A Japanese Character Flick-Input Interface for Entering Text in VR

Ryota Takahashi* Shizuka Shirai† Jason Orlosky ‡ Yuki Uranishi § Haruo Takemura ¶

Osaka University

ABSTRACT

This paper presents new flick input interfaces to improve the usability of Japanese character input in VR space. We designed three different interfaces called TouchFlick, EyeFlick, and RoundFlick, which make use of various controller interactions and eye gestures. To investigate the effectiveness of these methods, we compared them with a conventional VR QWERTY keyboard with ray-based selection. We found that TouchFlick was significantly faster and RoundFlick had a significantly lower error rate for experienced users. On the other hand, the input efficiency was the same as that of the conventional method for those who had no experience with flick input. Regarding subjective evaluation, there was no significant difference in usability and mental workload.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction techniques—Text input; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality;

1 INTRODUCTION

In recent years, head-mounted displays (HMDs) have become more affordable as technology has developed. Virtual reality (VR) is expected to be used not only in the entertainment field but in the educational field. However, one significant barrier to adopting these technologies is the speed, comfort, and accuracy of text input. Most platforms such as Oculus Desktop and SteamVR use the input interface with pointing (ray-casting) at a virtual QWERTY keyboard displayed in space. Unfortunately, this input method is inferior to a physical keyboard in terms of input speed and accuracy because even a slight shake of the hand can cause a significant error in pointing accuracy when selecting characters. Previous studies have proposed new methods [3, 6] to address this issue, while new approaches have a high learning curve, and it is hard for novice users.

In this study, we proposed a Japanese input interface based on the flick input system to improve the usability of Japanese input in VR space. The flick input system is widely used in smartphones, tablets, and other systems used by Japanese students [5], so we expected that users could make use of their experiences with these interfaces. However, one of the challenges with building an interface is how to select the consonant portion of a "kana," or Japanese consonant-vowel phonetic character. These characters are used as sources for all other character conversions in Japanese, such as the switch to Chinese Kanji characters or other syllabaries. To examine the appropriate way to select the consonant for the input interface in VR, we created Japanese input interfaces using three different character selection methods: "input using moving controllers (TouchFlick)," "input using gaze (EyeFlick)," and "input completed on a trackpad (RoundFlick)." We investigated the performance of these three methods through user testing.

*e-mail: takahashi.ryota@lab.ime.cmc.osaka-u.ac.jp
†e-mail:shirai@ime.cmc.osaka-u.ac.jp
‡e-mail:orlosky@lab.ime.cmc.osaka-u.ac.jp
§e-mail:uranishi@ime.cmc.osaka-u.ac.jp
¶e-mail:takemura@ime.cmc.osaka-u.ac.jp

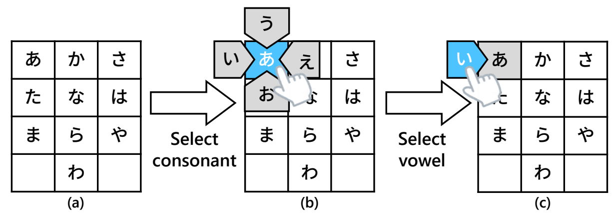


Figure 1: Standard flick input, in which one of the 10 keys from the pad is selected to specify a consonant, and the directional flick (up/down/left/right/center) is used to select the corresponding vowel.

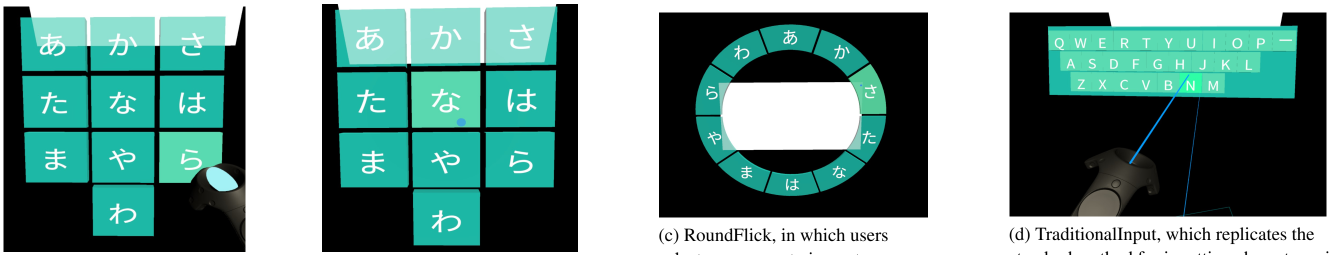
2 PROPOSED FLICK-INPUT METHODS

Flick input is an input method used in Japan for devices with touch panels, such as smartphones. The Flick input interface has 10 Japanese keys that correspond to consonants placed on a software keyboard (the 12-key layout), as shown in Fig. 1. First, users press one of the keys to determine the consonant (Fig. 1 (b)). Then, users flick the key to decide the vowel for the previously selected consonant. The vowels are selected by the position at which the depressed finger is released (Fig. 1 (c)). All of our proposed methods employed this flick method, but key placement and consonant determination methods were different among each method.

TouchFlick uses this 12-key layout and allows users to determine the character by flicking the controller while directly touching the keyboard. This method is the closest operation to the flick input method in a smartphone, so that we expected that users would be able to transfer some of their existing skills to this flick input interface. As shown in Fig. 2 (a), users can select a consonant by touching the key with the controller. Vowel selection follows the same procedure for the flick input method on smartphones. We calculated the size and direction of the vector from the coordinates of the finger touching and leaving the trackpad. Other proposed methods use the same vowel determination method, and we omit the explanation of vowel determination in the following sections.

EyeFlick uses the same key layout and same vowel determination as TouchFlick, but the consonant determination is different. This method uses eye gaze to determine the consonants. EyeFlick interface is shown in Fig. 2 (b). The small blue dot shown over the center character circle represents the line of sight, and the user touches the controller's trackpad to select letters while keeping their line of sight aligned with the key. We designed the system so that the selection of the current consonant does not change even if the user moves his or her eyes while touching the trackpad, thus improving the stability of input. The user can hold the controller in any position, so that this method was expected to reduce hand fatigue during long sentence input.

RoundFlick has original key placement. The keys for consonants are arranged in a circle around the input field, as shown in Fig. 2 (c). In this method, the input interface and the input field exist in the field of view simultaneously, so the user does not need to move their head significantly. The method uses two controllers: one is to



(a) Touchflick, in which users touch a target key to select a consonant.

(b) EyeFlick, in which users gaze at a key to select a consonant.

(c) RoundFlick, in which users select a consonant via a rotary gesture along the edge of the touchpad.

(d) TraditionalInput, which replicates the standard method for inputting characters via a QWERTY keyboard and controller rays.

Figure 2: Appearances of the three interfaces we designed, including (a) TouchFlick, (b) EyeFlick, and (c) RoundFlick, and the standard method for entering text in VR (d), referred to as TraditionalInput.

determine consonants and the other for vowels. The user selects the consonant based on the finger position on the controller's trackpad in one hand. After determining the consonant, the character is input by flicking with the other controller. Although EyeFlick does not cause arm fatigue, it does cause eye and neck fatigue due to the movement of the eyes. In contrast, this input is performed only with the controller, so the degree of physical fatigue is likely to be the lowest.

3 EXPERIMENTS

To investigate the effectiveness of the three proposed methods, we measured their performance and compared them, as well as the conventional QWERTY keyboard input method. We used the VIVE Pro Eye with the gaze tracking function and a dedicated controller as the VR environment. There is a trackpad in the center of the controller, which can read the position of a finger touch. Unity 2019.4.11f1 was used to build the virtual environment. As a conventional input method using QWERTY arrays, we created the interface shown in Fig. 2 (d) (hereafter referred to as TraditionalInput). The size of the entire keyboard was the same as that of the proposed methods to match the general conditions of practical use.

The experiment was conducted on 24 undergraduate and graduate students between the ages of 18 and 25 ($Mean$: 21.3, SD: 1.95) because the interfaces are targeted at students and individuals with flick-input experience. All participants could fluently speak and write Japanese. To investigate the participants' experiences with HMDs, Roman character input with QWERTY keyboard, and Flick input, we used a 5-point rating scale from 1 (strongly disagree) to 5 (strongly agree). Table 1 shows the result. We counted 1 and 2 as "no experience" and 3 to 5 as "experience" and used them for analysis.

We used a within-subjects design and all participants input test with all methods. For the dataset of the questions, we used the typing practice in the P-ken-ICT Proficiency Test [2], administered by the Benesse Corporation. An example screenshot of the experiment viewed through the HMD is shown in Fig. 3. In the center of the screen, the sentence to be input is displayed. On the bottom, the Japanese phonetic character, or *hiragana*, notation of the sentence, and the current input characters are displayed. We measured the number of input characters achieved and the error rate over five minutes to evaluate the performance. The number of input characters per minute (CPM) was calculated by dividing the number of input characters entered in the 5 minutes by 5. The error rate is calculated for the entire 5 minutes. Moreover, participants took two questionnaires about mental workload using the Japanese simplified versions of the NASA Task Load Index (NASA-TLX) [4] and usability using the System Usability Scale (SUS) [1] after each input test with methods.

For each experimental condition, one set of tasks consisted of 3 minutes of practice, 5 minutes of input task, 5 minutes of rest, and a questionnaire response. The participant repeated this process

Table 1: Breakdown of experience with HMDs, Roman character input, and flick input.

	HMD	Roman character input	Flick input
Experience	5	21	19
No Experience	19	3	5

(a) Flick input (b) Roman character input

Figure 3: Screenshots of the text input tasks from the experiment.

with all four methods. In addition, participants performed the four methods in different orders using counter-balanced to alleviate ordering effects. We explained the input method verbally during each practice session, and participants were allowed to practice inputting freely. For the practice phrases, we used other sentences extracted from the same dataset used for the experiment, taking into account the balance of complexity and long sounds. TouchFlick can have two-handed input, so we told them that they could input with both hands if desired. In addition, we calibrated the eye gaze before the EyeFlick experiment. We did not use the sentences used for the practice in the experimental trials. After practice, the participant can start the input experiment at any time. After 5 minutes, the input experiment automatically ended, and the participant took a break. During the break, they removed the HMD and answered the subjective evaluation questionnaire.

4 RESULTS

The average CPM of each input method is as follows: TraditionalInput was 36.76 characters (SD: 8.28), TouchFlick was 42.46 characters (SD: 17.51), EyeFlick was 39.10 characters (SD: 12.16), and RoundFlick was 31.28 characters (SD: 8.19). We conducted a three-way ANOVA with factors "input method", "HMD-using experience", and "Flick-input experience". As a result, there was a main effect of "Flick-input experience" ($F(1,20) = 5.10, p < .05$), and a significant interaction between "input method" and "Flick-input experience" ($F(2.27, 45.3) = 3.73, p < .05$). Thus, we tested the simple main effect and found a significant difference in "Flick-input experience" among the three proposed methods (TouchFlick: $F(1,20) = 10.68, p < .01$; EyeFlick: $F(1,20) =$

$6.06, p < .05$; RoundFlick: $F(1,20) = 5.49, p < .05$). In addition, there was a simple main effect of "input method" among those with Flick-input experience was significant ($F(3,20) = 12.14, p < .001$). Using Bonferroni correction, multiple comparisons showed that there was a significant difference ($p < .05$) between TraditionalInput and RoundFlick for those who had no experience with flick input, and significant differences between TraditionalInput and TouchFlick ($p < .05$), TouchFlick and RoundFlick ($p < .01$), and EyeFlick and RoundFlick ($p < .01$), respectively, for those who had experience with flick input.

The average error rates for each input method were as follows. TraditionalInput was 24.36% (*SD*: 13.27%), TouchFlick was 24.55% (*SD*: 9.02%), EyeFlick was 20.14% (*SD*: 7.63%), and RoundFlick was 14.21% (*SD*: 6.43%). A three-way ANOVA was performed on the error rate for each method, including "input method," "HMD-using experience," and "Flick-input experience". The result of ANOVA reveals that a significant difference in "input method" ($F(2.44, 48.83) = 4.51, p < .05$) . Multiple comparisons using Bonferroni correction showed that there was a significant difference between TraditionalInput and RoundFlick ($p < .05$), TouchFlick and RoundFlick ($p < .001$), and EyeFlick and RoundFlick ($p < .05$).

The results of the NASA-TLX score were as follows. TraditionalInput was 5.97 (*SD*: 1.29), TouchFlick was 5.64 (*SD*: 1.61), EyeFlick was 5.96 (*SD*: 1.60), and RoundFlick was 5.36 (*SD*: 1.41). We conducted a three-way ANOVA on the NASA-TLX scores with factors "input method,", "experience of using HMD," and "experience of using flick input". As a result, a significant difference ($F(1,20) = 4.61, p < .05$) was found for "Experience of using HMD". There were no significant differences in other factors or interactions.

The results of SUS were as follows. TraditionalInput was 57.71 (*SD*: 16.50), TouchFlick was 64.38 (*SD*: 18.19), EyeFlick was 50.31 (*SD*: 19.27), and RoundFlick was 58.02 (*SD*: 16.14). We run a three-way ANOVA with factors "input method," "HMD-using," and "Flick-input". There were no significant main effects or interaction effects.

5 Conclusions

In this study, we proposed three types of text input interface in VR space that can be used for the flick-based text input used in smartphones: "Input using moving controllers (TouchFlick)," "Input using gaze (EyeFlick)," and "Input completed via trackpad (RoundFlick)".

As a result of our experiments, we found that users who had experience with flick input could input text with 47.73 CPM using TouchFlick, which was 10 CPM faster than the conventional method. Furthermore, the rate of input errors for RoundFlick was 14.21%, which is more than 10% lower than that of the conventional method, suggesting that RoundFlick might become the most efficient input method as users become accustomed to it. EyeFlick was not significantly different from the conventional method in terms of input speed and error rate, but the results may differ depending on the task.

Mental workload and subjective satisfaction with each input method were also investigated, but no significant difference was found between the input methods, meaning that the subjective evaluation was similar for all of them.

In the future, it will be necessary to examine the evaluations for different tasks, such as correcting input errors or transcribing speech are assigned. In this experiment, the evaluation was conducted by copying the characters displayed on the screen, but it will be necessary to conduct an evaluation based on a situation in which the user is generating words or new sentences.

Another issue is the construction of a language input interface other than Japanese. Non-native speakers of Japanese may not be familiar with flick input, and they may not be able to utilize the advantage of 5-direction flick input since it is unique to the separation of consonants and vowels in the Japanese Kana system.

Acknowledgments

This work was funded in part by the Office of Naval Research Global, grant N62909-18-1-2036. We would like to thank Benesse Holdings, Inc. for providing the typing dataset used in the experiment.

References

[1] J. Brooke. Sus: A quick and dirty usability scale. *Usability evaluation in industry*, p. 189, 1996.

[2] B. Corporation. P-ken-ict proficiency test association. `https://www.pken.com/`. (Accessed on 01/15/2021).

[3] J. Fashimpaur, K. Kin, and M. Longest. Pinchtype: Text entry for virtual and augmented reality using comfortable thumb to fingertip pinches. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, pp. 1–7. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3334480.3382888

[4] S. Miyake and M. Kumashiro. Subjective mental workload assessment technique-an introduction to nasa-tlx and swat and a proposal of simple scoring methods. *The Japanese journal of ergonomics*, 29(6):399–408, 1993. doi: 10.5100/jje.29.399

[5] N. Nagasawa. How japanese university students type on smartphone and pc. *Computer and Education*, 43:67–72, 2017. doi: 10.14949/konpyutariyoukyouiku.43.67

[6] N. Yanagihara, B. Shizuki, and S. Takahashi. A text input method using 3d strokes. In *IPSJ SIG Technical Reports*, vol. 2019-HCI-181, pp. 1–6, Jan. 2019.