EyeTTS: Evaluating and Calibrating Eye Tracking for Mixed-Reality Locomotion

Satyam Awasthi*[†] Vivian Ross*[‡] Michael Beyeler[§] Tobias Höllerer[¶]
University of California, Santa Barbara
Santa Barbara, CA, USA


Head-constrained Body-constrained Screen-Stabilized Walking World-Stabilized Walking Body-Stabilized Walking Hallway

Figure 1: Tasks from our user-study performed on augmented reality (AR) headsets to test the in-built eye-trackers' accuracy

ABSTRACT

Eye tracking is an increasingly popular method for interacting with AR and measuring user attention. However, implementing and evaluating eye tracking across multiple platforms and use cases can be challenging due to the lack of standardized metrics and measurements. Additionally, existing calibration methods and accuracy measurements often do not account for the common scenarios of walking and scanning in mobile AR settings. To test and compare different eye tracking devices on various AR tasks and metrics, we developed EyeTTS, an eye tracking test suite specifically designed for scenarios involving head movement and locomotion in AR. We conducted user studies on the Magic Leap (n=36, 1 trial per task) and HoloLens 2 (n=54, 2 trials per task) devices to collect data and compare the precision and accuracy of each headset under different movement and reference frame conditions.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality; Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User studies; Human-centered computing—Ubiquitous and mobile computing—Ubiquitous and mobile devices—Mobile devices;

1 INTRODUCTION

Recent advancements in eye-tracking technology have made it more accessible and affordable for a wider range of applications [4]. As such, it is important that developers and users understand the capabilities and limitations of their eye tracking devices during use. With the rise of AR head-worn displays such as the HoloLens 2, Magic Leap, Meta Quest Pro, and the recently-announced Apple Vision Pro, eye trackers have been integrated into various platforms and use cases. However, optimally utilizing eye-tracking technology on multiple platforms and standardizing their performance is a difficult and complex problem.

Specific metrics, such as accuracy and precision, are often used to evaluate the performance of eye-tracking systems [1, 4, 5]. Although existing work has used these metrics to examine the accuracy of eye trackers while stationary, there has been little research into how locomotion affects eye tracking, even though there is considerable interest in mobile eye tracking for specific applications, such as assessing behaviors for shopping, navigation, or wayfinding [2, 6]. Current eye tracking device calibration methods rely on users keeping their heads stationary, without incorporating any head or body movement. It is unclear whether these methods are effective when a user intends to use eye tracking while walking.

As a step towards more unified eye tracking assessment while walking, we developed EyeTTS, a test suite designed to assess eye tracking performance in various AR scenarios that involve movement and locomotion. Using EyeTTS, we conducted two user studies, with the HoloLens 2 and Magic Leap headsets respectively. We analyzed eye tracking performance in different movement conditions in pursuit of a possible recalibration procedure that may improve eye tracking accuracy during locomotion. Our analysis revealed that 1) eye tracking accuracy tended to decrease as head and body motion increased, and 2) our two hitherto tested AR headsets yielded significantly different results on several of our tasks

2 DEVELOPMENT OF THE EYETTS TEST SUITE

The EyeTTS test suite was initially developed for the Magic Leap headset and later expanded to include support for the HoloLens 2 and the Oculus Quest Pro headsets. Our test suite included tasks that required participants to maintain their gaze on stimuli, which is common in eye gaze studies [3, 7], but we expand the focus to include walking tasks. The stimuli could be stationary or moving relative to the user and could be presented on a world-stabilized, head-stabilized, or body-stabilized reference frame. This allowed for data collection from conditions representative of various situations in AR.

EyeTTS was created using Unity 2019.4.30f1, with each task designed as a Unity scene. These scenes are adaptable to each headset to ensure an equitable comparison of their performances. Some scenes are tailored for each device, taking into account their individual capabilities. Notably, the body-stabilized task was omitted for the Hololens 2 due to its absence of a controller that could serve as a body frame of reference.

3 USER STUDY

Two separate user studies were carried out to evaluate the performance of eye tracking on different devices. The initial user study (n=36) was conducted utilizing the Magic Leap One headset. The subsequent user study (n=54) was performed using the HoloLens 2 headset, with slight modifications based on feedback and results from the first user study. While in-depth comparisons between the two headsets are limited due to the differences between the studies, some broad comparisons are elucidated in Sect. 5. We outline the tasks presented in the two studies below.

Recalibration (R) Users rest their head on a chin rest and watch a tracking stimulus on the screen, which shifts to random positions after a brief delay.

---

*Awasthi and Ross contributed equally to this work

[†] e-mail:satyam@ucsb.edu

[‡] e-mail:vivianross@ucsb.edu

[§] e-mail:mbeyeler@ucsb.edu

[¶] e-mail:holl@cs.ucsb.edu

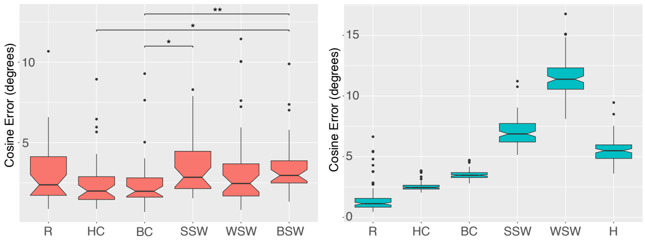

Figure 2: Cosine error per task for the Magic Leap (left) and HoloLens 2 (right). Notches show 95% confidence interval, * denotes $p < 0.05$, and ** denote $p < 0.01$. Significance bars for right omitted as all pairs differ significantly from each other.

Head-constrained (HC) Users rest their head on a chin rest and watch a tracking stimulus, which moves along a path on the screen inside their field of view (FOV).

Body-constrained (BC) Users sit and turn their heads to watch a tracking stimulus, which moves along a path in the world frame of reference (FOR) and wider than their FOV.

Screen-Stabilized Walking (SSW) Users walk in circles around a table while watching a tracking stimulus, which moves along a path on the screen inside their FOV.

World-Stabilized Walking (WSW) Users walk in circles around a table while watching a tracking stimulus, which moves along a path in the world FOR above the table.

Body-Stabilized Walking (BSW) Users walk in circles around a table while watching a tracking stimulus, which moves along a path constrained to a controller strapped to their body.

Hallway (H) Users walk down a hallway while watching a tracking stimulus, which moves down the hallway in front of them.

4 RESULTS

Results for both the Magic Leap and the HoloLens 2 study are summarized here. Similar to [1], we employed the arccos of the dot product between the gaze and stimulus vectors to establish the cosine error. Fig. 2 shows the graphs of cosine error split by task. Magic Leap data is shown on the left, HoloLens 2 data is shown on the right.

For the HoloLens 2, an ART ANOVA with task as an independent variable reveals a significant difference in cosine error between tasks ($F = 1877, p < 2.22 * 10^{-16}$), and a Bonferroni-corrected Mann-Whitney U test reveals a difference in cosine error between every pairwise combination of tasks.

For the Magic Leap, an ART ANOVA with task as an independent variable reveals a significant difference in cosine error between tasks ($F = 11.04, p = 2.95 * 10^{-9}$). A Bonferroni-corrected Mann-Whitney U test reveals a difference in cosine error between the head-constrained and body-stabilized walking tasks ($p = 0.019$), the body-constrained and the screen-stabilized walking tasks ($p = 0.031$), and the body-constrained and the body-stabilized walking tasks ($p = 0.00012$).

5 DISCUSSION

For the HoloLens 2, the error between tasks follows a very clear trend in which head-constrained tasks have the lowest error, followed by the body constrained task, then walking tasks. Latency between the head position and the eye position in all walking tasks may have resulted in the lack of accuracy for the walking tasks, in spite of some latency correction that we applied. In contrast, the Magic Leap data displayed a generally higher accuracy than the data collected from the HoloLens 2, with a limited loss of accuracy due to locomotion or head motion. The differences between the accuracy of the Magic Leap and the HoloLens 2 may be explained by the fit of the headsets. The Magic Leap rests on the nose, forming a secure fit around a user's head. In contrast, the HoloLens 2 has a visor that can be flipped up, which results in a lack of contact between the display and the face. The lack of stability of the HoloLens 2 on a user's head could explain why the eye tracker is significantly more accurate when the head is kept still, whereas the Magic Leap headset shows more consistent eye tracking accuracy throughout.

The only task for which it seemed the HoloLens 2 exhibited better eye tracking performance than the Magic Leap was the recalibration task. This can potentially be explained by the larger distance between the display and the user's eyes on the HoloLens 2, which may have allowed the eye tracking cameras to better capture eye gazes at the periphery of the display. Differences in the built-in calibration procedure may also have an effect here.

6 CONCLUSION AND FUTURE WORK

In this work we developed the Eye Tracking Test Suite, EyeTTS, which is designed to test the eye tracking accuracy of an AR device in various user scenarios. We conducted user studies on the Magic Leap and HoloLens 2 to analyze their eye tracking performance in different locomotion conditions. Our analysis revealed overall lower eye tracking accuracy while walking compared to stationary poses. Our work provides future researchers with a protocol and test suite for investigating the accuracy of the eye tracker of their AR device before use. A future goal is to identify a convenient custom recalibration method effective for AR during locomotion. A first analysis when testing each of our individual user study tasks for recalibration of the walking tasks suggests that we can indeed enact improvements to eye tracking performance while walking when a calibration technique involving stipulated head motion is employed.

ACKNOWLEDGMENTS

This work was supported in part by NSF grant IIS-2211784 as well as ONR grants N00014-19-1-2553 and N00014-23-1-2118. Thanks also go to Dr. Tom Bullock for helpful discussions and loaning of a chin rest.

REFERENCES

[1] S. Aziz and O. Komogortsev. An assessment of the eye tracking signal quality captured in the HoloLens 2. In *2022 Symposium on Eye Tracking Research and Applications*. ACM, June 2022. doi: 10.1145/3517031.3529626

[2] J. de Winter, P. Bazilinskyy, D. Wesdorp, V. de Vlam, B. Hopmans, J. Visscher, and D. Dodou. How do pedestrians distribute their visual attention when walking through a parking garage? An eye-tracking study. *Ergonomics*, 64(6):793–805, June 2021.

[3] A. T. Duchowski. A breadth-first survey of eye-tracking applications. *Behav. Res. Methods Instrum. Comput.*, 34(4):455–470, Nov. 2002. doi: 10.3758/BF03195475

[4] A. T. Duchowski. *Eye tracking methodology: Theory and practice*. Springer, 2017.

[5] S. Kapp, M. Barz, S. Mukhametov, D. Sonntag, and J. Kuhn. ARETT: Augmented reality eye tracking toolkit for head mounted displays. *Sensors*, 21(6), 2021. doi: 10.3390/s21062234

[6] P. Kiefer, I. Giannopoulos, M. Raubal, and A. Duchowski. Eye tracking for spatial research: Cognition, computation, challenges. *Spat. Cogn. Comput.*, 17(1-2):1–19, Jan. 2017.

[7] A. Poole and L. Ball. *Eye tracking in human-computer interaction and usability research: Current status and future prospects*, pp. 211–219. 01 2006.