Title

Evaluation of Gaze-to-Object Mapping Algorithms for Use in “Real-World” Translatable Neuropsychological Paradigms

Journal
Psychology & Neuroscience, 16(4)

ISSN
1984-3054

Authors
Liu, Weichen
Andrade, Gianna
Schulze, Jürgen
et al.

Publication Date
2023-12-01

DOI
10.1037/pne0000324

Peer reviewed

HHS Public Access
Author manuscript
Psychol Neurosci. Author manuscript; available in PMC 2024 December 01.

Published in final edited form as:
Psychol Neurosci. 2023 December ; 16(4): 339–348. doi:10.1037/pne0000324.

Evaluation of Gaze-to-Object Mapping Algorithms for Use in "Real-World" Translatable Neuropsychological Paradigms

Weichen Liu, BS[1], **Gianna Andrade, BS**[2], **Jurgen Schulze, PhD**[1], **Kelly E. Courtney, PhD**[2]

[1]Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, USA

[2]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA

Abstract

Objective: Eye-tracking technology is commonly used for identifying objects of visual attention. However, applying this technology to virtual reality (VR) applications is challenging. This report analyzes the performance of two gaze-to-object mapping (GTOM) algorithms applied to eye-gaze data acquired during a "real-world" VR cue-reactivity paradigm.

Methods: Two groups of participants completed a VR paradigm using an HTC Vive Pro Eye. The gazed objects were determined by the reported gaze rays and one of two GTOM algorithms − naïve ray-casting (n=18) or a combination of ray-casting and Tobii's G2OM algorithm (n=18). Percent gaze duration was calculated from 1-second intervals before each object interaction to estimate gaze accuracy. The object volume of maximal divergence between algorithms was determined by maximizing the difference in Hedge's G effect sizes between small and large percent gaze duration distributions. Differences in percent gaze duration based on algorithm and target object size were tested with a mixed ANOVA.

Results: The maximum Hedge's G effect sizes differentiating large and small target objects was observed at an 800cm^3 threshold. The combination algorithm performed better than the naïve ray-casting algorithm (p=.003, η_p^2=.23), and large objects (>800cm^3) were associated with a higher gaze duration percentage than small objects (≤800cm^3; p<.001, η_p^2=.76). No significant interaction between algorithm and size was observed.

Conclusions: Results demonstrated that Tobii's G2OM method outperformed naïve ray-casting in this "real-world" paradigm. As both algorithms show a clear decrease in performance for detecting objects with volumes <800cm^3, we recommend using gaze-interactable objects >800cm^3 for future HTC Vive Pro Eye applications.

Keywords

virtual reality; eye tracking; gaze-to-object mapping; visual attention; gaze accuracy

Correspondence concerning this article should be addressed to Kelly E. Courtney, UC San Diego, Department of Psychiatry, 9500 Gilman Drive, La Jolla, CA 92093, USA. kecourtney@health.ucsd.edu; Tel: 858.534.5157; Fax: 858.534.4989.

Introduction

Eye-tracking has a long history in neuropsychological research. Eye-movement and eye-gaze data are often collected to help elucidate mechanisms involved in visual information processing and attention (Deubel & Schneider, 1996; Duchowski, 2002; Hoffman & Subramaniam, 1995) and has been used as indirect markers of a variety of behavioral and psychological outcomes such as covert cognitive processes or traits [e.g., information processing patterns, cognitive effort (Ke et al., 2021)], patterns of social media interaction (Hussain et al., 2019), clinical diagnoses (Yaneva et al., 2020), and driver drowsiness recognition (Zhao et al., 2018). The methods for detecting and monitoring eye movements have changed considerably over the previous 100 years or so, with most modern approaches relying on video-based systems and computer vision techniques (Clay et al., 2019). Given the tremendous technical advances, decreasing costs, and increasing convenience [e.g., integrated virtual reality (VR) head-mounted displays (HMDs)], eye-tracking has become a fruitful tool for the investigation of wide variety of neuropsychological processes (Clay et al., 2019).

Psychological research frequently capitalizes on visual attention metrics to inform on a broad set of psychological phenomena. Within these endeavors, mapping eye-gaze fixation points to specific semantic or target object in the visual field is a common goal. For example, studies on infant language development (Yu et al., 2019), attentional bias to dysphoric and threatening stimuli associated with depression and anxiety (Armstrong & Olatunji, 2012), and new advances in the use of eye-tracking to improve attention management via live biofeedback (Toreini et al., 2020), all employ some form of eye-gaze fixation mapping. Gaze-path visualization estimates such as gaze duration towards an object [also referred to as dwell time or fixation time (e.g., Liu et al., 2022; Paulus & Remijn, 2021)], number of object fixations (e.g., Pluciennicka et al., 2016), orienting bias towards objects [or the probability of the initial object fixation (e.g., Liang et al., 2017)], and break frequencies [or the number of times distractor stimuli are gazed at (e.g., Qureshi et al., 2018)] are conventional measures of interest in these endeavors. These metrics rely on accurate identification of the attentional target object, a process termed gaze-to-object mapping (GTOM), to ensure their validity. GTOM algorithms attempt to provide information on the object of intended attentional gaze (the "what") despite frequent discrepancies from actual recorded eye-gaze location (the "where"). Rapid involuntary eye-movements (i.e., micro-saccades), hardware limitations [e.g., accuracy and precision (Valtakari et al., 2021)], individual biological differences [e.g., eye shape, color, pupil size and reflectivity (Wang et al., 2017)], and imperfect calibration processes all contribute to error in the acquired eye-tracking data and complicate successful attentional target object identification.

Paradigms using 2D graphical interfaces, such as traditional free-viewing tasks (e.g., Liang et al., 2017), allow for a relatively simple and accurate extraction of GTOM metrics as the target stimuli remain mostly static and can be based predominately on aligning objects with recorded eye-gaze position in 2D space (Duchowski, 2017; Paulus & Remijn, 2021). However, extraction of 2D gaze points cannot be adapted to 3D immersive environments (Stellmach et al., 2010). Due to the relative novelty of eye-tracking acquisition in 3D applications, the literature remains sparse as to the utility of various GTOM algorithms under the additional complexities imposed by the 3D environment. For instance, in 3D scenarios, gaze may be directed towards multiple objects at various depths, and dynamic scene changes and movement of scene objects or cameras can cause occlusions (Bernhard et al., 2014). Thus, the identification of "where" (i.e., eye-gaze position) an individual is looking becomes less informative (Sundstedt et al., 2013). Ray-casting (Mine, 1995), one of the most commonly used object selection algorithm for 3D applications due to its ease of implementation (Starker & Bolt, 1990), involves casting rays from the tracked body part (i.e., hands or pupils) into the virtual environment and identifying any object that is intersected by the rays as a target. However, ray-casting becomes problematic and inefficient when the target objects are small, or at least appear small due to distance from the individual (Poupyrev et al., 1998; Wang & Kopper, 2021).

Attempts have been made to account for some of the 3D-related complexities in GTOM by leveraging the multitude of information contained in the 3D scenes to identify the objects of attention. Techniques based on identifying dynamic areas of interest (Bernhard et al., 2014; Papenmeier & Huff, 2010; Sundstedt et al., 2013), "active" approaches utilizing Bayesian inferential models to predict attentional targets under the assumption that gaze follows attention (Bernhard et al., 2014), and progressive refinement (including user input) to allow for improvements in accurate 3D selection (Wang & Kopper, 2021) have shown initial promise in addressing some of these complexities. However, to our knowledge, these techniques have yet to be tested within immersive 3D VR environments which are much more dynamic, result in increased participant engagement, and encourage head and body movement within the scenes. VR environments which attempt to mimic the intricacy of real-world scenarios further add to this complexity by encouraging additional in-scene movement and interaction, making the implementation of these techniques more difficult. Tobii, a leader in eye-tracking technology, developed a machine-learned selection algorithm named G2OM which is compatible with various publicly available VR headsets (including HTC's VIVE Pro Eye), thus increasing its popularity and potential for use in commercial and research settings (Shadiev & Li, 2023). According to Tobii (https://developer.tobii.com/xr/solutions/tobii-g2om/), their G2OM algorithm has been trained using millions of data points and purportedly improves upon naïve ray-casting, yet to our knowledge, has not been previously validated in an independent study with eye-tracking metrics derived from a 3D or VR environment.

Given the paucity of research on GTOM performance in immersive VR environments, the present report sought to analyze the performance of two of the more common GTOM algorithms as applied to eye-tracking data acquired from adult community participants during a "real-world" VR nicotine cue-reactivity paradigm. We hypothesized that a combination of Tobii's G2OM and naïve ray-casting would perform better than the use of naïve ray-casting alone due to Tobii's inclusion of machine learning which is purported to facilitate object selections in the VR environment. Tobii's G2OM was not tested in isolation due to its detrimental effect on paradigm performance. Given the error sources described above, we further hypothesized that both algorithms would perform better when the gazed objects were larger in size.

Materials and Methods

VR Hardware and Software

The HTC Vive Pro Eye VR headset (HTC, Taoyuan City, Taiwan) was used to enable VR capabilities and collect eye-related data. The headset's built-in eye tracker allows a trackable field of view of 110 degrees in perfect conditions with a frequency of 120Hz and accuracy of 0.5 − 1.1 degrees (when the pupils are within a 20-degree field of view). HTC's SRanipal SDK (https://developer-express.vive.com/resources/vive-sense/eye-and-facial-tracking-sdk/) was used in conjunction with Tobii XR SDK (https://vr.tobii.com/sdk/) to gain access to various data from the eye tracker. The SRanipal SDK provided access to raw eye tracking data, such as pupil positions and pupil diameters, while Tobii's XR SDK provided the G2OM algorithm. Two hand-held HTC Vive controllers were also used to allow for participant interaction with select scene objects.

NTP Cue VR Paradigm Composition

The Nicotine and Tobacco Product (NTP) Cue VR paradigm was developed by the authors for use in an ongoing study on improving the assessment of the nicotine craving construct through immersive VR and eye-tracking (Liu et al., 2022). The paradigm was built using Unity and contains six different scenes; three scenes include the presence of nicotine cues (e.g., tobacco cigarettes, e-cigarettes, lighters) and three scenes include only control cues (e.g., cell phone, pens, candy). Across the six scenes, there are 348 interactable objects with volume ranging from 2 cm^3 to 7720 cm^3 (mean=965.00, SD=1729.04). Of those, 230 objects have a volume of less than 800 cm^3. Most of the nicotine related objects belong to the small object group. Examples of small and large interactable objects can be seen in Figure 1 (large objects are highlighted in green and small objects are highlighted in red).

Study participants are free to move around (via teleportation) within limited areas of the scenes and interact with various cue objects within the VR environment using the controllers. None of the interactable objects are in motion by default, and only move when interacted with by the participant. Participants are instructed to "Just explore everything around you until the scene changes."

Data Collection

During each scene, regular time series data including timestamp and raw gaze intersection point (GIP) is collected every 10 milliseconds (100Hz), independent of the frame time. Button presses on the controller (including time, button pressed, and object of interaction, if applicable) are recorded when they occur. GTOM identified target objects and their timestamps are recorded when the eye-gaze switches to a new GTOM identified object.

Gaze Duration Calculation and Statistical Analysis

Two different GTOM algorithms are used in this experiment: naïve ray-casting and a combination of ray-casting and Tobii's G2OM algorithm. The naïve ray casting algorithm relies on PhysX to perform ray vs. collider intersection tests. A collider is any object of interest in the virtual environment. The invisible ray emanates from the user's eye position and points in the user's gaze direction (as reported by the HTC Vive Pro Eye). The software then checks for intersections of the ray with a collider in the virtual environment. The closest intersected object is then assumed to be the target of the user's attention.

The combination method was developed as a compromise between performance and quality. The G2OM algorithm provided by Tobii's XR SDK is a machine-learning based mapping algorithm that aims to improve small object and fast-moving object tracking (https://developer.tobii.com/xr/solutions/tobii-g2om/). The specifics of the G2OM algorithm and how it works are not provided by Tobii, except that it is a machine-learned selection algorithm that has been trained using vast amounts of data to determine precisely where the user is looking within any given scene. For our study, we used Tobii's G2OM algorithm as is, with no ability to modify any parameters or settings provided by Tobii (https://developer.tobii.com/xr/develop/xr-sdk/documentation/). Due to the complexity of the "real-world" scenes in the NTP Cue VR paradigm, considering all objects within the scenes as candidates for Tobii's G2OM algorithm detrimentally affects the framerate. Therefore, for the combination method, only the interactable objects are considered candidates and detected by Tobii's G2OM. The detection of background and other non-movable large objects are then demoted to relying on the naïve ray-casting algorithm. Thus, for each frame, given the eye-tracking data from the Vive Pro Eye, if Tobii's G2OM algorithm detects any object of interest, then such object is selected. Otherwise, the naïve ray-casting method is used. In both cases, the GTOM algorithms perform target object selection in real-time during the task performance.

Given that individuals typically gaze at objects between .09 and .64 seconds prior to directed object interaction, with substantial individual variability (Land & Hayhoe, 2001; Lavoie et al., 2018), relative performance was estimated as the percent of time an object is gazed at 1 second prior to each interaction in the scene. For these analyses, each 1 second interval prior to an interaction was considered as a trial. To calculate the total gaze duration on the target object during each trial, we summed the time differences between frames when the gaze was on the target object. This allowed us to estimate the percentage of time that participants spent gazing at the target object (identified by ray-casting or combined G2OM) before the interaction occurred by dividing the duration of gaze on the target object by the total trial duration (1 second). Participants conducted 1391 interactions in total. Out of the total number of interactions, the group using direct naïve ray-casting conducted 791 interactions, while the group using a combination of ray-casting and Tobii's G2OM conducted 600 interactions.

We then applied kernel density estimation over the percent gaze duration data across all trials for small objects and large objects separately to get two distributions for each GTOM algorithm. The object volume threshold of maximal divergence between the naïve ray-casting and combination algorithms was determined by adjusting the threshold of size classification to maximize the difference in Hedge's G effect sizes between the small and large percent gaze duration distributions. This allowed us to estimate the volume threshold at which the naïve ray-casting and combination algorithms performed the worst in correctly identifying target objects; A perfectly precise GTOM algorithm would result in similar large and small object percent gaze durations (e.g., Špakov, 2011) as individuals generally spend comparable amounts of time looking at any object, regardless of size, before interacting with them in novel environments. Therefore, a discrepancy in estimated gaze duration between small and large objects likely signifies inaccuracies in a GTOM algorithm. The resulting maximum Hedges g effect sizes differentiating between large and small target objects was achieved when the threshold was set to 800 cm^3, corresponding to Hedges g's of 0.37 for the naïve ray-casting method and 0.21 for the combination method. Thus, small objects were defined as those ≤ 800 cm^3 and large objects were defined as those > 800 cm^3 for the remainder of the analyses.

A mixed ANOVA model was used to test the effect of target object size (within-subjects effect: small, large) and GTOM algorithm (between-subjects effect: ray-casting, combination) on percent of object gaze duration. Demographic variables identified as statistically different (via chi-square or independent *t*-tests) between the algorithm groups, namely age and nicotine use days, were entered as covariates of no-interest into a second mixed ANOVA model to test the durability of the effects of interest. The threshold of statistical significance was set at $p < .05$ for all analyses. Based on empirical observation of participant eye-gaze performance during the task (selection of videos reviewed), all analyses were also conducted using a 5-second trial period prior to object interaction to ensure variability in eye-gaze duration was captured in the trial period, in addition to the primary analysis which used a 1-second trial period prior to object interaction.

Participant Recruitment and Screening Procedures

Participant eye-tracking data for this report was culled from an ongoing study investigating eye-tracking indices of nicotine cue-reactivity in daily and non-daily users of nicotine and tobacco products. The first 36 subjects from this study were included in the analyses. Inclusion criteria for the ongoing study are: 1) ages 18+, 2) non-daily (average use on 4–27 days per month, past 3 months) *or* daily nicotine use (average use on 7 days per week, past 3 months), and 3) nicotine use history ≥1 year. Exclusionary criteria include: 1) medical or psychiatric history affecting brain development (i.e., history and/or treatment of neurologic disorders, severe head trauma with loss of consciousness >2 minutes, or current severe DSM-5 psychiatric disorders other than tobacco use disorders), 2) non-fluency in English, 3) visual problems that may make task completion difficult (e.g., severe motion sickness, blindness, glasses).

Data from 18 of the earlier participants used the naïve ray-casting to perform GTOM, while the next 18 participants' data utilized the combination ray-casting and Tobii's G2OM algorithm.

Sample Demographics

Sample demographic information is provided in Table 1. Overall, the included sample was predominantly male (58.3%) and White (66.7%), and 58.3% had no or very limited (one time) previous experiences with VR. The average age of the sample was 32.47 years-old (SD=15.70) and participants used nicotine or tobacco products on 68.92 of the past 90 days (SD=30.51), on average. Differences between participants in the two algorithm groups were observed with respect to age ($t=-5.26$, $p<.001$) and nicotine use ($t=-2.76$, $p=.01$); no differences were observed between the groups on the other demographic variables (*p*s>.05).

Results

Results of the mixed ANOVA testing the effects of the algorithm (naïve ray-casting versus combination), target object size (small, large), and their interaction (group*size) on algorithm performance (percentage of target gaze duration before each interaction during the paradigm) revealed no significant interaction between algorithms and target object size (F=0.90, p=.35, η_p^2=.03; see Table 2 for means and standard deviations). However, a main effect of group was observed whereby the combination algorithm (mean=19.73%, SD=7.92%) outperformed the naïve ray-casting algorithm (mean=15.99%, SD=5.96%), regardless of size (F=10.06, p=.003, η_p^2=.23; Figures 2 and 3). A main effect of size was also observed, whereby large objects with volumes > 800 cm^3 (mean=23.15%, SD=6.17%) had a higher gaze duration percentage than small objects (mean=12.55%, SD=3.30%), regardless of GTOM algorithm (F=107.98, p<.001, η_p^2=.76; Figures 2 and 3). These analyses were rerun including age and nicotine use as covariates in the mixed ANOVA model with no changes in statistical significance or direction of effects observed – interaction (group*size): F=0.85, p=.36, η_p^2=.03; main effect of group: F=11.65, p=.002, η_p^2=.27; main effect of size: F=8.10, p=.01, η_p^2=.20. The 5-second trial period ANOVA analyses also did not differ in terms of direction or significance of effects - interaction (group*size): F=2.21, p=.15, η_p^2=.06; main effect of group: F=11.08, p=.002, η_p^2=.25; main effect of size: F=41.58, p<.001, η_p^2=.55.

Discussion

The present report describes the performance of two common GTOM algorithms used within a “real-world” VR nicotine cue-reactivity paradigm. The analyses support the use of the combined naïve ray-casting and Tobii’s G2OM algorithm, as opposed to relying on the naïve ray-casting alone, for eye-tracking investigations in virtual 3D scenes. We observed that the combination method generally outperformed the naïve ray-casting method in identifying target objects, yet still exhibited a clear decrease in performance for objects with volumes less than 800 cm^3. This suggests that, while the combination method may be more effective overall, it still has difficulty accurately identifying smaller objects. Thus, we recommend that researchers interested in using the HTC VIVE Pro Eye to acquire eye-tracking data for object mapping within VR paradigms avoid including small objects of interest in their virtual scenes until improved GTOM algorithms are available.

Despite the observed imprecision in the GTOM algorithms, pilot analyses of attentional target gaze duration from this same sample demonstrated effects in the hypothesized direction (Liu et al., 2022). Specifically, consistent with the literature on attentional bias in individuals who regularly use nicotine products (Bradley et al., 2004; Gamito et al., 2014; Kwak et al., 2007; Mogg et al., 2003), we observed greater gaze durations towards the nicotine cues in our pilot analyses, which were predominately small (<800 cm^3), as compared with the control cues, in two out of the three active cue VR scenes (Liu et al., 2022). Thus, although the present findings suggest a precision bias in the GTOM algorithms used in our 2022 study, the practical significance of this imprecision may be modest as the effect sizes of some psychological constructs (e.g., nicotine-related attentional bias) may exceed the magnitude of this imprecision.

Given the small sample and between subject nature of the GTOM comparison, we caution against over interpretation of the present study results; however, we believe the large effect size observed between GTOM algorithms (η_p^2=.23) and within-in subject object size comparisons (η_p^2=.76) warrant further investigation and careful consideration when designing future VR eye-tracking studies involving GTOM. Further, our metric of performance assumes that humans look at smaller versus larger objects for similar durations prior to interacting with them. Although intuitive, a direct test incorporating a ground truth is needed to confirm these results. Similarly, it is theoretically possible that the observed reduction in percent gaze-time for smaller objects is due to less time spent gazing at smaller objects overall. However, data from the parent study on nicotine attentional bias support that greater time was spent gazing at smaller, most often nicotine-related, objects on average as opposed to generally larger control objects (Liu et al., 2022), supporting that the effect we observed in this study is likely due to the GTOM algorithm performance and not due to shorter overall gaze times towards smaller objects. Lastly, we did not collect data on the angular size of the scene objects which would have allowed for more precise control of object distance in the calculations. Given that participants were necessarily within arm's reach of the objects during the gaze collection time, it is unclear whether the slight differences in object distances would have impacted the study results.

The potential utility of improving and testing GTOM algorithms for use in VR settings is considerable for commercial marketing/entertainment, education, clinical treatment, and research applications that rely on object selection (Adhanom et al., 2023). With respect to psychological research, conversion of traditional 2D and 3D paradigms frequently used to study visual attention (e.g., free-viewing and visual search tasks) to VR paradigms would allow for greater translatability of study findings to real-world settings, reduce the need for repetitive trial structures, and increase consistency and reliability of the data by allowing the researcher to control all sensory input received by the participant while still maintaining an immersive "real life" experience (Peeters, 2019). Limitations still exist as, similar to most 2D and 3D applications, objects for the most part must be labeled for GTOM to be accomplished and meaningfully interpreted. Practically, this should be done prior to data collection, although post-processing labelling is also possible with additional effort. This need may be eliminated in the future as algorithms relying on artificial intelligence for object identification become more accurate and widely available. Regardless, routine testing of algorithm performance is critical to ensure data accuracy before we can capitalize on the vast benefits VR paradigms have to offer.

To our knowledge, this is the first report on GTOM algorithm performance as applied to eye-tracking data acquired from a "real-world" immersive VR environment. The results highlight the complexities of acquiring accurate GTOM data in complex immersive 3D scenes including smaller sized objects and suggest higher-level algorithms are needed to enhance our mapping capabilities. Additionally, the analysis process presented can serve as a template for future analyses comparing performances of novel eye tracking devices and GTOM algorithms.

Acknowledgments

Research supported by the California Tobacco-Related Disease Research Grants Program Office of the University of California grant T30IP0962 and the National Institutes of Health, National Center for Advancing Translational Sciences grant UL1TR001442. Dr. Jurgen Schulze is now a paid employee at Qualcomm. The remaining authors declare no known conflicts of interest.

References

Adhanom IB, Macneilage P, & Folmer E (2023, 2023-01-18). Eye Tracking in Virtual Reality: a Broad Review of Applications and Challenges. Virtual Reality. 10.1007/s10055-022-00738-z

Armstrong T, & Olatunji BO (2012, 2012-12-01). Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. Clinical Psychology Review, 32(8), 704–723. 10.1016/j.cpr.2012.09.004 [PubMed: 23059623]

Bernhard M, Stavrakis E, Hecher M, & Wimmer M (2014, 2014-10-28). Gaze-to-Object Mapping during Visual Search in 3D Virtual Environments. ACM Transactions on Applied Perception, 11(3), 1–17. 10.1145/2644812

Bradley B, Field M, Mogg K, & De Houwer J (2004, Feb). Attentional and evaluative biases for smoking cues in nicotine dependence: component processes of biases in visual orienting. Behavioural Pharmacology, 15(1), 29–36. 10.1097/00008877-200402000-00004 [PubMed: 15075624]

Clay V, König P, & König S (2019, Apr 05). Eye Tracking in Virtual Reality. Journal of Eye Movement Research, 12(1). 10.16910/jemr.12.1.3

Deubel H, & Schneider WX (1996, Jun). Saccade target selection and object recognition: evidence for a common attentional mechanism. Vision Research, 36(12), 1827–1837.10.1016/0042-6989(95)00294-4 [PubMed: 8759451]

Duchowski AT (2002, 2002-11-01). A breadth-first survey of eye-tracking applications. Behavior Research Methods, Instruments, & Computers, 34(4), 455–470. 10.3758/bf03195475

Duchowski AT (2017). Eye Tracking Methodology: Theory and Practice (Third ed.). Springer. 10.1007/978-1-84628-609-4

Gamito P, Oliveira J, Baptista A, Morais D, Lopes P, Rosa P, Santos N, & Brito R (2014, Aug). Eliciting nicotine craving with virtual smoking cues. Cyberpsychology, Behavior and Social Networking, 17(8), 556–561. 10.1089/cyber.2013.0329 [PubMed: 24660864]

Hoffman JE, & Subramaniam B (1995, Aug). The role of visual attention in saccadic eye movements. Perception & Psychophysics, 57(6), 787–795. 10.3758/bf03206794 [PubMed: 7651803]

Hussain Z, Simonovic B, Stupple EJN, & Austin M (2019, Feb 18). Using Eye Tracking to Explore Facebook Use and Associations with Facebook Addiction, Mental Well-being, and Personality. Behavioral Sciences, 9(2). 10.3390/bs9020019

Ke F, Liu R, Sokolikj Z, Dahlstrom-Hakki I, & Israel M (2021). Using Eye Tracking for Research on Learning and Computational Thinking. In HCI in Games: Serious and Immersive Games (pp. 216–228). Springer International Publishing. 10.1007/978-3-030-77414-1_16

Kwak SM, Na DL, Kim G, Kim GS, & Lee JH (2007, Apr). Use of eye movement to measure smokers' attentional bias to smoking-related cues. Cyberpsychology & Behavior, 10(2), 299–304. 10.1089/cpb.2006.9953 [PubMed: 17474851]

Land MF, & Hayhoe M (2001). In what ways do eye movements contribute to everyday activities? Vision Research, 41(25–26), 3559–3565. https://doi.org/10.1016/S0042-6989(01)00102-X [PubMed: 11718795]

Lavoie EB, Valevicius AM, Boser QA, Kovic O, Vette AH, Pilarski PM, Hebert JS, & Chapman CS (2018). Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during object-interaction tasks. Journal of Vision, 18(6), 18–18. 10.1167/18.6.18

Liang CW, Tsai JL, & Hsu WY (2017, Mar). Sustained visual attention for competing emotional stimuli in social anxiety: An eye tracking study. J Behav Ther Exp Psychiatry, 54, 178–185. 10.1016/j.jbtep.2016.08.009 [PubMed: 27569741]

Liu W, Andrade G, Schulze J, Doran N, & Courtney KE (2022). Using virtual reality to induce and assess objective correlates of nicotine craving: Paradigm development study. JMIR Serious Games, 10(1), e32243. 10.2196/32243 [PubMed: 35166685]

Mine MR (1995). Virtual Environment Interaction Techniques. UNC Chapel Hill CS Dept.

Mogg K, Bradley BP, Field M, & De Houwer J (2003, Jun). Eye movements to smoking-related pictures in smokers: relationship between attentional biases and implicit and explicit measures of stimulus valence. Addiction, 98(6), 825–836. 10.1046/j.1360-0443.2003.00392.x [PubMed: 12780371]

Papenmeier F, & Huff M (2010, 2010-02-01). DynAOI: A tool for matching eye-movement data with dynamic areas of interest in animations and movies. Behavior Research Methods, 42(1), 179–187. 10.3758/brm.42.1.179 [PubMed: 20160298]

Paulus YT, & Remijn GB (2021). Usability of various dwell times for eye-gaze-based object selection with eye tracking. Displays, 67, 101997. 10.1016/j.displa.2021.101997

Peeters D (2019, 2019-06-01). Virtual reality: A game-changing method for the language sciences. Psychonomic Bulletin & Review, 26(3), 894–900. 10.3758/s13423-019-01571-3 [PubMed: 30734158]

Pluciennicka E, Wamain Y, Coello Y, & Kalénine S (2016, 2016-07-01). Impact of action primes on implicit processing of thematic and functional similarity relations: evidence from eye-tracking. Psychological Research, 80(4), 566–580. 10.1007/s00426-015-0674-9 [PubMed: 26077343]

Poupyrev I, Ichikawa T, Weghorst S, & Billinghurst M (1998, 1998-08-01). Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques. Computer Graphics Forum, 17(3), 41–52. 10.1111/1467-8659.00252

Qureshi A, Monk RL, Pennington CR, Wilcockson TDW, & Heim D (2018, Nov 24). Alcohol-related attentional bias in a gaze contingency task: Comparing appetitive and non-appetitive cues. Addict Behav, 90, 312–317. 10.1016/j.addbeh.2018.11.034 [PubMed: 30502741]

Shadiev R, & Li D (2023). A review study on eye-tracking technology usage in immersive virtual reality learning environments. Computers & Education 196, 104681. https://doi.org/10.1016/j.compedu.2022.104681

Špakov O (2011). Comparison of gaze-to-objects mapping algorithms. Proceedings of the 1st Conference on Novel Gaze-Controlled Applications,

Starker I, & Bolt RA (1990, 1990-01-01). A gaze-responsive self-disclosing display. Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI ‘90,

Stellmach S, Nacke L, & Dachselt R (2010, 2010-01-01). Advanced gaze visualizations for three-dimensional virtual environments. Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA ‘10,

Sundstedt V, Bernhard M, Stavrakis E, Reinhard E, & Wimmer M (2013). Visual Attention and Gaze Behavior in Games: An Object-Based Approach. In Game Analytics (pp. 543–583). Springer London. 10.1007/978-1-4471-4769-5_25

Toreini P, Langner M, & Maedche A (2020). Using Eye-Tracking for Visual Attention Feedback. In Information Systems and Neuroscience (pp. 261–270). Springer International Publishing. 10.1007/978-3-030-28144-1_29

Valtakari NV, Hooge ITC, Viktorsson C, Nyström P, Falck-Ytter T, & Hessels RS (2021, 2021-08-01). Eye tracking in human interaction: Possibilities and limitations. Behavior Research Methods, 53(4), 1592–1608. 10.3758/s13428-020-01517-x [PubMed: 33409984]

Wang D, Mulvey FB, Pelz JB, & Holmqvist K (2017, 2017-06-01). A study of artificial eyes for the measurement of precision in eye-trackers. Behavior Research Methods, 49(3), 947–959. 10.3758/s13428-016-0755-8 [PubMed: 27383751]

Wang Y, & Kopper R (2021, 2021-June-17). Efficient and Accurate Object 3D Selection With Eye Tracking-Based Progressive Refinement. Frontiers in Virtual Reality, 2(78). 10.3389/frvir.2021.607165

Yaneva V, Ha LA, Eraslan S, Yesilada Y, & Mitkov R (2020, 06). Detecting High-Functioning Autism in Adults Using Eye Tracking and Machine Learning. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28(6), 1254–1261. 10.1109/TNSRE.2020.2991675

Yu C, Suanda SH, & Smith LB (2019, 2019-01-01). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. Developmental Science, 22(1), e12735. 10.1111/desc.12735 [PubMed: 30255968]

Zhao L, Wang Z, Zhang G, Qi Y, & Wang X (2018, 2018-08-01). Eye state recognition based on deep integrated neural network and transfer learning. Multimedia Tools and Applications, 77(15), 19415–19438. 10.1007/s11042-017-5380-8

Public Significance Statement

This is the first report on gaze-to-object mapping (GTOM) algorithm performance as applied to eye-tracking data acquired from a “real-world” immersive VR environment. The results highlight the complexities of acquiring accurate GTOM data in complex 3D scenes including smaller sized objects and suggest higher-level algorithms are needed to enhance our mapping capabilities.

Figure 1.

Examples of small versus large interactable objects in the NTP Cue VR paradigm scenes. Large objects > 800cm^3 are highlighted in green, small objects ≤ 800cm^3 are highlighted in red (e.g., small-cigarette: 3.98 cm^3, large-ashtray: 1339.69 cm^3).

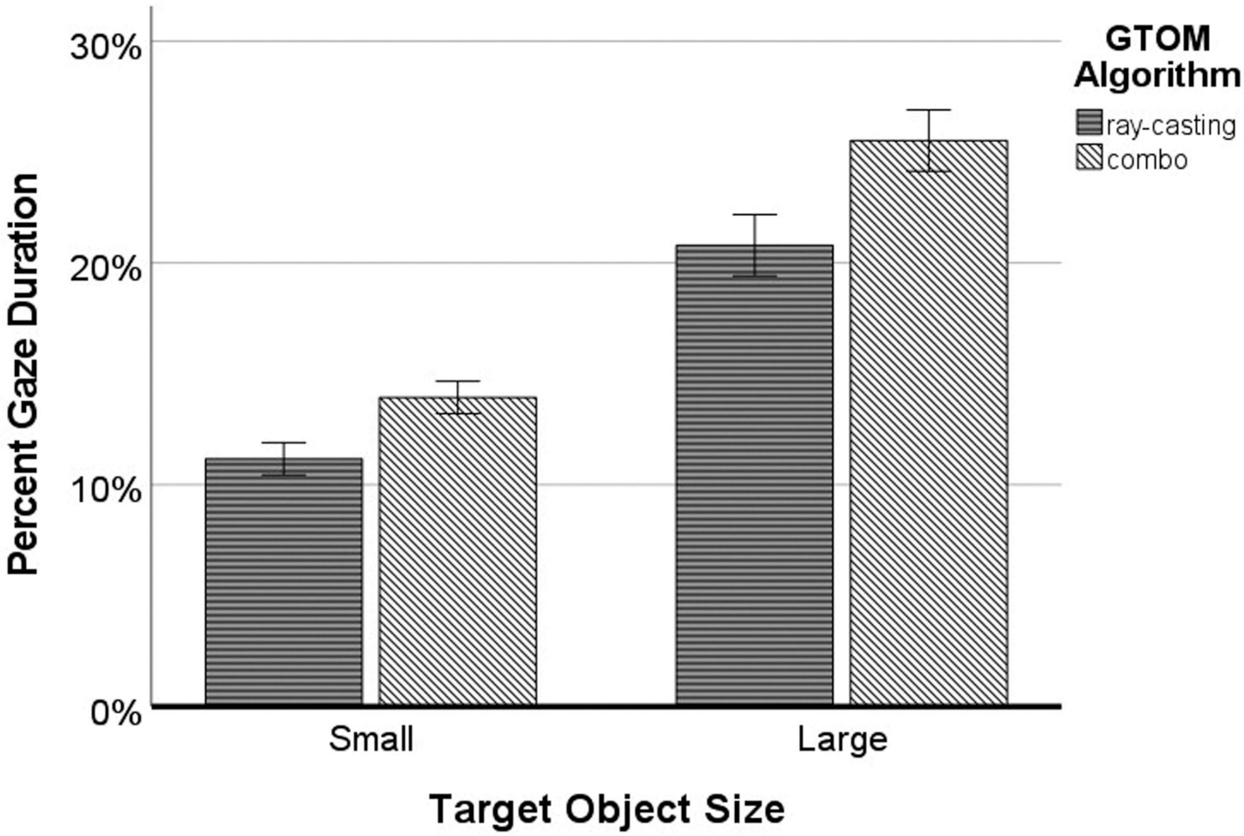

Figure 2.
Bar plot displaying percent gaze duration (metric of algorithm accuracy) means for the naïve ray-casting (ray-casting) and the combination of ray-casting and Tobii's G2OM (combo) gaze-to-object mapping (GTOM) algorithms plotted by object size (large = objects with volumes >800 cm^3; small = objects with volumes ≤ 800 cm^3). Error bars represent +/− 1 standard error.

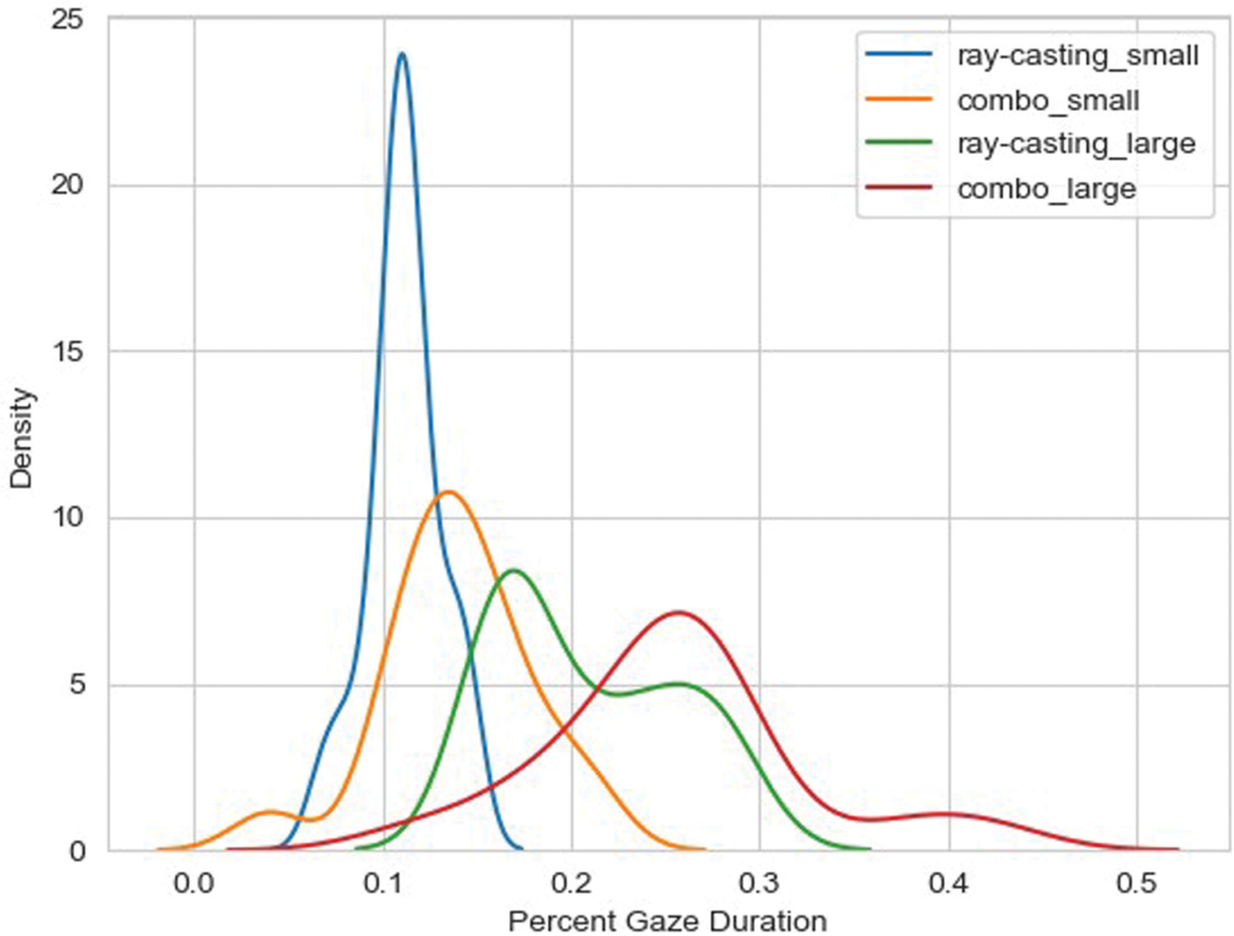

Figure 3.
Kernel density estimation of percent gaze duration (metric of algorithm accuracy) for the naïve ray-casting (ray-casting) and the combination of ray-casting and Tobii's G2OM (combo) gaze-to-object mapping algorithms plotted by object size (large = objects with volumes >800 cm^3; small = objects with volumes ≤ 800 cm^3).

Table 1.

Sample demographics (N=36).

Variable	% Frequency or Mean (SD)	
	Ray-Casting (n=18)	Ray-Casting + Tobii's G2OM (n=18)
Age *	22.11 (2.56)	42.83 (16.53)
Sex – % Male	61.1%	55.5%
Ethnicity – % White	50.0%	83.3%
Previous VR experience		
- Never	33.3%	44.4%
- Once	11.1%	27.8%
- A few times	44.4%	27.8%
- Many times	11.1%	0.0%
Nicotine use days (past 90 days) *	56.06 (35.69)	81.78 (17.10)

* $p<.05$

Table 2.

Percent gaze durations by object size and GTOM algorithm.

Percent Gaze Durations	**Mean (SD)**	
	Ray-Casting (n=18)	**Ray-Casting + Tobii's G2OM (n=18)**
Small object (≤ 800 cm^3volume)	11.17% (1.85%)	13.94% (3.81%)
Large object (>800 cm^3volume)	20.80% (4.61%)	25.52% (6.61%)