and target collision in 3D environments. In addition, pursuits have been used to select occluded targets in VR by Sidenmark et al. [47] in Outline Pursuits by inducing motions on target outlines. Target motions become too small for selection if the circumference of the target (i.e. size) is too small. Through Vergence Matching, we can detect the attention towards small objects while avoiding target collisions and minimising scene changes by moving objects relative to the user's perspective.

2.3 Vergence-based Interaction

Vergence eye movements are when the eyes move in opposite directions and naturally occur when users shift their visual focus between depths to maintain binocular vision [19]. Vision research usually classifies vergence movement into "fast" vergence that occurs in conjunction with saccades when switching focus between objects at different depths and "slow" vergence that is performed independently from saccades, usually when following a moving target which we leverage for interaction [10]. A key insight for our work is that these vergence movements are performed as long as users are able to focus on the target, meaning that targets can be smaller than the accuracy of pointing sensors.

Several works have investigated vergence-based interaction in binocular 3D environments to decide whether the user is looking at the world or an interface [28, 38, 59], or as X-ray vision to see occluded objects [18]. Yet, a limitation of these works is that without explicit 3D calibration, these systems can only distinguish 2 depths (i.e. near and far). We avoid this limitation by using relative vergence movements and can thus differentiate between multiple simultaneous objects. A closely related paper is Ahn et al.'s Vergeit [2], which provided a first feasibility assessment of modulated vergence movement by overlaying visual stimuli that moved horizontally to control objects in the scene, pointing to its potential for interaction. We build on and differentiate from this work by inducing back-and-forth movements on objects within the environment and using multiple concurrent motions to allow for a greater variety of applications and techniques. Further, we take a deeper look at vergence parameters and usability through two experiments.

3 VERGENCE MATCHING

Vergence Matching is defined by the following steps:

(1) A trigger from the user or system pre-selects a subset of candidate objects within the scene to minimise user distraction and maximise accurate detection;

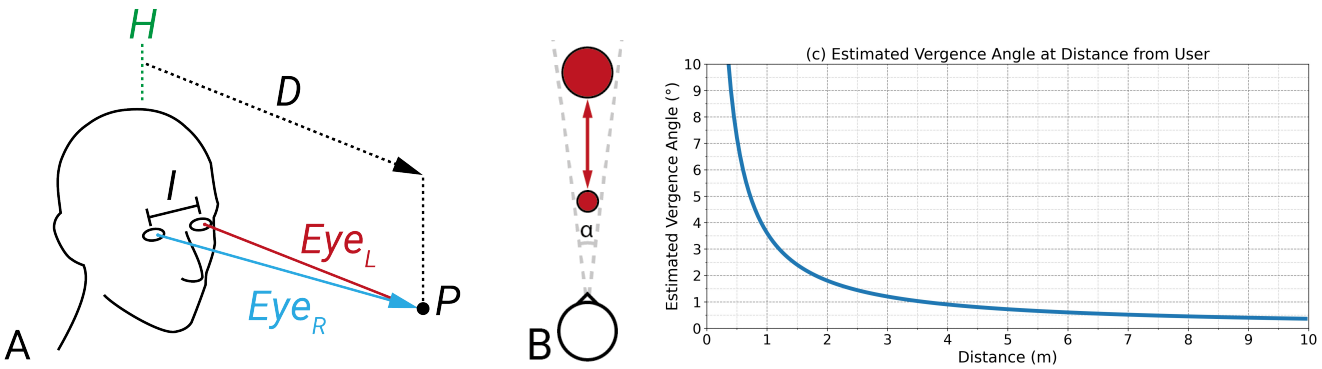


Figure 2: A: A user with interpupillary distance (I) focusing on a point (P) at a distance (D) from the centre of the head (H), with the two gaze vectors ($\vec{Eye}_L$ and $\vec{Eye}_R$) from the left eye and right eye respectively. B: The object is scaled when moving to retain the same angular width. C: Estimated vergence angle at distances from users based on Equation 1, showing how the vergence angle exponentially increases as the distance from the object decreases.

(2) Unique depth movements perpendicular to the user are induced on candidate objects;
(3) The user signals their intent by focussing on a candidate object with their gaze, and as a result, will intuitively perform vergence eye movements based on the depth movement induced on the objects. The system detects that the user is focussing on an object based on the correlation of the object's depth movement and the user's vergence movements.

These distinct steps can be broadly categorised into two distinct stages: (1) *candidate selection* where a subset of objects is selected and depth motions are induced, and (2) *inferring user attention* which involves correlating the object and user movements.

3.1 Candidate Selection

The first step of Vergence Matching is deciding candidate objects in the environment to induce movements that the user is likely to attend to. We incorporate this to minimise the likelihood that depth movements distract the user. In addition, we only consider movement in the depth axis for Vergence Matching, unlike smooth pursuit-based interaction techniques, which use movement in 2D space. This reduces the maximum number of unique trajectories and places greater importance on specifying potential objects of interest. Candidate selection can be performed by the system implicitly or explicitly by the user. System triggers can be started through models based on user context (e.g. location or proxemics). User triggers can be performed via pointing techniques (e.g. ray casting). A specific research question that we address in this work is how many candidate objects can be simultaneously displayed.

3.2 Motion Generation

For each candidate object, we induce a smooth motion on the depth axis perpendicular to the user. Each object exhibits a unique motion so the system can differentiate which object the user is focused on. Selecting appropriate values for these characteristics requires a deeper understanding of how slow vergence eye movements follow a moving object and has large implications for the success of Vergence Matching. Therefore, in Study 1 we systematically explore different values with a data-driven approach to understand the motion parameters that are best suited for vergence-based motion correlation. As vergence movements are one-dimensional, they are defined by the following characteristics:

Amplitude. The amplitude of vergence eye movements is not linearly correlated with the distance of the object from the user [60, 61]. As such, an object further away from the user must have a larger motion amplitude than a closer object to induce the same amount of vergence movement. In some scenarios, a designer may be able to control where the object movement starts and thus select the optimal distance, while in other scenarios movement may be induced on an object that is at a certain depth in the scene. To account for the nonlinearity and help downstream matching of movements, we translate object distances relative to the user into the expected vergence angle using Equation 1.

$$\alpha = arccos\left(\frac{\vec{Eye}_L \cdot \vec{Eye}_R}{||\vec{Eye}_L||||\vec{Eye}_R||}\right) \quad (1)$$

Where $\vec{Eye}_L$ and $\vec{Eye}_R$ are the directional gaze vectors of the left and right eye, respectively, towards a point P at a distance D (Figure 2a). To define the positions of the left and right eye, we assume an interpupillary distance of $I = 6.5cm$ based on [12]. The relationship defined by Equation 1 can be seen in Figure 2c.

Cycle Time. The object speed must be nonlinear to maintain a constant rate of change of vergence angle. As such, we define the speed of motion based on cycle time, i.e. the time it takes for an object to leave and return to its starting position. The duration of the cycle time should ensure that a full motion cycle can be performed in a reasonable amount of time for accurate detection while being robust against natural vergence behaviour. Also, the speed of the objects can be employed to distinguish object motions.

Phase. As the matching of object and eye movement is performed temporally, the phase in which the candidate objects move on the depth axis can be varied to induce unique movements. This has been leveraged in smooth pursuit eye movement interaction techniques

to distinguish between multiple concurrent motions and allow for the selection of one object among many [14, 47, 59]. The key to this is maximising the phase difference between the candidate object movements. The number of concurrent objects is then dependent on how well the eye movements can track the movement of the object for the system to be able to distinguish different trajectories.

Type of Movement. The type of depth movement induced on the candidate objects is also an important factor for both how well a user's eye movement tracks the movement of the candidate object and how well the system can match the two movements. Ideally, maximising the similarity of the object trajectory with the vergence movement of the eyes provides additional confidence and would allow the system to reject false positives and accidental activations.

Object Scaling. The required depth changes induced on the candidate objects must elicit enough change in the vergence angle for successful Vergence Matching. Changing the depth of objects, especially those in close proximity, can be problematic because occlusion may occur, and the changes in object size may affect the user's focal point. As a result, we chose to modify the size of the objects during the depth movement so that the object appears to be the same size (in visual degrees) to the user (Figure 2b).

3.3 Inferring User Attention

To infer user attention, the object movement for each candidate is correlated with the user's vergence eye movement. Similar to previous eye-based motion correlation techniques using smooth pursuits (e.g., [14, 42, 47, 59]), we use the Pearson correlation coefficient for the matching process. In our implementation, we use the pupil positions reported by the eye tracker for gaze input which, unlike the gaze position, does not require calibration. To translate the two pupil positions into a single value, $Dist_p$, we calculate pupil distance based on the horizontal pupil positions (Equation 2).

$$Dist_p = ((1 + x_R) - x_L) \times 0.5 \quad (2)$$

Where x_L and x_R are the pupil positions of the left and right eyes respectively. Note that the pupil distance will decrease with increased vergence angle resulting in a negative correlation (i.e., better matches are indicated by a negative correlation coefficient). For ease of understanding and for consistency with similar motion correlation work, we invert this relationship so that a positive correlation indicates a better match.

Vergence is a reactive process that we perform intuitively in reaction to an object moving in the depth axis. As a result, there is an inherent lag between the object and the user's vergence movements [63]. This lag must be taken into account to maximise the correlation and accuracy of Vergence Matching. Technical factors of the sensing and display equipment, such as discrepancies between the sampling of the HMD's display and the eye tracker, may also introduce additional lag in dynamic ways. To compensate, we select the best Pearson correlation coefficient based on a range of object delays between D_{min} and D_{max}. To ensure robust detection of attention, we use a sliding window of N_m frames to calculate the correlation and add the result to a sliding post-hoc buffer of length N_{phoc} as described by Velloso et al. [56]. Whether a detection is made depends on the criteria and the way in which Vergence Matching is implemented (e.g. correlation threshold or trigger).

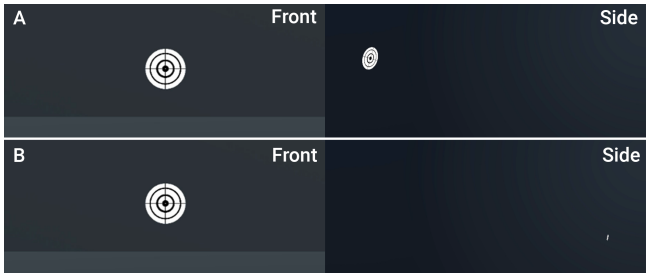


Figure 3: True positive task target. The target remains the visual angle size (A and B Front) but moves back and forth and scales to retain the visual angle size (A and B Side).

4 STUDYING VERGENCE MATCHING AND IDENTIFYING PARAMETERS

To validate that Vergence Matching can effectively and robustly detect attention, we first conducted a VR study to collect data on vergence eye movements while following moving targets. To direct our analysis, we define two main research questions of interest:

RQ1: Can we robustly detect vergence eye movements directed at single targets?

RQ2: Can we display multiple simultaneous targets without false detection?

We were also interested in understanding how motion generation parameters (i.e. motion speed and amplitude), and other Vergence Matching factors such as target scaling affect detection performance. To answer these questions and gain a deeper understanding of Vergence Matching, we collected data from two tasks. The first task studies how well users can perform slow vergence eye movements in an abstract task in which we vary the motion parameters of a 3D target (Figure 3). The resultant dataset provides insight into how well slow vergence eye movements can follow moving targets in VR using different motion parameters and, in turn, the optimal types of induced target movement to maximise correlation for interaction. We complement this with the second task, which collects data about vergence eye movements in a naturalistic environment, with which we can demonstrate the robustness of Vergence Matching to accidental activation due to "normal" gaze behaviour.

4.1 True Positive Task – Slow Vergence

For data collection of slow vergence eye movements, we presented participants with targets to focus on that oscillated back and forth while maintaining the same visual angle (Figure 3). We varied the motion parameters of the induced target movement to understand how well the user's eyes matched the target. We systematically explore all permutations of the following motion parameters:

Start distance from user (m): 1, 5, 10
Amplitude (°): 0.5, 1, 2
Cycle time (*s*): 1, 2, 4
Type of movement: Linear, Simple Harmonic

The targets appeared at the specified depth distance and were attached to the centre of the user's head direction. This ensures that we collect data where vergence occurs in the centre of the participant's FOV. The targets were then made to oscillate back and

forth relative to the participant at the specified amplitude, speed and type of movement. All targets started at the specified distance from the user and oscillated toward the user (i.e., the starting distance is the maximum distance of the target). All targets had a size that corresponds to 4° and were scaled as described in Section 3.1. The participants performed two blocks of data collection, one for each type of movement. Each target was shown for 5 seconds, with a 2-second pause between them. In total, each participant would gaze on 2 types of movement × 3 start distances × 3 amplitudes × 3 cycle times × 2 repetitions = 108 targets.

4.2 False Positive Task – Naturalistic Vergence

To gain naturalistic vergence eye movement data, participants were placed in a VR office environment (Figure 3b). We used the "Unity-JapanOffice" environment from the Unity Asset Store [1]. Participants were placed so that objects were at distances of 1 to 10 metres from the participants. The participants were then asked to perform search and comparative tasks (e.g., "Count the number of chairs." and "Are there more red or blue pillows?") to make the participants move their gaze between objects at different depths. Eye tracking data was recorded throughout the session.

4.3 Apparatus & Participants

We developed the study environment in Unity version 2017.4.3f1 and used an HTC Vive with an integrated Tobii Pro Eye Tracker (120Hz) to record eye movement in both tasks. The HTC Vive has a FOV of 100° in the horizontal plane, 110° in the vertical plane and a frame rate of 90Hz (correlation was therefore performed at 90Hz). We recruited 12 participants to take part in the data collection (6M/6F, 31.3±6.2). Three participants had normal vision, and nine wore glasses or corrective lenses. Ten participants had VR experience. Six participants have used an eye tracker previously.

4.4 Procedure

Upon arrival, participants completed a consent form and a short demographic questionnaire. They put on the HMD for a short testing session, where the researcher introduced the task and checked for any calibration issues. In the first part of the study, participants performed the true positive task in two sessions, one for each motion type (counterbalanced). After each condition, participants answered three 7-point Likert scales about their experience (ease, concentration, and strain) of focussing on the target. The questionnaire also included a field to enter comments freely. In the second part, participants performed the false positive task. Participants were calibrated to the HMD's eye tracking before beginning each task. Participants were allowed to take a break at any point during the study and were encouraged to take a break between each condition. The study took approximately 30–45 minutes.

4.5 Results

Our analysis explores the best target motion (motion type, amplitude, cycle time) and optimal parameters (target delay, threshold, window size) for attention detection while ensuring the system is robust to false positives. In addition, we explore how many unique targets could be displayed simultaneously to the user and distinguishable by the system. We first investigated whether we could accurately detect the user's vergence eye movements in response to the moving targets. Therefore, we systematically explore different target delays (0 to 500ms), correlation thresholds (c_t = 0.1 to 0.9 in 0.1 increments) and window sizes (N_m = 15, 30, 45, 90, 180 frames).

[1] https://assetstore.unity.com/packages/3d/environments/unityjapanoffice-152800

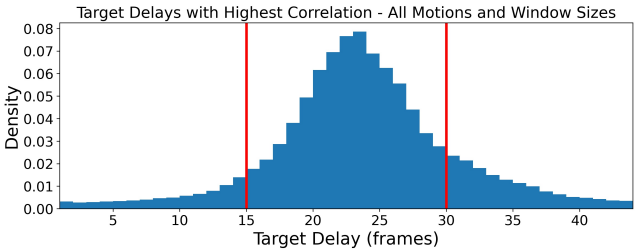


Figure 4: Target delay that resulted in the highest correlation for all window sizes. Red vertical lines represent D_{min} = 15 and D_{max} = 30 i.e. the range of delay accounted for.

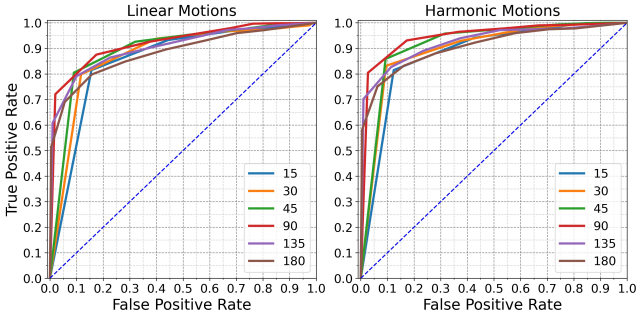


Figure 5: ROC curves for all linear and harmonic motions across different window sizes.

4.5.1 Target Delay. Since vergence movements are reactive, there is an inherent delay from when the target changes depth and the eyes converge or diverge to refocus. To understand this delay in the context of Vergence Matching, we counted the number of times each target delay between 0 to 45 frames (equivalent to 0 to 500ms) led to the highest correlation for each rolling window in all window sizes and trials. Figure 4 shows how the delay follows a normal distribution, with the peak occurring at around 23-24 frames (250-260ms). Existing literature reports that vergence movement usually starts with a latency of 160-180ms [63]. Taking into account the additional delay caused by eye tracking, we can see how our results align with this. As the delay is variable, we use a dynamic target delay for the remainder of the analysis, where D_{min} = 15 and D_{max} = 30 to calculate correlations.

4.5.2 Comparing Thresholds and Window Sizes. Next, we systematically investigate the true and false positive rates for each combination of threshold, window size, and motion parameters using the dynamic target delay range. We calculated the number of correct detections from the abstract task to find the true positive rate. We used a threshold-based approach to define successful detections.

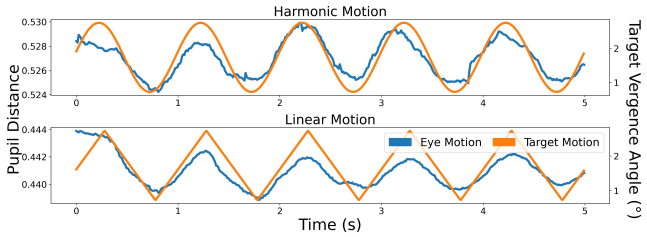


Figure 6: Example trials of a linear and harmonic trial of amplitude = 2°, cycle time = 1s and target start distance = 5m performed by the same participant. The phase of the target motion is adjusted to match the eye motion.

That is, a successful detection is reached when the post hoc moving window with $N_{phoc} = 45$ frames has reached a specified correlation threshold (c_t) for $N_{valid} = 30$ frames [47]. For the false-positive rate, we calculate the number of successful detections that are made using the false-positive task data and the same definition for detection. Since there are varying amounts of trials between true ($n = 54$) and false positive ($n = 1$) trials, we split the false positive task data into 54 overlapping 5-second sections. The corresponding target motion was simulated for each section. This is the equivalent of showing the target motion to the user during the false activation task to see if their eye movement would result in detection. We include target motions with all possible phase differences to ensure that any potential false positives are captured. A trial is classified as a false positive if a detection occurs. On average, we collected 122 seconds of false positive data per participant.

Figure 5 shows ROC curves for every window size for linear and harmonic motions separately. The results showed high true positive rates and low false positive rates. In addition, we found that harmonic motions lead to better performance than linear motions. Figure 6 illustrates why harmonic eye movement led to better performance. While participants appear to be able to follow linear and harmonic motions equally well, the change from convergence to divergence and vice versa appears to be more harmonic in nature. As a result, this leads to higher correlations for harmonic motions. Therefore, we focus the rest of our analysis on harmonic motions, and use a window size of 90 (N_m) which provided a good trade-off between true positive rate, false positive rate, and detection time.

4.5.3 Comparing Motion Characteristics. Next, we investigated which motion parameters are best suited for selecting with Vergence Matching. We performed the same analysis as in the previous section for each cycle time and motion amplitude combination. From our analysis, we saw that the target start distance did not show a difference and, as such, grouped motions with different start distances together. Figure 7 shows that faster motions with large amplitudes offer better performance. Presumably, this is because larger amplitudes are more likely to lead to more distinguishable vergence movements, in turn leading to higher true positive rates. Meanwhile, shorter cycle times caused more frequent shifts in the vergence direction, which may be less likely to occur in a natural environment, leading to lower false positive rates. These results show that motions with an amplitude of 2° and cycle time of 1s are most appropriate for target detection. Furthermore, we found that a

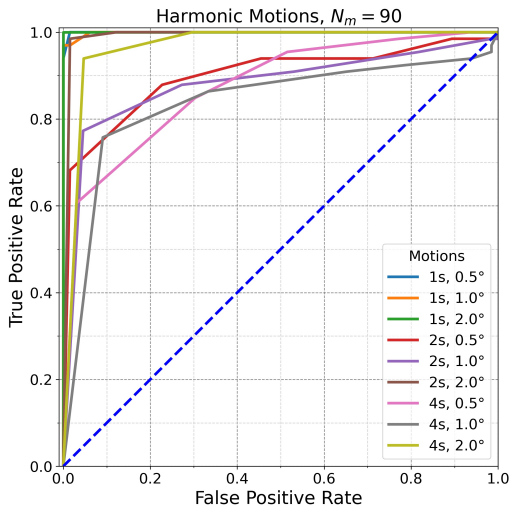


Figure 7: ROC curves for all linear and harmonic motions across different window sizes.

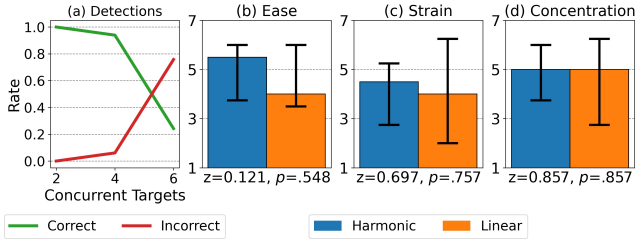


Figure 8: A: Detection performance based on number concurrent motions. B-D: Subjective responses and Wilcoxon signed-rank test results.

threshold (c_t) of 0.8 leads to a high true positive rate (1.0) and a low false positive rate (0.0). These results are significant because we can detect attention using Vergence Matching, yet do not trigger accidental detections caused by natural vergence movements (RQ1).

4.5.4 Number of Simultaneous Motions. Next, we were interested in how many concurrent motions we could display without leading to incorrect activations (RQ2). Based on our previous findings, we only considered motions with a cycle time of 1s and motion amplitude of 2°. We set $N_m = 90$ and $c_t = 0.8$ and used the same definition of detection as in previous sections. We incrementally increased the number of targets by generating motions of the same time and amplitude and setting them to the most distant phase relative to the followed target. We only consider the first detection. The trial is successful if a correct target is selected first and is labelled incorrect if the wrong target is selected first.

The results in Figure 8a imply that Vergence Matching can successfully differentiate the correct target from two or four simultaneous targets without significantly impacting detection performance. However, further targets result in deteriorating performance. These results are likely due to the 1-dimensional aspect of vergence movements where phase alone can only be used to distinguish motions

to a certain point before motions become too similar for accurate disambiguation. We also investigated if varying cycle times would positively affect performance but found that it had a detrimental impact on performance as it led to periods with significant motion overlap. We can surmise that Vergence Matching interaction should be limited to 4 concurrent motions (RQ2).

4.5.5 Subjective Responses. Finally, we analysed the questionnaire responses (Figure 8b-d) using Wilcoxon signed-rank tests to understand if users felt differently about the different target motions. We found no significant differences. Participants' comments showed that four participants found harmonic motions easier to follow (P1, P2, P4, P6), while only one preferred linear motions (P10). These results imply that there is little difference between motion functions perceptually, giving us the confidence to choose motion parameters based on correlation data. On the user experience of following moving targets, participants mentioned that they "did not do much, just looking at moving targets" (P4) and "sleepiness" (P8 and P11) due to the simplicity of the task. Participants also mentioned that following moving targets felt like a "reflex" and that if "I relax, my eyes follow the target automatically" (P1). Participants also mentioned that "faster movements were easier to focus on" (P4) but more "straining" (P5). Only one participant mentioned the target scaling and said that targets "appeared to create an optical illusion where the target appeared to be getting smaller or larger. This was more apparent when the animation speed was increased", but it was "still overall easy to focus on targets" (P2).

4.6 Summary

These results provide an understanding of slow vergence movements and demonstrate Vergence Matching as a viable attention detection technique. We demonstrate how Vergence Matching is very robust to accidental detection caused by natural gaze behaviour. Finally, we show that harmonic motions with high amplitude and low cycle are optimal for correlation and that subsequent explorations of concurrent motions should be limited to four simultaneous targets.

5 VERGENCE MATCHING SELECTION TECHNIQUES

Based on the results of study 1, we have shown that Vergence Matching can accurately detect the attention of users on targets through vergence movements (RQ1). We also found through post-hoc testing that multiple targets can be presented with high accuracy (RQ2). Based on these results, two Vergence Matching techniques, a threshold-based Vergence Matching technique which allows users to make hands-free selections, and a trigger-assisted Vergence Matching technique which uses a controller to provide the user with more control over the selection process.

5.1 Candidate Selection

Inspired by the *OutlinePursuits* interaction technique proposed by Sidenmark et al. [47], both techniques use cone-casting to decide which targets to select as candidates. This involves casting a ray into the scene and selecting the nearest N_c targets to the ray direction within a given radius of visual angle, r_c. The ray-casting can be performed by any type of pointing modality, including the head, a controller, or the gaze ray itself. The selection of pointing modality may be application-dependent or based on the inherent tracking accuracy and precision of the ray-casting. We show variants of Vergence Matching that use all these modalities.

5.2 Target Selection

5.2.1 Threshold-based Vergence Matching. The first technique variant is a threshold-based version of the Vergence Matching technique. From a user's perspective, this is comparable to a typical gaze fixation dwell technique in which the user must only focus on the target for selection. Users can select targets beyond their reach and their hands are free to perform other interactions. However, in contrast to gaze dwell, the proposed threshold-based Vergence Matching variant is able to select targets much smaller than the accuracy of an eye tracker. As such, the technique works as an alternative when targets are too small or when the eye tracker is too inaccurate for efficient selection.

For cone-casting, we use the head direction to ensure stable pre-selection and to make sure that small targets can be pre-selected, which may be difficult if relying on an inaccurate gaze pointer. For selection, if a given number of correlation values, N_{valid}, in the post-hoc buffer are above a given correlation threshold, c_t, we assume the motions are matched, and the corresponding target is automatically selected without further user input. The key to this technique is to ensure that the parameters selected for N_{valid} and c_t allow selection in a timely manner while being robust against false activations that may occur due to natural vergence eye movement. Therefore, we collect data to optimise these parameters in Study 1.

5.2.2 Trigger-assisted Vergence Matching. Inspired by Controller-based Outline Pursuits [47], we developed a second version of Vergence Matching to demonstrate how a trigger can be used to provide selection. Unlike threshold-based Vergence Matching, trigger-assisted Vergence Matching provides the user with greater control over the selection process. We use head-pointing for cone-casting again based on its stability and to provide a fair comparison across the techniques for Study 2. The candidate target with the highest mean correlation value within the post-hoc window is highlighted as the current selection candidate, and selection is confirmed when the user activates a simple trigger (in our case, we use a button press on a controller). This design avoids issues caused by users having to reach a specific threshold for selection which may result in no targets being selected, and also helps to reduce the Midas touch problem [21] as the user needs to confirm the selection explicitly.

6 SELECTION TECHNIQUE EVALUATION

In study 2, we investigate the user performance and perception of the proposed Vergence Matching techniques. In particular, we focus on Vergence Matching's ability to select very small targets (Figure 9). Due to the uniqueness of Vergence Matching that induces object movements relative to the user to minimise changes in the scene, we focus on a deeper exploration of how the technique works and factors such as the number of concurrent targets, target depth from the user, and target size. As a result, our study compares the threshold-based and trigger-assisted technique variants.

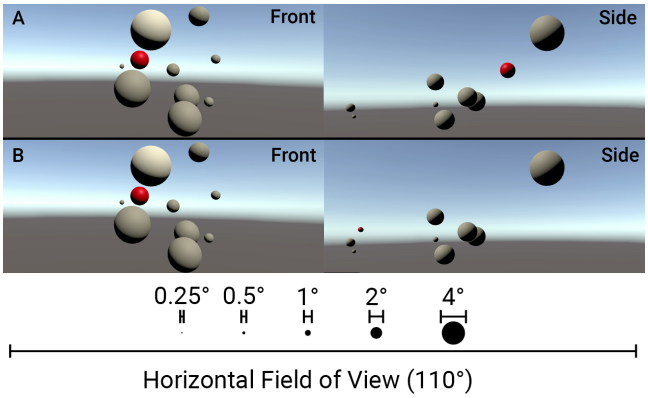


Figure 9: Top: Study 2 example trial showing the target layout. A and B show two candidate targets in motion. Bottom: Target widths in relation to the horizontal FOV.

6.1 Task

The task was based on other work that explored the selection of one target from many in VR [30, 39, 54]. Ten spherical targets were presented in random positions within a cone-based layout of 50 degrees (Figure 9) within a specified depth range. The targets were placed to ensure that no occlusion occurred from the participant's perspective. Participants were tasked with selecting a specific target (highlighted in red) as quickly and accurately as possible. The specified target had a specific target width, while distractor targets varied in widths (1-10° in diameter). Distractor targets were placed, so at least 3 targets were within 2° of the main target. Participants could not proceed to the next trial until the correct object had been selected or 10 seconds had elapsed. For each trial, the target width and target depth range were randomised. The independent variables of the study were:

Technique: Threshold-based with 2 candidates (TH2), Threshold-based with 4 candidates (TH4), Trigger-assisted with 2 candidates (TR2) and Trigger-assisted with 4 candidates (TR4).
Target width: 0.25, 0.5, 1, 2, 4° in diameter (Figure 9).
Target depth: Near (0.2-1m), Middle (1-5m), Far (10-30m).

6.2 Apparatus & Participants

We used the same equipment as in study 1. We were able to record the data at a mean gaze accuracy of 1.09 ± 0.26° and a mean gaze precision of 0.37 ± 0.16°. Based on the results of study 1 and internal pilot testing, we selected the following correlation parameters: $r_c = 10°$, $N_m = 90$, $N_{phoc} = 45$, $N_{valid} = 30$, $c_t = 0.8$. All target motions were harmonic, had an amplitude of 2°, and a cycle time of 1 second based on Study 1 results. We used the HTC Vive controller touchpad for the trigger-assisted technique. We recruited 12 participants for the study (7M/5F, 27.3±4.1). Six participants had normal vision, and six wore glasses or corrective lenses. Nine participants had VR experience. Six participants reported previous eye tracker experience. None participated in the first study.

6.3 Procedure

Participants were seated, signed a consent form, and answered a demographic questionnaire. The participants were then instructed to put on the HMD and perform an eye tracking calibration. Participants then performed a second calibration procedure to record eye tracking accuracy and precision using the GazeMetrics toolkit [6] by fixating on 9 calibration points in a square arrangement. Afterwards, the participants performed a training session with the designated technique (counterbalanced) before the test session. Participants performed 5 repetitions of each trial condition. After completing the task, the participants removed the HMD and completed a questionnaire consisting of 12 7-point Likert items based on usability factors from previous work [47] before moving on to the next technique. A semi-structured interview was conducted to extract technique rankings and opinions. In total, each participant performed 4 techniques × 5 target widths × 3 target depths × 3 repetitions = 180 selections. The study took 60 minutes to complete.

6.4 Results

The dependent variables of interest were trial completion time, pre-selection time, selection time, error rate, and perceived usability. Unless otherwise stated, the analysis was performed with a three-way repeated measures ANOVA (α=.05) with Technique, Target width, and Target depth as independent variables. When the assumption of sphericity was violated, as tested with the Mauchly test, Greenhouse-Geisser corrected values were used in the analysis. QQ-plots were used to validate the assumption of normality. Bonferroni corrected post-hoc tests were used when applicable. Effect sizes are reported as partial eta squared (η_p^2). Likert scale data and rankings were analysed using Friedman tests and Bonferroni corrected Wilcoxon signed-rank tests for post-hoc analysis.

6.4.1 Completion Time. We define the completion time as the time from trial start until correct selection (Figure 10a). We found no significant interactions. We found significant main effects for all independent variables. Tests on technique ($F_{3,33}$=4.79, p=.007, η_p^2=.304) showed that TH4 was significantly slower than TR2 (p<.001). For target depth ($F_{2,22}$=6.12, p=.008, η_p^2=.358), we found that completion times were significantly longer for the Near condition compared to Middle (p=.013). Finally, for target width ($F_{1.91,21.00}$=12.42, p<.001, η_p^2=.530), we found that participants were slower to complete trials with target widths of 0.25° and 0.5° than 2° and 4° (all $p \leq .037$).

6.4.2 Pre-selection Time. To investigate the impact of search and pre-selection on performance, we investigated the time taken from trial start until the correct target was pre-selected (Figure 10b). We found no interactions, but again found significant main effects for all independent variables. Post-hoc tests on techniques ($F_{3,33}$=37.35, p<.001, η_p^2=.773) showed that TR2 and TH2 were significantly slower than TR4 and TH4 (all p<.001), implying that pre-selection takes longer with fewer concurrent candidates. This effect is expected as less accuracy is required when more concurrent candidates can be pre-selected. For target depth ($F_{2,22}$=6.12, p=.002, η_p^2=.443), targets at near and middle conditions led to longer pre-selection times than far ($p \leq .034$). Finally, the width of the target also had a significant effect ($F_{2.03,22.32}$=9.98, p<.001, η_p^2=.476). Unsurprisingly, larger targets led to shorter pre-selection times. We

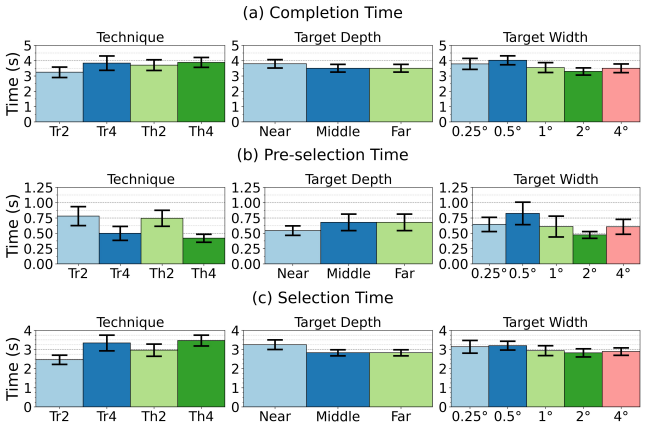


Figure 10: Completion, pre-selection, and selection times. Error bars represent the mean 95% confidence interval.

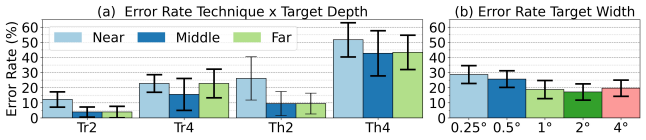


Figure 11: Error Rate. Error bars represent the mean 95% confidence interval.

found that 0.25° and 0.5° had higher pre-selection times than 1°, 2°, and 4° target widths (all $p \leq .025$).

6.4.3 Selection Time. We define the selection time as the time from the target was chosen as a candidate to successful selection (Figure 10c). We found no interactions but found main effects for technique ($F_{1.58,17.41}$=12.89, p<.001, η_p^2=.539), target depth ($F_{2,22}$=13.70, p<.001, η_p^2=.555) and target width ($F_{2.36,25.95}$=4.17, p=.022, η_p^2=.530). Post-hoc tests showed that vergence-based selection with Th4 was significantly slower than Tr2 (p<.001). The results also showed that participants were slower in selecting targets the near distance compared to middle (p=.013). Although far targets were also selected faster, no significance was found. Finally, post-hoc results of target width found no significant differences. These results imply that the target selection time is independent of target size.

6.4.4 Error Rate. We define an erroneous trial as when another target is selected before the correct selection or if a trial is timed out. The number of errors violated the assumption of normality of repeated measures ANOVA after usual transformations, and the Align Rank Transform technique [62] showed that the aligned responses did not sum up to ≈ 0. Using the number of errors as count data, we fit a Poisson regression model [34]. 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.

We included all main effects and all interactions that involved technique in the regression and found that the overall model was significant, $\chi^2(23, N=720)=281.67$, p<.001. Investigation revealed significant two-way interactions for technique × target width ($\chi^2(6)$= 9.56, p=.014), and technique × target depth ($\chi^2(6)$=15.89, p=.023).

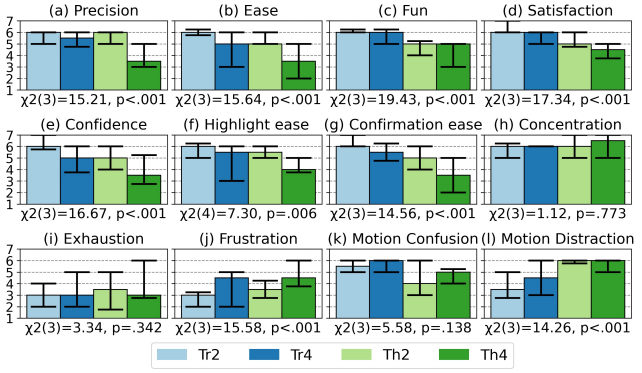


Figure 12: Median scores on the usability Likert scales with error bars representing interquartile ranges.

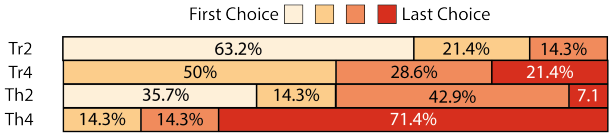
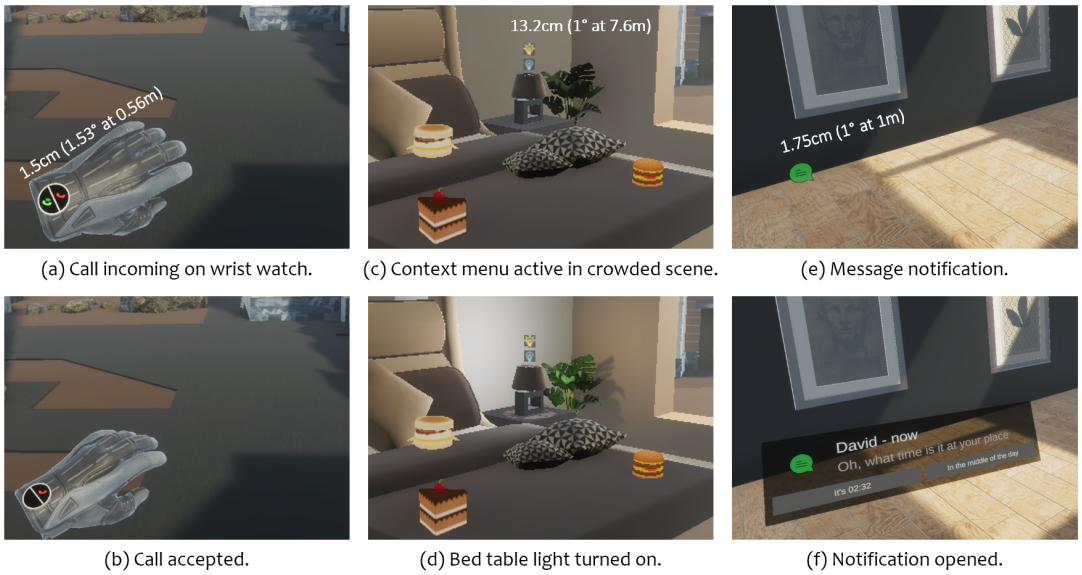


Figure 13: Technique rankings.

Post-hoc analysis showed no differences for the technique × target width interaction. For technique × target depth (Figure 11a), further analysis showed that participants performed more errors with Th2 at the near distance compared to the middle and far distances (all $p \leq .034$). No other techniques showed significant differences at different depths. Regarding technique differences, Tr2 had fewer errors at all depths than all other techniques (all $p \leq .004$), except for Th2 in middle and far. Th2 showed significantly less errors than Tr4 and Th4 under all conditions (all $p \leq .07$) except for the near depth. These results imply that it was difficult for users to follow the moving target at close distances. Th4 had significantly more errors at all depths compared to Tr4. Meanwhile, Th2 had significantly fewer errors than both Th4 and Tr4 (both $p \leq .002$). Finally, Tr4 had significantly fewer errors than Th4 (p<.001).

6.4.5 Subjective Preferences. Friedman tests on usability ratings (Figure 12) showed significant differences in Precision, Ease, Fun, Satisfaction, Confidence, Confirmation ease, Frustration and Motion Distraction. Post-hoc tests showed that the participants preferred Tr2 over Th4 in all scores (all $p \leq .004$). There were also significant differences for Satisfaction between Tr2 and Th2 (p=.046) and Tr4 and Th4 (p = .046). Figure 13 shows participants' technique rankings, in which Tr2 received the highest overall ranking. A Friedman test ($\chi^2(3)$=15.3, p=0.00158), and following post-hoc tests found significant differences between Tr2 and Tr4 (p=.009), Tr2 and Th4 (p=.036) and Th2 and Th4 (p=.033).

Interview results showed that the trigger-based technique gave them control over the selection, and selection could be performed as soon as the correct targets were highlighted, delaying triggering the selection until they were certain that the right target was registered. This was especially useful in situations where it was difficult to highlight the correct target. Participants also mentioned that



(a) Call incoming on wrist watch. (c) Context menu active in crowded scene. (e) Message notification.

(b) Call accepted. (d) Bed table light turned on. (f) Notification opened.

Figure 14: Application examples. (a) The virtual smart-watch UI presents users with two buttons, each involving modulated depth motion. (b) In the event of a call, the user can accept by focusing on the left button, reflected in the interface by hiding the accept button. (c) Contextual menus are used to interact with the virtual environment. (d) The user can pre-select an object to present actions performed by Vergence Matching, e.g. toggle the light. (e) Small notifications appear in the user's HUD. (f) Focusing on the target will activate the notification to get more information.

the extra control allowed faster selections, unlike threshold-based techniques, where they had to wait for the selection to be registered. Finally, the participants felt less pressure and more relaxed as the technique would not trigger a wrong selection, and it also required less concentration. Participants who preferred threshold-based techniques mentioned that they liked that they did not have to use their hands and that they felt accurate with the technique.

We further analysed the data based on the target characteristics. As expected, participants reported that either technique was easier to use with fewer targets. Additionally, having more targets was generally distracting, and participants indicated that they had to pay more attention to the correct target to avoid false activation. In general, participants mentioned that focussing on targets felt "natural" (P8) and "straight-forward" (P2). However, on target depth and width, most of the participants indicated that it was more straining when the targets were very close or small. In the former, a third of participants explicitly reported double vision when the target appeared close and therefore had issues focussing on the target. On the latter, participants reported that smaller targets were slightly more difficult to select than larger targets due to visibility or difficulty focussing in the presence of large distractors.

7 VERGENCE MATCHING APPLICATIONS

In Study 2, we showcased the capability of Vergence Matching techniques to select small targets. We further developed three applications that show the versatility and flexibility of the Vergence Matching techniques enabled by our approach of inducing movements on targets in the virtual environment (Figure 14). Specifically, the applications highlight the benefits of being able to select small targets be reducing the required screen real estate. Across the applications, we vary the modality used for candidate selection and placements of Vergence Matching targets (Table 1).

Our first application demonstrates Vergence Matching as an AR/VR smartwatch where participants interact with small widgets on the watch. Prior work on gaze-based smartwatch interaction used motion correlation to accurately select targets via smooth pursuit [14] but requires visual movement in a 2D plane, requiring more screen real estate. In contrast, we designed a virtual smartwatch using Vergence Matching that can display stationary targets from the user's perspective. The watch appears on the user's left forearm, providing two buttons (1.5cm width) that can be selected by Vergence Matching. Candidate selection is controlled by the hand by simply moving the watch into an orientation facing the user, highlighting how candidate selection can be performed by inverse the pointing the mechanism (pointing at the user). Once the watch is aligned, motion is induced on the buttons for interaction.

In the second example, we show how Vergence Matching can be leveraged for contextual menus for interaction with objects placed in the scene while minimising visual clutter (Figure 14c-d). Previous work has highlighted visual clutter as an issue and proposed models based on gaze behaviour or mental workload [16, 27]. In our application, head-based pointing presents interactable contextual menus only for objects close to the user's visual attention. The contextual menus include a small set of buttons for functions such as "turn on", or "close". The buttons are presented as icons and made small in size (1°) to further reduce visual clutter and induce motion for interaction with Vergence Matching.

Our final application shows Vergence Matching for head-up display (HUD) icons Figure 14e-f. Recent work has shown the utility

Table 1: Overview of Vergence Matching applications. The applications highlight how Vergence Matching techniques enable differences in how users choose candidates for selection and the position of targets relative to the user.

Application	Candidate Selection Modality	Candidate Pointing Direction	Target Reference
Smartwatch	Hand	Towards user	Hand-referenced
Contextual Menu	Head	Towards target	World-referenced
Head-up Display	Gaze	Towards target	Head-referenced

of notifications that dynamically appear to the user in 3D environments [28, 29]. Furthermore, HUD notifications have been proven to be beneficial in noticeability and avoiding users from missing important information [44]. However, they are equally disturbing and intrusive [44]. We consider a more subtle design that leverages Vergence Matching's ability to select tiny targets. The notifications are transparent so that users can see through, but sufficiently small to be avoidable and not distract (1°). Gaze pointing on the notification direction triggers the icon's visibility and induces motion that can be selected via Vergence Matching for further interaction. Gaze pointing allows a more implicit candidate selection where pre-selection and selection are linked (i.e., look and dwell).

8 DISCUSSION

Vergence Matching is a novel method of detecting user attention on very small objects that leverages the principle of motion correlation combined with the ability of the eyes to automatically converge or diverge based on an object's distance. Our results demonstrate how Vergence Matching can be used to accurately infer when a user attends to very small targets, significantly smaller than the recorded eye tracking accuracy and precision. Vergence Matching exploits the difference between fast and slow vergence movements [10], utilising object movement that induces controlled slow vergence movements which can be distinguished from naturally occurring vergence movements, such as when visually searching for objects. Due to the rarity of motions that cause slow vergence in natural environments we would expect our findings to hold true across a wide range of contexts and environments. Vergence Matching enables unique capabilities that are not possible with dwell-only techniques and contrasts smooth pursuit-based detection by requiring minimal screen real-estate for displaying objects.

Inferring when a user is attending to very small targets with minimal changes to the environment is a key challenge for gaze interaction, and Vergence Matching presents a significant contribution in extending the capabilities of gaze-based interactions in VR. We demonstrate how Vergence Matching can be used as a selection technique in virtual environments using only the relative movements of the pupil positions and the object, without the need for accurate gaze calibration. Evaluation of threshold and trigger-based selection techniques confirm how very small targets can be attended to and selected using Vergence Matching. For context, the smallest target width of 0.25° is only 4.4mm wide at a distance of 1m. This presents new opportunities for gaze interaction that are not currently available due to current eye tracking limitations, including more subtle and discreet gaze-based interfaces that minimally affect the interface in comparison to artificially increasing target width, zooming in on the view, or having large movement trajectories. Our applications provide a glimpse into what is possible when the necessity to compensate for gaze-sensing inaccuracies is removed without the need to use additional screen real estate.

One of the trade-offs with the ability to select such small targets is the completion time. For our implementation, there is a lower bound for a completion time of 1.33s due to the requirement to fill the moving windows (N_m = 90, N_{valid} = 30), however we observed mean completion times ranging from 3.24s in the two target trigger condition to 3.89s in the four target threshold condition. In addition, the mean selection times alone have implications for detecting attention in non-interactive contexts, suggesting users need to pay extended attention to objects of interest for reliable detection. Although Vergence Matching is slower for selection compared to techniques such as dwell, other motion correlation techniques suffer from similarly high completion times. Smooth pursuit-based selection on a 2D display takes approximately 1.88–3.99s (application-dependent median times) [59], increasing to 3.2–4.6s for selection with wearable eye trackers in a real-world environment [58]. The completion times for Vergence Matching are more comparable to Outline Pursuits [47], which also employs a candidate selection phase, resulting in 2.81s for trigger-based selection and 4.03s for threshold-based selection. Despite the relatively large completion times, Vergence Matching enables unique capabilities that are not possible with dwell-only or smooth pursuit-based techniques.

Study 2 also revealed large error rates in some conditions and these issues affect the scalability of Vergence Matching in comparison to other smooth pursuit-based motion correlation techniques which use 2D motions. In Vergence Matching, unique motions can only be differentiated in one dimension, meaning that movements must be more precise and correlated. If an accurate interaction is needed, fewer motions can be deployed. Here lies an interesting trade-off – fewer candidates in the candidate selection phase means the user has to be more precise with their pre-selection. As such, the modality choice should depend on the interaction needs and usage context. Our applications showed that the pre-selection modality can be diverse and extends to any generic pointing technique.

It is also possible that other technical and physiological factors may have contributed towards both the magnitude and observed variability of errors. Some participants reported that they experienced double vision when selecting small targets at a close distance, which may have contributed to the longer selection times and error rates for near targets, especially in threshold-based conditions. These factors could relate to ocular issues such as convergence insufficiency, which is one of the most common causes of muscular discomfort [26], and previous research has also shown that some people exhibit weakness in performing vergence eye movements [22]. Alternatively, or in addition, there are current limitations with modern head-mounted displays due to fixed focal planes, which have been shown to affect vergence movements [64] and could have exacerbated these issues. A potential way to address the issue of double vision and focus difficulties is by using slower movements with less amplitude. We picked a motion for our techniques with a high amplitude and low cycle time based on study 1 results. However, the results in Figure 7 found similar performance for other

motions with smaller amplitudes or slower cycle times. It is possible that it would be easier for users to follow smaller and slower motions without a noticeable effect on correlation performance.

Upon reflection, we also identified several areas that could result in improvements to both selection time and error rates by adapting the vergence movements and algorithms to the user. Firstly, we relied on a fixed IPD of 6.5cm to decide motion amplitude. Any difference between the participant's actual IPD and the fixed IPD would change the actual vergence movement the eyes perform and this may have made it more difficult to detect and correlate. Calibrating the system to an individual's IPD may help to avoid this. Similarly, we also used the same thresholds across all participants, which are used for highlighting (trigger-based) and selection (threshold-based). Individually tailoring thresholds to participants could further alleviate some issues, and Vergence Matching may be more sensitive to these parameters due to the exponential change in vergence angle as distance increases. These personalised optimisations and further improvements to the detection algorithm underpinning Vergence Matching (e.g., probabilistic frameworks [57]) could further reduce the selection times and error rates, making it a more viable alternative to select one among many targets.

Despite the large error rates in its current form, Vergence Matching provides unique advantages for gaze-based attention inference and object selection. Even when selection is limited to one or two candidate targets, there are unique opportunities to synergistically combine Vergence Matching with other gaze-based techniques to use its unique ability to select very small targets. For example, a system could account for the inherent accuracy and precision limitations of the eye tracker by automatically detecting the required gaze accuracy for selection of an object. In such a case, dwell-based (or similar) techniques can be used for basic, unoccluded selections, while Vergence Matching-based selection can be utilised for selection of objects beyond the accuracy of the gaze tracker, or when disambiguation is required due to close proximity with other targets. There are also other use cases in which the unique capabilities of Vergence Matching can be leveraged without the complexities of candidate selection. For example, the attention-sensing capability of Vergence Matching alone can be utilised for subtle and discreet prompt-based interactions, such as a confirmation pop-up, where selection from a limited number of options is common.

Although we have systematically investigated how slow vergence movements can be used to infer user attention and to underpin interaction, we note that the proposed techniques were studied in a lab-based context with abstract environments. More natural settings, such as the applications we propose, were not formally studied but may provide rich insights into the usage of Vergence Matching. In addition, it may be more challenging to perform slow vergence movements in more visually salient environments, and the vergence movements in our studies were all performed in the centre area of the user's field of view. Performance of Vergence Matching at significant eye-in-head angles, such as in the notification application, remains open for evaluation.

9 CONCLUSION

Vergence Matching is novel in addressing the challenge of attention detection and selection of targets much smaller than the accuracy of an eye tracker. We further demonstrated Vergence Matching as two selection techniques, a threshold-based approach for hands-free interaction and a trigger-based confirmation for more control of selection timing with the hands. In contrast to established AR/VR gaze techniques, we show that our techniques are size-independent, and users can interact with small targets as long as they can be focussed upon. This capability allows a novel way of tackling the challenge of interaction with small targets in 3D environments.

ACKNOWLEDGMENTS

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant No. 101021229, GEMINI: Gaze and Eye Movement in Interaction).

REFERENCES

[1] Sunggeun Ahn, Stephanie Santosa, Mark Parent, Daniel Wigdor, Tovi Grossman, and Marcello Giordano. 2021. StickyPie: A Gaze-Based, Scale-Invariant Marking Menu Optimized for AR/VR. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 739, 16 pages. https://doi.org/10.1145/3411764.3445297

[2] Sunggeun Ahn, Jeongmin Son, Sangyoon Lee, and Geehyuk Lee. 2020. VergeIt: Gaze Interaction for a Binocular Head-Worn Display Using Modulated Disparity Vergence Eye Movement. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3334480.3382908

[3] Rawan Alghofaili, Yasuhito Sawahata, Haikun Huang, Hsueh-Cheng Wang, Takaaki Shiratori, and Lap-Fai Yu. 2019. Lost in Style: Gaze-Driven Adaptive Aid for VR Navigation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300578

[4] Rawan Alghofaili, Michael S Solah, Haikun Huang, Yasuhito Sawahata, Marc Pomplun, and Lap-Fai Yu. 2019. Optimizing Visual Element Placement via Visual Attention Analysis. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 464–473. https://doi.org/10.1109/VR.2019.8797816

[5] Ferran Argelaguet and Carlos Andujar. 2013. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics* 37, 3 (2013), 121–136. https://doi.org/10.1016/j.cag.2012.12.003

[6] Isayas B. Adhanom, Samantha C. Lee, Eelke Folmer, and Paul MacNeilage. 2020. GazeMetrics: An Open-Source Tool for Measuring the Data Quality of HMD-Based Eye Trackers. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20 Short Papers)*. ACM, Article 19, 5 pages. https://doi.org/10.1145/3379156.3391374

[7] Marc Baloup, Thomas Pietrzak, and Géry Casiez. 2019. RayCursor: A 3D Pointing Facilitation Technique Based on Raycasting. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 101, 12 pages. https://doi.org/10.1145/3290605.3300331

[8] Pieter Blignaut. 2009. Fixation identification: The optimum threshold for a dispersion algorithm. *Attention, Perception, & Psychophysics* 71, 4 (01 May 2009), 881–895. https://doi.org/10.3758/APP.71.4.881

[9] Richard A. Bolt. 1981. Gaze-Orchestrated Dynamic Windows. *SIGGRAPH Comput. Graph.* 15, 3 (aug 1981), 109–119. https://doi.org/10.1145/965161.806796

[10] Kathleen E. Cullen and Marion R. Van Horn. 2011. The neural control of fast vs. slow vergence eye movements. *European Journal of Neuroscience* 33, 11 (2011), 2147–2154. https://doi.org/10.1111/j.1460-9568.2011.07692.x

[11] Gerwin de Haan, Michal Koutek, and Frits H. Post. 2005. IntenSelect: Using Dynamic Object Rating for Assisting 3D Object Selection. In *Proceedings of the 11th Eurographics Conference on Virtual Environments* (Aalborg, Denmark) *(EGVE'05)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 201–209. https://doi.org/10.2312/EGVE/IPT_EGVE2005/201-209

[12] Neil A. Dodgson. 2004. Variation and extrema of human interpupillary distance. In *Stereoscopic Displays and Virtual Reality Systems XI*, Mark T. Bolas, Andrew J. Woods, John O. Merritt, and Stephen A. Benton (Eds.), Vol. 5291. International Society for Optics and Photonics, SPIE, 36 – 46. https://doi.org/10.1117/12.529999

[13] Niklas Elmqvist and Philippas Tsigas. 2008. A Taxonomy of 3D Occlusion Management for Visualization. *IEEE Transactions on Visualization and Computer Graphics* 14, 5 (2008), 1095–1109. https://doi.org/10.1109/TVCG.2008.59

[14] Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015. Orbits: Gaze Interaction for Smart Watches Using Smooth Pursuit Eye Movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) *(UIST '15)*. ACM, New York, NY, USA, 457–466. https://doi.org/10.1145/2807442.2807499

[15] Anna Maria Feit, Shane Williams, Arturo Toledo, Ann Paradiso, Harish Kulkarni, Shaun Kane, and Meredith Ringel Morris. 2017. Toward Everyday Gaze Input: Accuracy and Precision of Eye Tracking and Implications for Design. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1118–1130. https://doi.org/10.1145/3025453.3025599

[16] Christoph Gebhardt, Brian Hecox, Bas van Opheusden, Daniel Wigdor, James Hillis, Otmar Hilliges, and Hrvoje Benko. 2019. Learning Cooperative Personalized Policies from Gaze Data. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 197–208. https://doi.org/10.1145/3332165.3347933

[17] Tovi Grossman and Ravin Balakrishnan. 2006. The Design and Evaluation of Selection Techniques for 3D Volumetric Displays. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology* (Montreux, Switzerland) *(UIST '06)*. ACM, New York, NY, USA, 3–12. https://doi.org/10.1145/1166253.1166257

[18] Teresa Hirzle, Jan Gugenheimer, Florian Geiselhart, Andreas Bulling, and Enrico Rukzio. 2019. A Design Space for Gaze Interaction on Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300855

[19] Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Jarodzka Halszka, and Joost van de Weijer. 2011. *Eye Tracking : A Comprehensive Guide to Methods and Measures*. Oxford University Press. 560 pages.

[20] Zhiming Hu, Andreas Bulling, Sheng Li, and Guoping Wang. 2021. FixationNet: Forecasting Eye Fixations in Task-Oriented Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2681–2690. https://doi.org/10.1109/TVCG.2021.3067779

[21] Robert J. K. Jacob. 1991. The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look at is What You Get. *ACM Trans. Inf. Syst.* 9, 2 (apr 1991), 152–169. https://doi.org/10.1145/123078.128728

[22] Stephanie Jainta, Maria Pia Bucci, Sylvette Wiener-Vacher, and Zoï Kapoula. 2011. Changes in vergence dynamics due to repetition. *Vision Research* 51, 16 (2011), 1845–1852. https://doi.org/10.1016/j.visres.2011.06.014

[23] Shahram Jalaliniya and Diako Mardanbegi. 2016. EyeGrip: Detecting Targets in a Series of Uni-Directional Moving Objects Using Optokinetic Nystagmus Eye Movements. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5801–5811. https://doi.org/10.1145/2858036.2858584

[24] 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* (Castiglione della Pescaia, Grosseto, Italy) *(AVI '18)*. ACM, New York, NY, USA, Article 18, 8 pages. https://doi.org/10.1145/3206505.3206522

[25] 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* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 81, 14 pages. https://doi.org/10.1145/3173574.3173655

[26] Judith B Lavrich. 2010. Convergence insufficiency and its current treatment. *Current opinion in ophthalmology* 21, 5 (2010), 356–360. https://doi.org/10.1097/ICU.0b013e32833cf03a

[27] David Lindlbauer, Anna Maria Feit, and Otmar Hilliges. 2019. Context-Aware Online Adaptation of Mixed Reality Interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 147–160. https://doi.org/10.1145/3332165.3347945

[28] Feiyu Lu, Shakiba Davari, and Doug Bowman. 2021. Exploration of Techniques for Rapid Activation of Glanceable Information in Head-Worn Augmented Reality. In *Symposium on Spatial User Interaction* (Virtual Event, USA) *(SUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 14, 11 pages. https://doi.org/10.1145/3485279.3485286

[29] Feiyu Lu, Shakiba Davari, Lee Lisle, Yuan Li, and Doug A. Bowman. 2020. Glanceable AR: Evaluating Information Access Methods for Head-Worn Augmented Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 930–939. https://doi.org/10.1109/VR46266.2020.00113

[30] Yiqin Lu, Chun Yu, and Yuanchun Shi. 2020. Investigating Bubble Mechanism for Ray-Casting to Improve 3D Target Acquisition in Virtual Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 35–43. https://doi.org/10.1109/VR46266.2020.00021

[31] Diako Mardanbegi, Christopher Clarke, and Hans Gellersen. 2019. Monocular Gaze Depth Estimation Using the Vestibulo-ocular Reflex. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (Denver, Colorado) *(ETRA '19)*. ACM, New York, NY, USA, Article 20, 9 pages. https://doi.org/10.1145/3314111.3319822

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

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

[34] P. McCullagh and J. A. Nelder. 1989. *Generalized Linear Models*. Chapman and Hall/CRC. 532 pages.

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

[36] Brad A. Myers, Rishi Bhatnagar, Jeffrey Nichols, Choon Hong Peck, Dave Kong, Robert Miller, and A. Chris Long. 2002. Interacting at a Distance: Measuring the Performance of Laser Pointers and Other Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Minneapolis, Minnesota, USA) *(CHI '02)*. Association for Computing Machinery, New York, NY, USA, 33–40. https://doi.org/10.1145/503376.503383

[37] Marcus Nyström and Kenneth Holmqvist. 2010. An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods* 42, 1 (01 Feb 2010), 188–204. https://doi.org/10.3758/BRM.42.1.188

[38] Yun Suen Pai, Benjamin Outram, Noriyasu Vontin, and Kai Kunze. 2016. Transparent Reality: Using Eye Gaze Focus Depth as Interaction Modality. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 171–172. https://doi.org/10.1145/2984751.2984754

[39] Soonchan Park, Seokyeol Kim, and Jinah Park. 2012. Select Ahead: Efficient Object Selection Technique Using the Tendency of Recent Cursor Movements. In *Proceedings of the 10th Asia Pacific Conference on Computer Human Interaction* (Matsue-city, Shimane, Japan) *(APCHI '12)*. Association for Computing Machinery, New York, NY, USA, 51–58. https://doi.org/10.1145/2350046.2350060

[40] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. 2014. GazeTouch: Combining Gaze with Multi-Touch for Interaction on the Same Surface. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 509–518. https://doi.org/10.1145/2642918.2647397

[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* (Brighton, United Kingdom) *(SUI '17)*. ACM, New York, NY, USA, 99–108. https://doi.org/10.1145/3131277.3132180

[42] Ken Pfeuffer, Melodie Vidal, Jayson Turner, Andreas Bulling, and Hans Gellersen. 2013. Pursuit Calibration: Making Gaze Calibration Less Tedious and More Flexible. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (St. Andrews, Scotland, United Kingdom) *(UIST '13)*. Association for Computing Machinery, New York, NY, USA, 261–270. https://doi.org/10.1145/2501988.2501998

[43] 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* (Brighton, United Kingdom) *(SUI '17)*. ACM, New York, NY, USA, 91–98. https://doi.org/10.1145/3131277.3132182

[44] Rufat Rzayev, Sven Mayer, Christian Krauter, and Niels Henze. 2019. Notification in VR: The Effect of Notification Placement, Task and Environment. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Barcelona, Spain) *(CHI PLAY '19)*. Association for Computing Machinery, New York, NY, USA, 199–211. https://doi.org/10.1145/3311350.3347190

[45] Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying Fixations and Saccades in Eye-Tracking Protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (Palm Beach Gardens, Florida, USA) *(ETRA '00)*. Association for Computing Machinery, New York, NY, USA, 71–78. https://doi.org/10.1145/355017.355028

[46] Greg Schmidt, Yohan Baillot, Dennis G. Brown, Erik B. Tomlin, and J. Edward Swan. 2006. Toward Disambiguating Multiple Selections for Frustum-Based Pointing. In *3D User Interfaces (3DUI'06)*. IEEE, 87–94. https://doi.org/10.1109/VR.2006.133

[47] Ludwig Sidenmark, Christopher Clarke, Xuesong Zhang, Jenny Phu, and Hans Gellersen. 2020. Outline Pursuits: Gaze-Assisted Selection of Occluded Objects in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376438

[48] Ludwig Sidenmark and Hans Gellersen. 2019. 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* (New Orleans,

LA, USA) *(UIST '19)*. ACM, New York, NY, USA, 1161–1174. https://doi.org/10.1145/3332165.3347921

[49] 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 *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20 Full Papers)*. Association for Computing Machinery, New York, NY, USA, Article 8, 9 pages. https://doi.org/10.1145/3379155.3391312

[50] Ludwig Sidenmark, Mark Parent, Chi-Hao Wu, Joannes Chan, Michael Glueck, Daniel Wigdor, Tovi Grossman, and Marcello Giordano. 2022. Weighted Pointer: Error-aware Gaze-based Interaction through Fallback Modalities. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (2022), 3585–3595. https://doi.org/10.1109/TVCG.2022.3203096

[51] Ludwig Sidenmark, Dominic Potts, Bill Bapisch, and Hans Gellersen. 2021. Radi-Eye: Hands-Free Radial Interfaces for 3D Interaction Using Gaze-Activated Head-Crossing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 740, 11 pages. https://doi.org/10.1145/3411764.3445697

[52] Oleg Špakov and Päivi Majaranta. 2012. Enhanced Gaze Interaction Using Simple Head Gestures. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (Pittsburgh, Pennsylvania) *(UbiComp '12)*. ACM, New York, NY, USA, 705–710. https://doi.org/10.1145/2370216.2370369

[53] Frank Steinicke, Timo Ropinski, and Klaus Hinrichs. 2006. *Object Selection in Virtual Environments Using an Improved Virtual Pointer Metaphor*. Springer Netherlands, Dordrecht, 320–326. https://doi.org/10.1007/1-4020-4179-9_46

[54] Lode Vanacken, Tovi Grossman, and Karin Coninx. 2007. Exploring the Effects of Environment Density and Target Visibility on Object Selection in 3D Virtual Environments. In *2007 IEEE Symposium on 3D User Interfaces*. IEEE. https://doi.org/10.1109/3DUI.2007.340783

[55] Eduardo Velloso, Marcus Carter, Joshua Newn, Augusto Esteves, Christopher Clarke, and Hans Gellersen. 2017. Motion Correlation: Selecting Objects by Matching Their Movement. *ACM Trans. Comput.-Hum. Interact.* 24, 3, Article 22 (April 2017), 35 pages. https://doi.org/10.1145/3064937

[56] Eduardo Velloso, Flavio Luiz Coutinho, Andrew Kurauchi, and Carlos H. Morimoto. 2018. Circular Orbits Detection for Gaze Interaction Using 2D Correlation and Profile Matching Algorithms. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications* (Warsaw, Poland) *(ETRA '18)*. ACM, New York, NY, USA, Article 25, 9 pages. https://doi.org/10.1145/3204493.3204524

[57] Eduardo Velloso and Carlos H Morimoto. 2021. A Probabilistic Interpretation of Motion Correlation Selection Techniques. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 285, 13 pages. https://doi.org/10.1145/3411764.3445184

[58] Eduardo Velloso, Markus Wirth, Christian Weichel, Augusto Esteves, and Hans Gellersen. 2016. AmbiGaze: Direct Control of Ambient Devices by Gaze. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (Brisbane, QLD, Australia) *(DIS '16)*. ACM, New York, NY, USA, 812–817. https://doi.org/10.1145/2901790.2901867

[59] 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* (Zurich, Switzerland) *(UbiComp '13)*. ACM, New York, NY, USA, 439–448. https://doi.org/10.1145/2493432.2493477

[60] Rui I. Wang, Brandon Pelfrey, Andrew T. Duchowski, and Donald H. House. 2014. Online 3D Gaze Localization on Stereoscopic Displays. *ACM Trans. Appl. Percept.* 11, 1, Article 3 (April 2014), 21 pages. https://doi.org/10.1145/2593689

[61] Martin Weier, Thorsten Roth, André Hinkenjann, and Philipp Slusallek. 2018. Predicting the Gaze Depth in Head-Mounted Displays Using Multiple Feature Regression. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications* (Warsaw, Poland) *(ETRA '18)*. Association for Computing Machinery, New York, NY, USA, Article 19, 9 pages. https://doi.org/10.1145/3204493.3204547

[62] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The xAligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. ACM, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[63] Qing Yang, Maria Pia Bucci, and Zoï Kapoula. 2002. The Latency of Saccades, Vergence, and Combined Eye Movements in Children and in Adults. *Investigative Ophthalmology & Visual Science* 43, 9 (09 2002), 2939–2949.

[64] Shunnan Yang and James E. Sheedy. 2011. Effects of vergence and accommodative responses on viewer's comfort in viewing 3D stimuli. In *Stereoscopic Displays and Applications XXII*, Andrew J. Woods, Nicolas S. Holliman, and Neil A. Dodgson (Eds.), Vol. 7863. International Society for Optics and Photonics, SPIE, 78630Q. https://doi.org/10.1117/12.872546

[65] Shumin Zhai, William Buxton, and Paul Milgram. 1994. The "Silk Cursor": Investigating Transparency for 3D Target Acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, Massachusetts, USA) *(CHI '94)*. ACM, New York, NY, USA, 459–464. https://doi.org/10.1145/191666.191822

[66] Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 246–253. https://doi.org/10.1145/302979.303053