Gaze tracker accuracy and precision measurements in virtual reality headsets

Jari Kangas, Olli Koskinen and Roope Raisamo

ABSTRACT

To effectively utilize a gaze tracker in user interaction it is important to know the quality of the gaze data that it is measuring. We have developed a method to evaluate the accuracy and precision of gaze trackers in virtual reality headsets. The method consists of two software components. The first component is a simulation software that calibrates the gaze tracker and then performs data collection by providing a gaze target that moves around the headset's field-of-view. The second component makes an off-line analysis of the logged gaze data and provides a number of measurement results of the accuracy and precision. The analysis results consist of the accuracy and precision of the gaze tracker in different directions inside the virtual 3D space. Our method combines the measurements into overall accuracy and precision. Separately, visualizations of the measurements are created to see possible trends over the display area. Results from selected areas in the display are analyzed to find out differences between the areas (for example, the middle/outer edge of the display or the upper/lower part of display).

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; **Virtual reality**; **Pointing**.

KEYWORDS

Eye Tracking; Virtual Reality; Accuracy measurement; Precision

ACM Reference Format:

Anonymous Author(s). 2018. Gaze tracker accuracy and precision measurements in virtual reality headsets. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

It is important to know how accurate is the gaze data that is being measured with a certain gaze tracker. There has been some effort to raise more interest on measuring and reporting the gaze tracker quality (see, for example, Holmqvist *et al.* [5] and Nyström *et al.* [7]), but that hasn't led to much action on real measurements. Some tools have also been developed to enable easy practice of getting quality measurements (for example, Akkil *et al.* [1] and Blignaut and Beelders [3]). Lohr *et al.* [6] used an eye tracking HMD based on HTC Vive and SMI's eye tracker to study the tracker signal quality.

For gaze data quality it has been a common practice to refer to the average numbers given by the tracker manufacturers. However, the manufacturer specifications have usually been measured in a well controlled environment and they don't necessarily translate to practical use cases in varying environments. It is important to measure the tracking quality achievable in the specific situation and using the equipment employed in the final use case.

Lately, many virtual reality (VR) headsets (head mounted displays, HMDs) have been equipped with gaze trackers (e.g. VIVE Pro Eye, [9]). These provide an attractive tool for trying out and studying all kinds of gaze related phenomena as the virtual environment is highly immersive and can be customized at will thus providing a relatively easy way to arrange a test environment where conditions can be flexibly changed. However, little is known of the gaze data quality in VR headsets. As the data quality is an important parameter in interpreting the results we proceeded to create an analysis method for measuring accuracy and precision of a gaze tracker.

Akkil *et al.* [1] listed several reasons to collect eye tracking quality measures (e.g. to compare between calibration methods or to exclude certain participants). In the present paper the main goal was to get a basic understanding of the expected gaze accuracy to help make informed decisions on gaze enabled user interfaces[1], and to understand possible differences in gaze tracking quality in different directions in the headset's field-of-view.

2 GAZE DATA QUALITY MEASURES

The two most important measures of gaze data quality are the spatial accuracy and spatial precision [7]. *Accuracy* is a measure of systematic errors, a measure of a kind of bias that may cause a systematic difference between the measured results and the true value. The systematic error may be due, for example, to the visual system or the technical equipment. *Precision* is a measure of random errors in measurements, a measure of variability, how much the values are spread during the repeated measuring of a single target. In Figure 1 we show an example of what the measures mean and how the measures are related to each other.

3 PHASE 1: GAZE DATA COLLECTION

Our gaze data analysis method consisted of two separate phases. In the first phase we collected the gaze data to participant-specific log files. In the second phase we made an off-line analysis of the log file data. The basic idea was to collect static gaze data (gaze direction data) from a number of participants while they are shown a set of small objects, one at a time, at known locations in a VR environment. By that manner, we can observe how well the tracker's

[1]For example, to use big enough targets that one can select by gaze, but avoid using unnecessarily large targets not to waste display area.

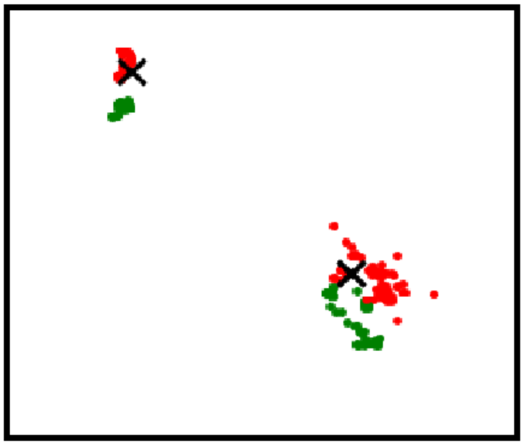

Figure 1: An illustration of the measures of accuracy and precision. The cross marks show the targets that the participant was asked to look at. Gaze samples are shown as red (left eye) and green (right eye) dots. In the upper left corner the left eye samples have high accuracy (=close to target) and high precision (=highly concentrated to one place), while right eye samples have high precision, as well, but lower accuracy (=slightly away from the target). In the right bottom corner both the left and the right eye samples have lower precision (=spread apart from each other) and lower accuracy.

estimated gaze direction matches the known object direction in the implementation.

The log files were self-contained, i.e. they had in them all the necessary information to compute the accuracy and precision. For each target session we first logged the position of the target in the VR environment. Then the gaze data, the gaze directions, were logged using a 60 Hz sampling rate. The gaze data was collected separately for both eyes. For some gaze samples the tracker indicated an 'invalid' status (for example for closed eyes) and we just skipped that sample.

3.1 Collecting hardware

The method can be used with any head mounted display that contains a gaze tracker and is integrated with Unity development environment [8].

3.2 Collecting software

The gaze measurements were done using a custom software that we developed in the Unity development environment. The software consisted of an initialization phase where the VR device was set up and the gaze tracker was calibrated[2], and then of a measurement phase during which the participant was shown a ball moving from location to location in an otherwise empty VR environment. The participant was asked to follow the target ball by his/her gaze. The locations were fixed in the display coordinates and not in the VR environment, which meant that they followed possible head turns. I.e., the participant had to turn his/her eyes and not head to see different targets.

People have a strong bias to look at the centre of the view [4], and we decided to concentrate on that area. However, we need to know of the the tracker quality on other areas, as well. In the end, we decided to use 29 different directions as shown in Figures 2 and 5. One location was in the very center of the viewing area, then 8 locations were on each of concentric circles 5, 10 and 20 degrees from the mid-point and the last 4 locations were on a circle 25 degrees from the mid-point. Nine middle points were repeated once to get more information of the center area. Overall, we were then collecting data of 38 different object occurrences.

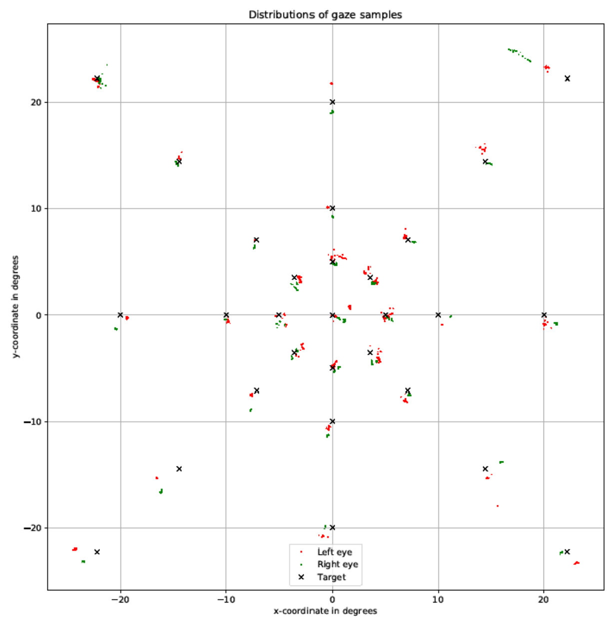

Figure 2: Gaze tracking data. The participant has been looking at the targets on shown locations (cross marks) and the collected data is visualized by colored dots. Red dots mean gaze positions by the left eye and the green dots by the right eye. The coordinates are the gaze direction angles in degrees from the middle point (vertical and horizontal). (Figure 1 is a closeup of similar data from near the center.)

The measure that we were most interested on, the distance between the target and the gaze direction was calculated as follows (see Figure 3): The target position was defined in the VR environment. A plane that was perpendicular to the line from the viewer to the target position was defined, after which the vector between the gaze point and the target position was automatically given by the Unity as the hit point of gaze vector and the plane. For completeness we separately collected the left eye position and the right eye position.

3.3 Collected data

In the data collection phase we were saving all the gaze samples that were flagged as valid, separately for left and right eyes. The

[2] In practice that was a call to the calibration routine provided by the gaze tracker manufacturer. We relied on manufacturer components as often as possible.

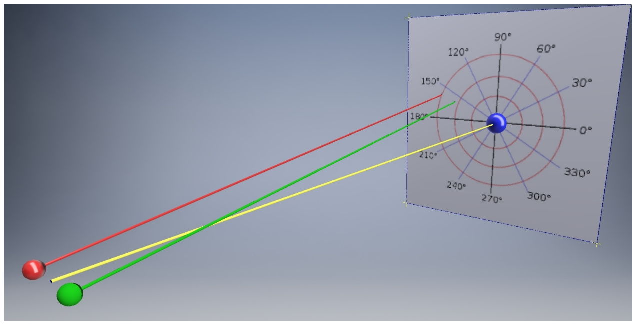

Figure 3: A gaze tracking measurement. The target is shown as a blue ball. A yellow line from the virtual reality coordinate center is drawn to the ball, and a perpendicular plane to that is defined. Green and red balls represent the right and left eyes, respectively. The gaze directions are shown as green and red lines. The locations where the gaze direction lines hit the plane determine the error in gaze tracking, both the magnitude of error and direction. (As shown here, the tracker estimate of right and left gaze directions do not usually hit the same location.)

collected data was in a separate log file for each participant. The data consisted of 38 different target sessions, where the location of the target was first logged and then the gaze data was logged using 60 Hz sampling frequency. For each gaze sample we saved the following data for both eyes:

(1) Position of the eye relative to the origin of the VR environment (the red and green balls in Figure 3).
(2) Gaze direction in the VR environment as a unit vector.
(3) The distance between the target and the gaze line hitting the target plane (see Figure 3). That distance was converted to view degrees.

In our initial analysis we were interested in only the item number 3, the distance between the target and gaze line.

4 PHASE 2: GAZE DATA ANALYSIS

In the second phase of our gaze data analysis we computed some quantitative measures and visualizations of the gaze tracker data. The analysis software was implemented with Python 3.4. In the preprocessing phase we removed all such samples where data of either eye was missing, i.e. we kept only data where measurement for both eyes were available.

4.1 Accuracy and precision calculations

We calculated the two measures, accuracy and precision, using the equations by Akkil et al. [1]. The accuracy was calculated as

$$\text{accuracy} = \sqrt{(x_{true} - x_{mean})^2 + (y_{true} - y_{mean})^2}, \quad (1)$$

where x_{true} and y_{true} are the x- and y-coordinates of the target point, and x_{mean} and y_{mean} are the coordinates of the mean gaze point in the collected data. The precision was calculated as

$$\text{precision} = \sqrt{\sigma_x^2 + \sigma_y^2}, \quad (2)$$

where σ_x and σ_y are the standard deviations on the sets of x- and y-coordinates over the data collection. As the input values are all in degrees of angle the results are also given as degrees of angle.

4.2 Gaze data selection

The gaze data is collected throughout the study. The focus was in determining of the gaze direction only when the target was not moving. Even then, there was a waiting time of 500 ms (same as in [1]) for the gaze to reach the target and be still, i.e. a short stabilization time was allowed for the eye. After that gaze samples were collected for 500 ms and the rest were discarded. In Figure 4 we show gaze accuracy measurements for short 100 ms intervals from the beginning of target onset. There is a clear stabilization period that lasts for about 300-400 ms. We then had (at most) 30 gaze samples for each target and for each participant.

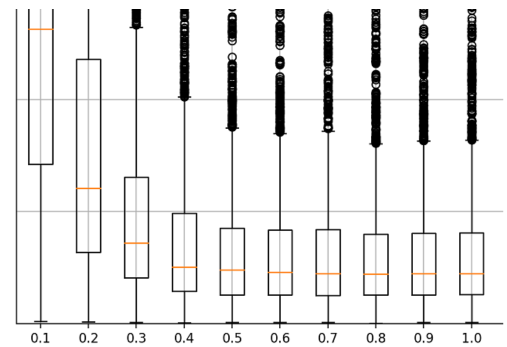

Figure 4: The development of the gaze accuracy measurements for short 100 millisecond intervals from the placement of the target object. The red line indicates the median value for each segment and the box contains values between first and third quartile.

5 ANALYSIS RESULTS

After the data selection we can visualize gaze data as points around the target directions (see Figure 2). In this specific data the gaze samples form rather tight clusters around most targets and the clusters are relatively close to the targets.

5.1 Quantitative results

A set of gaze points was collected around the true target position for each target and for each participant, from which we computed the median accuracy and precision numbers separately for each target and for each participant. These numbers can be combined in different ways to get a better understanding of the quality of the gaze tracker.

For overall quantitative analysis we computed the median values for both accuracy and precision over a selected set of directions and over all participants, as well as the median-absolute-deviation (MAD) value for accuracy and precision[3]. In Table 1 we show an

[3]We decided to use the median and MAD values as these are more robust against a small number of potential outlier values than the more common mean and standard deviation. Then we do not need a criteria for outlier removal.

	Median	MAD
Accuracy:	0.88	0.37
Precision:	0.15	0.04

Table 1: Example accuracy and precision measures, median and MAD values in degrees.

example of the results of one measurement session where all the target directions were analyzed.

As the main measures the accuracy and precision were computed over all 38 directions (29 separate directions, 9 directions repeated), and as secondary measures the same values were computed over 'inner' and 'outer' directions. By 'inner' directions we meant the innermost 17 directions, the very center and two circles around the center at 5 and 10 degrees. The 'outer' directions then contain the other 12 directions at 20 and 25 degrees from the center. Other divisions can be easily done. For example, it might be of interest to study the accuracy on the upper part versus the lower part of display.

5.2 Data visualizations

The quantitative results can be visualized in various ways. In Figure 2 we showed a single participant's gaze data. From that we could see, for example, if some targets have not been looked at.

In Figure 5 we visualize the distribution of the median accuracy values. From that one may observe if there are some differences between areas of display, like in the specific example there seems to be a difference between the top and bottom parts in accuracy.

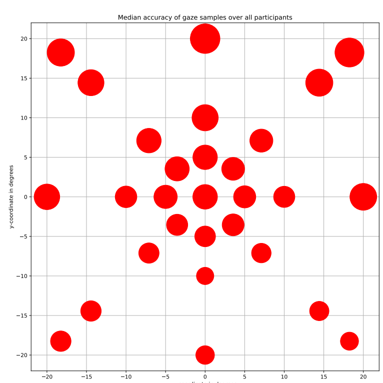

Figure 5: The accuracy visualization over the display. The area of each dot is relative to the median accuracy on a given location computed over the participants. One may notice that the accuracy is slightly better (smaller dots) on the lower part of the display than on the top part of the display.

In Figure 6 we show the distributions of the median accuracy values separately for left and right eye trackers. In this specific example there seems to be no real difference between the eyes.

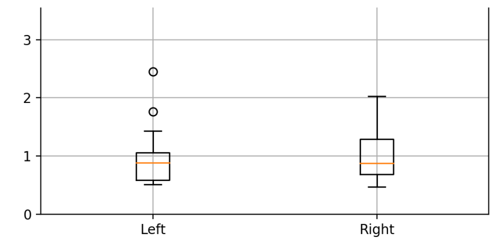

Figure 6: An example of a box-plot shows the distributions of the median accuracy values over all the targets and all participants on left and right eyes. The tracker results look very similar for both eyes.

6 DISCUSSION AND CONCLUSIONS

We have developed a straightforward method to collect and analyze the accuracy and precision of a gaze tracker in a virtual reality headset. The method consists of Unity software to collect the data and data analysis tools to compute quantitative statistics and visualizations.

The target of the method is to help make informed decisions on gaze enabled user interfaces based on accuracy and precision expectations. We were also interested in seeing if there were significant differences in accuracy measurements between display areas, to take those into account in the interface design. In parallel with our work Adhanom *et al.* [2] have developed a tool for accuracy and precision measurements in HMD gaze trackers. They were more interested on evaluating the overall quality of the tracker, while we visualized the accuracy on different parts on the viewing area.

The Unity software has to be tailored for different HMDs if/when the gaze tracker interface protocol changes, but other than that all the same components can be used between devices. We are planning to release the software for general use in a Github page.

The method enables its user to make an analysis of the gaze tracker quality in a head mounted display device when implementing gaze enabled interfaces. This means that the interface components can be designed for this particular device and the user will have a good user experience. The developer avoids costly surprises in the implementation that may arise if problems in the gaze tracker are noticed only in the deployment. The method makes it easy to compare the gaze tracker accuracy and precision in different head mounted displays and prepare for possible adaptation needs when a system is transferred to another device.

REFERENCES

[1] Deepak Akkil, Poika Isokoski, Jari Kangas, Jussi Rantala, and Roope Raisamo. 2014. TraQuMe: a tool for measuring the gaze tracking quality. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, 327–330.

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

[3] Pieter Blignaut and Tanya Beelders. 2012. TrackStick: A Data Quality Measuring Tool for Tobii Eye Trackers. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12)*. ACM, New York, NY, USA, 293–296. https://doi.org/10.1145/2168556.2168619

[4] Alasdair DF Clarke and Benjamin W Tatler. 2014. Deriving an appropriate baseline for describing fixation behaviour. *Vision research* 102 (2014), 41–51.

[5] Kenneth Holmqvist, Marcus Nyström, and Fiona Mulvey. 2012. Eye tracker data quality: what it is and how to measure it. In *Proceedings of the symposium on eye tracking research and applications*. ACM, 45–52.

[6] Dillon J Lohr, Lee Friedman, and Oleg V Komogortsev. 2019. Evaluating the Data Quality of Eye Tracking Signals from a Virtual Reality System: Case Study using SMI's Eye-Tracking HTC Vive. *arXiv preprint arXiv:1912.02083* (2019).

[7] Marcus Nyström, Richard Andersson, Kenneth Holmqvist, and Joost Van De Weijer. 2013. The influence of calibration method and eye physiology on eyetracking data quality. *Behavior research methods* 45, 1 (2013), 272–288.

[8] Unity. 2020. *Unity Real-Time Development Platform.* https://unity.com/

[9] HTC VIVE. 2019. The New VIVE Pro With Precision Eye Tracking. https://www.vive.com/eu/pro-eye/