Pinch, Click, or Dwell: Comparing Different Selection Techniques for Eye-Gaze-Based Pointing in Virtual Reality

Aunnoy K Mutasim
amutasim@sfu.ca
School of Interactive Arts &
Technology (SIAT)
Simon Fraser University
Vancouver, BC, Canada

Anil Ufuk Batmaz
abatmaz@sfu.ca
School of Interactive Arts &
Technology (SIAT)
Simon Fraser University
Vancouver, BC, Canada

Wolfgang Stuerzlinger
w.s@sfu.ca
School of Interactive Arts &
Technology (SIAT)
Simon Fraser University
Vancouver, BC, Canada

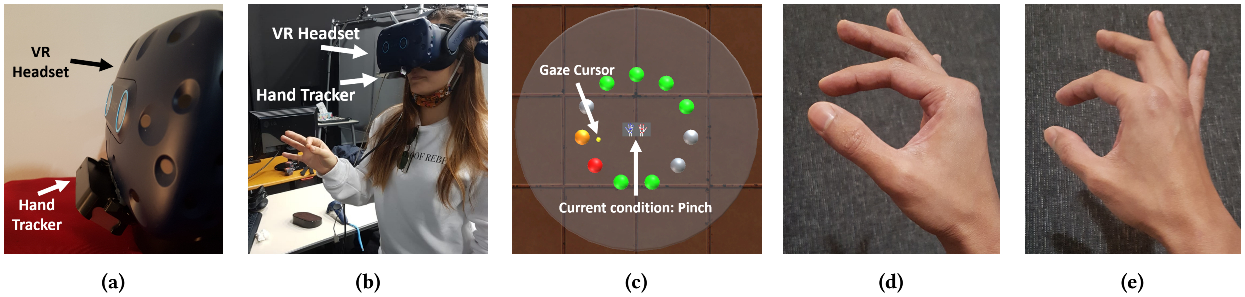

Figure 1: Experimental setup and apparatus. (a) A downward-facing Leap Motion attached to the bottom of the headset at an appropriate angle. (b) A participant performing the experiment. (c) The VR scene. In the pinch condition, subjects pinched with the palm either (d) facing horizontal towards their non-dominant side, (e) at approximately 45° with the floor, or somewhere in between. Figures (d) and (e) illustrate approximate views of the hand from the Leap Motion (cropped for illustration purposes).

ABSTRACT

While a pinch action is gaining popularity for selection of virtual objects in eye-gaze-based systems, it is still unknown how well this method performs compared to other popular alternatives, e.g., a button click or a dwell action. To determine pinch's performance in terms of execution time, error rate, and throughput, we implemented a Fitts' law task in Virtual Reality (VR) where the subjects pointed with their (eye-)gaze and selected / activated the targets by pinch, clicking a button, or dwell. Results revealed that although pinch was slower, made more errors, and had less throughput compared to button clicks, none of these differences were significant. Dwell exhibited the least errors but was significantly slower and achieved less throughput compared to the other conditions. Based on these findings, we conclude that the pinch gesture is a reasonable alternative to button clicks for eye-gaze-based VR systems.

CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**; *Pointing devices*; *HCI theory, concepts and models.*

KEYWORDS

Virtual Reality, Eye-Gaze Tracking, Fitts' Law, Throughput, Pinch, Gesture, Dwell, Click

ACM Reference Format:
Aunnoy K Mutasim, Anil Ufuk Batmaz, and Wolfgang Stuerzlinger. 2021. Pinch, Click, or Dwell: Comparing Different Selection Techniques for Eye-Gaze-Based Pointing in Virtual Reality. In *ETRA '21: 2021 Symposium on Eye Tracking Research and Applications (ETRA '21 Short Papers), May 25–27, 2021, Virtual Event, Germany*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3448018.3457998

1 INTRODUCTION

The addition of hand- and eye-trackers to Virtual Reality (VR) headsets has opened up alternative ways of interaction in VR. A hand tracker built into a VR headset [HTC 2019] or attached to it [Batmaz et al. 2020b] obviates the need to hold an extra device, such as a VR controller or keyboard [Pfeuffer et al. 2017]. This option then enables users to move more freely in the environment as well as increases their sense of presence and embodiment [Chiu et al. 2019].

Similarly, pointing with one's eye-gaze (referred to as *gaze* in the rest of this paper) has benefits [Jacob 1991]. Using gaze allows VR interaction designers to take advantage of faster pointing actions for selection tasks [Blattgerste et al. 2018], requiring less muscle movement and therefore energy [Sidenmark and Gellersen 2019a]. Gaze also provides an alternative way of interaction for patients with limited muscle control [Kumar et al. 2017]. Yet, an issue with gaze-only systems is the reliability of selecting / activating a target.

ETRA '21 Short Papers, May 25–27, 2021, Virtual Event, Germany
© 2021 Association for Computing Machinery.
ACM ISBN 978-1-4503-8345-5/21/05…$15.00
https://doi.org/10.1145/3448018.3457998

While gaze can move very fast (up to $900^\circ/s$ [Bahill et al. 1975]), it does not provide a direct way to indicate selection / activation of a target. The most commonly-used form of gaze-only selection / activation technique is *dwell*, i.e., fixating ones' gaze on the target for a certain *dwell time* [Hansen et al. 2018; Majaranta et al. 2009; Mott et al. 2017; Schuetz et al. 2019]. Beyond dwell, gaze has also been supplemented by button clicks, speech, gaze gestures, and other methods for the selection / activation step [Esteves et al. 2020; Pai et al. 2019; Piumsomboon et al. 2017]. Previous work [Choe et al. 2019; Esteves et al. 2020; Hassan et al. 2019; Rajanna and Hansen 2018; Zhang and MacKenzie 2007] identified that the button click typically outperforms all other selection techniques.

Hand gestures have also been proposed as a means to select / activate targets [Esteves et al. 2020; Jimenez and Schulze 2018; Pfeuffer et al. 2017]. With the goal of moving towards more natural forms of interaction [Mine 1995], researchers have experimented with different types of hand gestures to manipulate a virtual object [Canare et al. 2018; Ryu et al. 2019; Speicher et al. 2018]. As it is easy and comfortable to perform and reliable to recognize, the pinch gesture has frequently been proposed and used for selection and manipulation of both 2D and 3D virtual targets [Chatterjee et al. 2015; HoloLens 2019; Jimenez and Schulze 2018; Kosunen et al. 2013; Pfeuffer et al. 2017; Surale et al. 2019; Wilson 2006].

However, even with its increasing popularity, there is no study on *how well the pinch gesture performs for selection tasks compared to other popular alternatives for gaze-based systems, specifically, button click or dwell.* To answer this research question, we implemented a gaze-based ISO 9241-411 Fitts' law task [Bækgaard et al. 2019; ISO 9241-411:2012 2012; Pai et al. 2019] in VR. Our main contribution here is a rigorous performance comparison for pinch, button click, and dwell in terms of pointing execution time, error rate, and throughput. In the process, we also evaluated how effective the pinch gesture is for selecting / activating targets compared to the other two conditions.

2 LITERATURE REVIEW

2.1 Fitts' Law and Throughput

Fitts' law has been used to model the human movement time in human-computer interaction (HCI) studies. In our study, we used the Shannon formulation [MacKenzie 1992] of Fitts' law, as in Equation 1:

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

Here, A and W represent the target distance and size, and a and b are empirically derived via linear regression. The logarithmic term in Fitts' law is known as the index of difficulty ID and represents the task difficulty.

We use throughput, also known as the index of performance, to quantify the overall user performance. We use the measures proposed in ISO 9241-411:2012 [2012] to assess the participants' throughput performance.

$$Throughput = \left(\frac{Effective\ Index\ of\ Difficulty}{MovementTime}\right) = \left(\frac{ID_e}{MT}\right) \quad (2)$$

According to ISO 9241-411:2012 [2012], the effective index of difficulty ID_e in Equation 2 represents the precision achieved by the users and is calculated as:

$$ID_e = \left(\frac{A_e}{W_e} + 1\right) = \left(\frac{A_e}{4.133 * SD_x} + 1\right) \quad (3)$$

Here, A_e represents the effective target distance, i.e., the distance traveled between two selection points. W_e represents the effective target width, where the width of the distribution of the selection points is calculated as $4.133 \times SD_x$, with SD_x being the standard deviation of the selection coordinates. SD_x measures participants' accuracy [MacKenzie 1992].

2.2 Fitts' Law and Eye-Gaze Tracking

Gaze-based selection has been studied and analyzed with Fitts' law-like tasks on 2D screens [Chatterjee et al. 2015; Isomoto et al. 2018; Schuetz et al. 2019] as well as in head-mounted displays (HMDs) [Bækgaard et al. 2019; Esteves et al. 2020; Hansen et al. 2018; Pai et al. 2019; Qian and Teather 2017]. Since gaze moves very quickly, directly comparing it with other input modalities, such as a mouse, is challenging [Schuetz et al. 2019; Sibert and Jacob 2000]. Gaze as an interaction modality for selection can only be compared to other input modalities when it is supplemented with appropriate selection / activation methods. Nonetheless, Fitts' law still applies to gaze [Teather and Stuerzlinger 2014; Wu et al. 2010; Zhang and MacKenzie 2007], especially when one considers that after a main target-directed saccade, targets are still subject to secondary, "corrective" movements of the gaze [Schuetz et al. 2019].

2.3 Interaction with Virtual Content using Hand Tracking

In recent VR pointing studies [Batmaz et al. 2020b; Mutasim et al. 2020], participants used their dominant hand's index finger to press / select virtual buttons in a VR-based eye-hand coordination task. A similar technique was employed by Speicher et al. [2018] for typing on a virtual keyboard with both hands. Esteves et al. [2020] instructed their subjects to make a finger circle gesture to trigger a selection. A grab and release hand gesture (making / opening a fist) has also been proposed to drag-and-drop 2D objects [Canare et al. 2018]. Chatterjee et al. [2015] used both gaze and pinch gestures to improve pointing on a 2D screen. In their approach, the (approximate) position of the gaze cursor could be manipulated more precisely by pinching and pointing with the hand. Once the hand-controlled cursor was inside a target, releasing the pinch gesture activated it. Similarly, pinch, hold and drag, and un-pinch was also investigated to select, drag, and release virtual 2D [Kosunen et al. 2013] and 3D [Pfeuffer et al. 2017; Velloso et al. 2015] objects. Pinch was also used to simply activate a target in VR both with [Pfeuffer et al. 2017; Velloso et al. 2015] or without [HoloLens 2019; Jimenez and Schulze 2018] gaze. However, it is still unknown how well pinch performs as a selection / activation method in terms of execution time, error rate, and throughput compared to other popular alternatives.

2.4 Activation Methods for Head- and Eye-Gaze-Based Systems

Although dwell is the most common form of gaze-only activation method in the literature [Hansen et al. 2018; Mott et al. 2017; Schuetz et al. 2019], it has several disadvantages [Mott et al. 2017; Sidenmark and Gellersen 2019b], including being slow. Thus, researchers have previously investigated ways to improve this technique [Isomoto et al. 2018; Mott et al. 2017] or find alternatives to it. A comparison of activation methods such as button clicks, hand gesture, dwell, and speech found that button clicks were the fastest method while dwell made the least errors [Esteves et al. 2020]. Similar findings were reported by Choe et al. [2019] and also apply to mobile touchscreen buttons [Yu et al. 2017]. Recently, Lu et al. [2020] studied eye blinks as an alternative to dwell for *head-gaze* pointing and found that blinks performed better than dwell. Several types of gaze gestures have also been explored in the past as a means to activate a target [Feng et al. 2014; Hyrskykari et al. 2012; Møllenbach et al. 2010; Patidar et al. 2014; Piumsomboon et al. 2017; Sidenmark et al. 2020]. Further, similar studies on activation techniques can also be found in the literature, e.g., [Hansen et al. 2018; Hassan et al. 2019; Pai et al. 2019; Rivu et al. 2019]. The fact that many different activation methods have been investigated by researchers underlines the need for a well-performing alternative to button clicks in gaze-based systems.

3 USER STUDY

3.1 Participants and Apparatus

12 participants (6 female) took part in the study. Their average age was 30.42 ± 4.66 years and all of them were right-handed. 9 and 3 participants mentioned that their right and left eye was their dominant one respectively.

The system was developed in Unity3D on a computer with i7-4790 processor, 16 GB RAM, and a GTX 1060 graphics card. An HTC VIVE Pro Eye VR headset was used in the experiment, which has a resolution of 2880×1600 pixels, 90 Hz refresh rate, and 110° (diagonal) FOV. The built-in Tobii eye-tracker in the headset transmits data at the rate of 120 Hz and is accurate to 0.5-1.1°. To track hand movements, a Leap Motion was attached below the headset (see Figure 1a).

3.2 Conditions and Implementation

- *Pinch*: We detect pinch actions through the Leap Motion. The system was designed in a way so that, to perform two consecutive pinches, one has to pinch, un-pinch (release the pinch), and pinch again. In other words, holding a single pinch even when the gaze moved was classified as only one action. Whenever the system detected a pinch, auditory feedback was given to the user. Subjects pinched with their dominant hand. They were asked to do the experiment while standing and to keep their elbow reasonably close to their body with the arm extended away from the body (see Figure 1b). Based on our pilot studies, this resulted in better hand tracking as there was less confusion with other body parts, e.g., the legs while sitting. Also, with the support of the body for the elbow, this was the least tiring position, minimizing the *Gorilla Arm* issue [Jang et al. 2017; Velloso et al. 2015]. To enable this posture, the Leap Motion was attached below the headset facing downward at an angle to clearly capture the extended arm and hand (see Figure 1a). Subjects pinched with the palm either facing horizontal towards their non-dominant side (see Figure 1d), at approximately 45° to the floor (see Figure 1e), or somewhere in between, whichever resulted in better pinch recognition. For both hand postures, subjects were instructed to keep their other three fingers (reasonably) wide open so that the hand tracker did not, e.g., confuse the middle finger for the index finger (due to palm occlusion).
- *Click*: For the click condition, participants simply pressed the trackpad on the Vive VR controller. We chose the trackpad over the trigger as the trackpad has a shorter travel distance.
- *Dwell*: Here, the user's gaze has to dwell on a target for 300 ms to activate it. This also means that if the user dwelled on a non-target for 300 ms, they would end up making an incorrect selection. We used 300 ms, as this value has been identified as one of the most feasible dwell times (in terms of speed and avoiding the Midas touch problem) in several studies [Bækgaard et al. 2019; Hansen et al. 2018; Hassan et al. 2019; Majaranta et al. 2009; Shakil et al. 2019]. Upon activation of a target, auditory feedback was given to the participants.

3.3 Hypotheses

Several past studies reported technical and physical limitations of the Leap Motion and therefore, sometimes poor hand and finger tracking, resulting in pinch recognition issues [Canare et al. 2018; Pfeuffer et al. 2017; Speicher et al. 2018; Surale et al. 2019]. Thus we were not surprised that we noticed the same issue in our pilots, where participants, at times, had to pinch more than once to activate a target. Optimizing the user posture in our final design avoided pinch detection issues as far as possible (without instrumenting the hand). As we could not guarantee 100% reliable pinch detection, we still hypothesize that **H1. button click will perform significantly better than pinch**. Based on the findings of Esteves et al. [2020] and Choe et al. [2019], we also hypothesize that **H2. the dwell condition will have the least amount of errors**.

3.4 Procedure

All participants experienced all three conditions, presented in counterbalanced order using a Latin square design. They initially signed a consent form, followed by filling a demographic questionnaire. After an experimenter explained the task, participants were given practice trials to familiarize themselves with each of the three techniques until they felt ready. Participants then started the main experiment by first performing Tobii Eye Tracking's 5-point calibration. Then, they were instructed to perform the pointing tasks as quickly and accurately as possible. Participants stood at the center of the tracking area during the experiment. At the end of all the three conditions, participants filled out a short questionnaire where they were asked to share their feedback for each condition. Whenever participants struggled to accurately point to a target, we gave them the option to calibrate the eye-tracker again. This

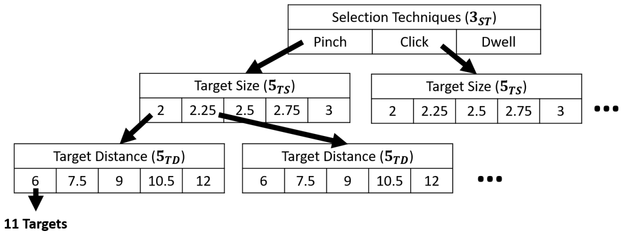

Figure 2: For each condition, a single round of trials comprised of one target size at one target distance with 11 targets.

happened only once during the experiment for each of (just) four participants.

Subjects performed an ISO 9241-411 pointing task. Similar to previous work [Batmaz et al. 2020a], the stimulus comprised of 11 targets which appeared in a circular arrangement equally distant from each other. The first target was chosen at random and the next targets alternated across the center of the circle. This alteration of targets was either in a clockwise or anti-clockwise direction for each "round" of trials (see Figure 2), determined randomly by the software. The experiment comprised of five **target sizes** (2°, 2.25°, 2.5°, 2.75°, or 3°) each of which was repeated for five **target distances** (6°, 7.5°, 9°, 10.5°, or 12°). As we use a distal pointing paradigm, we specified all sizes and distances through angular measures to make the results independent of the actual dimensions and how far the targets were away from the user. Based on our pilot studies, the maximum target distance was restricted to 12° as we wanted to minimize participants' head movements to make the experiment less tiring. Each condition took about 7 minutes. In total the whole experiment lasted about 40 minutes, including calibration, practice trials, and the pre- and post-questionnaires.

At the beginning of each round, all target spheres were grey except for the orange colored one denoting the first (current) target (see Figure 1c). The color of the target was changed to blue whenever the cursor sphere came in contact, i.e., we used highlighting [Teather and Stuerzlinger 2014]. Upon correct selection / activation of a target, its color was changed to green. Similarly, for an incorrect selection / activation, the target sphere's color was changed to red. In this case, auditory error feedback was provided to the participants. To keep subjects informed about the selection condition for the current round, an image (of a controller, hand, or eyes for dwell) was also shown at the center of the target grid (see Figure 1c). For consistency, the target grid's center was placed at the eye level of participants.

3.5 Experimental Design

We designed a within-subjects study where the subjects performed the task with three **selection techniques** (3_{ST}, pinch, click, and dwell) for each of the five **target sizes** (5_{TS}), all of which were presented in counterbalanced order following a Latin square design. Our experiment used five **target distances** (5_{TD}), for a total of ($5_{TS} \times 5_{TD} =$) 25 *ID*s (22 of which were unique) between 1.5 and 2.9. As dependent variables, participants' movement time (s), error rate (%), and effective throughput ($bits/s$) [ISO 9241-411:2012 2012] were measured. There were 11 targets per round of trials yielding $11 \times 5_{TS} \times 5_{TD} \times 3_{ST} = 825$ data points per participant.

Table 1: RM ANOVA results for Time, Error Rate, and Throughput across Selection Techniques and ID.

	Selection Techniques	ID	Selection Techniques × ID
Time	$F(2, 22) = 45.95$ ***, $\eta^2 = 0.81$	$F(21, 231) = 9.15$ ***, $\eta^2 = 0.45$	$F(42, 462) = 1.75$ **, $\eta^2 = 0.14$
Error rate	$F(1.25, 13.71) = 65.37$ ***, $\eta^2 = 0.86$	$F(21, 231) = 4.42$ ***, $\eta^2 = 0.29$	$F(42, 462) = 1.61$ *, $\eta^2 = 0.13$
Throughput	$F(2, 22) = 32.85$ ***, $\eta^2 = 0.75$	$F(21, 231) = 3.53$ ***, $\eta^2 = 0.24$	$F(42, 462) = 1.58$ *, $\eta^2 = 0.13$

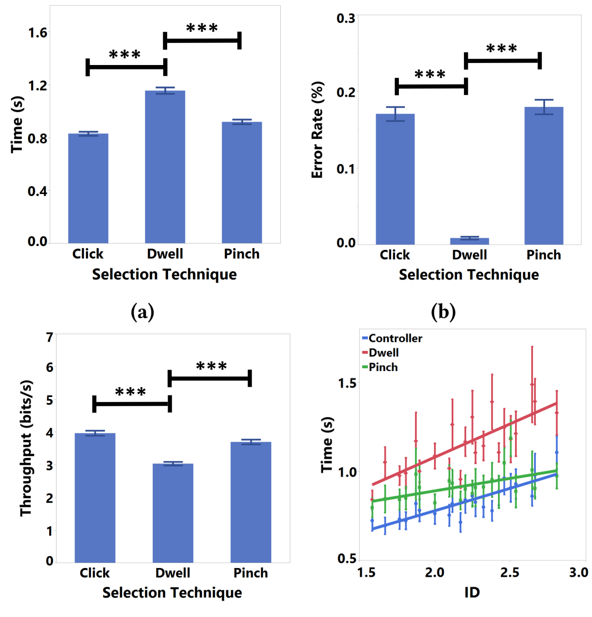

Figure 3: (a) Time, (b) Error Rate, (c) Throughput, and (d) Fitts' Law results for Selection Techniques.

4 RESULTS

We used repeated measures (RM) ANOVA with $\alpha = 0.05$ in SPSS 26. The data was considered to be normal when Skewness (S) and Kurtosis (K) values were within ±1.5 [Hair Jr et al. 2014; Mallery and George 2003]. Upon violation of the sphericity assumption (according to Mauchly's sphericity test), Huynh-Feldt correction was used where ϵ was less than 0.75. If the data was not normally distributed, log-transforming the data resulted in a normal distribution. Only statistically significant results are mentioned here. We used the Bonferroni method for post-hoc analyses. Significance levels are shown as *** for $p < 0.001$, ** $p < 0.01$, and * $p < 0.05$. Figures represent means and standard error of means.

4.1 Time, Error Rate, Throughput, and Fitts' Law Analysis

Results of the RM ANOVA for the selection techniques are presented in Table 1. As shown in Figure 3a, dwell was the slowest among all the conditions. However, according to the error rate data shown in Figure 3b, it experienced the least errors. The results for throughput in Figure 3c shows that participants' overall performance was significantly better in both the click and the pinch conditions relative to dwell. Although click was slightly faster, exhibited fewer errors, and achieved higher throughput than pinch, none of these differences were significant.

Fitts' law analysis according to Equation 1 shows that the selection time can be modeled as $MT = 0.57 + 0.15ID, R^2 = 0.32$ for pinch, $MT = 0.28 + 0.25ID, R^2 = 0.70$ for click, and $MT = 0.39 + 0.34ID, R^2 = 0.60$ for dwell, see Figure 3d.

4.2 Subjective Measures

In the post-experiment questionnaire, 6 out of the 12 participants preferred the button click selection technique, 4 preferred dwell, and only 2 pinch. However, 6 and 4 subjects ranked the button click or pinch as their second-most preferred selection method, respectively. Reasons given for the choice of button click as the most preferred technique were "*felt a lot faster*", "*more robust, which means less error on executing the trigger*", "*I could control [better] when I wanted to select targets (compared to dwell) and it was less physically tiring (compared to pinch)*", "*The controller was still somewhat familiar, the pinch got tiring in the forearm after a little while*", and "*Did not have to do double clicks to select one target*". The subjects who chose dwell mentioned "*was just easier to [do] one thing, rather do too many things*", "*it did not involve moving any part of the body except for the eyes*", and "*it was more accurate and easier*". Reasons for choosing pinch were "*pinch was easy and fun, it makes sense*" and "*using my hand and not being a passive actor allowed me to get more involved with the process*".

When queried about the level of frustration associated with each technique on a 7-point Likert scale (1: very frustrating, 7: very satisfied), button clicks were the least frustrating ($\mu = 5.58, median = 6$), followed by dwell ($\mu = 4.75, median = 5$) and pinch ($\mu = 3.83, median = 4$). Similarly, for physical fatigue (1: very fatiguing, 7: very relaxing), pinch was rated as the most fatiguing technique ($\mu = 2.67, median = 3$) compared to a button click ($\mu = 4.16, median = 4$) and dwell ($\mu = 3.83, median = 4$). For mental fatigue, pinch was again classified as the most tiring (pinch: $\mu = 3.58, median = 3$, button click: $\mu = 4.75, median = 4$, and dwell: $\mu = 3.92, median = 4$).

5 DISCUSSION

In this work, we used a standardized ISO 9241-411 Fitts' law task to identify how well the pinch gesture performs as a selection / activation technique compared to click and dwell in terms of execution time, error rate, and throughput.

As we did not find any significant difference between the two techniques, we find pinch to be comparable to button clicks. These findings do not match previous work [Canare et al. 2018; Pfeuffer et al. 2017; Speicher et al. 2018; Surale et al. 2019] where researchers reported poor performance of systems that involved hand tracking. One possible reason for this can be the position and angle at which we attached the hand tracker to the VR headset (see Figure 1a), which seems to have worked substantially better than other alternatives. Previous studies [Batmaz et al. 2020b; Mutasim et al. 2020; Pfeuffer et al. 2017; Speicher et al. 2018; Surale et al. 2019] mounted the hand tracker on the front of the VR headset. This option increases fatigue and thus suffers from the *Gorilla Arm* [Jang et al. 2017; Velloso et al. 2015] effect, as the subjects always have to keep their hands directly in front of their face. Also, our experiment required no or only very little head movement from the participant when pointing at the targets. This is important because whenever the user turns their head, they need to move their hands along with it to not loose hand tracking (a problem well discussed in the literature [Chiu et al. 2019]), which then increases task execution time and with it also fatigue, and thus significantly reduces performance. Our placement of the hand tracker underneath the VR headset addressed both these issues. Not only was the hand and elbow position more natural and (much) more comfortable but (limited) head movements also did not affect the hand tracking in a notable manner. We believe that recent advances of built-in external world-view cameras for hand tracking in VR headsets [Oculus 2019] has great promise and will potentially get rid of the need to attach a separate downward-facing hand tracking device to VR headsets.

Nonetheless, we still have to acknowledge that we did face some pinch recognition issues during the study. Yet, once a participant found a sweet spot for the angle of their palm (as discussed above), together with a comfortable posture, the pinch gesture was reliably recognized by the system. We observed that some subjects struggled with this at the beginning of the experiment but once they found a good palm angle, the system was quite efficient is recognizing the pinch gesture. Unfortunately, we observed a few instances where the user's hand drifted from the ideal zone over time and thus participants had to sometimes change their palm angle to regain reliable pinch recognition. Even with these issues, pinch was still competitive with button clicks. Nonetheless, adopting recently presented advanced hand tracking algorithm [Smith et al. 2020] should address the restricted hand posture in our pinch condition, potentially allowing our findings to be generalized across different applications.

Even though they achieved results comparable with button clicks, subjects did not prefer pinch as a selection / activation technique. Beyond the pinch recognition issue, we identified that frustration and fatigue also influenced this decision. We believe that both these factors are caused by a limitation of the Leap Motion. For the system to recognize consecutive pinches, subjects had to (reasonably) fully release the pinch in between consecutive pinch actions by spreading the thumb and index finger relatively clearly apart, which was perceived to be unnatural. In other words, due to the limitation of the un-pinch gesture, subjects had to make (unnaturally) large pinches. This forced the subjects to, at times, pinch twice to activate one target which "broke" their rhythm. In contrast to the small travel distance of the trackpad button of the controller and its reliable activation, this limitation induced fatigue and frustration in subjects. We speculate that solving this issue might even reveal pinch to be better than button clicks. Although our quantitative results does not support our hypothesis **H1** that button clicks will have better performance than pinch, taking subjects' preferences into account we conclude that our findings partially support hypothesis **H1**.

Just like previous work [Choe et al. 2019; Esteves et al. 2020], our results also showed that the dwell condition experienced the least errors among all the three conditions (see Figure 3b), thus supporting our hypothesis **H2**. Yet, this technique had the worst performance in terms of selection time and overall throughput (see Figures 3a and 3c), making it not competitive. A widely accepted reason for this outcome is that dwell is a time consuming selection / activation method because one has to wait for a non-trivial amount of time, i.e., the dwell time, to activate a target, which seems to reverse any potential efficiency gains. Yet, making the dwell time (too) short increases the Midas touch problem [Isomoto et al. 2018; Mott et al. 2017; Sidenmark and Gellersen 2019b]. On top of this, dwell, along with other gaze gesture activation methods, e.g. [Feng et al. 2014], suffers also from the problem of having to alternate between two tasks with the eyes, i.e., pointing and selection / activation. In contrast, button clicks or pinch have the advantage of overlapping these two tasks to some extent, i.e., a user can mentally initiate the act of pressing a button even before their gaze has reached the target, if they anticipate that the cursor will reach the target by the time the button is pressed. We believe that this dual-task issue is a major drawback of gaze-only systems, one that needs to be addressed in order to make such systems more popular, especially important for patients with limited muscle control [Kumar et al. 2017].

Even though only 12 subjects participated in this study, we found high effect sizes for the significant differences. The minimum effect size for the selection techniques was 0.75 and 0.24 for ID, i.e., both large effects, commonly defined through a criterion of $\eta^2 > 0.14$. Based on these large effect sizes, we believe our findings to be robust.

6 CONCLUSION AND FUTURE WORK

In this work, we compared the performance of three selection / activation techniques, namely, pinch, click, and dwell, for eye-gaze-based interaction in VR. Our results revealed that dwell as a selection technique made the least errors. However, it had the worst performance in terms of execution time and throughput. Results also showed that although button clicks achieved the highest performance, this technique was not significantly different from pinch for each of the three performance metrics analyzed in this study. Still, participants preferred button click and dwell over pinch as pinch was sometimes frustrating due to recognition errors and seems to have induced more physical and mental fatigue.

Nonetheless, we recommend gaze and pinch for selection tasks when controllers are not a viable option and where hands can be tracked with sufficient reliability, as this combination allows practitioners, developers, and researchers to take advantage of its competitive performance. In the future, we plan to further explore the pinch gesture and its performance in other 3D virtual object selection and manipulation tasks.

REFERENCES

Per Bækgaard, John Paulin Hansen, Katsumi Minakata, and I. Scott MacKenzie. 2019. A Fitts' Law Study of Pupil Dilations in a Head-Mounted Display. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research and Applications* (Denver, Colorado) *(ETRA '19)*. Association for Computing Machinery, New York, NY, USA, Article 32, 5 pages. https://doi.org/10.1145/3314111.3319831

A. Terry Bahill, Michael R. Clark, and Lawrence Stark. 1975. The main sequence, a tool for studying human eye movements. *Mathematical Biosciences* 24, 3 (1975), 191 – 204. https://doi.org/10.1016/0025-5564(75)90075-9

Anil Ufuk Batmaz, Aunnoy K Mutasim, Morteza Malekmakan, Elham Sadr, and Wolfgang Stuerzlinger. 2020b. Touch the Wall: Comparison of Virtual and Augmented Reality with Conventional 2D Screen Eye-Hand Coordination Training Systems. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 184–193. https://doi.org/10.1109/VR46266.2020.00037

Anil Ufuk Batmaz, Aunnoy K Mutasim, and Wolfgang Stuerzlinger. 2020a. Precision vs. Power Grip: A Comparison of Pen Grip Styles for Selection in Virtual Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 23–28. https://doi.org/10.1109/VRW50115.2020.00012

Jonas Blattgerste, Patrick Renner, and Thies Pfeiffer. 2018. Advantages of Eye-Gaze over Head-Gaze-Based Selection in Virtual and Augmented Reality under Varying Field of Views. In *Proceedings of the Workshop on Communication by Gaze Interaction* (Warsaw, Poland) *(COGAIN '18)*. Association for Computing Machinery, New York, NY, USA, Article 1, 9 pages. https://doi.org/10.1145/3206343.3206349

Dominic Canare, Barbara Chaparro, and Alex Chaparro. 2018. Using Gesture, Gaze, and Combination Input Schemes as Alternatives to the Computer Mouse. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 62 (2018), 297 – 301.

Ishan Chatterjee, Robert Xiao, and Chris Harrison. 2015. Gaze+Gesture: Expressive, Precise and Targeted Free-Space Interactions. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (Seattle, Washington, USA) *(ICMI '15)*. Association for Computing Machinery, New York, NY, USA, 131–138. https://doi.org/10.1145/2818346.2820752

Pascal Chiu, Kazuki Takashima, Kazuyuki Fujita, and Yoshifumi Kitamura. 2019. Pursuit Sensing: Extending Hand Tracking Space in Mobile VR Applications. In *Symposium on Spatial User Interaction* (New Orleans, LA, USA) *(SUI '19)*. Association for Computing Machinery, New York, NY, USA, Article 1, 5 pages. https://doi.org/10.1145/3357251.3357578

Mungyeong Choe, Yeongcheol Choi, Jaehyun Park, and Hyun K. Kim. 2019. Comparison of Gaze Cursor Input Methods for Virtual Reality Devices. *International Journal of Human–Computer Interaction* 35, 7 (2019), 620–629. https://doi.org/10.1080/10447318.2018.1484054 arXiv:https://doi.org/10.1080/10447318.2018.1484054

Augusto Esteves, Yonghwan Shin, and Ian Oakley. 2020. Comparing selection mechanisms for gaze input techniques in head-mounted displays. *International Journal of Human-Computer Studies* 139 (2020), 102414. https://doi.org/10.1016/j.ijhcs.2020.102414

Wenxin Feng, Ming Chen, and Margrit Betke. 2014. Target Reverse Crossing: A Selection Method for Camera-Based Mouse-Replacement Systems. In *Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments* (Rhodes, Greece) *(PETRA '14)*. Association for Computing Machinery, New York, NY, USA, Article 39, 4 pages. https://doi.org/10.1145/2674396.2674443

Joseph F Hair Jr, William C Black, Barry J Babin, and Rolph E. Anderson. 2014. Multivariate data analysis.

John Paulin Hansen, Vijay Rajanna, I. Scott MacKenzie, and Per Bækgaard. 2018. A Fitts' Law Study of Click and Dwell Interaction by Gaze, Head and Mouse with a Head-Mounted Display. In *Proceedings of the Workshop on Communication by Gaze Interaction* (Warsaw, Poland) *(COGAIN '18)*. Association for Computing Machinery, New York, NY, USA, Article 7, 5 pages. https://doi.org/10.1145/3206343.3206344

Mehedi Hassan, John Magee, and I Scott MacKenzie. 2019. A Fitts' law evaluation of hands-free and hands-on input on a laptop computer. In *International Conference on Human-Computer Interaction*. Springer, 234–249.

HoloLens. 2019. Microsoft HoloLens: Mixed Reality Technology for Business. https://www.microsoft.com/en-us/hololens

HTC. 2019. HTC VIVE Hand Tracking SDK Early Access. https://developer.vive.com/resources/vive-sense/sdk/vive-hand-tracking-sdk/

Aulikki Hyrskykari, Howell Istance, and Stephen Vickers. 2012. Gaze Gestures or Dwell-Based Interaction?. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, California) *(ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 229–232. https://doi.org/10.1145/2168556.2168602

ISO 9241-411:2012. 2012. Ergonomics of human-system interaction – Part 411: Evaluation methods for the design of physical input devices. ISO. https://www.iso.org/standard/54106.html

Toshiya Isomoto, Toshiyuki Ando, Buntarou Shizuki, and Shin Takahashi. 2018. Dwell Time Reduction Technique Using Fitts' Law for Gaze-Based Target Acquisition. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research and Applications* (Warsaw, Poland) *(ETRA '18)*. Association for Computing Machinery, New York, NY, USA, Article 26, 7 pages. https://doi.org/10.1145/3204493.3204532

Robert JK Jacob. 1991. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems (TOIS)* 9, 2 (1991), 152–169.

Sujin Jang, Wolfgang Stuerzlinger, Satyajit Ambike, and Karthik Ramani. 2017. *Modeling Cumulative Arm Fatigue in Mid-Air Interaction Based on Perceived Exertion and Kinetics of Arm Motion.* Association for Computing Machinery, New York, NY, USA, 3328–3339. https://doi.org/10.1145/3025453.3025523

Janis G Jimenez and Jürgen P Schulze. 2018. Continuous-Motion Text Input in Virtual Reality. *Electronic Imaging* 2018, 3 (2018), 450–1.

Ilkka Kosunen, Antti Jylha, Imtiaj Ahmed, Chao An, Luca Chech, Luciano Gamberini, Marc Cavazza, and Giulio Jacucci. 2013. Comparing Eye and Gesture Pointing to Drag Items on Large Screens. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces* (St. Andrews, Scotland, United Kingdom) *(ITS '13)*. Association for Computing Machinery, New York, NY, USA, 425–428. https://doi.org/10.1145/2512349.2514920

Chandan Kumar, Raphael Menges, Daniel Müller, and Steffen Staab. 2017. Chromium Based Framework to Include Gaze Interaction in Web Browser. In *Proceedings of the 26th International Conference on World Wide Web Companion* (Perth, Australia) *(WWW '17 Companion)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 219–223. https://doi.org/10.1145/3041021.3054730

Xueshi Lu, Difeng Yu, Hai-Ning Liang, Wenge Xu, Yuzheng Chen, Xiang Li, and Khalad Hasan. 2020. Exploration of Hands-free Text Entry Techniques For Virtual Reality. arXiv:2010.03247 [cs.HC]

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

Päivi Majaranta, Ulla-Kaija Ahola, and Oleg Špakov. 2009. Fast Gaze Typing with an Adjustable Dwell Time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 357–360. https://doi.org/10.1145/1518701.1518758

Paul Mallery and Darren George. 2003. SPSS for Windows step by step: a simple guide and reference. *Allyn, Bacon, Boston,* (2003).

Mark R. Mine. 1995. *Virtual Environment Interaction Techniques.* Technical Report. USA.

Emilie Møllenbach, Martin Lillholm, Alastair Gail, and John Paulin Hansen. 2010. Single Gaze Gestures. In *Proceedings of the 2010 Symposium on Eye-Tracking Research and Applications* (Austin, Texas) *(ETRA '10)*. Association for Computing Machinery, New York, NY, USA, 177–180. https://doi.org/10.1145/1743666.1743710

Martez E. Mott, Shane Williams, Jacob O. Wobbrock, and Meredith Ringel Morris. 2017. Improving Dwell-Based Gaze Typing with Dynamic, Cascading Dwell Times. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2558–2570. https://doi.org/10.1145/3025453.3025517

Aunnoy K. Mutasim, Anil Ufuk Batmaz, and Wolfgang Stuerzlinger. 2020. Gaze Tracking for Eye-Hand Coordination Training Systems in Virtual Reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3334480.3382924

Oculus. 2019. Oculus Quest @ OC6: Introducing Hand Tracking, Oculus Link, Passthrough+ on Quest, and More. https://www.oculus.com/blog/oculus-quest-at-oc6-introducing-hand-tracking-oculus-link-passthrough-on-quest-and-more/

Yun Suen Pai, Tilman Dingler, and Kai Kunze. 2019. Assessing hands-free interactions for VR using eye gaze and electromyography. *Virtual Reality* 23, 2 (2019), 119–131.

Pawan Patidar, Himanshu Raghuvanshi, and Sayan Sarcar. 2014. Quickpie: An Interface for Fast and Accurate Eye Gazed based Text Entry. arXiv:1407.7313 [cs.HC]

Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze + Pinch Interaction in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 99–108. https://doi.org/10.1145/3131277.3132180

T. Piumsomboon, G. Lee, R. W. Lindeman, and M. Billinghurst. 2017. Exploring natural eye-gaze-based interaction for immersive virtual reality. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. 36–39.

Yuan Yuan Qian and Robert J. Teather. 2017. The Eyes Don't Have It: An Empirical Comparison of Head-Based and Eye-Based Selection in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 91–98. https://doi.org/10.1145/3131277.3132182

Vijay Rajanna and John Paulin Hansen. 2018. Gaze Typing in Virtual Reality: Impact of Keyboard Design, Selection Method, and Motion. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research and Applications* (Warsaw, Poland) *(ETRA '18)*. Association for Computing Machinery, New York, NY, USA, Article 15, 10 pages. https://doi.org/10.1145/3204493.3204541

Sheikh Rivu, Yasmeen Abdrabou, Thomas Mayer, Ken Pfeuffer, and Florian Alt. 2019. GazeButton: Enhancing Buttons with Eye Gaze Interactions. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research and Applications* (Denver, Colorado) *(ETRA '19)*. Association for Computing Machinery, New York, NY, USA, Article 73, 7 pages. https://doi.org/10.1145/3317956.3318154

Kunhee Ryu, Joong-Jae Lee, and Jung-Min Park. 2019. GG Interaction: a gaze–grasp pose interaction for 3D virtual object selection. *Journal on Multimodal User Interfaces* 13 (2019), 383–393. https://doi.org/10.1007/s12193-019-00305-y

Immo Schuetz, T. Scott Murdison, Kevin J. MacKenzie, and Marina Zannoli. 2019. 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* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300765

Asma Shakil, Christof Lutteroth, and Gerald Weber. 2019. CodeGazer: Making Code Navigation Easy and Natural With Gaze Input. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300306

Linda E Sibert and Robert JK Jacob. 2000. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 281–288.

Ludwig Sidenmark, Christopher Clarke, Xuesong Zhang, Jenny Phu, and Hans Gellersen. 2020. *Outline Pursuits: Gaze-Assisted Selection of Occluded Objects in Virtual Reality.* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi-org.proxy.lib.sfu.ca/10.1145/3313831.3376438

Ludwig Sidenmark and Hans Gellersen. 2019a. Eye, Head and Torso Coordination During Gaze Shifts in Virtual Reality. *ACM Trans. Comput.-Hum. Interact.* 27, 1, Article 4 (Dec. 2019), 40 pages. https://doi.org/10.1145/3361218

Ludwig Sidenmark and Hans Gellersen. 2019b. Eye&Head: Synergetic Eye and Head Movement for Gaze Pointing and Selection. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1161–1174. https://doi.org/10.1145/3332165.3347921

Breannan Smith, Chenglei Wu, He Wen, Patrick Peluse, Yaser Sheikh, Jessica K. Hodgins, and Takaaki Shiratori. 2020. Constraining Dense Hand Surface Tracking with Elasticity. *ACM Trans. Graph.* 39, 6, Article 219 (Nov. 2020), 14 pages. https://doi.org/10.1145/3414685.3417768

Marco Speicher, Anna Maria Feit, Pascal Ziegler, and Antonio Krüger. 2018. Selection-Based Text Entry in Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174221

Hemant Bhaskar Surale, Fabrice Matulic, and Daniel Vogel. 2019. Experimental Analysis of Barehand Mid-Air Mode-Switching Techniques in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300426

Robert J. Teather and Wolfgang Stuerzlinger. 2014. Visual Aids in 3D Point Selection Experiments. In *Proceedings of the 2nd ACM Symposium on Spatial User Interaction* (Honolulu, Hawaii, USA) *(SUI '14)*. Association for Computing Machinery, New York, NY, USA, 127–136. https://doi.org/10.1145/2659766.2659770

Eduardo Velloso, Jayson Turner, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2015. An Empirical Investigation of Gaze Selection in Mid-Air Gestural 3D Manipulation. In *Human-Computer Interaction – INTERACT 2015*, Julio Abascal, Simone Barbosa, Mirko Fetter, Tom Gross, Philippe Palanque, and Marco Winckler (Eds.). Springer International Publishing, Cham, 315–330.

Andy Wilson. 2006. Robust Computer Vision-Based Detection of Pinching for One and Two-Handed Gesture Input. In *UIST '06 Proceedings of the 19th annual ACM symposium on User interface software and technology.* ACM, 255–258. https://www.microsoft.com/en-us/research/publication/robust-computer-vision-based-detection-pinching-one-two-handed-gesture-input/

Chia-Chien Wu, Oh-Sang Kwon, and Eileen Kowler. 2010. Fitts's Law and speed/accuracy trade-offs during sequences of saccades: Implications for strategies of saccadic planning. *Vision Research* 50, 21 (2010), 2142 – 2157. https://doi.org/10.1016/j.visres.2010.08.008

Chun Yu, Yizheng Gu, Zhican Yang, Xin Yi, Hengliang Luo, and Yuanchun Shi. 2017. Tap, Dwell or Gesture? Exploring Head-Based Text Entry Techniques for HMDs. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 4479–4488. https://doi.org/10.1145/3025453.3025964

Xuan Zhang and I Scott MacKenzie. 2007. Evaluating eye tracking with ISO 9241-Part 9. In *International Conference on Human-Computer Interaction*. Springer, 779–788.