Decoder-Assisted Typing using an HMD and a Physical Keyboard

James Walker
Michigan Technological University
Houghton, MI, USA
jwwalker@mtu.edu

Scott Kuhl
Michigan Technological University
Houghton, MI, USA
kuhl@mtu.edu

Keith Vertanen
Michigan Technological University
Houghton, MI, USA
vertanen@mtu.edu

CHI'16 Workshop on Inviscid Text Entry and Beyond, 8 May 2016, San Jose, CA.

Abstract
In recent years, head-mounted displays (HMDs) have become increasingly affordable and popular. HMDs pose new challenges for text entry because they typically immerse people in a virtual world while preventing them from seeing the real world. As HMDs become widespread, we expect that some people will want to perform traditional computing tasks with these new devices, such as writing emails or browsing the web. This work explores new ways to support efficient text entry in HMDs by leveraging decoding techniques and ubiquitous physical keyboards. We describe an experiment where we measured the performance of our system on touch-typists who typed a series of short messages on a physical keyboard in a (1) keyboard visible condition, (2) occluded keyboard condition, and (3) head-mounted display condition. The results of our pilot study showed that users' speed and accuracy was considerably worse in the occluded and HMD conditions, that our decoder was able to correct a large number of errors, and that increasing compute time for the decoder increased the number of corrected errors.

Author Keywords
Text entry; text input; mobile interaction

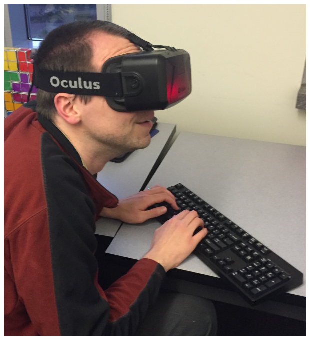

Figure 1: User entering text on a desktop keyboard while wearing a head mounted display (HMD).

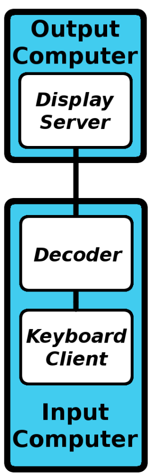

Figure 2: Main components of our system.

ACM Classification Keywords

H.5.2 [Information interfaces and presentation]: Input devices and strategies

Introduction

With the release of affordable head-mounted displays (HMDs), people will find new applications made possible by this technology. Although current trends suggest that HMDs will initially be used for entertainment applications, we believe that people will also wish to complete everyday tasks such as writing emails, documents, and short messages. Since QWERTY mechanical keyboards are also prevalent, this research aims to research ways to overcome problems introduced by using keyboards with an HMD. For example, HMDs typically prevent a user from seeing the real world including the location of the keyboard and their hands. Therefore, a key challenge of this work is to find ways to support efficient text entry under these circumstances. We addressed this challenge by adapting a state-of-the-art touchscreen decoder to correct errors in user input.

Related Work

There is little published research which implements and compares different text-entry systems for HMDs. A 2002 study by Bowman et al. [1] compared several different techniques including a one-hand chord keyboard, speech recognition (implemented with a person instead of software) and a virtual keyboard controlled with a handheld tablet and pen. They found that none of their approaches produced high levels of performance or usability. A more recent study in 2015 by McGill et al. [2] provides evidence that use of a mechanical keyboard with feedback does provide good performance. However, this study focused primarily on providing visual feedback in the form of a video of the user. In the present work, we focus on using a decoder to correct errors.

The current work builds on the VelociTap touchscreen decoder [4]. VelociTap takes a sequence of noisy touch locations and searches for the most probable sentence given those touches. The decoder has also been used to to research eyes-free touchscreen input [5].

System Implementation

Our prototype system consists of three components:

1. A keyboard client that is responsible for randomly selecting phrases to be typed, logging users' keystrokes, and sending that timestamped data to the decoder.
2. The decoding server which forwards phrases and keystrokes to the display server, computes WPM and accuracy, and runs the decoder.
3. A display server that accepts phrases and keystrokes and displays them to the user.

Each component is capable of being run on separate hardware or on the same piece of hardware. For our prototype system, the keyboard client and decoder were run on one computer while the display server ran on another. Each software component communicates with the others using TCP with Nagle's algorithm disabled in order to prevent transmission delays.

Recognition Details

To perform automatic correction of users' noisy keyboard typing, we modified the VelociTap touchscreen decoder [4]. Normally VelociTap searches for the most probable sentence given a time-ordered sequence of x- and y-locations recorded by a touchscreen sensor. VelociTap's keyboard model uses two-dimensional Gaussians centered on each key of the onscreen virtual keyboard. In this work, we don't have a virtual keyboard, we have a real one. We measured a physical keyboard and created a keyboard map based on physical size and locations measured in millimeters.

During typing each key down event is mapped to the center x- and y-position of that key on the physical keyboard. We then simulate a touchscreen tap on this center position. This allow VelociTap to create a probability distribution over all possible keys with keys closer to the key actually pressed having higher probability.

In addition to the keyboard model, VelociTap also uses a letter and word language model. We trained a 12-gram letter language model (2.2 GB on disk) and a 4-gram word language model (3.8 GB on disk) on billions of words of data from Twitter, social media, blog, and movie subtitles.

Experimental Design

For our pilot study, we had five participants in three within-subject conditions:

- **VISIBLE** - Participants typed using a fully visible keyboard and a desktop monitor.
- **OCCLUDED** - Participants typed using a keyboard which is occluded from the user's vision and a desktop monitor. The keyboard was occluded by placing it underneath a cardboard container which had enough space inside to hold the keyboard and the participant's hands comfortably, while simultaneously preventing the participant from being able to see the keyboard while typing.
- **HMD** - Participants typed using a keyboard while wearing an Oculus Rift DK2 head-mounted display.

In each condition, participants were shown 30 short memorable phrases taken from the Enron mobile data set [3]. Participants received a random selection of sentences from a set of 189 phrases. Participants only saw a given phrase once. Participant were instructed to proceed “quickly and accurately”. Participants first completed a practice session with five phrases using a fully visible keyboard and desktop monitor. After this practice session, participants completed the conditions in random order. For the HMD condition, participants first adjusted the HMD for comfort. After the final condition, participants completed a short questionnaire. Participants were paid $10.

Condition	Entry rate (wpm)	Before CER (%)	After CER (%)
VISIBLE	57.5 ± 18.2	1.8 ± 2.0	1.4 ± 1.7
OCCLUDED	47.7 ± 26.6	6.8 ± 7.3	3.0 ± 2.8
HMD	39.0 ± 14.8	7.0 ± 5.5	3.5 ± 1.6

Table 1: Mean and standard deviation of participant entry rates and character error rates (before and after automatic correction).

Results

We report error rate before and after automatic correction using character error rate (CER). CER is the number of character insertions, substitutions, and deletions required to transform the entered text into the reference text divided by the number of characters in the reference (times 100).

We report entry rate in words-per-minute (wpm) with a word being defined as five characters including space. Our entry rate includes the time required for the long-press to finish a sentence, the time to send the keystroke data to the server, the time to perform the recognition, and the time to display the result.

As shown in Table 1, the fastest and most accurate entry was obtained in the VISIBLE condition. While all our participants reported to be touch typists, entry rate slowed and error rate rose substantially once the keyboard was occluded. The complete occlusion of vision by wearing an HMD even further impacted speed and accuracy. Comparing the before and after error rates, we see that our sentence-based decoding approach successfully corrected about half the errors in the OCCLUDED and HMD conditions.

We were curious if even more accurate results might be possible if more compute time was spent on decoding. We pooled all the HMD data from the participants and ran offline recognition experiments. VelociTap has a beam parameter controlling its tradeoff between speed and accuracy. With the beam used in the study, the average sentence CER was 7.77% (0.04 s per decode). Doubling the beam substantially lowered the CER to 6.30% (0.26 s per decode). Tripling the beam offered only a small additional reduction of CER to 6.27% (1.63 s per decode).

Based on our past experience with VelociTap and touchscreen typing data, such accuracy improvements based on additional compute is not typical. We conjecture for physical keyboards there are common error types (e.g. transpositions) that are not explicitly modeled by VelociTap. This suggests adding additional features to the decoder may offer improved accuracy without widening the search beam.

Future Work

The eventual goal of this work is to continue to explore users' typing habits in virtual environments, decoding input from a physical keyboard, and improving our decoder's ability to aid users' typing with physical keyboards and while using HMDs. In addition to the display server being configured to display immersive virtual environments, our system is integrated with a Vicon tracker which is capable of tracking the positions and orientations of objects (such as keyboards) in physical space, features which will be put to use in future studies.

Conclusions

Tentatively, our results suggest that even experienced touch-typists may experience reduced speed and accuracy when typing blind, and even more so when using an HMD. We also have preliminary evidence suggesting that our decoder can significantly improve users' typing capabilities under these conditions. This suggests that using our techniques to improve users' typing ability with physical keyboards and while wearing HMDs is a fruitful area of study.

Author biographies

James Walker received his M.S. in computer science from Michigan Tech in 2013. For his ongoing research in pursuit of his Ph.D, he is currently studying ways to improve users' ability to interact with physical input devices such as keyboards while immersed in virtual environments using head-mounted displays.

Scott Kuhl is an Associate Professor of Computer Science at Michigan Technological University. He received his PhD from the University of Utah in 2009. His research interests include head-mounted displays, human perception, and computer graphics.

Keith Vertanen is an Assistant Professor at Michigan Technological University. He specializes in designing intelligent interactive systems that leverage uncertain input technologies including input via speech, touch, and eye-gaze.

Workshop demo

We plan to demonstrate the system described in this paper at the workshop, including the decoding software and the head-mounted display.

References

[1] Doug A. Bowman, Christopher J. Rhoton, and Marcio S. Pinho. 2002. Text Input Techniques for Immersive Virtual Environments: An Empirical Comparison. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 46, 26 (2002), 2154–2158. DOI: http://dx.doi.org/10.1177/154193120204602611

[2] Mark McGill, Daniel Boland, Roderick Murray-Smith, and Stephen Brewster. 2015. A Dose of Reality: Overcoming Usability Challenges in VR Head-Mounted Displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2143–2152. DOI: http://dx.doi.org/10.1145/2702123.2702382

[3] Keith Vertanen and Per Ola Kristensson. 2011. A Versatile Dataset for Text Entry Evaluations Based on Genuine Mobile Emails. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '11)*. ACM, New York, NY, USA, 295–298. DOI: http://dx.doi.org/10.1145/2037373.2037418

[4] Keith Vertanen, Haythem Memmi, Justin Emge, Shyam Reyal, and Per Ola Kristensson. 2015. VelociTap: Investigating Fast Mobile Text Entry Using Sentence-Based Decoding of Touchscreen Keyboard Input. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 659–668. DOI: http://dx.doi.org/10.1145/2702123.2702135

[5] Keith Vertanen, Haythem Memmi, and Per Ola Kristensson. 2013. The Feasibility of Eyes-free Touchscreen Keyboard Typing. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '13)*. ACM, New York, NY, USA, Article 69, 2 pages. DOI: http://dx.doi.org/10.1145/2513383.2513399