



Eye-Tracking Beyond Peripersonal Space in Virtual Reality: Validation and Best Practices

Maurice Lamb^{1,2*}, Malin Brundin¹, Estela Perez Luque² and Erik Billing¹

¹School of Informatics, University of Skövde, Skövde, Sweden, ²Department of Engineering, University of Skövde, Skövde, Sweden

OPEN ACCESS

Edited by:

Vsevolod Peysakhovich,
Institut Supérieur de l'Aéronautique et
de l'Espace (ISAE-SUPAERO), France

Reviewed by:

Kaan Akşit,
University College London,
United Kingdom
Margarita Vinnikov,
New Jersey Institute of Technology,
United States
Anthony Ries,
United States Army Research
Laboratory, United States

*Correspondence:

Maurice Lamb
Maurice.Lamb@his.se

Specialty section:

This article was submitted to
Technologies for VR,
a section of the journal
Frontiers in Virtual Reality

Received: 28 January 2022

Accepted: 14 March 2022

Published: 08 April 2022

Citation:

Lamb M, Brundin M, Perez Luque E
and Billing E (2022) Eye-Tracking
Beyond Peripersonal Space in Virtual
Reality: Validation and Best Practices.
Front. Virtual Real. 3:864653.
doi: 10.3389/frvir.2022.864653

Recent developments in commercial virtual reality (VR) hardware with embedded eye-tracking create tremendous opportunities for human subjects researchers. Accessible eye-tracking in VR opens new opportunities for highly controlled experimental setups in which participants can engage novel 3D digital environments. However, because VR embedded eye-tracking differs from the majority of historical eye-tracking research, in both providing for relatively unconstrained movement and stimulus presentation distances, there is a need for greater discussion around methods for implementation and validation of VR based eye-tracking tools. The aim of this paper is to provide a practical introduction to the challenges of, and methods for, 3D gaze-tracking in VR with a focus on best practices for results validation and reporting. Specifically, first, we identify and define challenges and methods for collecting and analyzing 3D eye-tracking data in VR. Then, we introduce a validation pilot study with a focus on factors related to 3D gaze tracking. The pilot study provides both a reference data point for a common commercial hardware/software platform (HTC Vive Pro Eye) and illustrates the proposed methods. One outcome of this study was the observation that accuracy and precision of collected data may depend on stimulus distance, which has consequences for studies where stimuli is presented on varying distances. We also conclude that vergence is a potentially problematic basis for estimating gaze depth in VR and should be used with caution as the field move towards a more established method for 3D eye-tracking.

Keywords: eye tracking, virtual reality, gaze depth, vergence, validation

1 INTRODUCTION

Various methods and tools for systematically measuring eye-gaze behaviors have been around for nearly a century. Though for much of that history, the skill, time, and cost required to collect and analyze eye-tracking data was considerable. However, recent advances in low-cost hardware and comprehensive software solutions have made eye-tracking tools and methods more broadly accessible and easier to implement than ever (Orquin and Holmqvist, 2018; Carter and Luke, 2020; Niehorster et al., 2020). The proliferation of accessible eye-tracking systems has contributed to a corresponding increase in discussions around collection, analysis, and validation methods relating to these systems among both gaze/eye behavior researchers and other researchers who would benefit from eye-tracking data (Johnsson and Matos, 2011; Feit et al., 2017; Hessels et al., 2018; Orquin and Holmqvist, 2018; Carter and Luke, 2020; Kothari et al., 2020; Niehorster et al., 2020). The primary focus of these discussions is on eye-tracking in traditional 2D stimulus and gaze tracking contexts.

However, as the cost and intrusiveness of eye-tracking hardware decreases, eye-tracking data is being collected outside of more well-established eye-tracking contexts.

One rapidly growing segment of eye-tracking technology involves the integration of low-cost eye-tracking hardware into virtual reality (VR) head-mounted displays (HMDs), particularly in commercial and entertainment contexts (Niehorster et al., 2017; Wibirama et al., 2017; Clay et al., 2019; Iskander et al., 2019; Koulieris et al., 2019; Zhao et al., 2019). Independent of eye-tracking integration, VR creates many unique design opportunities for human subjects research (Marmitt and Duchowski, 2002; Harris et al., 2019; Brookes et al., 2020; Harris et al., 2020). These design opportunities include being able to control many aspects of the environment that are difficult to manipulate in the real world, rapidly reset stimuli, present stimuli in a range of quasi-naturalistic settings, instantly change environment or stimulus states, create physically impossible stimuli, and rapidly develop, repeat, and replicate experiments. As consumer VR hardware quality has improved and prices have decreased, these opportunities have enticed research labs engaged in a variety of human subjects research to include VR as one of their tools for piloting, research, and demonstration. Most VR systems include head and basic hand motion tracking, with some systems offering extended motion tracking options (Borges et al., 2018; Koulieris et al., 2019; van der Veen et al., 2019). Moreover, these VR HMDs can be paired with a wide range of experimental research hardware. Recently, a few consumer VR system manufacturers have begun offering VR HMDs with built-in eye-tracking hardware integration, greatly simplifying and reducing the cost of collecting eye-gaze data in VR. This recent inclusion of eye-tracking hardware in VR headsets further extends the potential of VR in research, but also introduces some notable methodological questions and challenges.

While the underlying eye-tracking hardware in modern VR systems is essentially the same as many non-VR embedded systems, the typical use cases often involve significant departures from well-established eye-tracking paradigms in both stimulus presentation format and constraints on participant behavior. Notably, VR is commonly deployed as a tool for studying behavior in 3D environments with the relevant stimuli presented at a variety of simulated distances from participants, including both participants' immediately 'reachable' peripersonal space (0.01–2.0 m) and beyond (2.0 m—infinity) (Previc, 1998; Armbrüster et al., 2008; Iorizzo et al., 2011; Naceri et al., 2011; Deb et al., 2017; Harris et al., 2019; Wu et al., 2020). Non-VR eye-tracking is most commonly deployed and validated in the context of peripersonal stimulus presentations near a participant, typically between 0.1 and 1.5 m (Land and Lee, 1994; Verstraten et al., 2001; Johnsson and Matos, 2011; Kowler, 2011; Holmqvist et al., 2012; Hessels et al., 2015; Larsson et al., 2016; Feit et al., 2017; Niehorster et al., 2018; Carter and Luke, 2020; Kothari et al., 2020). However, when gaze data is collected in VR for behavioral research, the relevant stimuli may be presented both in and beyond a participant's peripersonal space (Kwon et al., 2006; Duchowski et al., 2011; Wang et al., 2018; Clay et al., 2019; Harris et al., 2019; Vienne et al., 2020).

Moreover, many of the most common use cases for VR that might benefit from eye-tracking data involve relevant simulated stimuli distances outside of peripersonal space, e.g. gaming, marketing and retail research, architectural and vehicle design, remote vehicle operation, road user safety research, and workplace training and instruction. Both the simulated depth experience and the presentation of stimuli beyond peripersonal space create unique challenges and considerations for combined VR and eye-tracking studies' experimental design, data interpretation, and system calibration and validation. Even when gaze depth is not a central focus of analysis in a VR eye-tracking study, simulated stimulus depth may affect gaze data quality due to changes in head and eye movement behaviors, geometric relations between the participant and stimuli, hardware design, software assumptions, and features of the stimulus presentation (Kwon et al., 2006; Wang et al., 2014; Kothari et al., 2020; Niehorster et al., 2020).

While eye-tracking embedded VR systems have garnered interest from researchers in a wide range of areas, including research on human behavior (e.g. Binaee and Diaz, 2019; Zhao et al., 2019; Brookes et al., 2020; Wu et al., 2020), the designers of these systems tend to focus on entertainment and commercial use cases as opposed to research contexts. As such, detailed specifications and recommended methods relevant to human-subjects researchers are not widely available from manufactures or software development companies. While such information would be welcome, it has been recommended that validation of eye-tracking hardware and software in general should be carried out locally at the lab and/or experiment level even with well-documented research grade eye-tracking systems (Johnsson and Matos, 2011; Feit et al., 2017; Holmqvist, 2017; Hessels et al., 2018; Niehorster et al., 2018). We believe a similar approach is also required for research applications of eye tracking in consumer VR HMDs. However, there are some potentially significant differences in hardware, stimulus presentation, experimental design, data processing, and data analysis when performing eye-tracking research in VR. Therefore, it is necessary to discuss some of the general distinctive features of eye tracking in VR that may affect methods and data handling and best practices relating to those features. While the pace of change in consumer VR technologies and related eye-tracking hardware is rapid, discussions of current features and challenges, best practices, and validation data points that keep pace with the rapid development of VR and eye-tracking technologies is critical to ensure validity and reliability of research conducted with these tools. Further, specification of current best practices and validation provides context for future researchers attempting to interpret and/or replicate contemporary research.

With the aim of supporting experimental eye-tracking research in VR, we here discuss some distinctive features of eye tracking in VR and illustrate methods for investigating the effects of stimulus presentation distance on accuracy and precision in a naturalistic virtual setting. We include a pilot study that provides both a preliminary validation data point for a common commercial hardware/software platform and illustrates the relevant data processing techniques and general methods. This pilot is part of our own lab's internal validation

process in preparation for upcoming projects involving stimuli outside of peripersonal space. We also propose some best practices for collecting, analysing, and reporting the 3D gaze of VR users in order to support uniform reporting of study results and avoid conflation of VR eye-tracking methods and results with eye-tracking in other contexts. There are few works specifically focused on basic methods and best practices for eye-tracking in VR (Clay et al., 2019). However, to our knowledge there are none which discuss validation of VR HMD embedded eye-tracking systems where gaze data may be tracked in 3D beyond peripersonal space.

We will begin with a discussion of two core distinctive features of VR embedded eye-tracking systems: free-motion of the eye-tracking system (Section 2) and variable stimulus presentation depth (Section 3). We also include a brief overview of key eye-tracking validation definitions specifically as they relate to eye tracking in VR (precision, accuracy, fixation, and vergence depth) in Section 4.3. We then present an example validation pilot study of 3D eye-tracking using a current consumer VR HMD (Section 4). The purpose of this pilot is to provide both 1) a practical example of implementing basic VR eye-tracking methodology taking gaze depth into consideration and 2) a validation data point for a popular consumer oriented VR eye-tracker across multiple visual depth conditions. We will conclude with a discussion of the results of our pilot along with a practical discussion of research best practices and challenges for using VR HMD embedded eye-tracking systems (Section 6).

2 MOTION TRACKING AND EYE TRACKING IN VR

Eye-tracking experiments commonly limit head movement, often requiring participants to remain perfectly still and sometimes fixing head movement with bite plates or similar devices. One of the first eye-tracking systems to allow for some free head motion was introduced by Land (1993). This system allowed for more naturalistic eye-tracking during a vehicle driving task. While this early system provided insights into coordination between head and eye movements, data processing required considerable amounts of time and labor. Much of the innovation in free-motion eye tracking over the decades since Land's initial device has been focused on improving the accuracy of head tracking systems and improving coordination between head and eye signals (Hessels et al., 2015; Carter and Luke, 2020; Niehorster et al., 2020). While there have been recent developments in new eye-tracking hardware designs, eye-tracking embedded in consumer VR HMDs is based on the same principles as Land's early system (Chang et al., 2019; Li et al., 2020; Angelopoulos et al., 2021). In order for free-motion eye-tracking to be done well, the head's position and rotation must be accurately tracked, the timing of head and eye signals must be closely coordinated, and the coordinate systems of at least the head tracking, eye-tracking, and visual scene must be aligned. Many of these same goals are, not coincidentally, shared by VR developers even when no eye-tracking is involved. In order to create an immersive VR experience, VR motion tracking must

accurately track position and rotation and coordinate this information with presentation of a visual scene. This means that the motion tracking systems included with consumer VR HMDs are quite good, and often on par or better than the IMU based systems used in many contemporary head mounted eye-tracking systems (Niehorster et al., 2017; Borges et al., 2018; van der Veen et al., 2019). However, it is important to note that current VR motion tracking are still somewhat limited compared to more expensive research grade motion tracking systems. Though due to rapid consumer technology development, peer reviewed validation lags behind state-of-the-art. In order to collect and analyze VR gaze data it is important to understand how the constraints of VR motion tracking systems interact with eye-tracking data. In this section we will discuss general consideration regarding motion-tracking technologies used with VR, coordinate frame representations, the coordination of head- and eye-tracking signals and finally the potential conflict between free head motion and fixation definitions.

2.1 Motion-Tracking Systems

Much like VR HMD embedded eye-tracking, motion tracking systems included with VR systems hold considerable promise for researchers, but validation and communication of best practices are limited. Regarding VR eye-gaze data, the motion tracking system is critical as it provides data regarding both head position and orientation. There are several approaches to motion tracking in consumer VR (Koulieris et al., 2019). Currently the most validated system is the SteamVR 1.0 tracking system typically associated with the HTC Vive HMD family (Niehorster et al., 2017; Borges et al., 2018; Luckett et al., 2019; van der Veen et al., 2019). Unlike many other commercial VR tracking systems, the SteamVR tracking systems are a good candidates for research motion-tracking because they allow for a relatively large tracked volume and the addition of custom motion trackers. SteamVR tracking systems use a hybrid approach combining inertial (IMU) tracking systems in the HMD with external laser (lighthouse) system which when combined with an optical receiver can provide a corrective signal based on the angle of the laser light beams (Yates and Selan, 2016; Koulieris et al., 2019). The precision (i.e. variability of the signal) of the SteamVR 1.0 system is consistently stable. However, there are reasons to be concerned about the accuracy (i.e. the distance of an object's reported location from an object's actual location) of the system (Niehorster et al., 2017; Borges et al., 2018; Luckett et al., 2019; van der Veen et al., 2019). Accuracy in these studies is measured between VR tracked physical objects, e.g. a VR controller or HMD, and a ground truth system which is simultaneously used for localizing these objects in a physical space. As Niehorster et al. (2017) found, while the VR motion-tracking was systematically offset from a ground truth measure, all measurements were internally consistent with each other. This internal consistency is critical if all positions and orientations of stimuli are generated internally, without reference to an external motion tracking system. If visual stimuli are localized using an external motion-tracking system, then precautions should be taken to ensure that the reference frames of the motion-tracking systems are properly coordinated. Unfortunately, there are currently few

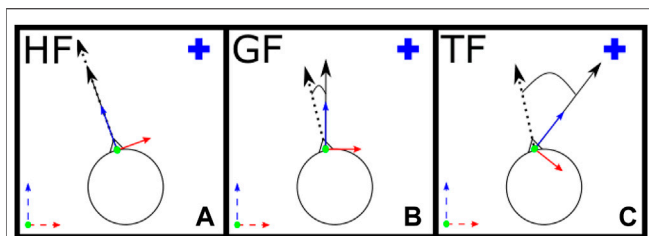


FIGURE 1 | Illustration of different frames of reference. In all cases, a head independent global coordinate system is defined by dashed lines in the lower left corner and for simplicity the gaze vector is assumed perpendicular to the ground plane. **(A)** a head forward frame of reference centered on the gaze center (HF_C) moves and rotates with the head, **(B)** A global forward frame of reference centered on the gaze center (GF_C) moves with the head but is always aligned with the global coordinate system and **(C)** a target forward frame of reference centered on the gaze center (TF_C) moves with the head but maintains a forward axis aligned with the vector from the gaze center to a specified target center.

systematic studies of within system accuracy, e.g. whether the distances between tracked objects is accurately represented in the virtual space (Luckett et al., 2019). Newer inside-out tracking systems are increasingly common in standalone HMDs and use computer vision techniques and camera arrays built into the HMD. Inside-out tracking systems do not depend on tracking hardware that is independent of the HMD, such as external cameras or lighthouses. The camera arrays used for HMD and controller localization are built into the HMD. As with external VR tracking systems, studies of precision and accuracy of inside-out tracking systems are lacking (Holzwarth et al., 2021).

2.2 Frames of Reference

For eye tracking in VR, the VR motion-tracking system provides the position and orientation of the HMD which then must be combined with measurements of tracked eye states, e.g. gaze vectors, to identify where in the visual scene a participant is looking. These measurement values are typically provided in different frames of reference, and may use different measurement units (Hessels et al., 2018). Thus, analysis and data presentation depend on coordinating these frames of reference and converting measurements to common units. In **Figure 1** we illustrate three frames of reference which are useful in both reporting VR embedded eye-tracking results and in developing custom VR experiments with an eye-tracking component.

Gaze vectors in 3D can be well-defined in a spherical coordinate system where we can define forward generically as an azimuthal angle (rotation around the zenith) of 0/deg in a given frame of reference. For the orientation of the head, the most intuitive definition of forward is illustrated in **Figure 1A**, where forward is wherever the head is directed. In this case, the frame of reference moves and rotates with the head. This head forward (HF) frame of reference is commonly used for precision measurements in eye-tracking validation or when reporting only eye-in-head angular rotation.

In many cases, the HF frame will be insufficient as it does not provide information about the head orientation which is required to localize gaze behavior in 3D space. One option is to use the global coordinate frame of reference, as illustrated in **Figure 1B**, where forward is fixed parallel to a selected global forward dimension. Using this global forward (GF) frame of reference, any two vectors with the same azimuth and polar angles will be parallel regardless of their location in the world. In the pilot study presented in **Section 4**, the GF frame of reference is used for some aspects of the programming and in the analysis of head and eye tracking synchronization. Alternatively, a target forward (TF) frame of reference, **Figure 1C**, can be used. Here, the forward dimension is defined by a vector starting at the relevant gaze origin point and terminating at a specified gaze target. In a TF frame of reference, a gaze that is directed forward is directed towards the specified gaze target. For changes in position of a participant or a gaze target two vectors with the same azimuth and polar angles will not be parallel. The TF frame of reference is commonly used when validating system accuracy, since it allows for easy comparison of values given distinct head and target positions.

For reporting eye-gaze values, the origin of the frame of reference should be located at the relevant gaze origin, e.g. the left or right eye origins. For example, if reporting the accuracy of the left eye in a TF frame of reference, then the origin of the frame of reference would be centered on the left-eye center. Likewise, for values relating to the head orientation the origin of the frame of reference should be located at the relevant head origin point. An average or combined eye origin may be defined between left and right eye origins when binocular eye data is available. For consumer VR systems the head origin as provided by the development software may be centered on the HMD, the user's forehead, or some other location. The head origin is likely fixed for most development contexts. However, different hardware and software systems may use different HMD relative origins and it is important to know where the origin is defined relative to the HMD.

2.3 Motion-Tracking Signal Synchronization

Along with spatial accuracy, synchronizing timing across tracking systems is extremely important for most eye-tracking research (Mardanbegi et al., 2019). Even when a participant's head is stabilized, the timing of the eye-tracking and visual scene data must be synchronized. Any offset between eye-tracking timing and scene data can result in lower quality results. In the case of free-motion eye tracking, head-tracking data must also be temporally synchronized with eye and scene data. Because an eye-gaze location in the world depends on the position and orientation of the head, timing offsets in the head-tracking signal relative to scene and eye-tracking signal can result in reduced data fidelity. One way to measure the offset between head and eye tracking is with a vestibulo ocular reflex (VOR) task. The VOR is a reflex of the visual system that stabilizes eye-gaze during head movement (Mardanbegi et al., 2019; Aw et al., 1996; Sidenmark and Gellersen, 2019; Feldman and Zhang, 2020; Callahan-Flintoft et al., 2021). In practice, the VOR can be observed when one fixates on a stable point while rotating

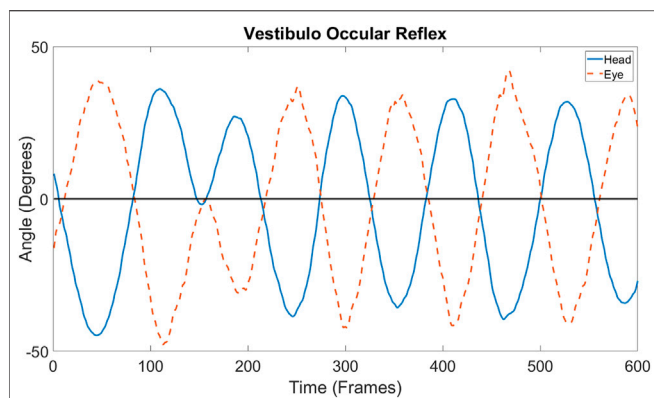


FIGURE 2 | Example of uncorrected measurement for a VOR task involving horizontal head rotation. The horizontal axis indicates time in frames recorded at 120 frames per second. The vertical axis represents a) angle relative to fixation target for the head (blue solid line), where 0° indicates the head is pointed directly at the fixation target and b) angle of combined eye gaze vector relative to the head (orange dashed line) where 0° indicates that the eye vector is directed parallel to the head vector.

their head along an axis of rotation, e.g. an axis orthogonal to a line drawn between the left/right pupils and centered on the head. As the head rotates, the eyes rotate in an opposite direction at the same speed as the head to maintain a stable gaze fixation. When the head and average eye angles are recorded and visualized during a VOR task, relative angles may be visualized as in **Figure 2**. Cross-correlation analysis can be used to measure if there is an offset between head-tracking and eye-tracking (Collewyn and Smeets, 2000). The VOR is expected to contribute ~10 ms delay between head and eye movements, thus any delay larger than 10 ms is likely due to delays in the eye-tracking system (Aw et al., 1996). For example, the Vive Pro Eye setup used in the pilot (**Section 4**) exhibited a stable delay of ~25 ms across multiple individuals and machines in our preliminary testing. When a study involves high accuracy or short duration measurements this ~25 ms delay may impact results. Note that while our system exhibited a consistent ~25 ms delay, it may be that a different hardware setup or software version could yield different results. Individual lab validation and reporting is always strongly advised if fast and accurate measurements are required.

2.4 Fixations and Head Movements

Fixations have become a standard unit of measure in most eye-tracking research (Salvucci and Goldberg, 2000; Blignaut, 2009; Kowler, 2011; Olsen, 2012; Andersson et al., 2017; Holmqvist, 2017; Steil et al., 2018). While there is some debate regarding the precise technical definition of the term, there are several generally accepted methods for identifying fixations in eye-tracking data (Andersson et al., 2017; Hessels et al., 2018). Many modern eye-tracking software packages include automatic fixation detection tools based on these generally accepted methods, making fixation counts and duration a relatively easy metric to include in eye-tracking studies (Orquin and Holmqvist, 2018). Methods for identifying fixations typically involve measures of eye velocity

and/or gaze dispersion (Hessels et al., 2018). In the case of velocity based methods, fixations are identified in data as periods of time where the eye moves relatively little. Dispersion based identification methods identify fixations as occurring when the point in the visual field where the eye is focused moves relatively little. When the head is held still and the subject is looking at objects that does not move, the two kinds of definitions typically provide the same results (Olsen, 2012; Larsson et al., 2016; Andersson et al., 2017; Steil et al., 2018). Because many eye-tracking systems and studies involve little or no head movement, the most widely used fixation detection algorithms are developed with they assumption of limited head movement and fixed visual plane depth.

The I-VT (velocity threshold) fixation detection algorithm is developed assuming that fixations involve relative eye stillness (Olsen, 2012). This eye-focused algorithm is typically applied to eye angles in the HF frame of reference so that head position and orientation are not taken into consideration. As the name suggests, fixations are defined as angular velocities of an eye below a specified threshold. One recommended threshold is 30°/second applied to filtered velocity data. However, because it assumes head stillness, it cannot be directly applied to gaze data collected when the head is allowed to freely move.

A method that might be better equipped to handle free head motion is the I-DT (dispersion threshold) fixation detection algorithm. It identifies fixation as involving a relative stabilization of a visual focus point (Salvucci and Goldberg, 2000; Blignaut, 2009), considering both eye and head motion. I-DT can be applied to eye angles or a gaze point projected on a fixed focus plane (typically a computer screen) at a single distance. In its simplest form I-DT identifies fixations as periods of data in which the dispersion of data is below a predefined threshold. Unlike I-VT there is no general threshold value that works as a starting point (Blignaut, 2009; Andersson et al., 2017). Typically dispersion is measured over a period of 100–200 ms and a threshold is identified by selecting windows where the pattern of candidate fixations matches expectations, e.g. when all task specified stimuli are included in at least one fixation (Blignaut, 2009). A variety of factors affect threshold value including stimulus target size and duration, though other factors including stimulus distance, lighting, and environmental clutter may also affect threshold selection. For these reasons, it is not straight forward to apply I-DT for analyzing eye-tracking data in VR, especially when targets are placed at multiple distances.

For the present work, we chose the *Minimum RMS* method for selecting fixation samples from collected data (Hessels et al., 2015; Holmqvist, 2017). With this method, a window of data with the smallest average RMS error in the combine eye angle within the HF frame of reference is selected, and the raw data within this window is used as our fixation sample. This approach is well suited for validation purposes since it does not introduce an explicit threshold but rather selects one window of data from each trial that is most likely to be identified as a fixation given a minimal set of theoretical assumptions (Holmqvist, 2017).

Several other fixation methods have been proposed, some providing refinements to I-VT or I-DT, others applying

insights from neural networks, cluster analysis, Bayesian analysis, and computer vision (Munn and Pelz, 2008; Larsson et al., 2016; Andersson et al., 2017; Sitzmann et al., 2018; Steil et al., 2018). As noted, there is still debate regarding what a fixation is and how best to identify them in data (Hessels et al., 2018). We believe it is critical that more research is done to understand how these algorithms can be implemented for VR embedded eye-tracking and how the results of fixation algorithms applied in VR contexts compare to fixations identified in other contexts. When fixations are identified in VR eye-tracking studies it is important to specify clearly both the algorithm used and relevant parameters or parameter identification methods.

3 GAZE BEHAVIORS IN 3D

Vision involves coordination between visual and motor systems in order to provide information about objects in 3D space, often resulting in an experience of 3D visual perception (Gibson, 1979; Erkelens et al., 1989; Inoue and Ohzu, 1997; Kramida, 2015; Wexler and Van Boxtel, 2005; Held et al., 2012; Blakemore, 1970; Callahan-Flintoft et al., 2021). While several systems are known to affect 3D gaze information, there is no agreed upon single mechanism that provides a primary or necessary source of depth information for the visual system (Lambooj et al., 2009; Reichelt et al., 2010; Naceri et al., 2011; Wexler and Van Boxtel, 2005; Held et al., 2012; Blakemore, 1970; Vienne et al., 2018). Understanding how visual behaviors and experiences are affected by stimulus depth is further complicated by a variety of technical challenges related to collecting 3D gaze data (Elmadjian et al., 2018; Kothari et al., 2020; Pieszala et al., 2016). In current consumer VR HMDs the visual depth is simulated and the experience of 3D is achieved by the use of binocular and motion parallax cues. The actual stimulus presentation occurs on one or more 2D screens located a few millimeters from the users eyes. The current methods of simulating 3D visual experiences in VR HMDs lead to the well known vergence-accommodation conflict, where it is assumed that the rotation of the individual eyes adjust to a simulated distance while at the same time pupils and eye lens shape adjust to the distance of the physical screen (Kramida, 2015; Vinnikov and Allison, 2014; Vinnikov et al., 2016; Iskander et al., 2019; Hoffman et al., 2008; Lanman and Luebke, 2013; Clay et al., 2019; Naceri et al., 2011; Koulieris et al., 2019). For examples of recent approaches that attempt to counter this conflict, see Kim et al. (2019); Akşit et al. (2019); Kaplanyan et al. (2019); Koulieris et al. (2019) and Lanman and Luebke (2013). The vergence-accommodation conflict is generally thought to contribute to visual fatigue but not have a significant impact on visual experiences of depth. Notably, in almost all discussions of the vergence-accommodation conflict, vergence is generally presented as accurate relative to the simulated depth of visual stimuli or non-VR conditions. For example, see **Figure 2** in Clay et al. (2019). While there is considerable discussion of the vergence-accommodation conflict in connection to consumer VR HMDs, there are almost no eye-tracking studies verifying the phenomena in these HMDs (Iskander et al., 2019). Thus, while it is known that stimulus distance affects eye behaviors, and

particularly eye angles, it is still unclear the extent to which the stimulus presentation format of VR, including hardware configuration and simulated depth cues, affect eye vergence and thus gaze depth estimates. It is therefore critical to have some insight into the reliability and validity of vergence measurements in contemporary VR systems for stimuli also beyond peripersonal space.

3.1 Gaze Depth and Data Variability

A large proportion of eye-tracking studies involve stimuli placed within 0.01–1.5 m from the participant. This placement is due, in part, to limitations in available hardware. It is also due to the fact that for depths beyond 300cm, gaze angles asymptote such that large changes in distance correspond to small changes in eye angle (Viguiet et al., 2001; Mlot et al., 2016). Moreover, there is some evidence that when the head is constrained, an individual's ability to estimate depth accurately is impaired for objects greater than 300 cm away, perhaps due to the reduction in resolution of information from eye vergence (Tresilian et al., 1999; Viguiet et al., 2001). The focus on peripersonal gaze behavior and head stabilized data collection means that there is little discussion of how eye-tracking data in free-motion and naturalistic contexts is affected when stimuli are farther away from participants (Blakemore, 1970; Erkelens et al., 1989; Viguiet et al., 2001; Naceri et al., 2011; Held et al., 2012; Vinnikov et al., 2016). This is particularly important because when the head is able to move freely, head motions may play a role in stabilizing stimuli in the visual field (Gibson, 1979; Wexler and Van Boxtel, 2005; Callahan-Flintoft et al., 2021). Geometrically this relationship entails that as the distance to the stimulus increases, smaller changes in both eye angle and head position and orientation are required to maintain fixations. Depending on how head stability changes with stimulus depth the amount of eye angle variability required to maintain fixation may change significantly with stimulus depth. Thus, the change in geometric relations may impact data quality. Without data on how head and eye behaviors adapt across stimulus presentation distances, it is unclear what to expect from eye-tracker data when stimuli are presented at a wide range of distances beyond peripersonal space. As such, given that VR often involves the possibility of presenting visual stimuli at multiple simulated depths within a single study trial, it is important to both report the range of simulated stimulus depths and consider how simulated stimulus depth may affect collected data. This may require estimating participant gaze depth, i.e. at what distance their gaze fixation is focused.

Unfortunately, estimating the actual gaze depth for specific fixations is not a straight forward process and several methods for gazed depth estimation have been proposed. These methods can be split into two relatively distinct categories: geometric (Tresilian et al., 1999; Kwon et al., 2006; Wang et al., 2014; Mlot et al., 2016; Weber et al., 2018; Wang et al., 2019; Lee and Civera, 2020) and heuristic (Duchowski et al., 2002; Clay et al., 2019; Mardanbegi et al., 2019) methods. Geometric methods primarily depend on binocular vergence and typically require gaze vectors from both eyes. For geometric methods, accuracy and precision of stimulus depth estimations for stimulus presentations at distances greater than 300 cm may be limited as vergence angles asymptote with

farther depth. Heuristic methods primarily involve a process defining a ray based on eye angle and position. Monocular or binocular eye-tracking signals may be used for heuristic estimates depending on the specific heuristics used. For both distance estimation methods, the accuracy of the estimates may be affected by both the accuracy and precision of the eye-tracking data, which may in turn be affected by the stimulus distance. Further, the accuracy of both geometric and heuristic methods may be limited if the vergence system is affected by the nearness of the 2D display in a VR HMD.

3.2 Geometric Estimation: Vergence

In simplest terms, vergence typically refers to left/right rotations of an individual's eyes. Convergence is vergence rotation of the eyes towards a single point in space. Divergence is rotation of one or both eye(s) away from a single point in space. When focusing on a stimulus in an ideal case, an individual's eyes converge to focus on the stimulus. Estimating the location of a convergence point in space in an ideal case is a simple matter of triangulation (Mlot et al., 2016; Wang et al., 2019). Thus, on the surface, vergence is perhaps the most obvious choice for estimating gaze depth. Unfortunately for researchers, eyes rarely act in an ideal way (Tresilian et al., 1999; Duchowski et al., 2002; Duchowski et al., 2014; Mlot et al., 2016; Hooge et al., 2019; Mardanbegi et al., 2019; Wang et al., 2019; Lee and Civera, 2020).

An estimation of a vergence point begins with a gaze vector or angle for each eye. In the ideal case, these values can be used to define lines intersecting at a point of focus in 3D space (Wang et al., 2019; Mlot et al., 2016). However, because both eye vectors are projected in 3D space and each eye can move independently, the gaze vectors may never actually intersect (see **Figure 4**). Small differences in vertical and/or horizontal gaze origins and angles can result in gaze vectors that only pass near one another in 3D space. As a result, vergence depth estimates typically depend on mid-point estimation methods whereby the gaze point is estimated as the midpoint of the shortest line segment of a line which runs perpendicular to both gaze rays (Mlot et al., 2016; Wang et al., 2019; Lee and Civera, 2020). However, even with a robust midpoint estimation solution, inaccuracy in both the human visual system and the eye-tracking hardware may result in larger than desired gaze depth estimation inaccuracies (Wang et al., 2019). In order to avoid these inaccuracies, while also simplifying analysis, ray casting is sometimes introduced as an alternative means of gaze depth estimation.

3.3 Heuristic Methods: Ray Casting

Ray casting is the process of simulating a ray directed along an eye-gaze vector and calculating the intersection of that ray with a visible object in the visual scene (Duchowski et al., 2002; Clay et al., 2019; Mardanbegi et al., 2019). Ray casting is analogous to identifying a gaze target as the visible object located at the fixation point in a 2D eye-tracking context. In 3D contexts, the distance of the expected gaze target from the eye origin can be used to indicate a probable gaze focus depth. As with analogous 2D methods, 3D ray casting methods can be implemented in both binocular and monocular data collection contexts. When using binocular eye-tracking hardware, ray casting methods allow for

additional validation checks, reduce data loss, and may improve accuracy because each eye can independently provide a gaze ray and corresponding estimated focus point.

While geometric methods can provide gaze depth estimates independent of the visual scene, heuristic methods require that gaze targets can be identified and localized reliably and accurately. When stimuli are presented at a variety of distances from a participant, stimuli must be scaled relative to distance and eye-tracking data quality in order for ray casting to reliably measure gaze fixation points. For example, in a system with 1° accuracy, a circular fixation target must have a minimum diameter at least 1° visual angle at the stimulus presentation depth. Otherwise the ray cast method may fail to identify the stimulus as fixated when it is fixated. As the stimulus distance is increased, the minimum diameter in metric units specified by 1° visual angle increases. In situations where it is undesirable to adjust stimuli size to viewing distance, the stimulus can be padded such that its ray intractable size is within the eye-tracking system's accuracy limits while the visual presentation of the stimuli remains unchanged. When this visual padding approach is used, the amount of padding and method for specifying additional padding should be reported. When gaze targets cannot be readily identified and/or localized, ray casting may not be a viable depth estimation option.

4 METHODS

The purpose of this pilot is to gain insight into the validity and quality of our VR HMD embedded eye-tracking system to guide further experimental design. We are particularly interested in how stimulus depth presentation can affect eye-tracking gaze data beyond peripersonal space. The purpose of presenting the methods and results of the pilot here is to provide insight into some best practices for collecting, analysing, and reporting 3D eye-gaze collected in VR, including a validation data point for a common commercial hardware/software platform (HTC Vive Pro Eye).

To this end, we set up a small experimental study which was conducted at the InteractionLab, University of Skövde. Eight participants (female = 4, male = 4, mean age = 26) from the University of Skövde were included in this study. Participants were recruited by group email and university message systems. Participation was voluntary and no compensation was provided. The project was submitted to the Ethical Review authority of Sweden (#2020-00677, Umeå) and was found to not require ethical review under Swedish legislation (2003:615). This research was conducted in accordance with the Declaration of Helsinki.

4.1 Materials and Equipment

The virtual environment was developed in Unity3D 2018.4 LTS and presented from the Unity editor using SteamVR 1.14.16. The VR HMD is an HTC Vive Pro Eye using the SteamVR 2.0 tracking system. The computer used is a Windows 10 system with an i7 3.6ghz processor, 16 gb ram, and an NVidia RTX2080 graphics card. A bare virtual office with the size of 10 m × 5 m ×

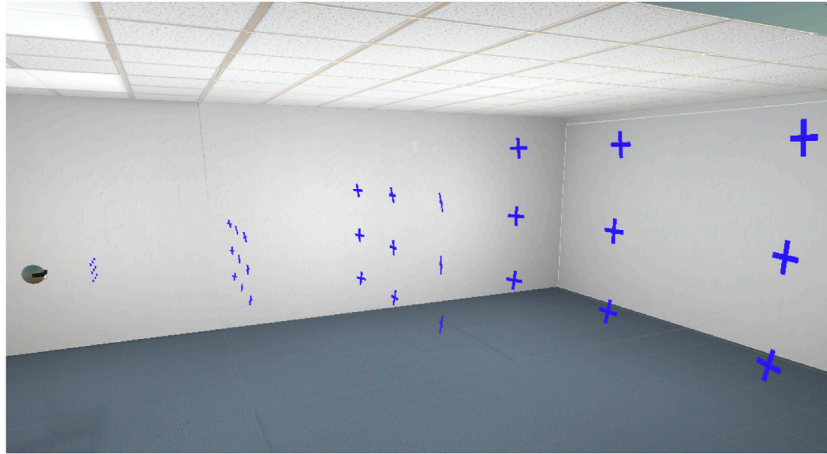


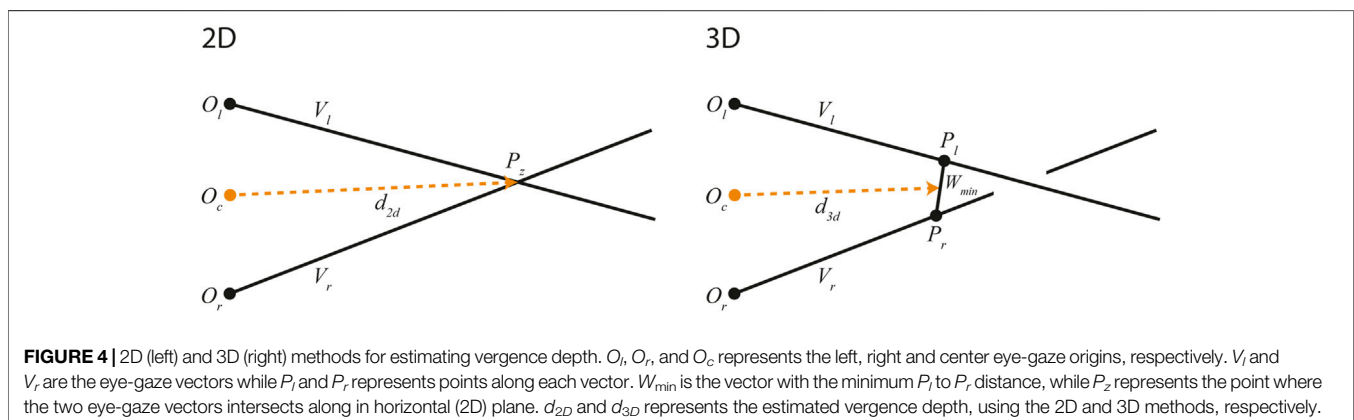
FIGURE 3 | The virtual office environment used in the study. Participant head illustrated on left. All possible stimulus targets are here presented at once. However, only a single stimulus was presented on a given trial during the study. Stimulus targets here shown 5x larger for visibility in figure.

15 m (W x H x D) was used as a naturalistic environment (see **Figure 3**). It was created using assets from the Unity asset pack *Office Megakit* (developed by Nitrousbutterfly). The participants had no visual representation of their own bodies in VR.

Eye data was collected using the SRanipal API 1.1.0.1 and eye data version 2. A custom solution was developed for ensuring data collection of both eye and position data at 120hz. To achieve this data rate for head tracking the OpenVR was queried directly instead of using the Unity provided camera values. The code for this solution, along with the entire experimental code base, has been archived on Github for reference (10.5281/zenodo.6368107). The screen frame rate of the HTC Vive was fixed at ~90fps, resulting in all Unity scene dependent visuals (including stimulus position and rotation) being presented to the participants at ~90hz. Because a lag of one to two frames can introduce small errors in gaze location we ran a preliminary investigation of system latency between SteamVR tracking system and the eye tracker using a VOR task as discussed in **Section 2**. The average latency between eye tracking and head tracking was approximately 25 ms (Collewijn and Smeets, 2000). The 120hz frame rate of the Vive Pro Eye tracking means that each eye-

tracked data frame is $8.3\overline{3}$ ms apart. Because we expected latency between head and eye movements in the VOR task to be ~10 ms, we would expect the eye tracking to always be lagged by approximately one data frame. The remaining ~17 ms of lag between eye and head data is likely attributed to internal processing time of the eye-tracker. Thus, we calculated all head and head-dependent gaze values based on head-tracking values lagged two data frames from the current eye-tracking frame (i.e. $16.6\overline{6}$ ms). This ensures that head and eye values were obtained as close in time as possible to one another. All relevant measures are reported in terms of left, right, and combined (average cyclopean) eye.

Unity and the Vive Pro Eye are designed primarily for developing engaging entertainment content. As a result several design decisions should be made explicit regarding the experimental setup. First, data collection was handled by a C# Task thread running at 120hz and independent of Unity's Update, FixedUpdate, and Coroutine loop features, which implement imprecise, variable, and (occasionally) simulated timing mechanisms. Second, most lighting effects in the virtual environment were pre-baked to ensure stability in the stimulus



presentation lighting across participants and to reduce the effects of lighting on the results (Feit et al., 2017). Following Feit et al. (2017), it is worth emphasizing that lab validation should occur using lighting conditions as similar to the experimental setup as possible because lighting can have a large effect on data validity. Third, while the stated field of view (FOV) for the HTC Vive Pro Eye is 110°, the functional FOV is constrained by both the location of the pupils relative to the VR lenses and the size of the virtual environment. Regarding pupil location, in lab pre-tests, gaze targets placed at the 110° limits (i.e. ±55° relative to gaze center) were typically visible in peripheral vision with a forward directed gaze. However, any shift in gaze towards the gaze target resulted in the target disappearing behind the HMD lens system. The difference between peripheral and direct gaze FOV was ~20–30°, with direct gaze producing a much narrower FOV than the stated FOV. Further, the wall farthest from the participant in the virtual office occupied a maximum possible visual angle of 50° × 30°. Any stimulus placed near the wall needed to be presented within those visual angle limits (i.e. ±25° horizontally and ±15° vertically relative to gaze center) or it would be placed outside of the room and not visible to the participant. Thus, we focused our validation on gaze angles within this limited FOV range.

4.2 Procedure

Upon arrival, participants were provided general information about the study and consent was obtained before continuing with data collection. After consent, the experimenter demonstrated how to adjust the VR HMD for proper fit and visual clarity. They were also shown how to adjust the interpupillary distance (IPD) of the HMD, in case they were required to do so during calibration. After task instructions were provided, participants sat in a chair and put on the HMD. Once fit and focus were adjusted, eye-gaze calibration was initiated. The SRanipal calibration, which is standard for the Vive Pro Eye HMD, was used. The calibration validated HMD positioning on the participants head and the IPD settings. Then a standard 5-point calibration sequence was presented at a single (unspecified) depth. After calibration, the participant was asked to focus on a blue cross 1 m in front of the HMD while accuracy and precision were checked in order to ensure proper calibration. If windowed average precision was consistently greater than 0.25° or if windowed average accuracy was consistently greater than 3° then calibration would be re-run. However, there were no poor calibrations according to these criteria. In the event of poor calibration, the plan was to run a second calibration, note any persistent excessive deviation and collect data with “poor” calibration. The moving data windows used for this check were 20 data samples long.

After calibration, participants were presented a virtual office environment (Figure 3). The camera, i.e., virtual head position, was initiated at 2.5 m above the floor and located so that the front wall was ~12.5 m away. The task involved fixating on 36 stimulus crosses presented one by one. Because the participant was seated, they remained roughly in this location for the duration of the study. The stimuli were arranged in three by three grid patterns at four radial distances from the participants (0.625, 2.5, 5, and

10 m). The four stimulus grids were defined with vertical columns at ±20° and 0° visual angle (i.e. azimuthal angle) and rows at ±10° and 0° visual angle (i.e. polar angle). For each trial, a randomly selected stimulus presentation position was calculated relative to the participant’s head position and orientation upon trial initialization. Each stimulus presentation position was specified by a head relative vertical and horizontal visual angle and stimulus distance. The height and width of the stimulus cross each subtended a 2° visual angle based on the participant’s head position at trial initialization and stimulus distance. This manipulation ensured that accuracy and precision were not affected by changes in visual size of the object due to distance. Notably, it also obscures some depth information such as distance dependent stimulus size differences, but not others, including information from parallax motion.

The trial was initialized using a two stage process involving two squares projected on the wall in front of the participant, each subtending ~5° visual angle. One of these squares (yellow) was aligned with the HMD orientation and placed on the wall directly in front of the HMD based on the orientation of the head. The other square (green) was fixed to the center of the wall. Initialization required first aligning one’s head so that it was oriented towards the center of the wall and the yellow square at least partially overlapped the green square. This ensured that stimulus positions would be relatively stable and remain inside the virtual room. The trial was then initiated by maintaining head forward alignment and focusing on the green square for 0.75 s, during which time both squares would fade. Prior to the start of the experiment, participants were instructed that they would initialize each trial by aligning two squares and then focusing on the green square. Participants were not told how to align the squares, in order to reduce any implications that they should consciously attempt to stabilize or center their head or eye motion during a trial.

Upon trial initialization, a blue cross was presented at a randomly selected stimulus distance and grid position (see Figure 3). During the initial instructions, participants were told that when the initialization squares disappeared, a blue cross would appear somewhere in front of them and that they were to focus on that blue cross until it disappeared. No additional instructions were provided regarding head movement in order to limit conscious stabilization of head or eye movements. If a participant exhibited a pattern of orienting their head in order to center the stimulus in their visual field, the researcher asked them to focus on the stimuli with their eyes. Only one participant received this additional instruction. The goal of this approach was to minimize the impact of participants becoming overly aware of head stabilization and also ensures that participants do not orient their heads in a way that makes all targets like the central target. Such a centering would limit the validity of the results for extreme gaze angles.

Stimulus presentation order was randomized across distance and position, and one stimulus was presented per trial. Each stimulus was presented for 3 s. Stimuli were presented in blocks comprising all 36 stimulus locations ($3_{width} \cdot 3_{height} \cdot 4_{depth}$). Participants observed three blocks with 20 s break in between each block.

TABLE 1 | Validation definitions.

Measure	Definition	General formulation
Precision	The data signal variability measured by the root mean squared difference between data sample	$\sqrt{(s_n - s_{n+1})^2}$
Accuracy	The difference between the expected state of the system and the recorded state of the system	$s_n - t_n$

n = current data sample count; s = data sample state; t = target state.

Following the experiment, participants were asked several questions regarding their experience with VR in general and with this experiment in particular. The survey included questions regarding participant comfort during the experiment and their experience of stimulus distance and size. If the participant consented, audio recordings of participant responses were made. If the participant declined audio recording, they were given the opportunity to respond to the questions in writing. Participants were then debriefed and provided further information on the purpose and structure of the study.

4.3 Definitions for Analysis

Over the past decade, there has been an increasing effort to formalize the measurements used in reporting and validating eye-tracking equipment and results (Johnsson and Matos, 2011; Feit et al., 2017; Holmqvist, 2017; Hessels et al., 2018; Niehorster et al., 2018). Because there is still a lack of consensus around some key terms, we define the relevant terms for our analysis here (Hessels et al., 2018). For all values reported in 5, vertical and horizontal components of the angles/positions are combined into a total angle/position value. The frames of reference used throughout are illustrated in **Figure 1** and described in **Section 2.2**. All eye/gaze measures are reported in terms of left, right, and combined (average cyclopean) eye. The states of the left, right, and combined eye are directly reported by the Vive SRanipal API at 120hz. Head position and orientation values were recorded from the Vive HMD through the SteamVR API at 120hz. Eye angle precision values are reported in the HF frame of reference (see **Figure 1A**). Unless specified, the remaining angular values are reported in a TF frame of reference. For these gaze values an angle of 0° indicates that the participant is looking directly at the center of the stimulus target.

Precision is a measure of variability of in the data signal. For gaze point measurements, precision is typically calculated in terms of the root mean square (RMS) of inter-sample distances in the data, see **Table 1** and Holmqvist et al. (2012) and Johnsson and Matos (2011) for details. It is important to note that while precision is often a measurement of eye-tracking quality, at least some measured imprecision is due to actual variability in eye (and head) movements (Johnsson and Matos, 2011). For VR embedded eye-tracking systems, precision can be analyzed for several different measurement variables: 1) The precision of the eye gaze independent of head position and orientation in the HF frame of reference provides insight into the hardware precision of the eye-tracker, including eye-movement variability, we refer to this as *eye angle precision*. 2) Because eye-movement variability may be influenced by changes in head position and orientation relative to the target we also report precision in a TF frame of reference including all head and

eye movements relative to the target, we refer to this as *gaze precision*. While not strictly precision, we also report the RMS error of the HMD position and orientation in a GF frame of reference and the combined HMD position and orientation angular offset relative the target in a TF frame of reference. These should provide further insight into motion tracking precision and potential additional sources of measured eye movement variability.

Accuracy is a measure of the difference between the actual state of the system and the recorded state of the system, see **Table 1** and Holmqvist et al. (2012); Johnsson and Matos (2011) for details. All accuracy validation values are reported in a TF frame of reference with the origin set at the relevant gaze origin position (e.g. left eye, right eye, combine eye). In angular units, gaze accuracy measures the angular distance of the eye gaze vector from a vector originating at the pupil and terminating at the center of a stimulus target. In metric units, gaze accuracy measures the metric distance from the center of the stimulus target to the point at the intersection of the gaze vector and stimulus target.

Fixation samples are a subset of a participant's gaze data that is directed towards the target given the task instructions. As discussed in **Section 2.4**, fixations are often provided automatically in many modern eye-tracking software packages. However, the SRanipal API used in the data collection does not provide a fixation identification algorithm. Because there is a lack of validation of fixation identification algorithms in VR, where the head is free to move and stimulus depth is not fixed, it is not clear what the impact of fixation identification method and parameterization will have on the validation values. Further, using all fixations selected by a given method may result in a subset of participants being over/under-represented in the validation results. In order to keep the focus of the pilot limited to validity of raw data the *Minimum RMS* method was used for selecting fixation samples to be used in further analysis (Hessels et al., 2015; Holmqvist, 2017). A 175 ms window of data with the smallest average RMS error in the combine eye angle in the HF frame of reference was extracted from each trial. The first 200 ms of data was skipped to give time for the participant to find the stimulus target. The window length of 175 ms follows the suggestion of Holmqvist (2017) and is also within the window size recommended for I-DT fixation identification methods (Blignaut, 2009). Within this window the eye is expected to be as still as it gets which should provide insight into the best case precision of the eye-tracking hardware, the focus of the validation pilot. By only selecting one 175 ms window per trial we ensure that all participants contribute relatively equally to the full validation data set, avoiding biases introduced by participants who fixate more often or longer than others. The short sample

window and best case precision value should also be particularly useful for setting expectations and thresholds in VR eye-tracking studies using a ray casting method in order to identify discrete gaze behaviors in a manner similar to area of interest methods in other screen based eye-tracking contexts (Duchowski et al., 2014; Alghofaili et al., 2019; Clay et al., 2019; Mardanbegi et al., 2019).

Vergence depth estimates provide an estimate of the distance from a participant to a point in space that they are focusing on. Gaze depth is still not a common metric for eye-tracker validation, but in naturalistic visual environments with stimuli presented at varying distances, it is becoming an increasingly important metric. For the current study we used a heuristic distance estimation method for real-time measurement of depth dependent variables, e.g. gaze position in the global coordinate space. However, we also investigate the relationship between stimulus presentation distance and two common geometric estimation techniques, one estimating gaze depth in 3D and one using a simpler, 2D estimate (see e.g., Duchowski et al., 2014; Mlot et al., 2016; Lee and Civera, 2020). The 3D method was proposed by Hennessey and Lawrence (2009) (Figure 4, right). With this method, the estimated vergence depth is the distance from the combined gaze origin to the mid-point of vector (W_{\min}) (W_{\min}) is the vector with the minimum Euclidean distance between any two points (P_L) and (P_R) along the left and right gaze vectors, respectively. We also considered a 2D estimate (Figure 4, left) where gaze depth was calculated on a plane defined by the forward vector of the HMD and the vector between the left and right eyes. This plane is aligned with the HMD and eye orientation, including rotation around the HMD's forward axis. Using this method, the estimated vergence depth is the distance from the combined gaze origin to the intersection point (in 2D) of the left and right gaze vectors. Both methods use a HF frame of reference.

4.4 Exclusion Criteria

For analysis, two participants were excluded, both due to a high number of trials with excessive head rotations. One of these participants was provided an additional instruction to focus on the target with their eyes as indicated in Section 4.2. Excessive head rotations could occur for targets placed at a visual angle not equal to zero degrees on either the vertical or horizontal task axes (i.e. targets placed at $\pm 10^\circ$ and $\pm 20^\circ$ respectively). In these cases, excessive head rotation towards the target has the effect of centering the target in HMD view, reducing eye angle required for focusing on the target and making it more like the central target. During identification of fixation samples for the non-central target, the average and max angle of the HMD in the TF frame of reference was calculated. Fixation samples were excluded if the absolute average angular distance from HMD to target for a given axis exceeded 25% of angular placement (e.g. 5° on the horizontal axis) or had a max absolute angular distance greater than 50% of the angular placement (e.g. 10° on the horizontal axis) at any point in the candidate fixation samples. Participants who had more than 25% of their total trials excluded due to no valid windows according to this criteria were excluded. For the remaining six participants less than 1% of all trials were excluded according to this criteria. Additionally, frames in which the

combine eye measurements were marked as invalid by SRanipal, indicating a loss of tracking or blink, were excluded from analysis. Invalid frames constituted 1.67% of data frames. Of the remaining participants (female = 4, male = 2, mean age = 24) are included in analysis.

5 RESULTS

Gaze fixation points with 95% confidence intervals for each target grid location and target distance are presented in Figure 5. Target accuracy for each eye and target distance is reported in Table 2 and mean accuracy values for each distance are visualized in Figure 6. Accuracy was calculated according to equations found in Table 1.

A within subjects 3 (eye: left, right, combine) x 4 (target distance: 0.625, 2.5, 5, 10 m) repeated measures ANOVA revealed a significant main effect of target distance ($F(3, 15) = 9.451, p < 0.001, \eta^2 = 0.530$). No significant main effect for eye ($p = 0.069$) or interaction between eye and target distance ($p = 0.838$) was indicated. Bonferroni post-hoc tests indicated that gaze angle estimates were significantly more accurate for the non-peripersonal target distances, 2.5 m ($M = 1.139, SD = 0.299; p = 0.003$), 5 m ($M = 1.204, SD = 0.448; p = 0.007$), and 10 m ($M = 1.096, SD = 0.345; p = 0.002$) than the gaze estimates at the peripersonal target distance of 0.625 m ($M = 1.984, SD = 0.937$). No significant accuracy differences were indicated between non-peripersonal target distances.

Precision values for both eye angle (eye-in-head in a HF frame of reference) and gaze (head + eye angle in a TF frame of reference) are presented in Table 3. No significant differences were found for either precision measure across eye or distance. Head movement variability is presented in Table 4. No significant differences were identified between target presentation distances.

The mean and median vergence depth estimates, including the interquartile range, are illustrated in Figure 7 (2D method) and Figure 8 (3D method). Estimates are presented for individual target locations (lower plots) and all targets (upper plot) of each figure. For some data frames, no valid distance solution could be calculated during the selected fixation. The percentages of excluded data frames at each target distance are overlaid in both figures. Invalid solutions were due to left and right eye angles summing to greater than 180° for the 2D estimates and for the 3D method when the minimum distance between gaze rays was at the gaze origin. This occurs when gaze vectors are parallel or divergent. Excluded cases can be due to individual eye physiology and/or noise in the eye data recordings.

According to the 2D estimate, the mean and median estimated vergence depths were beyond the actual target distance across all targets at 0.625 m ($M = 1.134, \text{Median} = 0.677, Q1 = 0.571, Q3 = 1.101$). At 2.5 m the mean estimate was beyond the target distance, but the median was nearer ($M = 3.194, \text{Median} = 1.876, Q1 = 1.257, Q3 = 2.987$). At both 5 m ($M = 3.890, \text{Median} = 2.443, Q1 = 1.567, Q3 = 4.481$) and 10 m ($M = 4.279, \text{Median} = 2.547, Q1 = 1.700, Q3 = 5.144$) mean and median estimates were nearer than the actual target distance.

According to the 3D method, both the mean and median estimated vergence depths were beyond the actual target

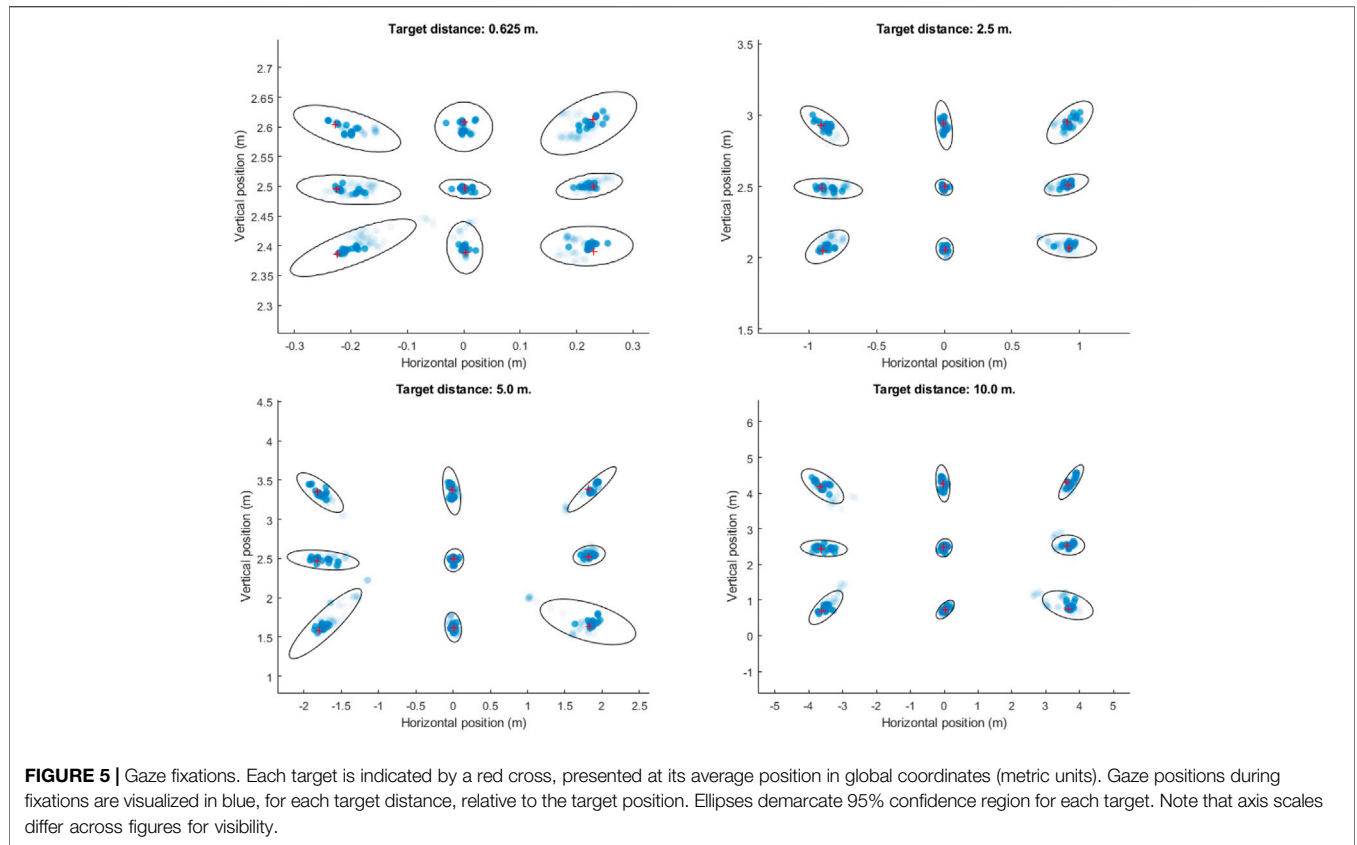


TABLE 2 | Gaze accuracy.

Distance	Eye	Angular accuracy ^{a,b}		Position accuracy ^{a,c}	
0.625	Combine	1.788	(1.525)	1.837	(1.566)
	Left	1.994	(1.608)	2.053	(1.657)
	Right	2.104	(1.702)	2.167	(1.770)
2.5	Combine	0.998	(0.766)	4.327	(3.302)
	Left	1.296	(0.988)	5.633	(4.301)
	Right	1.123	(0.739)	4.884	(3.178)
5	Combine	1.053	(1.076)	9.235	(9.482)
	Left	1.316	(1.224)	11.559	(10.813)
	Right	1.180	(0.983)	10.376	(8.654)
10	Combine	0.952	(0.694)	16.738	(12.189)
	Left	1.249	(0.931)	22.024	(16.479)
	Right	1.090	(0.705)	19.214	(12.334)

^aMean and (SD)

^bGaze angle (Eye angle + HMD, relative to target) in degrees. 0° gaze is directed exactly at target center. Target subtends 1° visual angle

^cGaze position estimate at target distance in cm. 0cm is a gaze at the center of the target. Target size in cm scales with distance.

distance across all targets at 0.625 m (M = 0.941, Median = 0.678, Q1 = 0.583, Q3 = 1.035). Mean and median estimated vergence depths were nearer than the actual target distance for targets at 2.5 m (M = 2.042, Median = 1.608, Q1 = 1.113, Q3

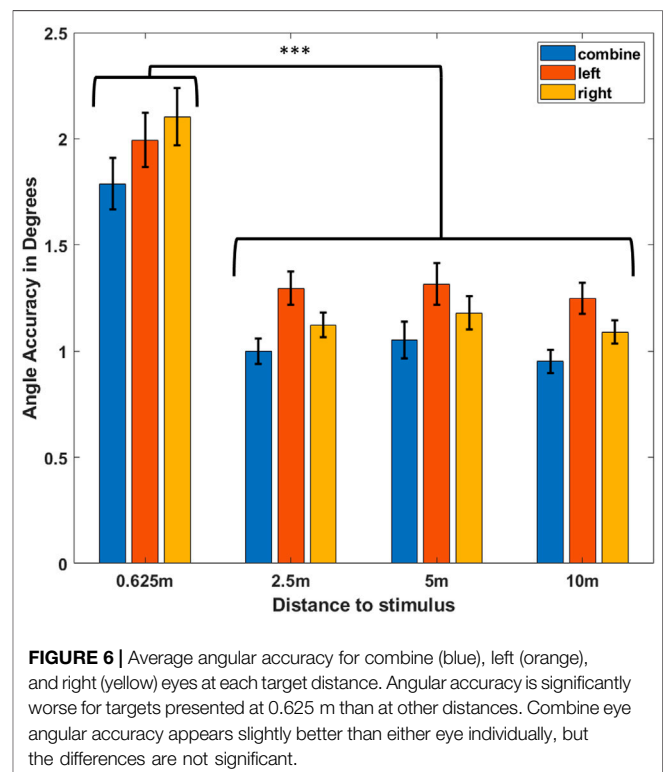


TABLE 3 | Eye and gaze precision values.

Distance	Eye	Eye angle precision ^a		Gaze precision ^b	
0.625	Combine	0.108	(0.198)	0.111	(0.214)
	Left	0.108	(0.194)	0.106	(0.189)
	Right	0.114	(0.216)	0.112	(0.208)
2.5	Combine	0.088	(0.169)	0.085	(0.157)
	Left	0.092	(0.167)	0.092	(0.166)
	Right	0.100	(0.180)	0.100	(0.176)
5	Combine	0.083	(0.161)	0.082	(0.154)
	Left	0.095	(0.176)	0.096	(0.175)
	Right	0.102	(0.192)	0.102	(0.186)
10	Combine	0.084	(0.167)	0.083	(0.164)
	Left	0.088	(0.163)	0.088	(0.160)
	Right	0.101	(0.197)	0.101	(0.193)

All values in degrees. Mean and (SD).

^aHF_C, frame of reference, eye-in-head movement independent of target distance.

^bTF_C, frame of reference includes both eye and head movement, relative to target distance.

= 2.235), 5 m (M = 2.447, Median = 1.802, Q1 = 1.165, Q3 = 2.936) and 10 m (M = 2.703, Median = 1.743, Q1 = 1.124, Q3 = 3.310).

In the post-study interview, three participants indicated that they had no experience with VR in the previous year, two indicated 2–10 h and one indicated more than 50 h of experience with VR in the previous year. No participants indicated discomfort or sickness with the VR environment in the current pilot, though the task was considered boring by some participants.

6 DISCUSSION

Eye tracking in VR shares many features of eye tracking in other contexts. However, the relatively unconstrained nature of moving in VR and the simulated 3D nature of visual stimuli introduce some important considerations for using this technology. We've discussed some of the core differences throughout this article and proposed some best practices for collecting, analysing, and reporting the eye-gaze in VR. As a part of this work we conducted a small pilot study focusing specifically on the effects of stimulus presentation distance and its effect on accuracy and precision of collected data. The results from the

pilot study indicate that the quality of the data collected in VR is in line with that of many other eye-tracking techniques available to researchers (Holmqvist, 2017). Overall, the eye-tracking equipment used in the present pilot study (HTC Vive Pro Eye) performed well, though there are clear compromises in terms of data accessibility that come from working with a non-research system. Specifically, we identified two larger challenges that should be considered when working with eye-tracking in VR:

1. The accuracy and precision of collected data depends on stimulus distance and
2. Vergence is a problematic method for estimating gaze depth, and should be used with caution.

Both of these challenges are outlined in more detail below. In many cases we believe it would be sufficient to report accuracy and gaze precision for the minimum, average, and maximum stimulus distances while describing equipment in the methods section. While there are some unique challenges in developing for and analyzing data from eye-tracking systems embedded in consumer VR HMDs, we do believe that the collected data is good enough for many research contexts, and we hope that the present work will support further use of eye tracking in VR.

6.1 Accuracy and Precision in Depth

In the pilot, stimulus presentation distance affected the overall accuracy of the eye-tracking results. Perhaps contrary to intuitions, angular accuracy improves for stimulus distances beyond peripersonal space. This creates an interesting challenge for studies that involve stimuli presented both within and beyond peripersonal space. Any resulting significant differences in gaze data at different presentation depths may be a result of differences in data quality and not gaze behavior. Studies that involve multiple stimulus depths should report hardware accuracy across the presentation depth range as well as note any significant or potential differences in eye-tracking quality at different depths. It is critical in these cases that accuracy validation is not based on a single depth or average of multiple depths. Whenever there is a significant difference in accuracy of the eye-tracking system across presentation depths, it should be accounted for in study design and/or data analysis and interpretation. When possible, it may be best to avoid comparisons across certain distances, particularly when

TABLE 4 | HMD movement.

Distance	Angular target offset variability ^{a,b,c,d}	HMD position variability ^{a,c,e}	HMD Orientation variability ^{a,c,d}
0.625	0.107 (0.202)	0.0094 (0.0164)	0.025 (0.095)
2.5	0.046 (0.094)	0.007 (0.0127)	0.008 (0.057)
5	0.037 (0.080)	0.0074 (0.0124)	0.021 (0.084)
10	0.036 (0.078)	0.0073 (0.0119)	0.030 (0.096)

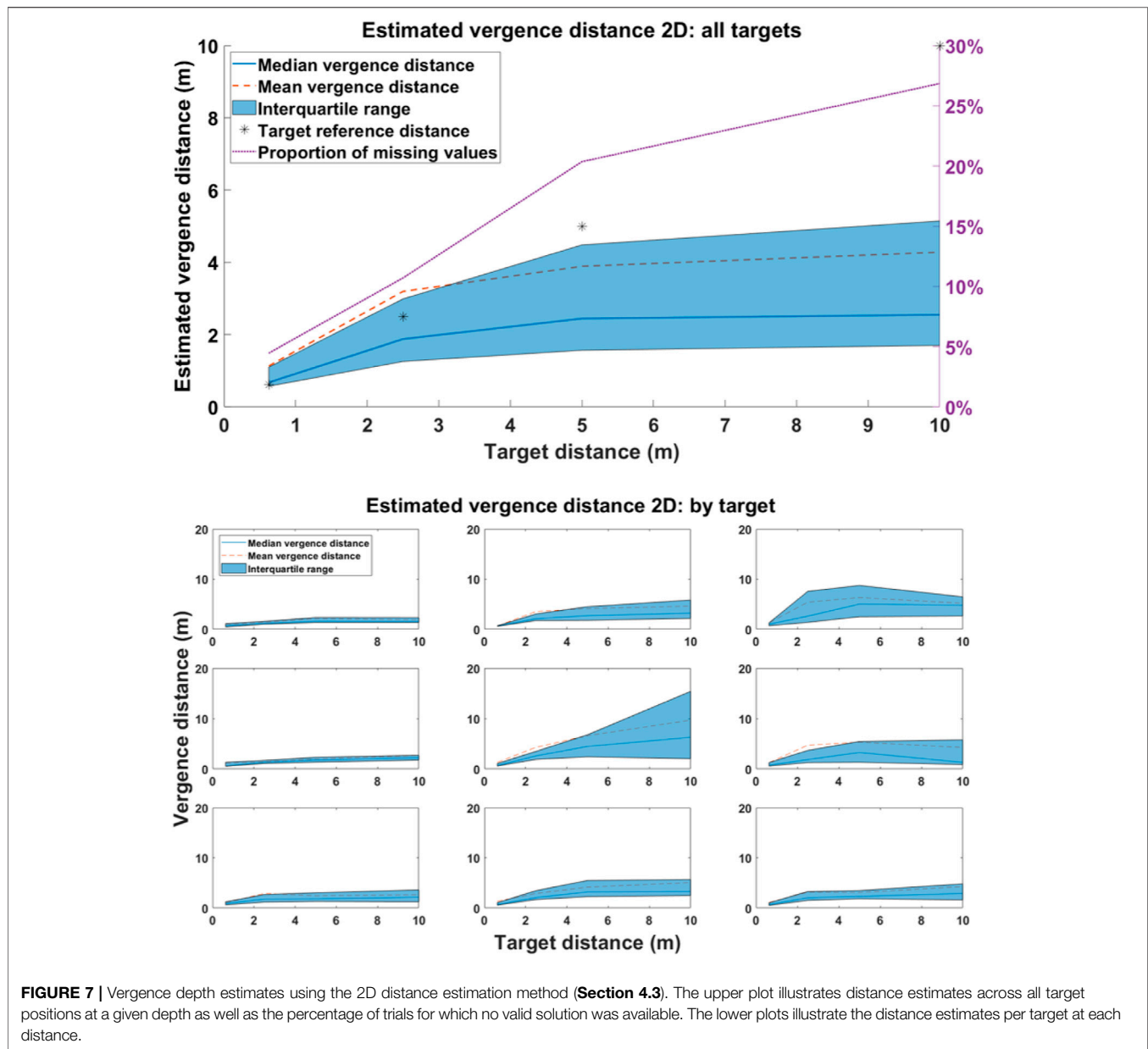
^aVariability is measured as mean sample to sample RMS, difference during a fixation sample. Mean and (SD)

^bTF, frame of reference

^cGF, frame of reference

^dvalues in degrees

^evalues in cm.

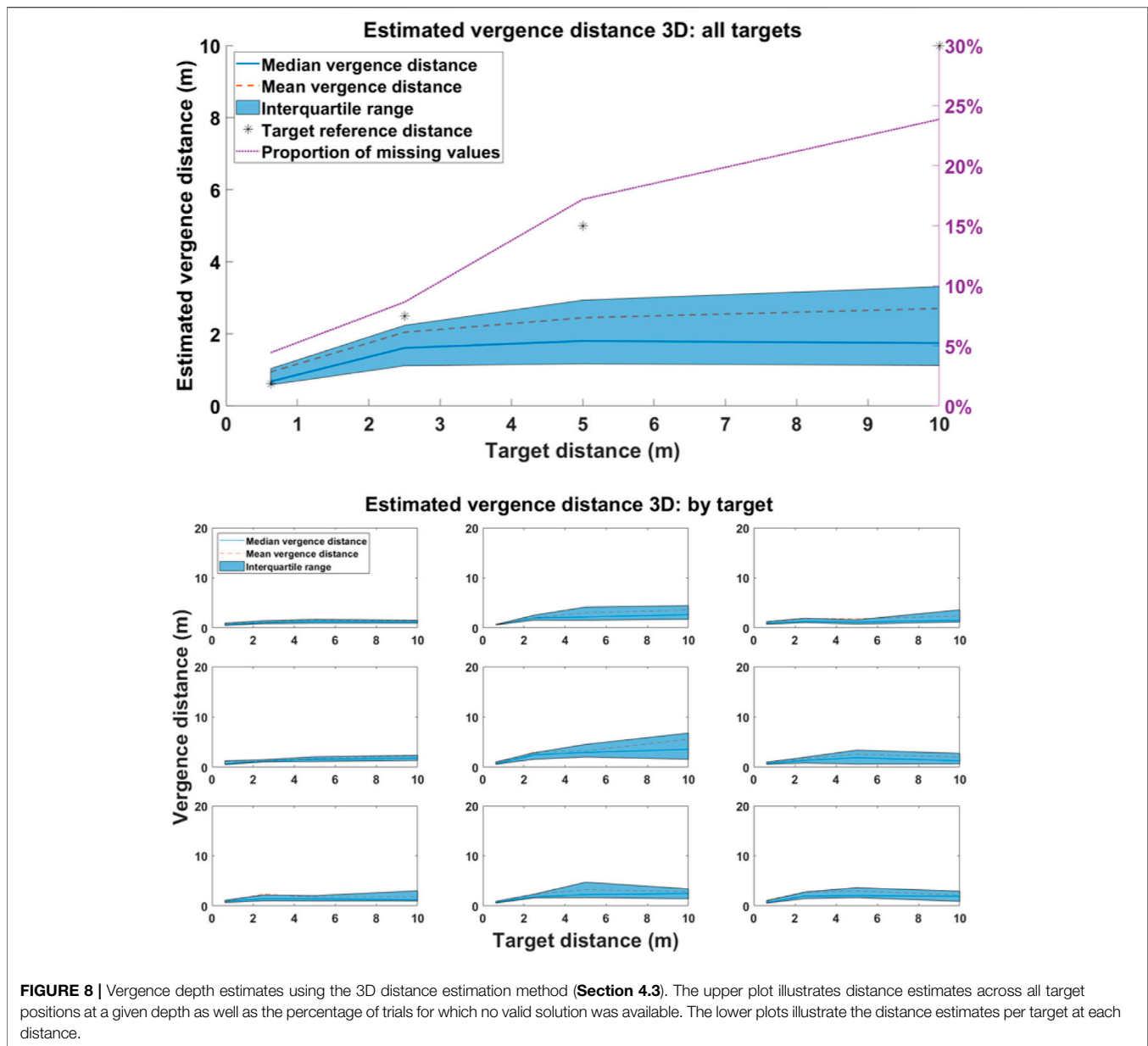


making comparisons between gaze targets within peripersonal space and beyond peripersonal space. A future study should take a more granular approach to presentation depths, as the current pilot focused on a wide range of depths beyond peripersonal space and may obscure important insights.

In the current pilot study, we used the default eye tracker calibration because this is the most likely calibration method to be used in typical lab contexts. It is possible that a modified calibration process could eliminate the observed accuracy differences across stimulus depth while yielding better overall data quality (Nyström et al., 2013; Wibirama et al., 2017; Elmadjian et al., 2018; Weber et al., 2018). The default calibration may be limited for at least two reasons. First, it occurs at a single stimulus presentation depth. Because the assumed calibration depth is not reported in the hardware or

software documentation, it is not clear if stimuli near the calibration depth had better or worse accuracy. A second factor to consider is that the calibration stimuli were fixed to the head so that head movement does not seem to be considered in the calibration process. Allowing the calibration targets to be independent of the head may provide better information for more robust calibration methods. Future work should aim to improve the calibration process for multi-depth paradigms to reduce data differences due to presentation depth. Custom or modified calibration processes should be reported and when possible documented and shared to ensure reproducibility.

While the pilot indicated no significant difference between validation values across eyes (left, right, combine), there is a clear, but non-significant, tendency for the combine eye to provide better accuracy values. Lab validations should pay attention to



this tendency in case it is reduced or amplified in specific system and stimulus contexts or given a larger sample size. While our pilot study indicated no significant difference between left and right eye accuracy and combine eye accuracy, there may be a greater difference in some setups. Eye accuracy differences may be influenced further by the amount of inaccuracy in the eye vergence relative to stimulus presentation depth.

When possible, it is likely best to choose a specific eye for all data collection before starting the study in order to ensure data consistency. In our setup based on the pilot, the combine eye seems preferable. The combine eye is calculated based on both eye gaze vectors, so combine eye values requires data from both eyes simultaneously. In our pilot, the combine eye suffered relatively little data loss e. g due to single eye blinks or occlusion. There may be reasons for preferring individual eye measurements in specific

setups, e.g. when looking at dominant eye behaviors or monocular stimulus presentations. In the default SRanipal code example, the combine eye data is used by default. However, if combine eye data is not available, the system automatically attempts to use single eye data (left or right), before returning no gaze values. While this is a good solution in some contexts, e.g. games and user interactions, it could introduce unaccounted variation in the data when used for research. Again, it is important that such failsafe assumptions are identified and accounted for when reporting results of VR based eye-tracking studies.

6.2 Vergence Depth Estimates

While there have been indications that vergence is not a good estimate of depth in at least some cases (Tresilian et al., 1999;

Hooge et al., 2019; Wang et al., 2019), vergence is often presented in VR eye-tracking literature as at least informative of gaze depth (Hoffman et al., 2008; Lanman and Luebke, 2013; Vinnikov and Allison, 2014; Kramida, 2015; Mlot et al., 2016; Clay et al., 2019; Iskander et al., 2019). However, in our pilot, vergence was not only an unreliable indicator of likely gaze depth, it drastically underestimated gaze depth for stimuli outside of peripersonal space. While this pilot was small, the effect was quite clear and robust across participants. Further research is required to get a better picture of distance related vergence behavior in VR. However, these results suggest that vergence behaviors or their measurement in VR may not generalize well to 3D gaze behavior in non-virtual spaces.

Given the size of the current pilot and the lack of non-VR data for comparison, the poor depth estimates cannot be definitively attributed to either a data quality issue or actual eye vergence behavior. One potential source of error due to data quality could be the exponentially increasing effect of eye angle variability on depth estimates for increasingly distant stimulus targets. However, the overall interquartile range of distance estimates does not appear to grow exponentially across peripersonal presentation depths, indicating that the inaccuracy is not due solely to greater eye angle variability. The lack of multi-depth calibration and the suppression of size-distance depth cues may have been contributing data quality factors in reducing the accuracy of vergence in the pilot. Both of these factors deserve deeper investigation. Moreover, it seems that, at least in the current context, the VR format could cause vergence to be less accurate than in non-VR presentation contexts. Perhaps due to a biasing of vergence towards the screen, which causes vergence to “max out” prematurely in VR. If this is case, it may require a more serious investigation of the vergence-accommodation conflict, with a particular focus on how similar vergence behavior in VR is to vergence behavior outside of VR. It may at least be important to be more cautious when presenting the vergence-accommodation conflict as it relates to VR in order to avoid implying that vergence mechanisms function similar to non-VR contexts. The inaccuracy of vergence may even hint at a novel source of visual fatigue dependent only on vergence inaccuracy.

Given the observed inaccuracy and variability of vergence in the current pilot, we would urge caution in assuming that vergence-based gaze depth estimates could be reliable in VR or similar technologies. While improved calibration techniques might improve the results, the availability and ease of implementing ray-casting techniques may generally render vergence based gaze depth estimates unnecessary in most cases. When possible, ray casting and similar heuristic methods should be used to estimate gaze depth in VR.

6.3 Limitations

We have attempted to be upfront with the many limitations of the presented pilot study, however we feel it is important to end with clear specification of the limitations along with related considerations. First, the number of participants is quite low for an eye-tracking study, and recommendations are to have at least 50–70 participants (King et al., 2019). For

internal lab validation and testing of smaller experimental setups may pass with fewer participants when paired with manufacturer’s specifications and insights from other validation studies. A second limitation is that participants were not allowed to fully move their heads, as their data was eliminated if it included excessive head movements (as defined in **Section 4.4**). This limitation served the purpose of validating extreme gaze angles in the VR environment but may obscure affects of large or continuous head movements on data validity (Callahan-Flintoft et al., 2021; Gibson, 1979; Niehorster et al., 2018; Hessels et al., 2015). Along with the recommendation that eye-tracking hardware and software should be validated in the local lab environment, it is also important to consider specific experimental design consequences on data validity (Orquin and Holmqvist, 2018). This may involve many factors, including the amount of head and body motion the participant is allowed during a trial. It may be relevant for some studies to report the percentage of gazes taking place near the center of the visible space and how much the head was utilized during fixation. Finally, the current discussion and presentation is focused on precision and accuracy of gaze data. Saccade and fixation identification is often considered important in eye-tracking studies (Kowler, 2011). These metrics can be extremely useful when used correctly and well defined in the specific research context. However, as indicated in **Section 2.4**, technical definitions of these behaviors and the algorithms for detecting them in data are not yet established and validated for likely VR eye-tracking use cases. In order to maintain a manageable scope, we did not include a measurement and validation of saccade behavior and chose to include only a limited discussion of fixation identification. Saccade behavior can also be quite complex and, as with fixation, there is ongoing debate as to what counts as a saccade and what frame rates are best suited to capture them (Kowler, 2011; Andersson et al., 2017; Hessels et al., 2018). There is a critical need for theory, definitions, algorithms, and guidance to support analysis of fixation and saccade behaviors in head mounted eye-tracking systems used in naturalistic environments as well as virtual environments such as those presented in VR.

7 CONCLUSION

In this paper, we have discussed some of the key issues and terms related to eye-tracking in VR, with a focus on recent consumer VR hardware and comparisons to more traditional 2D eye-tracking paradigms. Eye tracking with VR systems allows for less constrained participant motion and involves simulated 3D stimuli which allow for greater flexibility and novel experimental design. These features also mean that the data produced by VR embedded eye trackers is not directly comparable to eye tracking results collected in the majority of non-VR behavioural research contexts. Rapid changes in VR and eye-tracking hardware have made it difficult for validation and best practices to keep up with current

technology. We have focused on general insights as opposed to a deep dive into a specific hardware platform. However, we have looked at one of the more common current platforms in order to illustrate how individual labs should proceed in validating their local hardware and software context. Future discussions of specific hardware validation and research contexts may provide clearer insights into general hardware and software quality. For now, behavioral researchers hoping to leverage new low-cost VR embedded eye trackers should proceed with caution, validate their specific lab setup and report relevant validation results and related experimental design accommodations.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. The software used for data collection in this project can be found at <https://doi.org/10.5281/zenodo.6368107>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethical Review authority of Sweden. The patients/participants provided their written informed consent to participate in this study.

REFERENCES

- Aksit, K., Chakravarthula, P., Rathinavel, K., Jeong, Y., Albert, R., Fuchs, H., et al. (2019). Manufacturing Application-Driven Foveated Near-Eye Displays. *IEEE Trans. Visualization Comput. Graphics* 25, 1928–1939. doi:10.1109/TVCG.2019.2898781
- Alghofaili, R., Solah, M. S., and Huang, H. (2019). "Optimizing Visual Element Placement via Visual Attention Analysis," in Proceedings of the 26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings (Institute of Electrical and Electronics Engineers Inc.), Osaka, Japan, August 2019, 464–473. doi:10.1109/VR.2019.8797816
- Andersson, R., Larsson, L., Holmqvist, K., Stridh, M., and Nyström, M. (2017). One Algorithm to Rule Them All? an Evaluation and Discussion of Ten Eye Movement Event-Detection Algorithms. *Behav. Res. Methods* 49, 616–637. doi:10.3758/s13428-016-0738-9
- Angelopoulos, A. N., Martel, J. N., Kohli, A. P., Conrath, J., and Wetzstein, G. (2021). Event-Based Near-Eye Gaze Tracking beyond 10,000 Hz. *IEEE Trans. Visualization Comput. Graphics* 27, 2577–2586. doi:10.1109/TVCG.2021.3067784
- Armbrüster, C., Wolter, M., Kuhlen, T., Spijkers, W., and Fimm, B. (2008). Depth Perception in Virtual Reality: Distance Estimations in Peri- and Extrapersonal Space. *Cyberpsychology Behav.* 11, 9–15. doi:10.1089/cpb.2007.9935
- Aw, S. T., Haslwanter, T., Halmagyi, G. M., Curthoys, I. S., Yavor, R. A., and Todd, M. J. (1996). enThree-Dimensional Vector Analysis of the Human Vestibuloocular Reflex in Response to High-Acceleration Head Rotations. I. Responses in normal Subjects. *J. Neurophysiol.* 76, 4009–4020. doi:10.1152/jn.1996.76.6.4009
- Binaee, K., and Diaz, G. (2019). Movements of the Eyes and Hands Are Coordinated by a Common Predictive Strategy. *J. Vis.* 19, 1–16. doi:10.1167/19.12.3
- Blakemore, C. (1970). The Range and Scope of Binocular Depth Discrimination in Man. *J. Physiol.* 211, 599–622. doi:10.1113/jphysiol.1970.sp009296

AUTHOR CONTRIBUTIONS

All authors made critical and significant contributions to this paper and were central to its development. ML, EB, MB, and EPL contributed to research methods design and development and text editing and revision. ML, MB, and EPL contributed to data collection. ML and EB contributed to data processing and analysis. ML contributed to primary text generation. EB and ML contributed to project coordination.

FUNDING

Funding of this project was provided through the Knowledge Foundation as a part of both the Recruitment and Strategic Knowledge Reinforcement initiative and within the Synergy Virtual Ergonomics (SVE) project (#20180167).

ACKNOWLEDGMENTS

We want to thank the Knowledge Foundation and the associated INFINIT research environment at the University of Skövde for support through funding of both the Recruitment and Strategic Knowledge Reinforcement initiative and within the Synergy Virtual Ergonomics (SVE) project. This support is gratefully acknowledged.

- Blignaut, P. (2009). enFixation Identification: The Optimum Threshold for a Dispersion Algorithm. *Attention, Perception, & Psychophysics* 71, 881–895. doi:10.3758/APP.71.4.881
- Borges, M., Symington, A., Coltin, B., Smith, T., and Ventura, R. (2018). "HTC Vive: Analysis and Accuracy Improvement," in IEEE International Conference on Intelligent Robots and Systems, Madrid, Spain, January 2019 (Piscataway, New Jersey, United States: IEEE), 2610–2615. doi:10.1109/IROS.2018.8593707
- Brookes, J., Warburton, M., Alghadier, M., Mon-Williams, M., and Mushtaq, F. (2020). Studying Human Behavior with Virtual Reality: The Unity Experiment Framework. *Behav. Res. Methods* 52, 455–463. doi:10.3758/s13428-019-01242-0
- Callahan-Flintoft, C., Barentine, C., Touryan, J., and Ries, A. J. (2021). A Case for Studying Naturalistic Eye and Head Movements in Virtual Environments. *Front. Psychol.* 12, 650693. doi:10.3389/fpsyg.2021.650693
- Carter, B. T., and Luke, S. G. (2020). Best Practices in Eye Tracking Research. *Int. J. Psychophysiology* 119, 49–62. doi:10.1016/j.ijpsycho.2020.05.010
- Chang, C., Cui, W., Park, J., and Gao, L. (2019). enComputational Holographic Maxwellian Near-Eye Display with an Expanded Eyebox. *Scientific Rep.* 9, 18749. doi:10.1038/s41598-019-55346-w
- Clay, V., König, P., and König, S. U. (2019). Eye Tracking in Virtual Reality. *J. Eye Move. Res.* 12 (1). doi:10.16910/jemr.12.1.3
- Collewijn, H., and Smeets, J. B. (2000). Early Components of the Human Vestibulo-Ocular Response to Head Rotation: Latency and Gain. *J. Neurophysiol.* 84, 376–389. doi:10.1152/jn.2000.84.1.376
- Deb, S., Carruth, D. W., Sween, R., Strawderman, L., and Garrison, T. M. (2017). Efficacy of Virtual Reality in Pedestrian Safety Research. *Appl. Ergon.* 65, 449–460. doi:10.1016/j.apergo.2017.03.007
- Duchowski, A. T., House, D. H., Gestring, J., Congdon, R., Świrski, L., Dodgson, N. A., et al. (2014). "Comparing Estimated Gaze Depth in Virtual and Physical Environments," in Proceedings of the Eye Tracking Research and Applications Symposium (ETRA), Safety Harbor Florida, March 2014, 103–110. doi:10.1145/2578153.2578168

- Duchowski, A. T., Medlin, E., Cournia, N., Murphy, H., Gramopadhye, A., Nair, S., et al. (2002). 3-D Eye Movement Analysis. *Behav. Res. Methods Instr. Comput.* 34, 573–591. doi:10.3758/BF03195486
- Duchowski, A. T., Pelfrey, B., House, D. H., and Wang, R. (2011). “Measuring Gaze Depth with an Eye Tracker during Stereoscopic Display,” in Proceedings of the APGV 2011: ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization, Toulouse France, August 2011, 15–22. doi:10.1145/2077451.2077454
- Elmadjian, C., Shukla, P., Tula, A. D., and Morimoto, C. H. (2018). “3D Gaze Estimation in the Scene Volume with a Head-Mounted Eye Tracker,” in Proceedings of the COGAIN 2018: Communication by Gaze Interaction, Warsaw Poland, June 2018, 1–9. doi:10.1145/3206343.3206351
- Erkelens, C. J., Steinman, R. M., and Collewijn, H. (1989). Ocular Vergence under Natural Conditions. II. Gaze Shifts between Real Targets Differing in Distance and Direction. *Proc. R. Soc. B: Biol. Sci.* 236, 441–465. doi:10.1098/rspb.1989.0031
- Feit, A. M., Williams, S., Toledo, A., Paradiso, A., Kulkarni, H., Kane, S., et al. (2017). “Toward Everyday Gaze Input: Accuracy and Precision of Eye Tracking and Implications for Design,” in Proceedings of the 2017 Conference on Human Factors in Computing Systems, Denver Colorado USA, 2017-May, 1118–1130. doi:10.1145/3025453.3025599
- Feldman, A. G., and Zhang, L. (2020). Eye and Head Movements and Vestibulo-Ocular Reflex in the Context of Indirect, Referent Control of Motor Actions. *J. Neurophysiol.* 124, 115–133. doi:10.1152/jn.00076.2020
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA: Houghton-Mifflin.
- Harris, D. J., Bird, J. M., Smart, P. A., Wilson, M. R., and Vine, S. J. (2020). A Framework for the Testing and Validation of Simulated Environments in Experimentation and Training. *Front. Psychol.* 11, 605. doi:10.3389/fpsyg.2020.00605
- Harris, D. J., Buckingham, G., Wilson, M. R., and Vine, S. J. (2019). Virtually the Same? How Impaired Sensory Information in Virtual Reality May Disrupt Vision for Action. *Exp. Brain Res.* 237, 2761–2766. doi:10.1007/s00221-019-05642-8
- Held, R. T., Cooper, E. A., and Banks, M. S. (2012). Blur and Disparity Are Complementary Cues to Depth. *Curr. Biol.* 22, 426–431. doi:10.1016/j.cub.2012.01.033
- Hennessey, C., and Lawrence, P. (2009). Noncontact Binocular Eye-Gaze Tracking for point-of-gaze Estimation in Three Dimensions. *IEEE Trans. Biomed. Eng.* 56, 790–799. doi:10.1109/TBME.2008.2005943
- Hessels, R. S., Cornelissen, T. H., Kemner, C., and Hooge, I. T. (2015). Qualitative Tests of Remote Eyetracker Recovery and Performance during Head Rotation. *Behav. Res. Methods* 47, 848–859. doi:10.3758/s13428-014-0507-6
- Hessels, R. S., Niehorster, D. C., Nyström, M., Andersson, R., and Hooge, I. T. (2018). Is the Eye-Movement Field Confused about Fixations and Saccades? A Survey Among 124 Researchers. *R. Soc. Open Sci.* 5, 180502. doi:10.1098/rsos.180502
- Hoffman, D. M., Girshick, A. R., Akeley, K., and Banks, M. S. (2008). Vergence-accommodation Conflicts Hinder Visual Performance and Cause Visual Fatigue. *J. Vis.* 8, 1–30. doi:10.1167/8.3.33
- Holmqvist, K. (2017). “Common Predictors of Accuracy, Precision and Data Loss In 12 Eye-Trackers,” in Proceedings of the 7th Scandinavian Workshop on Eye Tracking, Turku, Finland, 2017.
- Holmqvist, K., Nyström, M., and Mulvey, F. (2012). “Eye Tracker Data Quality: What It Is and How to Measure it,” in Proceedings of the Eye Tracking Research and Applications Symposium (ETRA), Santa Barbara California, March 2012, 45–52. doi:10.1145/2168556.2168563
- Holzwarth, V., Gisler, J., Hirt, C., and Kunz, A. (2021). “Comparing the Accuracy and Precision of SteamVR Tracking 2.0 and Oculus Quest 2 in a Room Scale Setup,” in Proceedings of the 2021 the 5th International Conference on Virtual and Augmented Reality Simulations (ACM), Melbourne, VIC, Australia, March 2021, 42–46. doi:10.1145/3463914.3463921
- Hooge, I. T., Hessels, R. S., and Nyström, M. (2019). Do pupil-based Binocular Video Eye Trackers Reliably Measure Vergence? *Vis. Res.* 156, 1–9. doi:10.1016/j.visres.2019.01.004
- Inoue, T., and Ohzu, H. (1997). Accommodative Responses to Stereoscopic Three-Dimensional Display. *Appl. Opt.* 36, 4509. doi:10.1364/ao.36.004509
- Iorizzo, D. B., Riley, M. E., Hayhoe, M., and Huxlin, K. R. (2011). Differential Impact of Partial Cortical Blindness on Gaze Strategies when Sitting and Walking - an Immersive Virtual Reality Study. *Vis. Res.* 51, 1173–1184. doi:10.1016/j.visres.2011.03.006
- Iskander, J., Hossny, M., and Nahavandi, S. (2019). Using Biomechanics to Investigate the Effect of VR on Eye Vergence System. *Appl. Ergon.* 81, 102883. doi:10.1016/j.apergo.2019.102883
- Johnsson, J., and Matos, R. (2011). *Accuracy and Precision Test Method for Remote Eye Trackers TestSpecification*. Danderyd Municipality, Sweden: Tobii Technology.
- Kaplanyan, A. S., Sochenov, A., Leimkühler, T., Okunev, M., Goodall, T., and Rufo, G. (2019). DeepFovea: Neural Reconstruction for Foveated Rendering and Video Compression Using Learned Statistics of Natural Videos. *ACM Trans. Graphics* 38, 212. doi:10.1145/3355089.3356557
- Kim, J., Jeong, Y., Stengel, M., Akşit, K., Albert, R., Boudaoud, B., et al. (2019). enFoveated AR: Dynamically-Foveated Augmented Reality Display. *ACM Trans. Graphics* 38, 1–15. doi:10.1145/3306346.3322987
- King, A. J., Bol, N., Cummins, R. G., and John, K. K. (2019). Improving Visual Behavior Research in Communication Science: An Overview, Review, and Reporting Recommendations for Using Eye-Tracking Methods. *Commun. Methods Measures* 13, 149–177. doi:10.1080/19312458.2018.1558194
- Kothari, R., Yang, Z., Kanan, C., Bailey, R., Pelz, J. B., and Diaz, G. J. (2020). Gaze-in-wild: A Dataset for Studying Eye and Head Coordination in Everyday Activities. *Scientific Rep.* 10, 1–18. doi:10.1038/s41598-020-59251-5
- Koulieris, G. A., Akşit, K., Stengel, M., Mantiuk, R. K., Mania, K., and Richardt, C. (2019). enNear-Eye Display and Tracking Technologies for Virtual and Augmented Reality. *Comput. Graphics Forum* 38, 493–519. doi:10.1111/cgf.13654
- Kowler, E. (2011). Eye Movements: The Past 25years. *Vis. Res.* 51, 1457–1483. doi:10.1016/j.visres.2010.12.014
- Kramida, G. (2015). Resolving the Vergence-Accommodation Conflict in Head-Mounted Displays A Review of Problem Assessments, Potential Solutions, and Evaluation Methods. *IEEE Trans. Visualization Comput. Graphics* 22, 1912–1931. doi:10.1109/TVCG.2015.2473855
- Kwon, Y.-M., Jeon, K.-W., Ki, J., Shahab, Q. M., Jo, S., and Kim, S.-K. (2006). 3D Gaze Estimation and Interaction to Stereo Display. *Int. J. Virtual Reality* 5, 41–45. doi:10.20870/ijvr.2006.5.3.2697
- Lambooi, M., IJsselstein, W., Fortuin, M., and Heynderickx, I. (2009). Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review. *J. Imaging Sci. Technol.* 53, 030201. doi:10.2352/j.imagingsci.technol.2009.53.3.030201
- Land, M. F. (1993). Eye-head Coordination during Driving. *Proc. IEEE Int. Conf. Syst. Man Cybernetics* 3, 490–494. doi:10.1109/icsmc.1993.385060
- Land, M. F., and Lee, D. N. (1994). Where We Look when We Steer. *Nature* 369, 742–744. doi:10.1038/369742a0
- Langman, D., and Luebke, D. (2013). Near-eye Light Field Displays. *ACM Trans. Graphics* 32, 1–10. doi:10.1145/2508363.2508366
- Larsson, L., Schwaller, A., Nyström, M., and Stridh, M. (2016). enHead Movement Compensation and Multi-Modal Event Detection in Eye-Tracking Data for Unconstrained Head Movements. *J. Neurosci. Methods* 274, 13–26. doi:10.1016/j.jneumeth.2016.09.005
- Lee, S. H., and Civera, J. (2020). “Triangulation: Why Optimize?,” in Proceedings of the 20th British Machine Vision Conference 2019, BMVC 2019, Cardiff, UK, September 2019.
- Li, R., Whitmire, E., Stengel, M., Boudaoud, B., Kautz, J., Luebke, D., et al. (2020). “Optical Gaze Tracking with Spatially-Sparse Single-Pixel Detectors,” in Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Porto de Galinhas, Brazil, November 2020, 117–126. doi:10.1109/ISMAR50242.2020.00033
- Lockett, E., Key, T., Newsome, N., and Jones, J. A. (2019). “Metrics for the Evaluation of Tracking Systems for Virtual Environments,” in Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, August 2019, 1711–1716. doi:10.1109/VR.2019.8798374
- Mardanbegi, D., Langlotz, T., and Gellersen, H. (2019). “Resolving Target Ambiguity in 3D Gaze Interaction through VOR Depth Estimation.” in Proceedings of the 2019 Conference on Human Factors in Computing Systems, Glasgow Scotland UK, May 2019, 1–12. doi:10.1145/3290605.3300842
- Marmitt, G., and Duchowski, A. T. (2002). “Modeling Visual Attention in VR: Measuring the Accuracy of Predicted Scanpaths,” in Proceedings of the EuroGraphics (Short Presentations), Saarbrücken, Germany, September 2002.
- Mlot, E. G., Bahmani, H., Wahl, S., and Kasneci, E. (2016). “3D Gaze Estimation Using Eye Vergence,” in Proceedings of the HEALTHINF 2016 - 9th

- International Conference on Health Informatics, Proceedings; Part of 9th International Joint Conference on Biomedical Engineering Systems and Technologies, Rome Italy, February 2016, 125–131. doi:10.5220/0005821201250131
- Munn, S. M., and Pelz, J. B. (2008). “3D point-of-regard, Position and Head Orientation from a Portable Monocular Video-Based Eye Tracker,” in Eye Tracking Research and Applications Symposium (ETRA), Savannah Georgia, March 2008, 181–188. doi:10.1145/1344471.1344517
- Naceri, A., Chellali, R., and Hoinville, T. (2011). Depth Perception within Peripersonal Space Using Head-Mounted Display. *Presence: Teleoperators and Virtual Environments* 20, 254–272. doi:10.1162/PRES_a_00048
- Niehorster, D. C., Cornelissen, T. H., Holmqvist, K., Hooge, I. T., and Hessels, R. S. (2018). What to Expect from Your Remote Eye-Tracker when Participants Are Unrestrained. *Behav. Res. Methods* 50, 213–227. doi:10.3758/s13428-017-0863-0
- Niehorster, D. C., Li, L., and Lappe, M. (2017). The Accuracy and Precision of Position and Orientation Tracking in the HTC Vive Virtual Reality System for Scientific Research. *i-Perception* 8, 1–23. doi:10.1177/2041669517708205
- Niehorster, D. C., Santini, T., Hessels, R. S., Hooge, I. T., Kasnecki, E., and Nyström, M. (2020). The Impact of Slippage on the Data Quality of Head-Worn Eye Trackers. *Behav. Res. Methods* 52, 1140–1160. doi:10.3758/s13428-019-01307-0
- Nyström, M., Andersson, R., Holmqvist, K., and van de Weijer, J. (2013). The Influence of Calibration Method and Eye Physiology on Eyetracking Data Quality. *Behav. Res. Methods* 45, 272–288. doi:10.3758/s13428-012-0247-4
- Olsen, A. (2012). *enThe Tobii I-VT Fixation Filter*. Danderyd Municipality, Sweden: Tobii Technology.
- Orquin, J. L., and Holmqvist, K. (2018). Threats to the Validity of Eye-Movement Research in Psychology. *Behav. Res. Methods* 50, 1645–1656. doi:10.3758/s13428-017-0998-z
- Pieszala, J., Diaz, G., Pelz, J., Speir, J., and Bailey, R. (2016). “3D Gaze Point Localization and Visualization Using LiDAR-Based 3D Reconstructions,” in Proceedings of the Eye Tracking Research and Applications Symposium (ETRA), Charleston South Carolina, March 2016 (New York, New York, USA: Association for Computing Machinery), 201–204. doi:10.1145/2857491.2857545
- Previc, F. H. (1998). The Neuropsychology of 3-D Space. *Psychol. Bull.* 124, 123–164. doi:10.1037/0033-2909.124.2.123
- Reichelt, S., Häussler, R., Fütterer, G., and Leister, N. (2010). “Depth Cues in Human Visual Perception and Their Realization in 3D Displays,” in *Three-Dimensional Imaging, Visualization, and Display 2010 and Display Technologies and Applications for Defense, Security, and Avionics IV*. Editors B. Javidi, J.-Y. Son, J. T. Thomas, and D. D. Desjardins (Bellingham, Washington, United States: SPIE), 7690, 76900B. doi:10.1117/12.850094
- Salvucci, D. D., and Goldberg, J. H. (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, November 2000 (New York, NY, USA: Association for Computing Machinery), 71–78. doi:10.1145/355017.355028
- Sidenmark, L., and Gellersen, H. (2019). Eye, Head and Torso Coordination during Gaze Shifts in Virtual Reality. *ACM Trans. Computer-Human Interaction* 27, 1–40. doi:10.1145/3361218
- Sitzmann, V., Serrano, A., Pavel, A., Agrawala, M., Gutierrez, D., Masia, B., et al. (2018). Saliency in VR: How Do People Explore Virtual Environments? *IEEE Trans. Visualization Comput. Graphics* 24, 1633–1642. doi:10.1109/TVCG.2018.2793599
- Steil, J., Huang, M. X., and Bulling, A. (2018). “Fixation Detection for Head-Mounted Eye Tracking Based on Visual Similarity of Gaze Targets,” in Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, Warsaw Poland, June 2018 (New York, NY, USA: Association for Computing Machinery), 1–9. doi:10.1145/3204493.3204538
- Tresilian, J. R., Mon-Williams, M., and Kelly, B. M. (1999). Increasing Confidence in Vergence as a Cue to Distance. *Proc. R. Soc. B: Biol. Sci.* 266, 39–44. doi:10.1098/rspb.1999.0601
- van der Veen, S. M., Bordeleau, M., Pidcoe, P. E., France, C. R., and Thomas, J. S. (2019). Agreement Analysis between Vive and Vicon Systems to Monitor Lumbar Postural Changes. *Sensors (Switzerland)* 19, 3632. doi:10.3390/s19173632
- Verstraten, F. A., Hooge, I. T., Culham, J., and Van Wezel, R. J. (2001). Systematic Eye Movements Do Not Account for the Perception of Motion during Attentive Tracking. *Vis. Res. (Pergamon)* 41, 3505–3511. doi:10.1016/S0042-6989(01)00205-X
- Vienne, C., Masfrand, S., Bourdin, C., and Vercher, J. L. (2020). Depth Perception in Virtual Reality Systems: Effect of Screen Distance, Environment Richness and Display Factors. *IEEE Access* 8, 29099–29110. doi:10.1109/ACCESS.2020.2972122
- Vienne, C., Plantier, J., Neveu, P., and Priot, A.-E. (2018). Disparity-Driven Accommodation Response Contributes to Perceived Depth. *Front. Neurosci.* 12, 973. doi:10.3389/fnins.2018.00973
- Viguier, A., Clément, G., and Trotter, Y. (2001). Distance Perception within Near Visual Space. *Perception* 30, 115–124. doi:10.1068/p3119
- Vinnikov, M., Allison, R. S., and Fernandes, S. (2016). Impact of Depth of Field Simulation on Visual Fatigue: Who Are Impacted? and How? *Int. J. Hum. Comput. Stud.* 91, 37–51. doi:10.1016/j.ijhcs.2016.03.001
- Vinnikov, M., and Allison, R. S. (2014). “Gaze-contingent Depth of Field in Realistic Scenes: The User Experience,” in Proceedings of the Eye Tracking Research and Applications Symposium (ETRA), Safety Harbor Florida, March 2014, 119–126. doi:10.1145/2578153.2578170
- Wang, R. I., Pelfrey, B., Duchowski, A. T., and House, D. H. (2014). Online 3D Gaze Localization on Stereoscopic Displays. *ACM Trans. Appl. Perception* 11, 1–21. doi:10.1145/2593689
- Wang, X., Holmqvist, K., and Alexa, M. (2019). The Mean point of Vergence Is Biased under Projection. *J. Eye Move. Res.* 12 (4). doi:10.16910/jemr.12.4
- Wang, X., Koch, S., Holmqvist, K., and Alexa, M. (2018). Tracking the Gaze on Objects in 3D: How Do People Really Look at the Bunny? *ACM Trans. Graphics* 37, 1–18. doi:10.1145/3272127.3275094
- Weber, S., Schubert, R. S., Vogt, S., Velichkovsky, B. M., and Pannasch, S. (20182015). Gaze3DFix: Detecting 3D Fixations with an Ellipsoidal Bounding Volume. *Behav. Res. Methods* 50, 2004. doi:10.3758/s13428-017-0969-4
- Wexler, M., and Van Boxtel, J. J. (2005). Depth Perception by the Active Observer. *Trends Cogn. Sci.* 9, 431–438. doi:10.1016/j.tics.2005.06.018
- Wibrama, S., Nugroho, H. A., and Hamamoto, K. (2017). Evaluating 3D Gaze Tracking in Virtual Space: A Computer Graphics Approach. *Entertainment Comput.* 21, 11–17. doi:10.1016/j.entcom.2017.04.003
- Wu, R., Pandurangaiah, J., Blankenship, G. M., Castro, C. X., Guan, S., Ju, A., et al. (2020). “Evaluation of Virtual Reality Tracking Performance for Indoor Navigation,” in Proceedings of the 2020 IEEE/ION Position, Location and Navigation Symposium (PLANS), Portland, OR, USA, June 2020 (Piscataway, New Jersey, United States: IEEE), 1311–1316. doi:10.1109/PLANS46316.2020.9110225
- Yates, A., and Selan, J. (2016). *enPositional Tracking Systems and Methods*. US20160131761A1.
- Zhao, H., Straub, D., and Rothkopf, C. A. (2019). The Visual Control of Interceptive Steering: How Do People Steer a Car to Intercept a Moving Target? *J. Vis.* 19, 1–20. doi:10.1167/19.14.11

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Lamb, Brundin, Perez Luque and Billing. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.