Gaze Gestures for Interaction in VR

Mark Tolchinsky
Clemson University
Clemson, SC, USA
mtolchi@g.clemson.edu

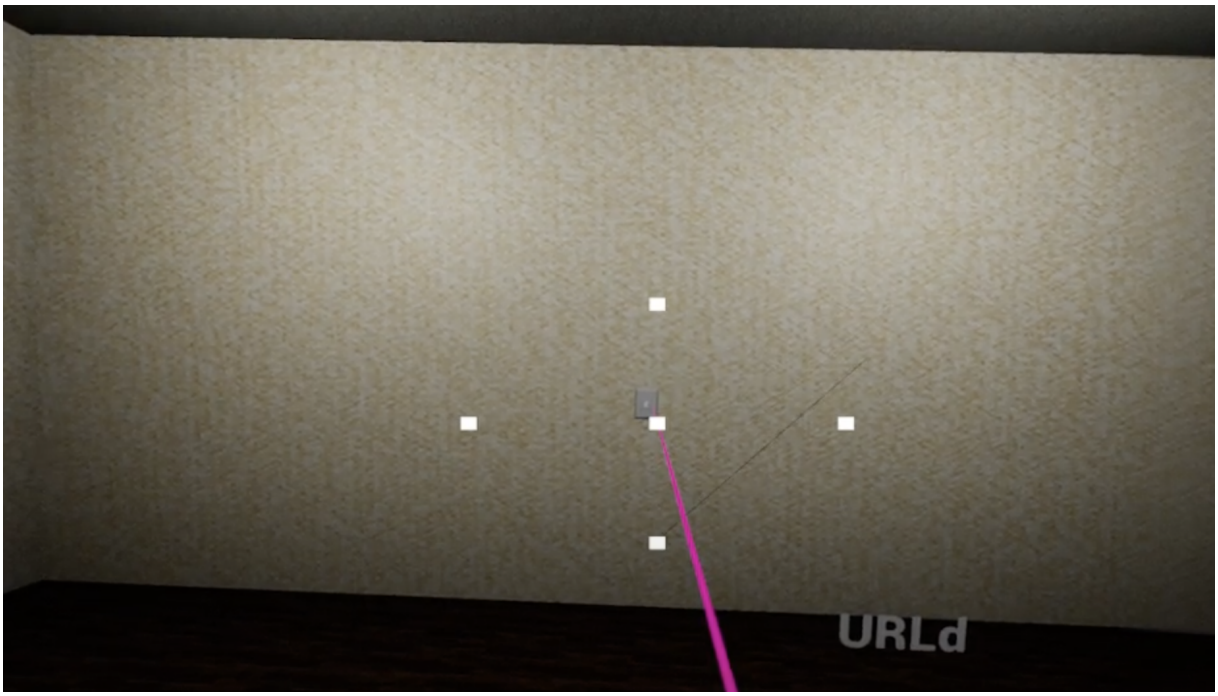

Figure 1: Gaze gesture system in use. Note the partial gesture written out in the bottom left.

ABSTRACT

This article proposes a system wherein gaze gestures are used to interact with a virtual environment. Compared to other gaze-based interaction metaphors, gaze gestures require less frequent calibration and may be faster or more comfortable to use. By utilizing head orientation for selection along with gaze gestures for action, the proposed system allows users to seamlessly access various actions quickly and from a distance. The development and functionality of a gaze gesture system is described, followed by suggestions for user studies and other future work.

CCS CONCEPTS

• **Computer systems organization → Embedded systems**.

ETRA '23, May 29–3 June, 2023, Tübingen, Germany

ACM ISBN 978-1-4503-8344-8/21/05…$15.00
https://doi.org/10.1145/3448017.xxxxxxx

KEYWORDS

eye tracking, gaze gestures, gaze control, virtual reality

ACM Reference Format:
Mark Tolchinsky. 2023. Gaze Gestures for Interaction in VR. In *2023 Symposium on Eye Tracking Research & Applications, May 29–3 June, 2023, April 23–28, 2023, Tübingen, Germany*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3448017.xxxxxxx

1 INTRODUCTION

As extended reality (XR) technologies develop, they will naturally become both more powerful and accessible. While XR head mounted devices (HMDs) are computer displays much in the way that monitors are, they facilitate dramatically different interaction techniques. With no need to rely on a desk-bound mouse and keyboard to provide input, XR systems have historically mainly used handheld controllers that can be manipulated in 3D space. However, as eye trackers are becoming a more common addition to HMDs, using the eye as an input device has been an increasingly popular prospect [Plopski et al. 2022].

Eye gaze based interaction systems have several benefits over more traditional methods, such as those that employ controllers or head

orientation. As compared to controller based methods, eye gaze interaction does not rely on dedicated external equipment to provide input, as eye trackers tend to be built into the HMD. Furthermore, individuals with motor disabilities may find it difficult or even impossible to effectively interact with an XR system using a controller. Surveys conducted by Mott et al. [2020] have suggested that eye gaze may serve as an effective alternate input method. When compared to purely head orientation based interaction, results are inconclusive. However, some studies have found that eye-based tracking is less exhausting [Blattgerste et al. 2018], faster [Kytö et al. 2018], or more comfortable [Qian and Teather 2017] depending on the task. While these factors do not conclusively establish eye gaze as a superior input method, they do suggest that it is worth further investigation.

Problems do arise in the use of eye gaze as input. Most notably, the 'Midas Touch' problem emphasizes the eyes' role as an information gathering tool, forcing developers to carefully consider selection criteria. Most commonly, eye gaze input systems use dwell time to confirm eye-based selections, which slows down the user's interactions [Jacob 1990]. Additionally, eye tracking systems have to be carefully calibrated, or else the user's gaze will not line up properly with the system's interpretation of it, hindering both speed and accuracy. Gaze gestures present a potential workaround to these issues. Because gestures are designed to be intentional and not overlap with eye movements associated with information gathering, the only limitation to interaction speed is how quickly the user can perform the gesture. Calibration also becomes less important, since gaze gestures rely on gaze movement (saccades or smooth pursuits) rather than the gaze position [Vidal et al. 2013]. With all of these factors in mind, gaze gestures present an attractive opportunity for hands-free XR interaction that sidesteps many of its siblings' common issues.

2 BACKGROUND

The gaze gesture system described in this paper uses head gaze and eye gaze in tandem to ensure intentional action from the user. This design was inspired by Stellmach and Dachselt, who coined the phrase "gaze suggests, touch confirms" in reference a similar paradigm [Stellmach and Dachselt 2012]. In their work, the researchers describe a system wherein eye gaze selects a general area in which to interact, while a hand-operated cursor actually performs the interactions. By using this dual-stage method, they found that common gaze interaction issues, such as Midas touch and tracker inaccuracy, could be greatly mitigated. This work has inspired the design of the system outlined later, which uses the head to "suggest" and eye gaze to "confirm".

While using gaze gestures to interact with a virtual environment is somewhat novel, there have been many implementations of similar systems in other contexts. For example, Istance et al. [2010] developed a gaze gesture-based system that was used to control a video game. Their system displayed 5 regions on the screen, and sets of fixations in these regions were interpreted as gaze gestures. Users had an easy time performing discrete actions using these gestures, whether they consisted of 2 or 3 consecutive fixations. However, such a method requires the regions to be displayed on screen at all times and may require dwell time to recognize fixations, and thus may not be suited to XR.

A different implementation by Drewes and Schmidt [2007] recognizes series of saccades as gestures instead. In this implementation, they eye tracking system qualifies all saccades as occurring in one of the 8 cardinal or ordinal directions. If the system recognizes a string of consecutive saccades as a predefined gesture, the action associated with it is performed. The results suggest that even complicated gestures consisting of 5 or 6 saccades can be reliable and comfortable to perform. Furthermore, the contents of the screen do not seem to significantly affect a user's ability to accurately perform the desired gesture. With these factors in mind, this saccade-based gesture system served as the inspiration for the one described in this article.

3 DESIGN

3.1 Apparatus

This system was built using HMD an HTC Vive Pro Eye. The integral eye tracker samples at a rate of 120 hZ, with a combined FOV of 110° and an accuracy of 0.5-1.1°.

3.2 System design

The proposed system allows users to interact with a virtual environment using gaze gestures. The environment used in this study situates the participant in a room in the presence of various interactable virtual devices. By leveraging gaze gestures, participants can issue commands to these devices quickly, from a distance, and without using their hands. Such a system is expected to be superior to traditional gaze-based interaction techniques in that it can reduce the amount of accidental input due to normal looking behavior, and leads to faster interactions because dwell time delays are unnecessary.

The devices in the environment are akin to typical household IoT devices, such as a smart TV and thermostat. The participant can select a device to interact with by using their head orientation. Once the head gaze is aligned with the position of a device, it is considered selected and the gaze gesture system activates. Navigating the head gaze away will almost immediately un-select the device and deactivate the gesture system. Because gaze gestures are difficult to preform on accident [Bâce et al. 2016], there is no need to confirm the selection of a device with head gaze. While an object is selected, 5 white dots appear in a "plus" pattern to guide the user's saccades.

Once a device is selected, the participant can perform one of several eye gaze gestures in order to interact with it. Gaze gestures are detected by recognition of consecutive saccade patterns. When a device is selected, the system can identify saccades in real time by comparing current gaze direction to recent gaze direction. If the gaze is considered to have moved far enough to constitute a saccade, the system can then assign it a direction. Akin to the system designed by Drewes and Schmidt [2007], each saccade most closely corresponds to a cardinal (up, down, left, right) direction. Ordinal

(diagonal) directions are ignored by the system to maintain simplicity of gestures. Once the system detects a pattern of saccades that matches a predefined gesture, the device performs the action corresponding to the gesture. For instance, the gesture to dim the lights using the lightswitch is "up, down". If the user performs saccades in this order while keeping their head gaze pointed at the light switch, the lights in the virtual environment will dim (see figure 2).

In order to ensure that regular looking behavior is not misinterpreted as gestures, the system takes several measures to ensure a gesture was intentional. Firstly, if head gaze deviates too far from a selected device, it becomes un-selected and any recent saccades are therefore ignored. There is a brief grace period when un-selecting a device to ensure that this action is also intentional. Additionally, fixations that exceed a certain time threshold cause the system to assume that any recent saccades are also part of normal looking behavior. Participants will therefore be expected to perform gestures reasonably quickly. If the participant is not attempting to gather information from the environment, consecutive saccades should be easy to perform; if the participant is using their eyes as normal to gather information, the fixations will prevent the system from recognizing their movements as a gaze gesture. Finally, there will be a minimum distance threshold for a saccade to be considered part of a gesture. As such, unconscious microsaccades will not contribute to gestures. Using these methods, the system ensures that accidental gestures are kept to a minimum.

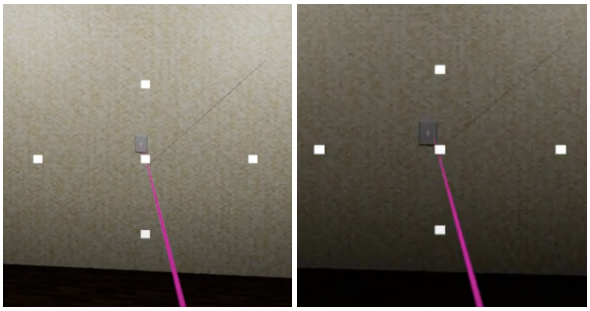

Figure 2: The gaze gesture system in use to dim the lights. Note the pink ray indicating head gaze, which must hit the light switch to activate the system.

3.3 Proposed study designs

3.3.1 Usability study. In order to test the effectiveness of the proposed system, participants are placed in a novel scenario in which they use gaze gestures to complete tasks within the environment. As described above, the environment contains various devices with which the participant can interact. Each device has its own set of commands corresponding to gaze gestures, but similar commands across different devices use the same gesture. For example, "turn on" uses an up, down gesture for all applicable devices, and "turn off" uses a down, up gesture.

The participant is first trained to use the gaze gesture system. They are informed how to select a device, and what gestures correspond to which actions. Once a participant has learned how to interact with the environment, they are presented with a randomly ordered series of tasks. These come in the form of spoken requests, from a "roommate" that is not visible to the participant, assumed to be elsewhere in the house. A typical request may be, "I'm trying to study; could you make less noise?" The participant would then be expected to use gaze gestures to turn down all noise-making devices in the environment. After the completion of each task, the environment resets to ensure the same starting conditions for each task. The system records the amount of time taken to complete the task and the number of accidental or incorrect gestures the participant performs. Furthermore, questionnaires and interviews investigating user comfort and perceived performance provide further insight into the efficacy of the system.

3.3.2 Comparative study. A comparative study may be used to evaluate the efficacy of gaze gestures as opposed to a more traditional gaze selection method. A monofactorial design is proposed, manipulating the gaze selection method between subjects. In the gaze gesture condition, a procedure similar to the usability study outlined above is performed, using the gaze gesture system to complete tasks in the virtual environment. The traditional gaze selection condition uses a similar environment but with a dwell time-based selection method, such as the Kuiper Belt system developed by Choi et al. [2022]. In this system, dwell time is used to confirm menu selections, but the menu items are placed at distant angles from the center of the user's field of vision in order to reduce accidental input. With such a safeguard, the actual dwell time used can be reduced. The proposed advantages of the Kuiper Belt system are similar to those of the gaze gesture system, so using them both for similar tasks and comparing users' speed and comfort will help assess the viability of gaze gestures as an interaction metaphor.

4 DISCUSSION

Over the course of designing and implementing the gaze gesture based system, several factors and considerations regarding this technique came to light. Firstly, when designing gestures for interaction, a balance must be struck between the complexity of the gesture as a whole and the complexity of individual movements. Because control of eye movement is not entirely voluntary [Fischer et al. 2000], it is imperative to compose gestures out of movements that are as simple as possible. At first, the system was designed to use saccades in both cardinal and ordinal directions, and with no visual aid. However, these complicating factors made it difficult for users to precisely compose gestures. As such, visual "anchors" were added when the gaze gesture system is active in order to guide user saccades along the directions in which they are classified. Furthermore, ignoring ordinal directions helps to avoid ambiguous saccades, as the cardinal directions are so distinct from one another that users could consistently input the correct directions. Only using cardinal directions limits the number of unique gestures composed of a specific number of saccades, so this solution may not be viable for systems that require access to many actions simultaneously. Reusing gestures with similar functionalities in different contexts can help keep each one short. For example, actions associated with increasing intensity (such as increasing light brightness or television audio volume) were assigned an "up, down" gesture whenever

possible, even if the actions were not identical across all contexts. Gestures that are too long may be harder to remember and cause fatigue in users, so making use of shorter gestures can increase the usability of the system.

The limitations of virtual reality apply some constraints to gaze gestures as well. While the Vive Pro's eye tracker is reasonably accurate within a narrow FOV, as gaze travels to more extreme angles the tracking begins to lose consistency. At first, the system used gestures that would send gaze far from the center of the display (for instance, "up, left"). While such motions were generally not difficult for users, the eye tracker would fail to recognize gaze angles that were too far from center, and as such saccades that ended in such gaze angles were detected inconsistently. As such, all of the gestures return the eye gaze to center before requiring additional motions (for instance, "up, down, right" rather than "up, right, down"). An additional consideration for use of gaze gestures in VR is cybersickness [Bruck and Watters 2011]. Use of gaze gestures over an extended period of time may cause eye strain, especially due to the exaggerated size of gesture saccades. While no formal study has been conducted here to evaluate the effect of gaze gestures on cybersickness, it may be a factor worth considering, particularly if the system also causes fatigue in users.

5 CONCLUSION AND FUTURE WORKS

In this work, a gaze gesture system for interaction in virtual reality was described, as well as its advantages, disadvantages, and some design considerations. Gaze gesture systems require careful design in order to ensure that they are both effective and comfortable to use, as they may cause users discomfort if they are poorly designed. The system detailed above prioritized simplicity and ease of use, to ensure that users could quickly become accustomed to it and leverage its advantages in speed and comfort. Furthermore, suggestions for validation of this system have been provided. Namely, procedures for a usability study and a comparative study to evaluate the gaze gesture system's efficacy were outlined. In the future, in addition to validation, further work may involve expanding upon the system or transferring it for use in different extended reality media. For now, the system provided minimal visual feedback, and cannot rigorously handle blinking; these limitations can be overcome in future iteration. Afterwards, a similar gaze gesture system could be implemented in augmented reality in order to interface with IoT devices in the real world.

ACKNOWLEDGMENTS

This work is supported in part by the U.S. National Science Foundation (grant IIS-1748380), and the . . . Any opinions, findings and conclusions or recommendations expressed in this material are the author(s) and do not necessarily reflect those of the sponsors.

REFERENCES

Bâce, M., Leppänen, T., de Gomez, D. G., and Gomez, A. R. (2016). Ubigaze: Ubiquitous augmented reality messaging using gaze gestures. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications*, SA '16, New York, NY, USA. Association for Computing Machinery.

Blattgerste, J., Renner, P., and Pfeiffer, T. (2018). Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. In *Proceedings of the Workshop on Communication by Gaze Interaction*, pages 1–9.

Bruck, S. and Watters, P. A. (2011). The factor structure of cybersickness. *Displays*, 32(4):153–158.

Choi, M., Sakamoto, D., and Ono, T. (2022). Kuiper belt: Utilizing the "out-of-natural angle" region in the eye-gaze interaction for virtual reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–17.

Drewes, H. and Schmidt, A. (2007). Interacting with the computer using gaze gestures. In *Human-Computer Interaction–INTERACT 2007: 11th IFIP TC 13 International Conference, Rio de Janeiro, Brazil, September 10-14, 2007, Proceedings, Part II 11*, pages 475–488. Springer Berlin Heidelberg.

Fischer, B., Gezeck, S., and Hartnegg, K. (2000). On the production and correction of involuntary prosaccades in a gap antisaccade task. *Vision Research*, 40(16):2211–2217.

Istance, H., Hyrskykari, A., Immonen, L., Mansikkamaa, S., and Vickers, S. (2010). Designing gaze gestures for gaming: An investigation of performance. In *Proceedings of the 2010 Symposium on Eye-Tracking Research Applications*, ETRA '10, page 323–330, New York, NY, USA. Association for Computing Machinery.

Jacob, R. J. (1990). What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 11–18.

Kytö, M., Ens, B., Piumsomboon, T., Lee, G. A., and Billinghurst, M. (2018). Pinpointing: Precise head-and eye-based target selection for augmented reality. association for computing machinery, new york, ny, usa, 1–14.

Mott, M., Tang, J., Kane, S., Cutrell, E., and Ringel Morris, M. (2020). "i just went into it assuming that i wouldn't be able to have the full experience": Understanding the accessibility of virtual reality for people with limited mobility.

Plopski, A., Hirzle, T., Norouzi, N., Qian, L., Bruder, G., and Langlotz, T. (2022). The eye in extended reality: A survey on gaze interaction and eye tracking in head-worn extended reality. *ACM Comput. Surv.*, 55(3).

Qian, Y. Y. and Teather, R. J. (2017). The eyes don't have it: an empirical comparison of head-based and eye-based selection in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, pages 91–98.

Stellmach, S. and Dachselt, R. (2012). Look & touch: gaze-supported target acquisition. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2981–2990.

Vidal, M., Bulling, A., and Gellersen, H. (2013). *Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets*, page 439–448. Association for Computing Machinery, New York, NY, USA.