Evaluating Ray Casting and Two Gaze-Based Pointing Techniques for Object Selection in Virtual Reality

Tomi Nukarinen
TAUCHI, University of Tampere
tomi.nukarinen@uta.fi

Jari Kangas
TAUCHI, University of Tampere
jari.kangas@uta.fi

Jussi Rantala
TAUCHI, University of Tampere
jussi.e.rantala@uta.fi

Olli Koskinen
TAUCHI, University of Tampere
olli.koskinen@uta.fi

Roope Raisamo
TAUCHI, University of Tampere
roope.raisamo@uta.fi

ABSTRACT

Selecting an object is a basic interaction task in virtual reality (VR) environments. Interaction techniques with gaze pointing have potential for this elementary task. There appears to be little empirical evidence concerning the benefits and drawbacks of these methods in VR. We ran an experiment studying three interaction techniques: ray casting, dwell time and gaze trigger, where gaze trigger was a combination of gaze pointing and controller selection. We studied user experience and interaction speed in a simple object selection task. The results indicated that ray casting outperforms both gaze-based methods while gaze trigger performs better than dwell time.

CCS CONCEPTS

• **Human-centered computing → Pointing devices**; **Pointing**; *Empirical studies in HCI*;

KEYWORDS

Object selection in virtual reality, gaze pointing, controller pointing

ACM Reference Format:
Tomi Nukarinen, Jari Kangas, Jussi Rantala, Olli Koskinen, and Roope Raisamo. 2018. Evaluating Ray Casting and Two Gaze-Based Pointing Techniques for Object Selection in Virtual Reality. In *VRST 2018: 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18), November 28-December 1, 2018, Tokyo, Japan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3281505.3283382

1 INTRODUCTION

Research in gaze-based interaction techniques has demonstrated that gaze could have potential as an input method in VR. However, integrating gaze-based interaction into VR systems has progressed at a slow pace [3]. The main motivations for using gaze in VR include faster, more accurate pointing and freeing the hands for other tasks.

The aim for the current work was to study user experience and interaction speed differences between ray casting, dwell time and gaze trigger in VR. The gaze trigger method was a combination of gaze pointing and controller selection, positioning itself between the two other techniques. While ray casting and dwell time have been compared previously [1], no one appears to have investigated gaze trigger in VR before. The closest parallel seems to be a study by Fono and Vertegaal [2] who reported that eye tracking with key activation was faster than a mouse-keyboard combination, and also preferred by most participants. Studying gaze trigger together with dwell time allows us to differentiate the usability issues arising from the pointing and the selection phases of gaze-based interaction. Further, ray casting was included to allow a comparison to a technique that has gained broad acceptance in VR systems. The current work shows that gaze pointing with trigger selection received worse user feedback and was substantially slower than ray casting.

2 METHOD

We compared three interaction methods in an object picking act. In *HandTrigger*, the participant pointed an object by casting a visual beam (Figure 1, left) from the controller, and pressed the trigger below the controller. In *GazeTrigger*, the participant looked at an object, and pressed the trigger button as in *HandTrigger*. In *GazeDwell*, the participant looked at an object, and kept the gaze on the object until 1000 ms (the dwell time) had elapsed. There was a separate visualization (Figure 1, right) to indicate the passage of the dwell time. We investigated if there are differences in user experience or interaction speed between the methods.

We recruited 18 volunteer participants (6 females, 12 males, mean age 21, range 19-28 years), all undergraduate students. 10 out of the 18 participants had at least some earlier experience of VR technology. One participant was left-handed, and 17 were right-handed. Two out of 18 wore eyeglasses during the experiment. We used a laptop PC, an HTC Vive VR headset with an inbuilt Tobii gaze tracker, an HTC Vive hand controller and Unity Virtual reality development environment to set up the experiment.

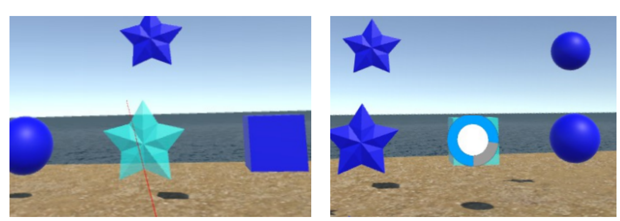

Figure 1: *HandTrigger* pointing (left), and the indicator for *GazeDwell* selection timer (right).

VRST '18, November 28-December 1, 2018, Tokyo, Japan

ACM ISBN 978-1-4503-6086-9/18/11.
https://doi.org/10.1145/3281505.3283382

Table 1: The typical participant comments per experiment condition and quality (positive/negative). The number of mentions is shown in parenthesis. The rightmost column shows the ratio between positive and negative comments.

Condition	Positive comments	Negative comments	Ratio
HandTrigger	fast (10), accuracy (5), easiness (4), controllability (4), naturalness (4), freeing the eyes (4)	instability of interaction (5), conventionality (2)	37:7
GazeTrigger	fast (6), freeing the hands (4), confirmation before selection (3), naturalness (3)	interaction effortfulness (4), tying the gaze (3), clumsiness (2), problems with calibration (2)	16:11
GazeDwell	fast interaction speed (4), easiness (4), least physical movement (2)	accidental selections (8), interruptions on eye-blinks (6), slow interaction speed (5), mental demand (3), requiring head turns (2), unnatural (2)	10:26

The experimental task was to select one to eight objects in each trial. All the target objects, round, star-shaped, and square, were randomly scattered in a 4x6 grid and visible at a glance. The participants began a trial by pressing the touchpad button and proceeded to select the objects. After selecting all the specified objects the participant confirmed the selections to end the trial. If there were any missing or extra selections, the participant would not be allowed to complete before correcting them. After successful completion, the system showed the next target shape, and the process would repeat. Altogether, each participant did three practice trials and 24 actual trials with one input method, after which they filled a questionnaire about the condition. The participants evaluated the interaction methods on several attributes using a scale from -4 to 4. The above procedure was repeated for the two other methods. After the third condition, the participant filled a separate questionnaire about the experiment as a whole and gave a preference order for the methods.

The participants repeated each trial length three times for each input method in a random order. Thus, an average trial had 4.5 ((1+2+...+8) / 8) object selections. In total, the study consisted of 5832 object selections (18 participants × 3 input methods × 24 trials × 4.5 selections per trial). We balanced the conditions between the participants using Latin squares.

3 RESULTS

We analyzed the questionnaire attributes (task success, easiness, pleasantness, speed, realism, and control) with Friedman tests, and found statistically significant ($p < 0.05$) differences in each attribute: interaction control ($p < 0.001$, $\chi^2(2) = 18.53$), easiness ($p < 0.001$, $\chi^2(2) = 17.83$), speed ($p = 0.001$), pleasantness ($p = 0.007$), realism ($p = 0.018$), and task success ($p = 0.037$). Further Bonferroni-corrected Wilcoxon tests showed significant differences between the conditions *HandTrigger* and *GazeDwell* in the attributes easiness ($p = 0.002$), control ($p = 0.003$), speed ($p = 0.003$), pleasantness ($p = 0.015$), and success ($p = 0.039$). *HandTrigger* was rated easier, more controllable, faster, more pleasant and successful than *GazeDwell*. We also found statistically significant differences between the conditions *HandTrigger* and *GazeTrigger* in the attributes control ($p = 0.021$), easiness ($p = 0.024$), success ($p = 0.045$), and speed ($p = 0.048$). *HandTrigger* was rated more controllable, easier, more successful, and faster than *GazeTrigger*.

In a preference order ranking, the most preferred methods were HandTrigger (50 %), *GazeTrigger* (33 %) and GazeDwell (17 %). The least preferred methods were GazeDwell (61 %), GazeTrigger (28 %), and HandTrigger (11 %). We analyzed the differences with a Friedman test and found a significant difference ($p = 0.042$). Bonferroni-corrected Wilcoxon tests showed a significant difference on the preference order between the conditions *HandTrigger* ($p = 0.037$) and *GazeDwell*. *HandTrigger* was significantly more preferred than *GazeDwell*. We also asked the participants how did they choose their most preferred, the middlemost, and the least preferred method. What were the pros and cons? We then clustered all of the participants' comments into thematically similar categories. We further separated the categories into positive and negative. The categories with at least two comments are listed in Table 1.

The average completion times for the conditions (all 24 trials) were 114.5 seconds for *HandTrigger*, 155.0 for *GazeTrigger* and 225.8 for *GazeDwell*. We analyzed the data with one-way repeated measures ANOVA and Bonferroni-corrected post-hoc pairwise comparisons. We also estimated the effect size. ANOVA showed statistically significant differences between all the conditions ($F_{2,16} = 43.51$, $p < 0.001$, $\eta_p^2 = 0.85$). The participants were faster with *HandTrigger* than with *GazeTrigger* ($t_{17} = 5.01$, $p < 0.001$, $d_{av} = 1.12$) or *GazeDwell* ($t_{17} = 9, 45$, $p < 0.001$, $d_{av} = 2.62$). Further, the participants were faster with *GazeTrigger* than *GazeDwell* ($t_{17} = 6.06$, $p < 0.001$, $d_{av} = 1.42$).

4 CONCLUSION

We demonstrated that ray casting performed better than two gaze-based pointing methods in a simple VR object selection task. The results contrast with findings in a desktop environment [2], where gaze pointing with key activation outperformed a mouse-keyboard combination. We relate the better performance of ray casting in part to faster object identification with peripheral attention and suggest that this may be a human factor issue and not only a question of eye-tracking precision.

ACKNOWLEDGMENTS

This work has received funding from the VARPU project (decision 1444/31/2016).

REFERENCES

[1] Nathan Cournia, John D Smith, and Andrew T Duchowski. 2003. Gaze-vs. hand-based pointing in virtual environments. In *CHI'03 extended abstracts on Human factors in computing systems*. ACM, ACM, New York, NY, USA, 772–773.

[2] David Fono and Roel Vertegaal. 2005. EyeWindows: Evaluation of Eye-controlled Zooming Windows for Focus Selection. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. ACM, New York, NY, USA, 151–160. https://doi.org/10.1145/1054972.1054994

[3] J. Jankowski and M. Hachet. 2015. Advances in Interaction with 3D Environments. *Computer Graphics Forum* 34, 1 (2015), 152–190. https://doi.org/10.1111/cgf.12466 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.12466