Effects of 3D Rotational Jitter and Selection Methods on 3D Pointing Tasks

Anil Ufuk Batmaz* Wolfgang Stuerzlinger†
School of Interactive Arts + Technology (SIAT)
Simon Fraser University, Vancouver, Canada

Abstract
3D pointing is an integral part of Virtual Reality interaction. Typical pointing devices rely on 3D trackers and are thus subject to fluctuations in the reported pose, i.e., jitter. In this work, we explored how different levels of rotational jitter affect pointing performance and if different selection methods can mitigate the effects of jitter. Towards this, we designed a Fitts’ Law experiment with three selection methods. In the first method, subjects used a single controller to position and select the object. In the second method, subjects used the controller in their dominant hand to point at objects and the trigger button of a second controller, held in their non-dominant hand, to select objects. Finally, subjects used the controller in their dominant hand to point to the objects and pressed the space bar on a keyboard to select the object in the third condition. During the pointing task we added five different levels of jitter: no jitter, ±0.5°, ±1°, and ±2° uniform noise, as well as White Gaussian noise with 1° standard deviation. Results showed that the Gaussian noise and ±2° of jitters significantly reduced the throughput of the participants. Moreover, subjects made fewer errors when they performed the experiment with two controllers. Our results inform the design of 3D user interfaces, input devices and interaction techniques.

Index Terms: Human-centered computing—Virtual Reality; Human-centered computing—Keyboards; Human-centered computing—Pointing devices

1 Introduction
Most current virtual reality (VR) systems use three major components to immerse the user in a virtual environment (VE): a head-mounted display (HMD) system, to show the view of the VE to the user, controllers/wands to allow user to interact with the VE, and a 3D tracking system to detect the pose of the headset and controllers. The performance of the 3D tracking system varies according to its design and is typically subject to some level of fluctuations, called jitter [18], regardless of the sensor technology used, such as infrared cameras, depth sensors, color cameras, and/or inertial measurement units.

The raw data collected from the 3D tracking system’s sensors is not suitable for end-user applications due to the noise in the signal. This noise can be seen as additional error signals that interfere with the real world pose data. To isolate and reconstruct the real world pose data, the raw data coming from the sensors is typically processed through various signal processing methods, either in hardware or software. Even after such signal processing, there is often residual noise, which can easily be observed when the tracker is idle on a stable surface. An example of such idle noise signals are shown in Figure 1(a). These plots show rotational jitter data for an immobilized HTC VIVE Pro controller, using two V2 (version 2) lighthouses as optical emitters.

Figure 1: Two exemplary rotational jitter signal samples collected while (a) a HTC Vive Pro controller was placed on a desk or (b) a subject was holding the device stable in mid-air. The average absolute jitter was 0.15°, 0.03° and 0.07° for the x, y and z-axes, respectively, as shown in (a), with a maximum observed deviation of 0.614° along the x-axis. In (b), the average absolute jitter was 0.21°, 0.24° and 0.21° for the x, y and z-axes, respectively, with a maximum observed deviation of 1.114° along the z-axis.

During interaction with the VE, the signal acquired from the 3D tracker in a VR controller is also affected by natural user behaviors, including hand tremor, breathing and the associated body sway, and task errors, such as fatigue. An example of a signal affected by such user behaviours is presented in Figure 1(b), where the subject was holding the controller stable in mid-air. All of these natural behaviours increase the controller jitter, which may require users to adapt their movements to accurately position a cursor during a 3D pointing task.

Furthermore, while a subject uses a controller, they often need to physically interact with the controller itself to issue an input command, such as pushing a button or pulling a trigger. The effect of such interactions on the 3D pose data are typically not filtered by the software and can create temporarily unstable outputs, e.g., instant rotations (or re-positioning) of the real and virtual controller [5]. In addition to jitter introduced by measurement errors and noise, controller/cursor position also temporary varies with these physical controller interactions.

The aim of this work is to explore how user performance is affected by different levels of rotational jitter and how different selection methods affect user performance during a pointing task. To investigate these topics in a 3D pointing task, we designed a Fitts’ law experiment with five levels of jitter and three different selection methods.

2 Related Work
Here we first review previous work on 3D pointing, then Fitts’ law, other work on the effect of jitter, and selection methods.

*e-mail: abatmaz@sfu.ca
†e-mail: wa@sfu.ca
2.1 3D Pointing in Virtual Environments

While selection with a virtual hand metaphor is easy in VR, it is more challenging to select targets that are further away [10]. Ray casting is the preferred choice for the selection of distant objects in many scenarios. Still, as it requires accurate pointing, ray casting does not perform so well for small and/or distant targets [16], similar to how a laser pointer behaves in the real world.

One proposed solution is to use a conic volume to facilitate selection of distant targets [11], but this method does not support high-precision selection within distant dense object groups. Thus, there is still a need to explore how rotational jitter affects the user performance and to understand the limitations imposed by the ray casting paradigm.

2.2 Fitts’ Law


$$\text{Movement Time} = a + b \cdot \log_2 \left( \frac{A}{W} + 1 \right) = a + b \cdot ID$$ \hspace{1cm} (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 $W$ is the target width. The logarithmic term in the equation 1 represents the task difficulty, and is called the index of difficulty, $ID$.

2.3 The Effect of 3D Tracking Noise on Pointing

Since pointing techniques based on a 6 DoF controller mainly depend on the rotational data to determine the pointing direction, they are more susceptible to tracking noise compared to other pointing techniques, such as a virtual hand [9]. For instance, hand tremors are well-known to cause oscillations between 4 and 8 Hz [20], which naturally limits the accuracy that can be achieved in any human pointing task.

Previous studies on the effect of positional jitter in a pointing task showed that a condition with 0.3 mm average positional jitter does not show a significant performance difference to a condition without jitter [19]. Yet, larger levels of positional jitter can significantly reduce user performance for smaller targets [15].

To our knowledge, we are not aware of any systematic investigation of different amounts of 3D rotational jitter on user pointing performance.

2.4 Selection methods

3D pointing has been explored by various studies that investigated different devices, see also a recent survey [1]. Researchers have also evaluated different selection methods, e.g., [13]. In more recent work, Brown and Stuerzlinger compared two different interaction styles and two different selection styles for mid-air pointing [6], but found no significant difference between them.

During the selection process, if the user has to physically interact with the controllers to activate the selection command, such as pressing a button, turning a wheel, pushing a grip, or pulling a trigger, this can easily change the pose of the controller and cause an error called the “Heisenberg effect” of spatial interaction [5], which occurs when the action of pressing the button causes the cursor to “slip off” the target object. Ray casting is especially prone to this effect, as any small rotation is magnified along the selection ray. One of the easiest ways to avoid the Heisenberg effect is to use different hands for different actions. For instance, while the users control the pose of the controller with their dominant hand, they can activate selection with their non-dominant hand. This approach is referred to as asymmetric bi-manual interaction and has been used in recent VR research, e.g., [17].

3 Motivation & Hypotheses

Previous work showed that spatial jitter can affect user performance negatively [15, 18, 19]. Based on these results, we hypothesized that subjects’ performance also decreases with increased levels of rotational jitter, especially for selection tasks with a larger index of difficulty, i.e., larger distances to the target or smaller targets. Thus, we expect that participants will exhibit increased error rates and lower throughput with increased rotational jitter. Additionally, we hypothesized that it is possible to observe the Heisenberg effect on spatial interaction during ray-based 3D object selection and that user performance does not change with varying levels of jitter, since the Heisenberg effect can also be considered to be a source of noise.

4 USER STUDY

4.1 Participants

Nine subjects (5 female), average age 22.2 ±2.9 years, participated in our experiment. All subjects were right-handed and they used their dominant hand to execute the task. All participants measured normal when tested for stereo viewing capability. The headset was adjusted to match the interpupillary distance of each individual. All participants were familiar with 3D environments from video games and 3D CAD systems; 77.7% played and used 3D CAD systems between 0-5 hours/week, and 22.3% 5-10 hours. 5 of them reported that their dominant eye is their left one.

4.2 Apparatus

We used a PC with an Intel (R) Core (TM) i7-5890 CPU with 16 GB RAM and a NVIDIA GeForce RTX 2080 graphics card. For the VR Headset, we used a HTC Vive Pro with two V2 Lighthouses. Subjects used two HTC Vive Pro controllers and a Logitech© keyboard as input devices.

4.3 Procedure

Participants were seated in a comfortable chair in front of the PC. They were first asked to fill a pre-questionnaire on their demographics. After the questionnaire, the experimenter explained and demonstrated the task to each participant. Before starting the experiments, subjects were allowed to perform practice trials for a few minutes to get used to the VR system and the VE. At the end of the experiment, subjects filled a post-questionnaire to choose their preferred selection method.

To assess 3D pointing performance, we used a variation of ISO 9241-411 task [8]. In our version of the task, two targets are placed along a single lateral axis. Target were visible alternatively, i.e., when one was selected, it disappeared and the other one appeared, and participants were asked to select these targets as fast and as precise as possible.

In the VE, subjects were placed in an empty room with adequate depth cues (Figure 2). A grey sphere appeared 50 cm in front of the center of the participant’s eye positions, i.e., at eye level. For selection, we provided a 30 cm long ray originating at the pivot point of the controller. A yellow sphere was placed at the end of the ray and acted as a cursor. We did not alter the ray length to avoid the potential confound of different distances to the targets, which could affect our experimental outcomes. This helped us to limit any potential issues with different control-display gains and also reduces ambiguities in terms of visual depth and visibility – after all the current view might not enable the user to see if the ray hits at the side of the target that is facing away from the user. Furthermore, in “standard” raycasting, the ray intersects with the surface of objects, which turns the 3D pointing task into a 2D one. Based on these considerations, we asked subjects to use the small yellow sphere as a cursor and to place it into the center of the gray target.

During the experiment, we added artificial rotational jitter at the starting point of the ray. Thus, both the yellow sphere and the ray were jittered together relative to the pivot point of the controller.
We applied five different levels of jitter to all three rotational axes of the controller as shown in Figure 3. In the no jitter condition, there was no additional jitter. In $\pm 0.5^\circ$, $\pm 1^\circ$, $\pm 2^\circ$ levels of jitter, we applied a uniform distribution to the given range. In the $\pm 1^\circ$ jitter condition we added random deviations between $-1^\circ$ to $+1^\circ$ to all three Euler axes. In the White Gaussian Noise (WGN) condition, we used a standard normal distribution generator, called Marsaglia Polar Method [14], to generate random values with a mean of 0 and a standard deviation of $1^\circ$. We deliberately added the white Gaussian noise condition since it is used to model random processes in information theory. We did not flatten or discard any random values generated by this method.

Subjects used three different selection methods to interact with the targets. In the one-controller condition, subjects needed to position the cursor with the controller they were holding with their dominant hand into the target and then press the trigger button on the same controller to select it. In the two-controllers condition, subjects positioned the cursor with the controller held in their dominant hand and then pressed the trigger button on the second controller, which they were holding with their non-dominant hand, to select objects. In the keyboard condition, subjects selected objects by changing the cursor position with the HMD speakers and changing the color of the target to red for visual feedback. Afterwards, the next target appeared and subjects continued their task.

Subjects selected 11 consecutive targets in each individual trial. We used three different target sizes by changing the diameter of the target sphere and three different target distances by changing the distances between to targets for each trial. The sequence of target sizes and target distances was random. The three different selection methods were counterbalanced between subjects.

4.4 Experimental Design

The 9 participants performed 11 trials in 135 experimental conditions: three Selection Methods (SM: one-controller, two-controllers and keyboard), five different Jitter Ranges (JR: $0^\circ$, $\pm 0.5^\circ$, $\pm 1^\circ$, $\pm 2^\circ$, and WGN), three Target Distances (TD: 10, 20, and 30 cm) and three Target Sizes (TS: 1.5, 2.5, and 3.5 cm), which results in a $3 \times 5 \times 3 \times 3$ within-subject design. Subjects movement time (ms), error rate (%) and effective throughput (bit/s) were measured as dependent variables. Based on the different values for TD and TS in equation 1, we evaluated 9 unique ID3s between 1.94 and 4.39. Thus, each subject performed 1485 trials ($3 \times 5 \times 3 \times 1 \times 11$ trials). Overall, we collected 13365 data points for each dependent variable.

5 RESULTS

The data was analyzed by repeated measures ANOVA with SPSS 24. Data was normally distributed. Before the ANOVA, we removed data for double clicks (82 instances, 0.62% of the data). The one-way ANOVA results are shown in Table 1, with **** for $p < 0.0001$, *** for $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, and n.s. for non-significant results. We used the Sidak method for post-hoc analyses.

<table>
<thead>
<tr>
<th>Table 1: RM ANOVA results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection Method</td>
</tr>
<tr>
<td>Movement time</td>
</tr>
<tr>
<td>Error rate</td>
</tr>
<tr>
<td>Effective throughput</td>
</tr>
</tbody>
</table>

For the Jitter Ranges, Mauchly’s sphericity test was violated for the Selection Methods, Mauchly’s sphericity test held for time ($\chi^2(2) = 5.79$, n.s), throughput ($\chi^2(2) = 1.01$, n.s) and error ($\chi^2(2) = 4.22$, n.s). The index of difficulty, Mauchly’s sphericity test was violated for time ($\chi^2(35) = 72.96$, $p < 0.001$) but not for throughput ($\chi^2(35) = 51.813$, n.s) and error ($\chi^2(35) = 22.92$, n.s). We used Greenhouse-Geisser correction to address the sphericity violations, since $\varepsilon < 0.75$ held for all the dependent variables.

5.1 One-Way RM ANOVA Results

5.1.1 Time results

The results for time are shown in Table 1, and in Figure 4(a) for the selection methods and Figure 4(b) for jitter ranges. According to these results, there was no significant difference between selection methods, but subjects were significantly slower with $\pm 2^\circ$ jitter.

5.1.2 Error rate results

The error rate results are shown in Table 1, and in Figure 4(c) for the selection methods and Figure 4(d) for jitter ranges. According to these results, the error rate was lower when subjects used two controllers, compared to their results with a single controller. Further, the error rate started to significantly increase above the $\pm 1^\circ$ jitter range.

5.1.3 Throughput results

The throughput results are shown in Table 1, and in Figure 4(e) for the selection methods and Figure 4(f) for jitter ranges. According
to these results, the selection method did not affect user throughput, but performance of subjects significantly decreased in terms of throughput when $\pm 2^\circ$ jitter or WGN were artificially added.

Figure 4: Movement time results for (a) selection methods and (b) jitter ranges; error rate results for (c) selection methods and (d) jitter ranges; effective throughput results for (e) selection methods and (f) jitter ranges.

5.2 Two-way RM ANOVA Results

We also performed a two-way RM ANOVA to detect interactions, but report only significant interactions for the dependent variables. The only two significant interactions were between jitter range and index of difficulty for throughput $F(32,256) = 2.569, p < 0.05$ and error rate $F(32,256) = 8.09, p < 0.001$. In further analysis of the error rate data, we found a significant interaction between jitter range and target width $F(2,504, 20.03) = 21.06, p < 0.0001$ (the sphericity assumption was violated $\chi^2(35) = 95.49, p < 0.001$, and a Greenhouse-Geisser correction was applied $\epsilon < 0.75$), as shown in Figure 5(a). Similarly, in further analysis for throughput, we found a significant interaction between jitter range and target width $F(3.49,27.93) = 3.371, p < 0.05$. The sphericity assumption was violated $\chi^2(35) = 106.93, p < 0.00$, and a Greenhouse-Geisser correction applied $\epsilon < 0.75$, as shown in Figure 5(b).

The interaction between target distance and jitter range was not significant for throughput $F(3.703,29.624) = 0.889$, n.s. (the sphericity assumption was violated $\chi^2(35) = 82.76, p < 0.001$, and a Greenhouse-Geisser correction applied $\epsilon < 0.75$), nor for error rate $F(8,64) = 1.254$, n.s.,

According to results shown in the Figure 5, subjects throughput decreased with larger target widths and they made more errors with smaller targets.

6 DISCUSSION

In this work, we explored how different levels of jitter affect user performance and the effect of different selection methods. When we look at the results for time, we can see that subjects are slowing down with an increase in the range of jitter, which means that instabilities in the cursor can affect user performance negatively. Similarly, when we plot the ID against the movement time according to equation (1) for different amounts of jitter, the a and b values are notably different for $\pm 2^\circ$ ($R^2 = 0.234$ with a=-146 and b=465) of jitter. See Figure 6 for an illustration of the corresponding Fitts’ law models.

Figure 5: Two-way RM ANOVA results for jitter range and target width for (a) error rate, (b) throughput.

Figure 6: Fitts’ law model for movement time for five different jitter levels

The error rate results showed that subjects started to make significantly more errors beyond the $\pm 1^\circ$ jitter condition. Subjects also made more errors for smaller targets, which is not surprising. Furthermore, our results for rotational jitter also shows similarity with results from previous work on spatial jitter [15, 19].
Moreover, effective throughput results also confirm that subjects’ performance decreases for conditions with larger amounts of jitter. Similar to the results for error rate, we again observed that subjects throughput performance decreased especially for small targets, which we see as support for one of our hypotheses.

Even though we did not find significant differences between different selection methods for time, when we plot the ID vs. movement time plot according to equation (1), the a and b values appear to be potentially different for a single controller ($R^2 = 0.228$ with a=46.76 and b=357), see Figure 7.

In the experiment reported here, we chose the target size and target distance according to previous studies, e.g., [2, 3]. Further, we chose the jitter ranges according the samples we collected from rotational data for a static controller 1. With the highest jitter deviation ($+2\°=+\pi/90$ or $-2\°=-\pi/90$) on the controller, the cursor can be dislocated by up to 1.05 cm ($\pm \sin(\pi/90) \times 30 \text{ cm} = \pm 1.05 \text{ cm}$), which is larger than the target width. These results also indicate that designers must consider the potential effect of such jitter on the selection error rate for objects of such size. Thus, we can state that for high-precision tasks a standard 6 DoF controller such as the HTC Vive controller might not be sufficiently accurate.

In the experimental design, jitter levels were completely independent and jitter signals were randomly generated within the given limits, except for the white Gaussian noise condition. Further, when we look at the time and effective throughput results, we can see a decrease in both dependent variables. This means that, even though the user spent more time to position the controller and select the object in conditions with added jitter, this did not permit users to benefit from a time-accuracy tradeoff, which would have potentially increased the results for effective throughput. Given that we did not observe this, we believe that users are effectively unable to compensate for larger amounts of jitter.

Previous work by Batmaz et al. [4] showed that for 3D pointing performance, there is no major significant difference between state-of-the-art AR and VR headsets. They used the same target sizes and distances as the studies by Barrera and Stuerzlinger [2, 3]. According to their results, stereo deficiencies negatively affect interaction in both AR and VR headsets, so we can hypothesize that the results we found here might also generalize to an AR 3D pointing task. However, such speculations should be validated by further experiments.

In this experiment, we used HTC Vive Pro and its’ controllers with two V2 lighthouses. This setup includes one of the best tracking systems currently available on the market, especially considering its price. Even though the system works well enough to enable many applications, the data still contains some level of jitter caused by a combination of measurement errors, human errors, signal processing errors and other noise sources (1(a)). We deliberately chose this VR setup, because it has a relatively low level of noise. Yet, to our knowledge, there is no system that can measure the exact pose of an input device in mid-air for the same tracking volume with at least an order of magnitude less jitter relative to our setup. Also, the jitter conditions here do not refer to absolute levels of jitter. Thus, designers and engineers can apply our results to their devices by using a HTC Vive controller as a baseline, i.e. if the proposed controller and HTC Vive Pro controller have a difference of $\pm 2\°$ jitter, one can expect this to result in significantly lower 3D pointing performance.

Another outcome of our study are insights for selection methods. As expected, separating the device controlling the selection ray pose from the device used for selection, resulted in a lower error rate of the subjects. On the other hand, for each dependent variable there was no significant interaction between selection methods and other independent variables, including the jitter range. This result shows that the negative effects of jitter cannot be compensated by using both hands. Moreover, in the post-questionnaire, we asked subjects for their preferred selection method. A third of our participants selected the keyboard, another third chose the single controller, and the remaining third picked the two-controller condition as their preferred input method. In their explanation, they all said their preferred selection method was easier to use compared to other the methods. Thus, we can say that the selection method is clearly a matter of user preferences.

7 CONCLUSIONS AND FUTURE WORK

In the work presented here, we measured subjects’ performance under five different jitter levels and with three different selection methods. Our results showed that subjects’ 3D pointing performance significantly degrades in conditions with high jitter; they were slower, made more errors and their performance in terms of throughput decreased. Moreover, while using two controllers had a positive effect, as it decreased the error rate, it did not ameliorate user performance with different levels of jitter.

Since we chose a constant ray length, all our analysis are valid for ray-casting interaction in the space close to the user. The subjects’ performance might change with more distant targets and/or with different ray lengths. Thus, the effect of different ray lengths and their effect on user performance should also be investigated in the future.

Moreover, the target position was stable in this work. If a real-world target object was tracked by a 3D tracking system, the corresponding spatial jitter of the target object might also affect the virtual target position. Therefore, we can point out that the combination of 3D spatial target jitter with 3D controller rotational jitter should be explored as well.

Last but not least, other 3D tracking errors caused by mis-calibration, hardware problems, software errors, poorly designed interaction methods, mismatched ergonomics, or other potential sources of noise should also be considered in future work. We plan to explore some of these factors in the future.

REFERENCES


