Finger Contact in Gesture Interaction Improves Time-domain Input Accuracy in HMD-based Augmented Reality

Seo Young Oh
KAIST UVR Lab
Daejeon, South Korea
seoyoung.oh@kaist.ac.kr

Boram Yoon
KAIST UVR Lab
Daejeon, South Korea
boram.yoon1206@kaist.ac.kr

Hyung-il Kim
KAIST UVR Lab
Daejeon, South Korea
hyungil@kaist.ac.kr

Woontack Woo
KAIST UVR Lab
Daejeon, South Korea
wwoo@kaist.ac.kr

CHI '20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6819-3/20/04.
http://dx.doi.org/10.1145/3334480.3383098

Abstract
This paper reports that the time-domain accuracy of bare-hand interactions in HMD-based Augmented Reality can be improved by using finger contact: touching a finger with another or tapping one's own hand. The activation of input can be precisely defined by the moment of finger contact, allowing the user to perform the input precisely at the desired moment. Finger contact is better suited to the user's mental model, and natural tactile feedback from the fingertip also benefits the user with the self-perception of the input. The experimental results revealed that using finger contact is the preferred method of input that increases the time-domain accuracy and enables the user to be aware of the moment the input is activated.

Author Keywords
Augmented Reality; 3D Gesture Interaction; Time-domain Input Accuracy; Passive Haptic Feedback; Finger Contact

CCS Concepts
•**Human-centered computing** → **Gestural input;** *Mixed / augmented reality;* User studies;

Introduction
While the increasing popularity of Head-mounted Displays (HMDs) have opened up new possibilities of developing more abundant interaction methods for Augmented Reality

(AR), little has been studied on what types of input interactions are efficient for HMD-based AR. Since the basic input method can affect user experience, user performance, mental model, and the form of applications, new computing environments like HMD-based AR especially should have well-established fundamental interactions.

Bare-hand gesture interactions have an advantage in that users do not need to hold handheld devices, and therefore can perform interactions whenever and wherever they want. However, most of the gestures currently used in HMD-based AR have a vague definition of the input activation point. For example, Air Tap, a staple bare-hand interaction method used for the Microsoft HoloLens [7], is activated when the user lowers one's index finger to a certain degree. This makes it hard for users to discern when exactly the input is activated. The existing gestures may seem sufficient for the current AR environment, but they will not be able to fully accommodate various interactions required for more dynamic and advanced AR content.

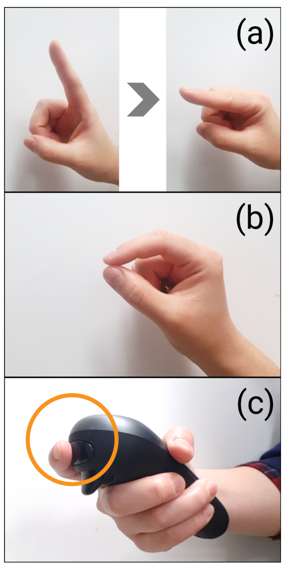

Figure 1: Experimental conditions. (a) Air Tap gesture, (b) Finger Tap gesture, (c) Physical Button.

In addition, the ambiguous input activation of existing gesture interactions is not suitable for the user's mental model. When there is no explicit cue of input activation while performing a mid-air gesture, users naturally try to compensate for the lack. Again with the example of Air Tap gesture, many users tend to bring their thumb and index finger together, even though the input is activated by lowering the index finger. This discrepancy between the user's mental model and the recognition algorithm causes a time-domain inaccuracy to occur when the user attempts an input through AR HMDs.

Validating and improving interactions have always been major issues in Human-Computer Interaction (HCI), but research has only recently started to cover basic interaction techniques for HMD-based environments. Moreover, most of them have focused on how to improve spatial pointing accuracy [2, 4, 8, 9] without considering temporal input accuracy. Latest studies also emphasized the importance of passive haptic feedback on interactions for HMD-based Virtual Reality (VR) [1, 10] and AR [11], regarding spatial accuracy. While Zhang et al. [12] proposed on-skin touch interaction for HMD-base AR/VR, they mainly focused on the touch detection technique.

In this study, we propose that gesture interaction using finger contact can improve time-domain input accuracy in HMD-based AR. By detecting the finger contact–when a finger contacts with another finger or surface on one's own hand–input can be activated at the precise moment. Through a user study, we demonstrate that using finger contact and detecting the moment of finger contact increase temporal input accuracy, user's self-perception of input, and user's preference.

User Study Design

We conducted a within-subject user study to validate that finger contact affects the temporal accuracy of input. In user interfaces, the selection technique consists of three steps: indication of an object, confirmation of selection, and feedback [5]. As we intended to compare confirmation methods in this study, we controlled the indication technique to be head pointing, which is most widely used in HMD-based AR and completely independent of confirmation methods. Visual and auditory feedback to the selection was also kept minimal as possible for all experimental tasks.

Three experimental conditions were set (Figure 1): the existing gesture interaction without finger contact (Air Tap, *AT*), a similar interaction but with finger contact (Finger Tap, *FT*), and using a physical button of a handheld controller (*PB*). For *AT*, we adopted the gesture definition from Microsoft HoloLens [7]: user points index finger upward to make ready posture, and lower the finger to perform input (Figure 1a). In this study, we set the input to be activated when the length of the index finger seen from the egocentric camera of HMD becomes 30% of that of ready posture.

In the Finger Tap gesture, the user also uses the index finger, but instead of lowering the finger, the user bends the finger to touch the thumb (Figure 1b). The input activation is defined by the contact of the two fingers, regardless of finger posture. For *FT* of this experiment, we set the input to be activated when the distance of the two fingertips becomes under 1.0 cm. In *PB*, we used a physical button of a handheld controller that can be pushed with the index finger to maintain three conditions as similar as possible (Figure 1c).

With the experimental conditions defined above, four hypotheses were set.

- Hypothesis 1. Task completion time will decrease in *FT* in comparison to *AT*.
- Hypothesis 2. Temporal accuracy of selection will increase in *FT* in comparison to *AT*.
- Hypothesis 3. User's self-perception on input activation will be more accurate in *FT* in comparison to *AT*.
- Hypothesis 4. User's mental load will decrease in *FT* in comparison to *AT*.

Hypotheses 1-3 were tested with experimental tasks 1-3 each, and hypothesis 4 was tested with a post-condition questionnaire. In all of the experimental tasks, the trial number was adjusted to be moderate so that high fatigue caused by mid-air interaction would not disturb the participants too much. All interface objects were augmented on the distance of 70.0 cm in front of the participant's eyes.

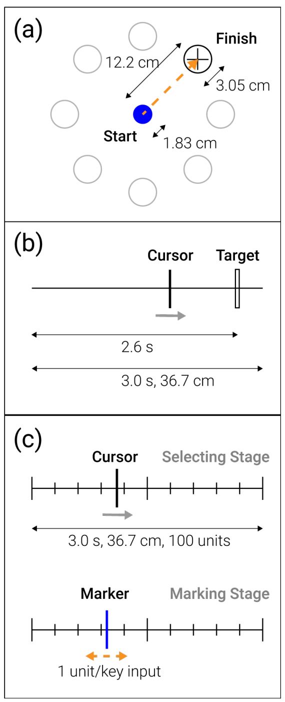

Figure 2: Experimental tasks. The represented graphics are only for explanation and are not in scale.

Task 1 was set to the typical 2D object selection task (Figure 2a). 2D selection task, not 3D, was used since head pointing is a 2D pointing technique around the user's head. After the cursor is positioned on the center, a target object appeared in one of 8 directions chosen in random order. Participants were asked to move the cursor to the target and make the selection gesture as fast and accurately as possible. The target disappeared shortly after the selection, regardless of the cursor's position. If the cursor was not on the target, the trial was marked as failed. Participants completed 24 trials for each condition. Task completion time was measured as a dependent variable. We expected that any time difference incurred will be on account of the duration of each interaction technique, not the pointer movement time since the pointing technique was controlled to be the same. Therefore, point linger time–the duration of time that the cursor remains on the target–was also measured.

The second hypothesis regarding temporal accuracy was tested with the temporal pointing task [6] (Figure 2b). While the cursor moved at a constant speed, participants were asked to make the selection gesture on the most precise moment as they can when the cursor was over the target. In this task, spatial pointing with head orientation was not considered. Participants had to anticipate the moment of overlap from the visual cue and perform the selection gesture accordingly. No visual feedback was given to the input, and each trial ended when the cursor proceeded to the end of the axis after passing the target. If the participant did not make an input until the cursor reached the end, the trial was marked as failed. Participants completed 15 trials for each condition. The time difference between the cursor and the target was measured as a dependent variable.

The third hypothesis about user's self-perception of input was tested with a newly designed marking task (Figure 2c).

Task 3 has two stages: selecting stage and marking stage. In the selecting stage, participants made the selection gesture whenever they wanted while a cursor moved at a constant speed along a scale bar. No visual feedback was given on the input, and head pointing was not considered likewise. In the marking stage, participants were asked to recall the cursor's position on the input as accurately as possible, and move the marker using a keyboard to the estimated position. No trial was marked as failed even if the participant did not make an input until the cursor reached the end; the cursor started over instead. Participants completed 15 trials for each condition. The distance between the cursor and the marker was measured as a dependent variable. This measurement was on the spatial domain because the task objective was to recall the position on the scale bar.

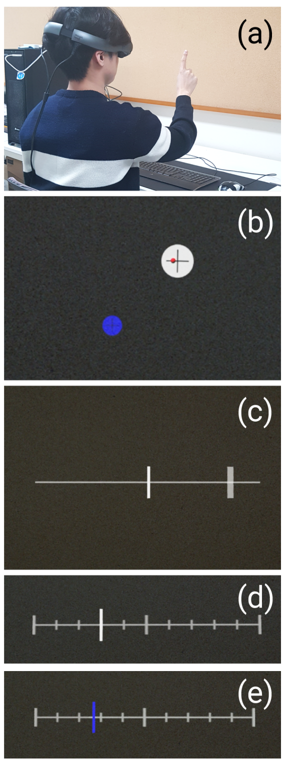

Figure 3: The implemented (a) experimental system and (b-e) tasks.

Participants completed a practice session before the experiment to make themselves familiar with the tasks. Mouse click was used as the selection technique in the practice session, and the mouse was immovable. Participants were also introduced to the gesture interactions beforehand. Participants completed three tasks and answered the raw-TLX questionnaire [3] for each condition. The orders of conditions and tasks were balanced using Latin Square to prevent the learning effect. After the experiment, participants were subject to a semi-structured written interview comparing *AT* and *FT*.

Results

We implemented the experimental system, as shown in Figure 3. We used optical see-through AR HMD with a horizontal field of view of 40°, connected to a desktop PC (Intel Core i7-6700K with NVIDIA GeForce GTX 1080). We did not particularly insist on using HoloLens in the system, because this experiment intended to compare the forms of gestures rather than the exact recognition algorithm, especially when the Air Tap gesture is becoming a standard in HMD-based AR environment going beyond the bounds of HoloLens. We recruited 18 participants (7 female, mean age 26.17 years, SD = 4.20) from our institute, and paid approx. 9 USD for roughly 1-hour experiment. To avoid bias, we sought participants with various levels of experience in HMD-based AR.

Figure 4 shows the summarized results. We excluded failed trials, system errors, and outliers from the analysis. All dependent variables, except raw-TLX score, were found to be non-parametric ($p < 0.05$). A significant difference was found in the task completion time ($\chi^2(2) = 59.406, p = .000$) through Friedman test. However, Wilcoxon signed-rank test with Bonferroni correction applied revealed that only *PB* had significantly shorter task completion time compared to *AT* and *FT*, while *AT* and *FT* had a statistically insignificant difference ($Z = -1.719, p = .086$) (Figure 4a). Thus, hypothesis 1 was rejected. Pointer linger time also showed a significant difference ($\chi^2(2) = 25.374, p = .000$), but the difference of pointer linger time between *AT* and *FT* was statistically insignificant ($Z = -.447, p = .655$) (Figure 4b). From Task 2, on the other hand, the target-cursor time difference showed a significant difference ($\chi^2(2) = 101.769, p = .000$) and *FT* had significantly smaller target-cursor time difference in comparison to *AT* ($Z = -2.566, p = .010$) (Figure 4c). A significant difference was also found in marker-cursor distance from Task 3 ($\chi^2(2) = 126.292, p = .000$), and *FT* had significantly shorter marker-cursor distance in comparison to *AT* ($Z = -3.751, p = .000$) (Figure 4d). Thus, both hypotheses 2 and 3 are accepted. Lastly, there was a statistically significant effect of conditions on the raw-TLX score ($F(2, 38) = 17.535, p = .000$) through repeated measures ANOVA, but it was revealed that only *PB* had a significantly

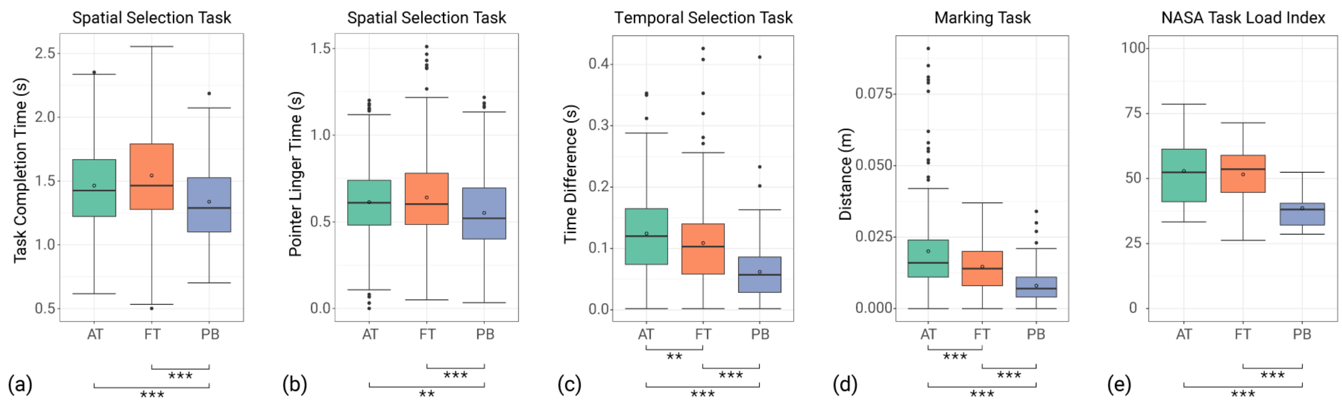

Figure 4: (a) Task completion time, (b) pointer linger time, (c) target-cursor time difference, (d) marker-cursor distance, (e) raw-TLX score.

lower score than *AT* and *FT*, rejecting hypothesis 4 (Figure 4e).

In the post-experimental interview, we asked the participants 4 questions. 17 participants (94.4%) answered that they felt the difference between *AT* and *FT*. Participants commonly mentioned physical load, the impression of accuracy, naturalness to be the difference. 14 participants (77.8%) answered that they preferred *FT* over *AT*. The repetitive keyword was confidence in the input, easiness, small and quick movement, and naturalness. *P16* said, "Finger movement was smaller with Finger Tap, and when my index finger tapped on the thumb, I could feel that I performed the input correctly." *P8* thought, "Finger Tap was a more natural and easy gesture to repeat." Again 14 participants (77.8%) answered that Finger Tap was more comfortable than Air Tap. Many participants mentioned lower physical burden, smaller movement, and familiarity as the reason. *P7* answered, "Air Tap made my wrist hurt, but Finger Tap was easy and light," and *P18* said that "Finger Tap was less demanding." Lastly, 10 participants (55.6%) answered that the input with Finger Tap felt more accurate than that with Air Tap. *P3* answered that, "With Finger Tap, I knew I made an input when fingers met each other, but I had no idea with Air Tap." *P15* mentioned that "the gesture I made was directly visible to me with Finger Tap." *P18* said, "Finger Tap has a smaller gap between fingers, and I think that makes shorter input delay," and *P14* said, "Air Tap had longer movement, making it difficult for me to guess the input moment."

Discussion

As we have demonstrated through the experiment, the finger contact benefits users when input accuracy in the temporal domain is required. Firstly, users can perform input on a moment closer to their intention, as the result of Task 2 revealed. The target-cursor time difference of *FT* was significantly shorter than that of *AT*, proving that *FT* outperforms *AT* in temporal accuracy. A finger contact activates input on the exact moment when fingers meet each other. We believe that this characteristic contributed most to the improvement of temporal accuracy. In the case of *AT*, input activation may be inconsistent, making it difficult for the user to perform input on the exact desired moment.

Secondly, users are more aware of the input when using finger contact, as the result of Task 3 verified. *FT* showed a significantly shorter distance between the cursor and the marker compared to *AT*. That is, the participants recognized their input activation more accurately in *FT*. The passive haptic feedback that naturally occurs would have played a major role, providing an explicit cue of the input and therefore making users more confident about the input. The importance of haptic feedback also has been emphasized in recent researches [1, 10, 11, 12].

Lastly, the finger contact decreases the user's load. Participants repeatedly mentioned that *AT* had a relatively higher physical load and felt unnatural. We also observed multiple participants unconsciously making finger contact even in the *AT* condition. While this behavior did not affect the experimental result since the recognition algorithm of *AT* was independent of finger contacts, it supports our assumption that *FT* will decrease the user's mental load than *AT*.

Gestures using finger contact will benefit not only tasks that require fast and accurate inputs, such as games, but also simple repetitive ones such as typing and multiple selections. In immersive remote collaboration, finger contact will deliver the collaborator's interaction more explicitly. We have demonstrated only one gesture using finger contact in this experiment, but the concept of finger contact is not limited to the presented Finger Tap gesture.

While our experiment revealed the advantages of gestures using finger contact, it also had some limitations. The AR HMD used in the experimental system had a small range of hand tracking. Participants, especially those with little AR experience, had difficulties with keeping the hand in the tracking frustum in both *AT* and *FT* conditions. We suppose that the fatigue from holding an arm up for mid-air gesture affected raw-TLX score more than the difference between two gestures.

In addition, the detection of finger contact relied solely on computer vision technology in the experimental system. To the egocentric camera of the HMD the fingers can easily be occluded, limiting the recognition accuracy of finger contact. A different recognition technique may better display the advantages of gestures using finger contact. On this account, a commercialized wearable sensor such as the smartwatch can be utilized in the following study, even enabling the interaction outside of the tracking range of the camera. Accurate detection of finger contact will also enable a variety of interactions, which was difficult with camera-based systems, for example, the double-tap gesture.

Conclusion

Starting from asserting the need for more accurate interaction techniques in the HMD-based AR environment, this paper verified that finger contact improves time-domain input accuracy through a user experiment. By using finger contact, users can activate input on a more precise moment and perceive better about the moment of input. Post-experimental interview responses also revealed that finger contact improves user experience. Participants felt that it is a more familiar and natural interaction method and were more confident about the exact point in time in which their input was activated. In our future study, we will improve the detection of finger contact with wearable sensors and explore new interaction vocabularies with it.

Acknowledgements
This work was partly supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT (NRF-2017M3C4A7066316), and Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2019-0-01270, WISE AR UI/UX Platform Development for Smartglasses).

REFERENCES

[1] M. Achibet, A. Girard, A. Talvas, M. Marchal, and A. Lécuyer. 2015. Elastic-Arm: Human-scale passive haptic feedback for augmenting interaction and perception in virtual environments. In *2015 IEEE Virtual Reality (VR)*. 63–68. DOI: http://dx.doi.org/10.1109/VR.2015.7223325

[2] A. U. Batmaz, M. D. B. Machuca, D. M. Pham, and W. Stuerzlinger. 2019. 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)*. 585–592. DOI: http://dx.doi.org/10.1109/VR.2019.8797975

[3] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139 – 183. DOI: http://dx.doi.org/10.1016/S0166-4115(08)62386-9

[4] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, Article Paper 81, 14 pages. DOI: http://dx.doi.org/10.1145/3173574.3173655

[5] Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. 2017. *3D User Interfaces: Theory and Practice*. Addison-Wesley Professional.

[6] Byungjoo Lee and Antti Oulasvirta. 2016. Modelling Error Rates in Temporal Pointing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1857–1868. DOI: http://dx.doi.org/10.1145/2858036.2858143

[7] Microsoft. 2019. Gaze and commit - Mixed Reality. (2019). Retrieved Feb 12, 2020 from https://docs.microsoft.com/en-us/windows/mixed-reality/gaze-and-commit

[8] 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 (SUI '17)*. Association for Computing Machinery, New York, NY, USA, 91–98. DOI:http://dx.doi.org/10.1145/3131277.3132182

[9] Ludwig Sidenmark and Hans Gellersen. 2019. 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 (UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1161–1174. DOI: http://dx.doi.org/10.1145/3332165.3347921

[10] C. Wang, C. Hsieh, N. Yu, A. Bianchi, and L. Chan. 2019. HapticSphere: Physical Support To Enable Precision Touch Interaction in Mobile Mixed-Reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 331–339. DOI: http://dx.doi.org/10.1109/VR.2019.8798255

[11] R. Xiao, J. Schwarz, N. Throm, A. D. Wilson, and H. Benko. 2018. MRTouch: Adding Touch Input to

Head-Mounted Mixed Reality. *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (April 2018), 1653–1660. DOI: http://dx.doi.org/10.1109/TVCG.2018.2794222

[12] Yang Zhang, Wolf Kienzle, Yanjun Ma, Shiu S. Ng, Hrvoje Benko, and Chris Harrison. 2019. ActiTouch: Robust Touch Detection for On-Skin AR/VR Interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1151–1159. DOI: http://dx.doi.org/10.1145/3332165.3347869