DepthText: Leveraging Head Movements towards the Depth Dimension for Hands-free Text Entry in Mobile Virtual Reality Systems

Xueshi Lu, Difeng Yu, Hai-Ning Liang*, Xiyu Feng, Wenge Xu

Xi'an Jiaotong-Liverpool University

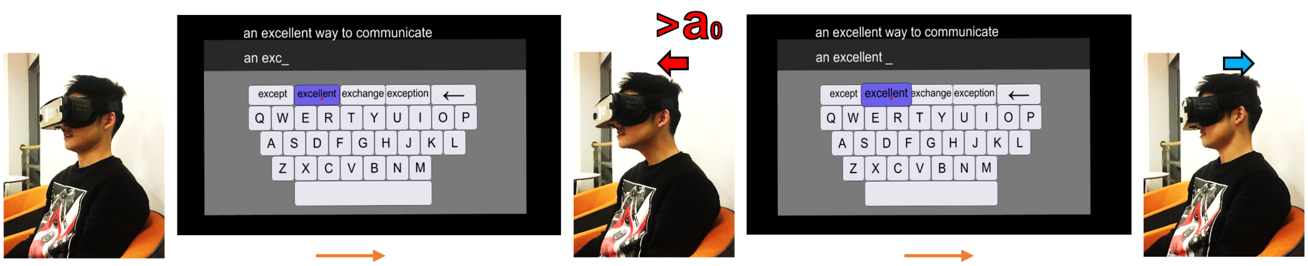

Figure 1: A user is entering some text using DepthText using a mobile virtual reality (VR) system. The user has entered the word "excellent" by performing short forward movements towards the depth dimension (z-axis) with an acceleration speed larger than a_0.

Abstract

Text entry is a common activity in virtual reality (VR) systems. However, there is a limited number of approaches available for mobile VR systems, where it might be inconvenient for users to carry an input device. We propose a novel hands-free text entry technique we call DepthText which leverages the acceleration sensing abilities of built-in IMU sensors of mobile VR systems. Users are able to enter text by moving their head forward. The results of a 5-day study indicate that users can achieve an average of 10.76 words per minute (wpm) on the last day with low errors. This performance is comparable to the dwell-based technique which is the most common way of entering text that is hands-free. One advantage of DepthText over the dwell-based technique is that users can have more control of the pace of selecting characters, rather than being pushed by a pre-set dwell time.

1 Introduction

Text entry is essential for virtual reality (VR) systems. For example, users often need to text short messages while playing a multiplayer game to talk with other players. In this case, a user's hands are occupied using the controllers. In addition, when a user is watching a video in VR at home or outside in a coffee shop or library, locating an input device to enter text may not be easy. In these scenarios, an efficient text entry technique that is hands-free and does not require any auxiliary handheld devices or trackers, could be very useful. Such a technique can be especially handy for mobile VR which enables users to perceive and interact with virtual environments anytime and anywhere [2]. Current solutions such as gaze tracking might not work for inexpensive mobile VR systems, and speech-based techniques do not work well in noisy environments and also have privacy issues in public places.

One solution to achieve cost-free text entry is to leverage the head motions that can be captured by the built-in IMU sensors of current mobile VR. The dwell-based technique is the most common ways that allow users to enter text by controlling a virtual pointer using head motions and let it hover on a target for a certain period of time (dwell time) to make a selection. However, determining the length of the dwell time can be problematic—a long dwell time decreases performance but shorter dwell times cause false positive selections and errors [4]. Moreover, because of the pre-set dwell time, users are always "pushed" to select a target key and quickly move to the next one. To perform well, the user needs to be very focused and act carefully to avoid selecting unwanted keys—in addition to being error-prone it could potentially lead to fatigue and increased motion sickness. When typing, it is usual that users often need to think for a while about the text and the characters they want to enter—these mental breaks are not likely supported by dwell-based techniques. In short, text entry techniques using dwell-time may not provide the most optimal experience to users.

This paper presents DepthText, a hands-free text entry technique in mobile VR systems that allows users to enter text in a controlled manner using the forward/backward movements of users' head towards the depth dimension (see Figure 1 above). We conducted a 5-day user study to evaluate the performance of DepthText. The results indicated that the users achieved an average of 10.76 words per minute (wpm) in the last day with low errors. This performance is comparable to the dwell-based technique with similar settings but without its disadvantages. In addition, user responses to questionnaires showed that participants would be willing to use DepthText in private spaces and in front of people familiar to them.

2 DepthText

DepthText is a hands-free technique that supports text entry by leveraging the forward (z-axis) acceleration movements that can be accurately captured by the IMU sensor of current mobile phones. To enter a character, a user will first move the virtual pointer to over the target key. The user then triggers the selection by performing a simple short forward motion. In this process, the IMU sensor first senses an acceleration towards the z-axis. Once the maximum forward acceleration speed is larger than the threshold a_0, the target character is selected. At the same time, a lock is set to avoid multiple unwanted entries. After, the IMU sensor senses an acceleration towards the negative z-axis (since the user slows down the forward movement). This allows the program to release the lock to allow selecting the next character. Note that we compensate gravitational acceleration sensed by the IMU sensor and can obtain the pure z-axis acceleration speed of the user (see Figure 2). The keyboard layout follows the work [6] to allow for a potential comparasion. We also have employed a statistical decoder by using a similar procedure as reported in [6].

* Corresponding author who can be contacted via haining.liang@xjtlu.edu.cn

2019 IEEE Conference on Virtual Reality and 3D User Interfaces
23-27 March, Osaka, Japan
978-1-7281-1377-7/19/$31.00 ©2019 IEEE

In our pilot study, we noticed that, during the process of searching the target (head orientation), there could be some cases where the acceleration speed would reach the threshold a_0. To prevent false selection, we added an extra condition that, if the IMU sensor detects fast movement along the y-axis (we set as 0.003m/s^2 acceleration speed), even the z-axis acceleration speed reaches a_0, the character would not be entered.

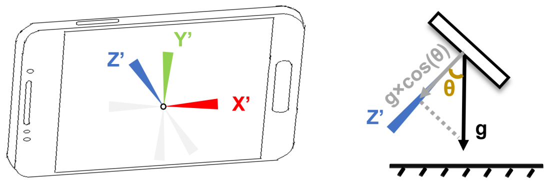

Figure 2: (Left) 3-axis of the IMU sensor equipped in a smartphone; (Right) gravity compensation of the IMU sensor.

3 User Study

We conducted a 5-day user study to evaluate the long-term usage of DepthText. The study aimed to measure user performance (in terms of speed and accuracy), assess the users' affective responses, and determine the level of simulator sickness.

Twelve unpaid participants (8 males and 4 females) were recruited from a local university campus. The Gear VR with a Samsung Galaxy S6 Edge+ G9280 mobile phone was used to immerse the user into the virtual world.

Before the experiment began, each participant was briefed with the details of DepthText and the VR devices they would use. They were instructed to perform the task as quickly and accurately as possible and were allowed to sit in their comfortable positions. For each experimental session, the participants would type 12 phrases, which were randomly generated from MacKenzie phrase set [3]. The 12 phrases were further divided into 3 blocks, with 4 phrases each. At the beginning of each block, participants were asked to try 5 different acceleration speeds and picked the one with which they could perform the fastest and most accurate (i.e., to allow some degree of personalization similar to [4]). The five different acceleration speeds (0.0005m/s^2, 0.001m/s^2, 0.005m/s^2, 0.01m/s^2, 0.02m/s^2) were determined from a pilot study with 4 participants. The text of the five acceleration speeds was not visible to participants, so they chose the speeds only based on their practices. After the experiment, participants completed several post-experiment questionnaires.

3.1 Results and Discussion

The results were based on 10 participants. One female participant was dropped from the final dataset because of the abnormal performance (with an average error of nearly 30%). One male participant quit the experiment after the first session. Therefore, the results were analyzed based on 5 (days) × 12 (sessions) × 10 (participants) = 600 timed trials. The evaluation matrics include word per minutes (WPM), the total error rate (TER), and the not corrected error rate (NCER) [5, 6].

Figure 3 (left) shows that the median text entry speed increased steadily across the five days. The average speed of DepthText for completely novice users in the first 12 trials was 6.76 wpm (s.e. = 0.23). The speed was then increased to an average of 10.76 wpm (s.e. = 0.23) in the last 12 trials after 48 trials of practices. With a similar setting, the text entry speed of DepthText is comparable to the dwell-based technique reported in [6] which has an average of 10.59 wpm in the last 8 trials after 40 trials of training. The repeated meausres ANOVA with Greenhouse-Geisser adjustment yielded a significant effect of Day on text entry speed ($F_{2.13, 19.18} = 24.426$, $p = 4\times 10^{-6}$). Pair-wise comparisons with Bonferroni corrections revealed significant differences between Day 1 and Day 2 ($p = 0.005$), Day 2 and Day 3 ($p = 0.011$), but not among Day 3, Day 4, and Day 5 ($p = 1$). This suggested that the learning curve became more stable from Day 3.

The mean NCER could be considered as low across the five days (from the lowest 1.96% to the highest 3.11%). The mean TER dropped from the first day of 16.34% (s.e. = 1.29%) to the fifth day of 7.79% (s.e. = 0.62%), with an overall decrease of 51.8% (see Figure 3 right). After some training, although the TER decreased a great deal because most participants seemed to get the knack of using DepthText, it was still a bit high (7.79%). Our observations pointed to two reasons. First, when the head turns, there could be some variances in the acceleration speed in the z-axis which triggered some unwanted selection. Second, a backward movement in a relatively fast speed could also lead to a high acceleration speed towards the z-axis which also induced some wrong selections. Future work could explore algorithms to better classify forward and backward movements.

The social acceptability questionnaire [1] demonstrated that most participants were willing to use DepthText alone (90%) and in their home (100%). More than half of them would accept using DepthText in front of the people who were familiar to them (families, friends, and partner). They usually would not prefer to use the technique in public areas or around strangers.

Overall, DepthText has comparable performance as dwell-based techniques [6], but give users more control of the text entry process.

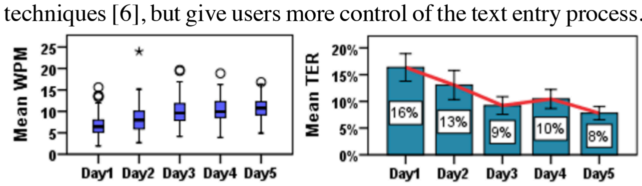

Figure 3: (Left) the boxplot of text entry speed across five days; and (Right) the mean total error rate (TER) across five days.

4 Conclusion

This paper presents DepthText, a hands-free text entry technique for mobile virtual reality (VR) systems. The technique leverages the built-in IMU sensors of current mobile VR devices and as such has no costs associated with external trackers and input devices. A user is able to enter characters through short bursts of forward head movements. According to our 5-day user study, this technique is able to support users in achieving an average of 10.59 words per minute (WPM). Responses to questionnaires also show that the technique is acceptable to users and they are willing to use it when they are alone or before people familiar to them. All in all, DepthText represents a viable alternative hands-free text entry approach to dwell-based techniques and offer users more control over the pace of selecting the characters.

References

[1] D. Ahlström, K. Hasan, and P. Irani. "Are you comfortable doing that?: acceptance studies of around-device gestures in and for public settings." In *Proc. MobileHCI'14*, pp. 193-202. ACM, 2014.

[2] J. Gugenheimer, E. Stemasov, H. Sareen, and E. Rukzio. FaceDisplay: enabling multi-user interaction for mobile virtual reality. In *CHI EA'17*, pp. 369-372. ACM, 2017.

[3] I. S. MacKenzie and R. W. Soukoreff. Phrase sets for evaluating text entry techniques. In *CHI EA'03*, pp. 754-755. ACM, 2003.

[4] M. E. Mott, S. Williams, J. O. Wobbrock, and M. R. Morris. Improving dwell-based gaze typing with dynamic, cascading dwell times. In *Proc. CHI'17*, pp. 2558-2570. ACM, 2017.

[5] R. W. Soukoreff and I. S. MacKenzie. Metrics for text entry research: an evaluation of MSD and KSPC, and a new unified error metric. In *Proc. CHI'03*, pp. 113-120. ACM, 2003.

[6] C. Yu, Y. Gu, Z. Yang, X. Yi, H. Luo, and Y. Shi. Tap, dwell or gesture?: Exploring head-based text entry techniques for hmds. In *Proc. CHI'17*, pp. 4479-4488. ACM, 2017.