My Eyes Hurt: Effects of Jitter in 3D Gaze Tracking

Moaaz Hudhud Mughrabi*
Department of Mechatronics Engineering
Kadir Has University

Aunnoy K Mutasim†
School of Interactive Arts & Technology
Simon Fraser University

Wolfgang Stuerzlinger‡
School of Interactive Arts & Technology
Simon Fraser University

Anil Ufuk Batmaz§
Department of Mechatronics Engineering
Kadir Has University

Abstract

Jitter, small fluctuations in the signal, is one of the major sources for a decrease in motor performance and a negative user experience in virtual reality (VR) systems. Current technologies still cannot eliminate jitter in VR systems, especially in the eye-gaze tracking systems embedded in many head-mounted displays. In this work, we used an HTC Vive Pro Eye, artificially added 0.5°, 1°, and 1.5° jitter to the eye-tracking data, and analyzed user performance in an ISO 9241:411 pointing task with targets at 1 or 2 meters visual distance using angular Fitts' law. The results showed that the user's error rate significantly increases with increased jitter levels. No significant difference was observed for time and throughput. Additionally, we observed a significant decrease in performance in terms of time, error rate, and accuracy for the more distant targets. We hope that our results guide researchers, practitioners, and developers towards better gaze-tracking-based VR applications.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction devices—Pointing devices; Human-centered computing—Human computer interaction (HCI)—Interaction devices—Graphics input devices

1 Introduction

For 3D interaction with a virtual environment and the 3D objects in it through a Virtual Reality (VR) systems, selection plays a critical role. For successful 3D selection, the user has to point accurately towards the desired target and the pointing device has thus to be in the correct position and orientation within the virtual world. To facilitate such pointing in virtual environments, the user is (usually) provided with some form of feedback, such as highlighting, when they correctly point to a target. Once the cursor is on the correct target, the user then confirms its selection, typically with a button click.

There are two aspects to pointing in virtual environments: first, the user points to the correct target, and the input system then transfers the pointing pose from the real world to the virtual world. During this transfer, a signal is generated by the input devices and sent to the application software. This generated signal also contains fluctuations, called jitter. Such jitter can be observed in all stages of pointing. When the pose data of the gaze is received by a software on a computer, this data is typically processed with a filtering algorithm, such as the One-Euro filter [11] or Kalman Filter. These filtering algorithms can also add (temporal) jitter due to the phase shift they introduce.

Apart from the different types of technical jitter that impact user interaction even when the device is stably hovering in mid-air, different interaction actions can also add additional jitter. For instance, blinking or moving the head can cause unintentional pose changes.

*e-mail: moaaz.mazen@stu.khas.edu.tr
†e-mail: amutasim@sfu.ca
‡e-mail: w.s@sfu.ca
§e-mail: aubatmaz@khas.edu.tr

Apart from all the different sources of jitter in input devices, jitter can be also observed in other parts of a VR system, such as the tracking of the pose of the head-mounted display.

When a 2D mouse is left stable on a table, the amount of jitter recorded by the system is usually non-existent [40], due to surface friction and well-tuned tracking methods. On the other hand, if a VR controller is left stable on a table, it is typically still possible to observe tracking jitter in its pose. This jitter is even more visible when the user points at a target on mid-air, where there is then substantial movement even outside the human tremor band (4-12 Hz) [19, 38].

One of the (many) ways to point virtual objects in 3D environments is using eye-gaze for pointing. Gaze tracking is an interesting input modality for VR systems, and can be beneficial especially when the user's hands are full or occupied with other tasks [1, 9, 37]. Moreover, gaze is a fast input method compared to hands, and gaze movement speed can reach up to 900°/s [1]. For example, surgeons typically hold instruments in their hands during an operation; in such a situation gaze tracking can come in handy. Another scenario where gaze might be useful is when the required reaction time is too short for a hand movement.

Jitter can greatly affect gaze tracking data. In current gaze tracking systems, there is always a certain level of jitter present, due to the accuracy and precision limitations of the current eye-tracking technologies [14] as well as natural eye movements [23]. However, when considering gaze tracking as an input modality for user interaction in the virtual world, such jitter will directly affect the performance of said interaction [17]. Even if the gaze tracking data is only used to the create "heat maps" of where the user looked, e.g., for product testing or prototyping [43], too much jitter can disrupt the data and make it less reliable.

With a series of experiments, previous work explored the effect of jitter on user pointing performance [6–8, 17]. In the current work, we extend the previous literature by exploring the effect of jitter in those eye-gaze tracking systems that are embedded in modern VR head-mounted displays (HMDs).

Our contributions in this work are as follows:

- *Investigating the effect of varying amounts of jitter on user performance and experience for gaze tracking devices.* Results showed that 0.5° additional jitter already significantly increases the error rate of the participants.
- *Investigating the effect of target depth on user performance and experience for gaze tracking devices.* Results showed that subjects are slower, less accurate, and that they made more errors with distant targets.

Towards our goal, we asked 12 participants to point to targets with their gaze (i.e., combination of head- and eye-gaze [9]), and then pressing the space bar on the keyboard to select them. To investigate the effect of jitter on gaze tracking devices, we used an ISO 9241:411 task with 11 targets.

2 PREVIOUS WORK

2.1 Fitts' Law

Fitts' law [16] models human movement time for pointing. Equation 1 shows the Shannon formulation [24] with Euclidean measures.

$$MovementTime = a + b * \log_2\left(\frac{A}{W} + 1\right) = a + b * ID \quad (1)$$

In Equation 1, a and b are empirical constants, typically identified by linear regression. A is the amplitude of the movement, which is the distance between two targets, and W the target width. The logarithmic term in Equation 1 represents the task difficulty and is called the *index of difficulty, ID*.

For pointing tasks in 3D environments, several variations that use an angular *ID* have been proposed in the literature [3, 12, 13, 21, 30, 39, 42]. Our work does not aim to analyze or propose a novel angular ID equation. For simplicity, we thus used Kopper et al.'s angular *ID* formula:

$$ID_{angular} = \log_2\left(\frac{\alpha}{\omega^k} + 1\right) \quad (2)$$

In Equation 2, α represents the angular distance between targets and ω represents the angular target width. The constant k represents a relative weight between α and ω. For simplicity, we set k = 1. To convert Euclidean distances to angular measures we used the same method as Kopper et al. [21].

We also use throughput (based on effective measures), as defined in the ISO 9241-411:2012 [20]:

$$Throughput = \left(\frac{ID_e}{MovementTime}\right) \quad (3)$$

In Equation 3, movement time is the time between initiation of the movement and the selection of the target. The effective index of difficulty (ID_e) is defined as in Equation 4:

$$ID_e = \log_2\left(\frac{A_e}{W_e} + 1\right) \quad (4)$$

In Equation 4, A_e represents the effective distance, the actual movement distance to the target position, and W_e is the effective target width, the distribution of selection coordinates, calculated as $W_e = 4.133 \times SD_x$, where SD_x is the standard deviation of selection coordinates along each task axis. SD_x represents the accuracy of the task performance [25, 26].

2.2 Jitter in 3D pointing

The adverse effect of jitter in VR systems was first analyzed by Teather et al. [40], showing that a small amount, only 0.3 mm, of spatial jitter in the input device already significantly decreased user performance. A further study identified that the negative impact of larger jitter levels increases with smaller targets [32].

Building on Teather et al. [40], Batmaz and Stuerzlinger [7] investigated the effects of jitter on user performance with a modern VR controller. This study used an HTC Vive Pro system, which uses one of the best tracking systems currently available on the market at that time. The authors generated artificial jitter using a uniform distribution. The results of this study identified that there is no significant difference between no jitter and $\pm 0.5°$ jitter in terms of execution time. On the other hand, the error rate significantly increased, and participants' throughput performance significantly decreased with $\pm 1°$ jitter or more. A follow-up study [8] eliminated the potential confound of the "Heisenberg effect" [10, 44] by adopting a bi-manual selection technique and analyzed the negative impacts of White Gaussian Noise (WGN) jitter, again using an HTC Vive Pro setup. WGN is used to model random processes in information theory and using it for jitter more closely models the cumulative impact of multiple sources of jitter on a controller in real-life. To generate WGN, the authors used the Marsaglia Polar Method [28], a standard normal distribution generator, to generate random values with a mean of 0 and a standard deviation of 1. The results of the follow-up study [8] were similar to the original work [7], as higher levels of jitter again increased the participants' execution time and error rate, while also decreasing effective throughput performance. As in the original work [7], the authors observed significant negative effects of jitter at and above $\pm 1°$ rotational jitter.

One of the interesting findings of this study concerned the speed-accuracy trade-off of the participants under the impact of jitter. Subjects were taking longer with an increased amount of jitter, but their error rate did not decrease, and effective throughput also did not increase. The authors observed that when the participants had to select a target with a VR controller with jitter, the participants were waiting for a "better moment" to select targets, i.e., when the cursor might have stabilized temporarily. Yet, since the jitter was generated continuously, the cursor never stabilized. Thus, it took participants longer to select targets, which explains why there were no performance improvements.

To investigate the negative impact of target depth and positional target jitter, Batmaz et al. [6] ran a study, again with a HTC Vive Pro setup. Different from the previous work [7,8], the authors tested with targets at three depth distances, 0.75, 1.5, and 2.25 m, to analyze the impact of the control-display ratio. The results revealed that, when the depth distance increases, user performance significantly decreases in terms of time, error rate, and throughput. Similarly, user performance decreases with increased target jitter.

VR HMDs have several deficiencies, including stereo deficiencies that affect the user performance [2–4]. Some of this work showed that user performance in stereoscopic VR HMDs is affected when the user interacts with targets at varying visual depths, i.e., distances from the head, in terms of time, error rate, and throughput [3]. Moreover, previous studies on rotational and positional jitter showed that the user performance significantly decreases with increasing target depth for rotational jitter on VR controllers and positional jitter on the targets themselves [6].

2.3 Gaze tracking

As state-of-the-art technology now allows VR headset producers to embed a gaze-tracking device inside a HMD, gaze tracking is now frequently included in VR systems. Since gaze tracking cameras integrated into the HMD are (approximately) stabilized relative to the user's head, which stabilizes the gaze even when the users move their head, these implementations allow convenient and reasonably reliable gaze data acquisition.

As the gaze tracking can be used as an input method while the user hands are otherwise occupied, various studies focused on improving the user performance and user experience with gaze tracking devices. For instance, Blattgerste et al. [9] studied different field of views with gaze tracking, Feit et al. [15] studied the accuracy and precision in gaze tracking devices, Mutasim et al. [31] studied different selection techniques, and Schuetz et al. [35] used psychophysics to model user selection performance.

The performance of gaze tracking systems has also been studied using Fitts' law, e.g., [31, 34]. For instance, Zhang and MacKenzie studied the performance of an desktop eye tracking device with the ISO 9241:411 task [45]. Their results showed that using a mouse affords substantially higher performance compared to pointing with gaze tracking.

However, current gaze tracking technology is not as accurate as VR controllers [36]. Even though artificial jitter has been studied in various works [22, 29, 36], none of them targeted VR HMDs.

Previous work also typically used Gaussian noise to analyze the effect of jitter on the gaze tracking. For instance, Graupner et al. [17] added Gaussian jitter to the gaze tracking signals and observed a significant increase in task execution time and error rate in a monocular see-thorough HMD. In this work, we similarly also used White Gaussian Noise to add jitter.

3 MOTIVATION & HYPOTHESES

Previous work on rotational and positional jitter showed that user performance decreases significantly with increased amount of jitter [6, 7]. The impact of jitter in gaze tracking systems was also investigated for 2D screens [22, 29, 36] and a prototype see-thorough HMD [17]. The results showed that the jitter significantly decreases the accuracy and precision of the participants. In our current work, we thus also hypothesize that we will get similar results for eye-gaze tracking in state-of-the-art VR HMDs, i.e., **H1: the user's motor performance and their experience decrease with increasing level of jitter**.

Moreover, because of the effect of the stereo-deficiencies in VR systems [2–4], we also know that 3D pointing performance of the users is not the same as in the real world [5]. However, we do not know how the jitter in gaze tracking impacts pointing at different visual depths. In this work, we speculate that **H2: the users' selection time, error rate, and throughput change for targets at different visual depths, even when targets take up the same visual angle**, i.e., as seen be the user.

4 USER STUDY

4.1 Participants

We recruited twelve participants (5 female, 7 male) aged between 18 and 28 years (mean = 20.8, SD = 2.5). Participation in our experiment was voluntary and no compensation was offered. All participants were from the local university. Ten participants had normal and two corrected-to-normal vision. None of them reported color blindness or other visual impairments.

4.2 Apparatus

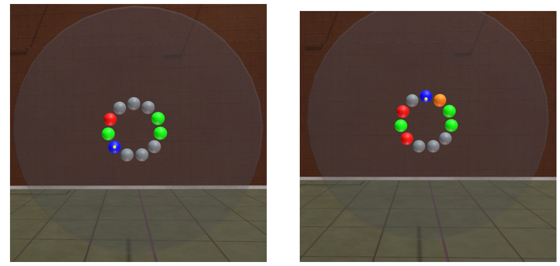
We conducted the experiment on an 11th Gen Intel(R) Core(TM) i7-11700F core 2.5 GHz, 32 GB RAM desktop PC with an NVIDIA GeForce GTX 3070 graphics card. We used an HTC VIVE Pro Eye headset with embedded Tobii eye tracking. For the virtual environment, we used Unity version 2020.3.21f1.

4.3 Procedure

After the participants had entered the room, we asked them to sign the consent form and fill the demographic questionnaire. We then explained the study to the participants and demonstrated how the setup works. The participants experimented with the VR system until they were comfortable with the gaze tracking system and target selection. Before starting the experiment, we calibrated the eye trackers for each individual using the built-in Vive Pro Eye Setup.

In the virtual environment, subjects were placed in an empty room with pictorial depth cues as shown in Fig. 1. To assess user performance with 3D pointing, we used an ISO 9241-411 task [20] with 11 targets distributed at equal distances in a circular arrangement, as in previous studies [31, 45].

The eleven (11) spheres were gray at the beginning of each round of trials, except for the target sphere. We indicated the target sphere by changing its color to orange. We recorded the position of the gaze and the looking direction from the Vive Pro system. With these two data, we calculated where the user is looking in 3D space. Based on this, we placed a small sphere, called cursor from now on, where the participants is looking. When the participant looked at the targets, we compared the distance between the target and cursor. If the cursor overlapped any sphere, we changed the color of that sphere to blue, i.e., highlighted it [41], to indicate the intersection. Participants then used the space bar on the keyboard to select the targets [45]. If the participant selected the correct target while the cursor overlapped



(a) (b)

Figure 1: Virtual environment used for the experiment. (a) Gaze cursor is over the target. (b) Gaze cursor slightly overlapping a non-target.

with the sphere, we changed the target's color to green and recorded a "hit" (Fig. 1(a)). However, if the cursor was outside of the target upon selection, we colored the target red, recorded a "miss", and played an error sound (Fig. 1(b). We asked participants to select targets as fast and as precisely as possible. Regardless if the user "hit" or "missed" the target, the next target appeared directly across the previous target on the other side of the circle of spheres. The target kept changing in the same pattern, i.e., alternating across the circle, until the user has made 11 selections. The software randomly chose a first target for each round. Similarly, the subjects experienced either a clockwise or counter-clockwise change in the sequence of targets, again selected randomly.

We applied four different levels of rotational jitter on all three rotation axes of the gaze direction. For the first level of jitter, we did not add any artificial jitter to the gaze direction, as the "no jitter" condition. For the second level of rotational jitter, we applied $\pm 0.5°$, for the third, $\pm 1°$, and for the fourth, we added $\pm 1.5°$ jitter to the gaze direction data received from the tracking system.

For jitter, we used the Marsaglia Polar Method [28] to generate White Gaussian Noise (WGN). We did not discard or cut off the values generated by this method. We simply multiplied the WGN result with 0.5 for the second condition, 1 for the third condition and 1.5 for the fourth one. Then, we added these values to all three angles of the gaze direction. For simplicity, we use only the coefficients for reporting the **Jitter Level**, i.e., 4_{JL} = **No jitter, $\pm$0.5°, $\pm$1°, and $\pm$1.5°**.

We also used two different **Depth Distances** (2_{DD} = **1 meter and 2 meters**). We chose these depth distances based on previous work. Further, we wanted to avoid targets within arm's length, i.e., closer than 70 cm [3], as participants might want to select them with arm movements. Also, we wanted targets to be closer than 2.25 meters, to be able to observe the impact of the jitter more reliably [6].

For angular target distance, i.e., the angular diameter of the "circle of targets", we used three **Angular Target Distances** (3_{TD} = **6°, 9°, and 12°**), and for each depth distance, we converted the angular measures to the Euclidean target sizes and distances for Unity. We used three different **Angular Target Sizes** (3_{TS} = **2°, 2.5°, and 3°**).

The use of angular rather than Euclidean measures keeps the target distance and size the same at different depth distances. Regardless if the targets appeared at 1m or 2m depth, using angular measures means that the targets will appear exactly the same size in both situations. The angular target distance and width conversions are shown in Table 1 and Table 2.

After the end of the tasks, we asked the participants to fill another questionnaire prompting them about their insights around jitter in gaze tracking.

Table 1: Angular to Euclidean Target Distance

Depth	Angular	Euclidean (cm)
1m	6°	10.4
	9°	15.7
	12°	21.0
2m	6°	20.9
	9°	31.4
	12°	42.0

Table 2: Angular to Euclidean Target Size

Depth	Angular	Euclidean (cm)
1m	2°	3.4
	2.5°	4.3
	3°	5.2
2m	2°	6.9
	2.5°	8.7
	3°	10.4

4.4 Experimental Design

We used a two-factor within-subjects design with four **jitter levels** (4_{JL} = no jitter, 0.5°, 1° and 1.5° jitter) and two **depth distances** (2_{DD} = 1 meter and 2 meter) comprising a $4_{JL} \times 2_{DD} \times$ design.

We counterbalanced jitter level conditions across subjects and randomized the depth distance to avoid learning effects. We collected movement time (s), error rate (%), accuracy SD_x, and effective throughput ($bits/s$) data as dependent variables to analyze user performance.

We also varied the index of difficulty (9_{ID}), by using three angular target sizes (3_{TS} = 2°, 2.5°, and 3°) and two angular target distances (3_{TD} = 6°, 9°, and 12°), which yields 6 unique ID between 1.5 and 2.9. Each subject performed (4_{JL} x 2_{DD} x 9_{ID} x 11 repetitions) = 792 trials. In total, we collected 14256 data points.

5 Results

To assess the effect of the factors on user performance we used Repeated Measures (RM) ANOVA in SPSS 24.0. As in previous work, we used Skewness (S) and Kurtosis (K) to determine the normality of data. We considered that the data is normally distributed if Skewness and Kurtosis are between ±1.5 [18, 27]. The data normality analysis showed that the error rate ($S = 0.008, K = -0.73$) and throughput ($S = 0.32, K = 0.028$) had a normal distribution. On the other hand, time ($S = 0.37, K = -0.03$) and SD_x ($S = 0.8, K = 1.3$) variables were only normally distributed after log-transformation. We used the Bonferroni method for post-hoc analysis. For brevity, we only focus on describing significant results here. The results are shown in Table 3 and Fig. 2.

Table 3: One-Way RM ANOVA results of Within Subjects Design

		DEPENDENT VARIABLES			
		Time	Error rate	Throughput	SD_x
INDEPENDENT VARIABLES	Jitter range	F(3, 33)=0.337 p=0.799, η^2= 0.03	F(3, 33)=94.431, p<0.001, η^2 =0.896	F(3, 33)=2.339, p=0.091, η^2= 0.175	F(2, 33)=2.764, p=0.057, η^2=0.201
	Depth	F(1, 11)=54.431, p<0.001, η^2= 0.896	F(1, 11)=45.558, p<0.05, η^2=0.798	F(1, 11)=0.261, p=0.619, η^2= 0.023	F(1, 11)=45.659, p<0.001, η^2=0.806
	ID	F(8, 88)=12.59, p<0.001, η^2= 0.534	F(8, 88)=4.531, p<0.001, η^2= 0.292	F(8, 88)=5.918, p<0.001, η^2=0.35	F(8, 88)=1.716, p=0.106, η^2=0.135

Time We did not observe a significant difference in selection time between jitter levels, see Table 3. However, based on Table 3 and Fig. 2, participants were faster with targets closer to them.

Error rate The error rate results in Table 3 and Fig. 2 show that participant made more errors with an increased amount of jitter. Moreover, the results in Table 3 and Fig. 2 illustrate that the error rate of the participants increased for distant targets.

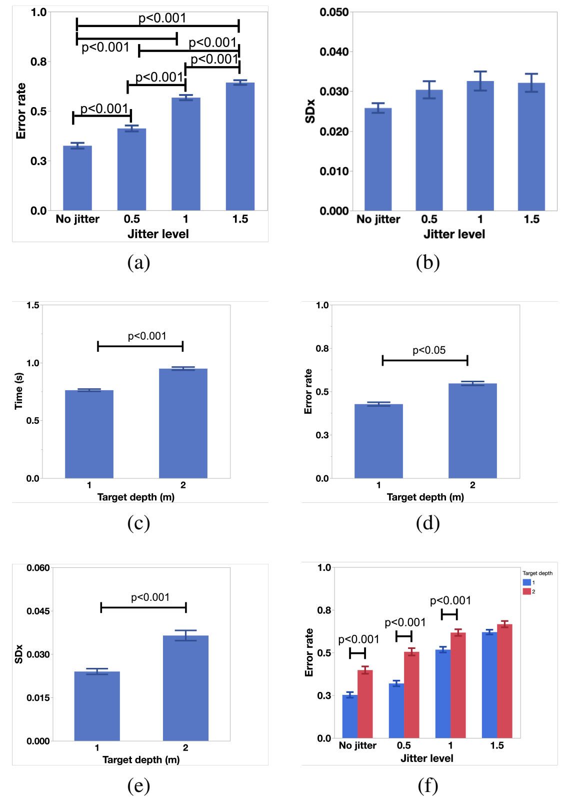


Figure 2: One way RM ANOVA results for (a) error rate and (b) SD_x for jitter level. Results varying across target depth for (c) time, (d) error rate, and (e) SD_x. (f) Interaction between jitter level and target depth results.

SD_x The accuracy results shown in Table 3 illustrate a marginally significant result for jitter levels. On the other hand, based on Table 3 and Fig. 2 we found that user accuracy significantly decreases with distant targets.

Throughput According to the results shown in Table 3, we did not observe a significant difference between jitter levels for throughput, nor for different target depths.

Interactions We found only one significant interaction, which was between jitter level and target depth for the error rate dependent variable, $F(3,33) = 3.507, p < 0.05, \eta^2 = 0.242$. According to the results shown in Fig. 2, participants' error rate did not change for ±1.5° of jitter between 1 and 2 meter.

Questionnaire After the experiment, we asked a series of questions to collect insights from the participants for the experimental conditions. 11 out of 12 participants preferred no jitter in gaze pointing. Overall, they commented: "*it makes it much harder to focus on the cursor*," "*I feel like I made more mistakes*," "*my eyes hurt*," "*doesn't feel comfortable*," and "*it was challenging and interesting*." The single participant who preferred jitter commented: "*jitter made it easy for my brain to focus on a certain point, since the jittering was provided in different levels, the faster the pace of jittering. It was difficult to focus on a certain point, where I had made lots of mistakes, however, it helped my brain stay focused on each point it was trying to track*." We also asked the participants if it was easy to select targets with jitter. According to our 7-point Likert scale results (1-I totally disagree, 7-I totally agree), participants did not agree that it was easy to choose the targets with jitter (mean = 2.5, median = 2, standard deviation = 1.08). Similarly, we asked participants if they prefer to select targets with jitter. The results showed that they do not prefer to select targets with jitter (mean = 2.25, median = 2, standard deviation = 1.05). When we asked about fatigue, the participants did not feel physically (mean = 3.9, median = 4.5, standard deviation = 1.85) nor mentally (mean = 2.7, median = 3, standard deviation = 1.15) fatigued after the experiment.

6 Discussion

In this paper, we investigated how three different levels of jitter and two target distances affect on gaze-based pointing performance and the user experience.

Our results match the findings of previous work on rotational jitter: user performance significantly decreases when there is jitter in the system [6–8]. Extending previous findings, the results in this work reveal that user error rate increases with a higher level of rotational jitter on gaze. The participants' error rate also already increased with $\pm 0.5°$ of jitter.

Previous work had shown that user performance can decrease with $\pm 1°$ rotational jitter on input devices [7]. Our current study indicates that participants' error rate already decreases with (only) $\pm 0.5°$ rotational gaze jitter. We believe that this result is also due to the fact that we conducted our study with angular measures in VR systems. Since we converted Euclidean distances to angular measures and applied WGN jitter to these angular measurements, we are able to correlate the amount of jitter to the target sizes and distances. This also confirms our hypothesis that **the user pointing performance and their experience decreases with increased amount of jitter**. The user experience results also support our findings on the effect of jitter in gaze tracking devices. To highlight these findings, we also included one of the participants' comments in the title of this manuscript.

Since we used angular sizes, we also increased the Euclidean target sizes and distances at farther distances. This essentially makes all targets appear at the same size, regardless of their distance. Thus, one could expect the same user performance for targets at different distances. However, previous work had identified that VR headsets suffer from various stereo deficiencies, such as the vergence and accommodation conflict, which has detrimental effects on user performance [2–4]. Apparently, this effect is also visible for gaze tracking, as our results show that both the participants' error rate and accuracy decrease with farther targets, which *partially* supports our hypothesis that **H2: the users' pointing time, error rate, and throughput change with target distance even though all targets subtend the same visual angle.** Even though we increased the size of the targets for farther targets, the user performance was negatively affected, likely due to the stereo deficiencies of the VR headsets or due to the gaze tracker working differently for different depth distances.

This negative impact was also observed in the interaction between depth distance and jitter level: the error rate of the participants was significantly higher for further targets. However, we did not observe a significant difference for $1.5°$ jitter at 2 meters. We hypothesize that at that distance, the amount of jitter was so high for the participants that they stopped attempting to select the targets precisely. Even though previous work had already investigated the effect of jitter in gaze tracking for 2D targets on 2D monitors, e.g., [22, 29, 36], our motivation was to extend this body of work to investigate the performance of 3D pointing with gaze tracking.

In this experiment, we used an HTC Vive Pro Eye HMD, with two V2 lighthouses. We deliberately chose this VR setup because it has a relatively low level of noise for head tracking. Previous work had also used only two V2 lighthouses [6, 8]. However, according to the manufacturer's specification, the gaze tracker built into the HTC Vive Pro Eye already exhibits 0.5-1.1° jitter. Our system added an appropriate range of additional jitter on top of this. We speculate that the fact that the base level of jitter is already substantial may be the reason why we did not find significant results for time and throughput. In essence, we believe that the level of system jitter was already too large to observe significant differences for additional jitter.

Another potential reason why we did not observe significant differences for time and throughput might be the factor levels used in our study. In Graupner et al. [17]'s work, the authors used 1° and 2° of jitter in gaze tracking and observed a significant effect on task execution time. Yet, in our study the highest level of jitter was 1.5°, which is lower than 2°. When looking at the results of previous work on rotational jitter, we can see that the task execution time seems to significantly increase with 2°, e.g., Batmaz et al. [7, 8]. Since we did not examine the effect of 2° jitter on user performance in this study, this may explain why we did not find a significant affect for execution time.

Another potential limitation of this work is the relatively low number of participants. Still, according to statistical effect size calculations, the minimum effect size we observed in this work is $\eta^2 = 0.201$, i.e., a large effect, commonly defined through a criterion of $\eta^2 > 0.14$. These large effect sizes are evidence that our research findings are robust and have practical significance.

We also acknowledge that previous studies on gaze tracking systems revealed that the user accuracy and precision varies with where the gaze is looking within the field of view [9, 15, 35]. This also means that error rate results here might vary with where the participant is looking, i.e., the error rate might be different for different target locations. Thus, beyond varying target size and target distance, there is a further need to study the impact of jitter for different target positions in different depths in the users' field of view.

We hope that our results help engineers, developers and practitioners who want to use gaze tracking as an input device for VR or AR systems, such as [33].

7 Conclusion

In this work, we studied the effects of rotational jitter on gaze-based pointing with four different jitter levels. Our results indicate that the error rate for pointing already significantly increases even with $\pm 0.5°$ added jitter. Based on our outcomes, we also suggest practitioners evaluate their VR gaze tracking devices and report user performance based on angular measures before releasing to the market.

We also observed that the user pointing performance with gaze tracking significantly decreases for more distant targets. This topic requires future research as there are multiple explanations, including the effect of stereo display deficiencies.

We believe that the our findings here are also useful for the design of novel gaze tracking hardware and inform the development of future, improved 3D gaze tracking systems for VR and AR applications.

References

[1] A. Bahill, M. R. Clark, and L. Stark. The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, 24(3):191 – 204, 1975. doi: 10.1016/0025-5564(75)90075-9

[2] M. D. Barrera Machuca and W. Stuerzlinger. Do stereo display deficiencies affect 3D pointing? In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, p. LBW126. ACM, 2018.

[3] M. D. Barrera Machuca and W. Stuerzlinger. The effect of stereo display deficiencies on virtual hand pointing,. In *2019 CHI Conference on Human Factors in Computing Systems*, p. 14 Pages. ACM, 2019.

[4] A. U. Batmaz, M. D. Barrera Machuca, D. M. Pham, and W. Stuerzlinger. Do head-mounted display stereo deficiencies affect 3D pointing tasks in AR and VR? In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, March 2019.

[5] A. U. Batmaz, A. K. Mutasim, M. Malekmakan, E. Sadr, and W. Stuerzlinger. Touch the wall: Comparison of virtual and augmented reality with conventional 2d screen eye-hand coordination training systems. In

Conference on Virtual Reality and 3D User Interfaces, VR '20, pp. 184–193. IEEE, Mar 2020. doi: 10.1109/VR46266.2020.1581205539914

[6] A. U. Batmaz, M. R. Seraji, J. Kneifel, and W. Stuerzlinger. No jitter please: Effects of rotational and positional jitter on 3D mid-air interaction. In *Future Technologies Conference*, vol. AISC 1289. Springer, Nov 2020. doi: 10.1007/978-3-030-63089-8_52

[7] A. U. Batmaz and W. Stuerzlinger. The effect of rotational jitter on 3D pointing tasks. Extended Abstract, May 2019. doi: 10.1145/3290607.3312752

[8] A. U. Batmaz and W. Stuerzlinger. Effects of 3D rotational jitter and selection methods on 3D pointing tasks. In *Workshop on Novel Input Devices and Interaction Techniques (NIDIT) at (IEEE) (VR) 2019)*, March 2019.

[9] J. Blattgerste, P. Renner, and T. Pfeiffer. 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*, pp. 1–9, 2018.

[10] D. Bowman, C. Wingrave, J. Campbell, and V. Q. Ly. Using pinch gloves (tm) for both natural and abstract interaction techniques in virtual environments, 2001.

[11] G. Casiez, N. Roussel, and D. Vogel. 1 € filter: A simple speed-based low-pass filter for noisy input in interactive systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, p. 2527–2530. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2207676.2208639

[12] Y. Cha and R. Myung. Extended fitts' law for 3d pointing tasks using 3d target arrangements. *International Journal of Industrial Ergonomics*, 43(4):350 – 355, 2013. doi: 10.1016/j.ergon.2013.05.005

[13] L. D. Clark, A. B. Bhagat, and S. L. Riggs. Extending fitts' law in three-dimensional virtual environments with current low-cost virtual reality technology. *International Journal of Human-Computer Studies*, 139:102413, 2020. doi: 10.1016/j.ijhcs.2020.102413

[14] V. Clay, P. König, and S. Koenig. Eye tracking in virtual reality. *Journal of Eye Movement Research*, 12, 04 2019. doi: 10.16910/jemr.12.1.3

[15] A. M. Feit, S. Williams, A. Toledo, A. Paradiso, H. Kulkarni, S. Kane, and M. R. Morris. Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design. In *Proceedings of the 2017 Chi conference on human factors in computing systems*, pp. 1118–1130, 2017.

[16] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381, 1954.

[17] S.-T. Graupner, M. Heubner, S. Pannasch, and B. M. Velichkovsky. Evaluating requirements for gaze-based interaction in a see-through head mounted display. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, pp. 91–94, 2008.

[18] J. F. Hair Jr, W. C. Black, B. J. Babin, and R. E. Anderson. Multivariate data analysis, 2014.

[19] H. Hefter, V. Hömberg, K. Reiners, and H.-J. Freund. Stability of frequency during long-term recordings of hand tremor. *Electroencephalography and clinical neurophysiology*, 67(5):439–446, 1987.

[20] ISO 9241-411:2012. Ergonomics of human-system interaction – part 411: Evaluation methods for the design of physical input devices. ISO, May 2012.

[21] R. Kopper, D. A. Bowman, M. G. Silva, and R. P. McMahan. A human motor behavior model for distal pointing tasks. *International journal of human-computer studies*, 68(10):603–615, 2010.

[22] K. Krejtz, A. Duchowski, H. Zhou, S. Jörg, and A. Niedzielska. Perceptual evaluation of synthetic gaze jitter. *Computer Animation and Virtual Worlds*, 29(6):e1745, 2018.

[23] M. Kumar, A. Paepcke, and T. Winograd. Eyepoint: Practical pointing and selection using gaze and keyboard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, p. 421–430. Association for Computing Machinery, New York, NY, USA, 2007. doi: 10.1145/1240624.1240692

[24] I. S. MacKenzie. Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction*, 7(1):91–139, 1992.

[25] I. S. MacKenzie and P. Isokoski. Fitts' throughput and the speed-accuracy tradeoff. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, p. 1633–1636. Association for Computing Machinery, New York, NY, USA, 2008. doi: 10.1145/1357054.1357308

[26] I. S. MacKenzie and A. Oniszczak. A comparison of three selection techniques for touchpads. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 336–343. ACM Press/Addison-Wesley Publishing Co., 1998.

[27] P. Mallery and D. George. SPSS for windows step by step: a simple guide and reference. *Allyn, Bacon, Boston,*, 2003.

[28] G. Marsaglia and T. A. Bray. A convenient method for generating normal variables. *SIAM review*, 6(3):260–264, 1964.

[29] D. Miniotas and O. Špakov. An algorithm to counteract eye jitter in gaze-controlled interfaces. *Information Technology and Control*, 30(1), 2004.

[30] A. Murata and H. Iwase. Extending fitts' law to a three-dimensional pointing task. *Human Movement Science*, 20(6):791 – 805, 2001. doi: 10.1016/S0167-9457(01)00058-6

[31] A. K. Mutasim, A. U. Batmaz, and W. Stuerzlinger. Pinch, click, or dwell: Comparing different selection techniques for eye-gaze-based pointing in virtual reality. In *ACM Symposium on Eye Tracking Research and Applications*, pp. 1–7, 2021.

[32] A. Pavlovych and W. Stuerzlinger. The tradeoff between spatial jitter and latency in pointing tasks. In *Proceedings of the 1st ACM SIGCHI symposium on Engineering interactive computing systems*, pp. 187–196. ACM, 2009.

[33] J. San Agustin, H. Skovsgaard, J. P. Hansen, and D. W. Hansen. Low-cost gaze interaction: ready to deliver the promises. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, pp. 4453–4458. 2009.

[34] I. Schuetz, T. S. Murdison, K. J. MacKenzie, and M. Zannoli. An explanation of fitts' law-like performance in gaze-based selection tasks using a psychophysics approach. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300765

[35] I. Schuetz, T. S. Murdison, and M. Zannoli. A psychophysics-inspired model of gaze selection performance. In *ACM Symposium on Eye Tracking Research and Applications*, pp. 1–5, 2020.

[36] A. Sesin, M. Ayala, M. Cabrerizo, and M. Adjouadi. Jitter reduction in eye gaze tracking system and conception of a metric for performance evaluation. *WSEAS Transactions on Computers*, (5):1268–1273, 2004.

[37] L. Sidenmark and H. Gellersen. Eye, head and torso coordination during gaze shifts in virtual reality. *ACM Trans. Comput.-Hum. Interact.*, 27(1), Dec. 2019. doi: 10.1145/3361218

[38] R. N. Stiles. Mechanical and neural feedback factors in postural hand tremor of normal subjects. *Journal of Neurophysiology*, 44(1):40–59, 1980.

[39] M. F. Stoelen and D. L. Akin. Assessment of fitts' law for quantifying combined rotational and translational movements. *Human factors*, 52(1):63–77, 2010.

[40] R. J. Teather, A. Pavlovych, W. Stuerzlinger, and I. S. MacKenzie. Effects of tracking technology, latency, and spatial jitter on object movement. In *3D User Interfaces, 2009. 3DUI 2009. IEEE Symposium on*, pp. 43–50. IEEE, 2009.

[41] R. J. Teather and W. Stuerzlinger. Visual aids in 3D point selection experiments. In *Proceedings of the 2nd ACM symposium on Spatial user interaction*, pp. 127–136. ACM, 2014.

[42] S. Vetter, J. Bützler, N. Jochems, and C. M. Schlick. Fitts' law in bivariate pointing on large touch screens: Age-differentiated analysis of motion angle effects on movement times and error rates. In *International Conference on Universal Access in Human-Computer Interaction*, pp. 620–628. Springer, 2011.

[43] B. Wkasikowska. The application of eye tracking in business. 2014.

[44] D. Wolf, J. Gugenheimer, M. Combosch, and E. Rukzio. Understanding the heisenberg effect of spatial interaction: A selection induced error for spatially tracked input devices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–10, 2020.

[45] X. Zhang and I. S. MacKenzie. Evaluating eye tracking with iso 9241-part 9. In *International Conference on Human-Computer Interaction*, pp. 779–788. Springer, 2007.