

Authors Biography

Junhyeok Kim is a graduate student in the HCI Lab at the University of Manitoba. His interest resides in input system for ubiquitous analytics. He is currently starting a PhD on text entry systems for ubiquitous computing.

Dr. William Delamare is a postdoc fellow in the HCI Lab at the University of Manitoba. His interests concern distant interactions offered by ubiquitous input devices.

Dr. Yumiko Sakamoto is an experimental psychologist who specializes in socio cognitive research. Currently, she is a professor at George Brown College, in Toronto, and her research interests center around interaction and perception.

Toward a Pool of Text-Entry Input Techniques

Junhyeok Kim

University of Manitoba
Winnipeg, MB, Canada
Kimj3415@cs.umanitoba.ca

William Delamare

University of Manitoba
Winnipeg, MB, Canada
delamarw@cs.umanitoba.ca

Yumiko Sakamoto

Psychology & English Literature
School of Liberal Arts and Sciences
George Brown College
Yumiko.Sakamoto@georgebrown.ca

Tony Havelka

Tek Gear, Inc.
Winnipeg, MB, Canada
tonyh@tekgear.com

Pourang Irani

University of Manitoba
Winnipeg, MB, Canada
irani@cs.umanitoba.ca

Abstract

Several text-entry techniques already exist, be it as research prototypes or commercial products. The number of text-entry input techniques will likely continue to increase in order to explore and leverage ubiquitous text-entry tasks with new technologies such as incoming wearable devices. In this era of increasing input techniques, we propose the creation of a pool of techniques, evaluated within a common experimental protocol, so that future work can (1) have a standardized set of baselines, and (2) easily contribute to the expansion of this pool of techniques. The contributions of this work are threefold. First, this work will unify the evaluation of text-entry techniques, which are difficult to compare in the current state of how research is done. Second, this work will set a standard regarding how future text-entry techniques will be evaluated. Third, future text-entry input techniques can be added to the pool of existing techniques.

Author Keywords

Text entry; Evaluation method; Experimental protocol.

ACM Classification Keywords

H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces;

Paste the appropriate copyright/license statement here. ACM now supports three different publication options:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single-spaced in Verdana 7 point font. Please do not change the size of this text box.

Each submission will be assigned a unique DOI string to be included here.

Authors Biography

Tony Havelka is President of Tek Gear, Inc. He specializes in body worn hardware for VR, AR and Wearable applications. His work focuses on creating I/O devices – such as the Twiddler – that make the human-computer interface more intuitive, interactive and immersive.

Pourang Irani, is Professor in Computer Science and Canada Research Chair in Ubiquitous Analytics. His work on ubiquitous text-entry has explored using tilt and pressure sensors for input.

Introduction

More than 20 years of research on text-entry exist [8]. It is already difficult to compare existing works due to divergences of experimental protocols used for the evaluations, or because of divergences of reported dependent measures [1,2,8]. In addition, the number of input-capable devices is increasing due to the emergence of new computing hardware – such as wearable devices (e.g. ring device [3]). Thus, as technology is pervading our everyday lives, there is a growing need to adapt existing and create new text-entry input techniques capable of the text-entry task to these new devices and method of use [5]. These future techniques will continue to add divergences making difficult to compare and draw conclusions between all these works.

We propose to fill the gap between existing techniques evaluations. More specifically, we propose an iterative approach in order to create a pool of techniques following a standardized evaluation (Figure 1). To do this, we first focus on text-entry direct input techniques specifically designed for small touch surfaces [4,6,7,9,10]. Indeed, we believe that the current trends are using smaller and smaller input touch-capable devices. The expansion of this pool of techniques to other modalities (e.g., mid-air finger gestures [11]), or with additional factors (e.g. direct vs indirect) will be the next step to cover (Figure 2, 4).

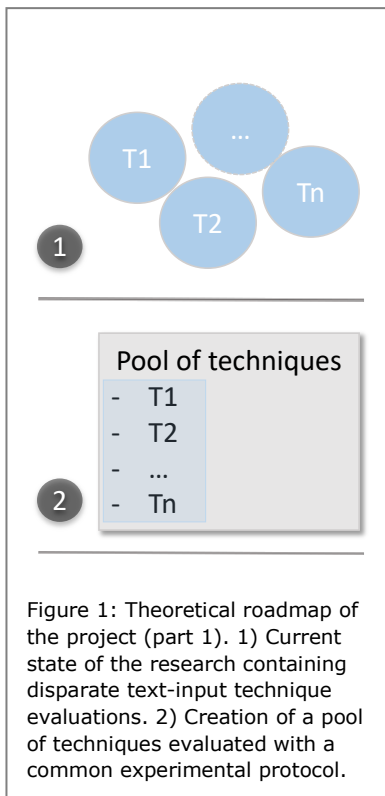
The main idea behind this work is to create and provide a tool – the pool of evaluated techniques – to researchers and practitioners that will be (1) used for future text-entry studies, (2) completed by future text-entry studies, and (3) provide a long-term structure to text-entry input techniques evaluations.

Related Work

We highlight how previous works report on their results and how the main differences make it challenging for future work to build on these results. We focus on two aspects to illustrate our point and motivate this position paper: experimental protocols and reported dependent measures.

Several experimental protocols are too different from each other to draw conclusion between two pieces of work. For instance, the evaluation of ZShift and Callout vs Zoomboard uses 1 session of 1h using 5 phrases by techniques [7]. H4-Writer's evaluation uses a longitudinal study with 10 sessions of 12 blocks of 5 phrases each [9]. On the contrary, SwipeBoard's evaluation uses an artificially accelerated novice-to-expert transition, using 2 sessions, with 4 blocks of 108 4-letters words [4]. Finally, ZoomBoard's evaluation had 4 sessions consisting of two 3-minutes trials [10]. Thus, even with high-level overviews and discussions, it is difficult to compare these techniques as they do not report on the same level of expertise and/or in the same conditions. The same can be discussed about experimental details, such as the way participants are notified [4] or instructed to correct errors [10] for instance.

A text-entry input technique can be evaluated according to several dependent measures [1,8]. Although it is completely understandable to not discuss all of them in a 10-pages research paper, previous works do not even report them. For instance, ZShift's evaluation reports word-per-minutes (WPM), key stroke per character (KSPC), character error rate (CER), corrected (resp. non corrected) error rate (Cerr, resp. Nerr), total error rate (Terr), and correction efficiency



(Ceff) [7]. However, SplitBoard’s evaluation reports only WPM, Terr and Nerr [6]. SwipeBoard’s evaluation reports only WPM and special error rates [4]. Finally, ZoomBoard’s evaluation reports WPM, KSPS and mean string distance (MSD) only [10]. Thus, the limited conclusions one can have because of experimental differences are also limited by the reported results.

Our goal is to (1) unify how the existing techniques have been evaluated, and (2) motivate the text-entry community to follow the same patterns – regarding the protocol and the results reporting (Figure 2, 3).

Challenges

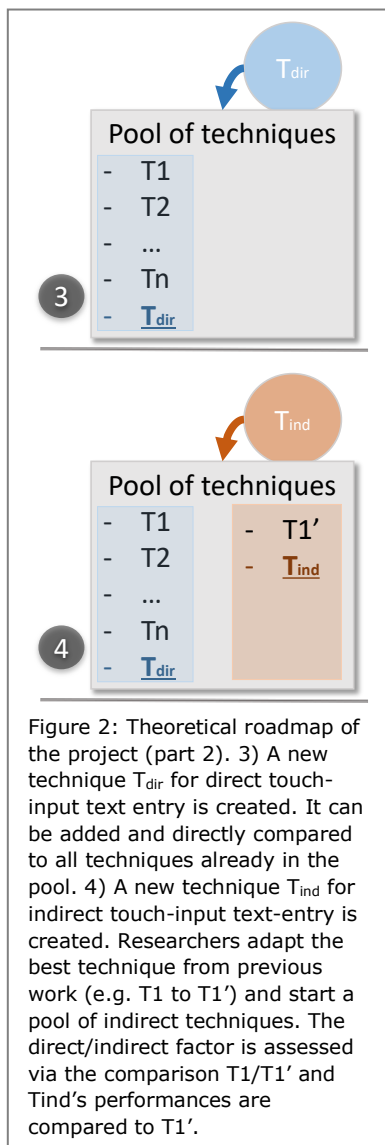
Although there are various challenges associated with the evaluation process, due to space constraints, we would like to focus on two of the major complexities linked to the challenge of collaboratively compiling data, and propose potential solutions for them.

First, we would like to address the issue concerning within-subject experimental designs. Many researchers in the HCI community prefer within-subject design to between-subject design, mainly because of a potential between-subject variability assumption issue (where researchers are inclined to believe that significant effects can be identified possibly due to outliers and imperfect random group assignment). Since within-subject designs require the same participants to be tested multiple times, this approach would make collaboration over time across multiple laboratories virtually impossible. Between-subject designs, on the other hand, would allow us to expand our data set. For instance, when a researcher tests *text-entry technique 1* as a means of data collection, the effectiveness of this technique can be evaluated in one laboratory.

Another researcher from a different laboratory can compile data with *text-entry technique 2*, and the evaluation of the effectiveness can be compared, *if we employ between-subject design* (i.e., with new sets of participants). Such data expansion is vastly useful for the evaluation of any text-entry technique.

In order to solve the between-subject variability assumption issue, we propose the inclusion of a baseline check, which can be accomplished by a simple on-screen QWERTY keyboard for example, the *de facto* standard of text-input. An experimenter could then make sure that each group exhibits similar performance levels regarding the baseline task. Thus, each evaluated technique could then be compared relative to the baseline, making the comparison legitimate from a statistical point of view.

The second challenge that we would like to focus on is the protocol for the evaluations. Participants are a limited and a critical resource when it comes to user studies. The evaluation procedure is also time consuming. Thus, the development of a strong protocol to follow would be beneficial. Such a protocol should be (1) justified, and (2) feasible. For instance, should we define a protocol involving ten 1-hour sessions with all the letters of the alphabet [9] or two 1-hour sessions with a subset of letters, to simulate expert behavior [4]? Researchers often have their preferences, and this is a topic that has been covered by a previous text-entry workshop [12]. We believe that our work can further improve on the results of previous workshops.



Finally, in order to ensure reproducibility, we will need to pay close attention to the details of the protocol. For instance, experimental materials (e.g., log templates, log parsers, statistical scripts and even oral experimental scripts) would need to be provided to ensure consistency across studies. We would like to discuss these issues so we can develop open-source scripts which could lay a basic foundation upon which metrics would be reported.

Workshop Relevance

We believe that this work is relevant to the CHI 2017 Workshop on Ubiquitous Text-Interaction for three reasons.

First, it fits the goal of the workshop by focusing on the challenges arising due to the increasing number of input techniques. This increase is likely to continue as more and more input devices will emerge, existing commercial text-entry systems iterate and update, and new text-entry input techniques will be needed to explore text-input on these new input devices. Instead of having disparate works that continue to spread into different directions, this work can define a common ground on which every future work can build on, and then contribute by joining the pool of techniques.

Second, our end-goal is the creation of a standardized text-entry input evaluation method. A workshop gathering text-entry experts and enthusiasts is the perfect place to discuss all aspects to cover. In addition, we believe that covering all of those aspects might not be realistically feasible in one day. Thus, short-, middle-, and long-term collaborations are likely to emerge from this workshop. Although these collaborations might gravitate around potential

publications, we also envision a focus on the synergy that this project can bring to the text-entry community: in the future, studies evaluating new techniques can share common core concepts and follow a practical guide to join the created pool of techniques.

Finally, we believe that this work, with its potential impacts in the text-entry research community and potential collaborations, provides an elegant concrete outcome expected from a workshop.

Future Work

This position paper introduces the idea of creating a pool of techniques following a standardized evaluation. If this position paper is accepted as a use-case for the CHI 2017 workshop, we propose to write an additional 8 pages paper reviewing previous work. This can summarize all challenges and potential solutions that could be used as a basis for discussions and brainstorming sessions during the workshop.

We also envision a tool that will be used by researchers and practitioners. Although publications are a good starting point for visibility and peer-acceptance purpose [2], a publication is static and cannot be updated. We hence envision an online tool, using a database of techniques satisfying both the common evaluation procedure, metrics reporting and a peer-review process. This online tool could then be kept updated by the text-entry community as new techniques appear.

References

1. a.S. Arif and W. Stuerzlinger. 2009. Analysis of text entry performance metrics. *Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference*: 100–105. <http://doi.org/10.1109/TIC-STH.2009.5444533>

2. Ahmed Sabbir Arif and Ali Mazalek. 2016. A survey of text entry techniques for smartwatches. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Lecture Notes in Computer Science), 255–267. http://doi.org/10.1007/978-3-319-39516-6_24
3. Liwei Chan, Rong-Hao Liang, Ming-Chang Tsai, et al. 2013. FingerPad: Private and Subtle Interaction Using Fingertips. *Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST '13*: 255–260. <http://doi.org/10.1145/2501988.2502016>
4. Xiang Anthony Chen, Tovi Grossman, and George Fitzmaurice. 2014. Swipeboard: A Text Entry Technique for Ultra-Small Interfaces That Supports Novice to Expert Transitions. *Proceedings of the 27th annual ACM symposium on User interface software and technology - UIST '14*: 615–620. <http://doi.org/10.1145/2642918.2647354>
5. James Clawson, Kent Lyons, Thad Starner, and Edward Clarkson. 2005. The impacts of limited visual feedback on mobile text entry for the twiddler and mini-QWERTY keyboards. *Proceedings - International Symposium on Wearable Computers, ISWC 2005*: 170–177. <http://doi.org/10.1109/ISWC.2005.49>
6. Jonggi Hong, Seongkook Heo, Poika Isokoski, and Geehyuk Lee. 2015. SplitBoard : A Simple Split Soft Keyboard for Wristwatch-sized Touch Screens. *Proc. CHI '15*: 1233–1236. <http://doi.org/10.1145/2702123.2702530>
7. Luis a Leiva, Alireza Sahami, Alejandro Catala, Niels Henze, and Albrecht Schmidt. 2015. Text Entry on Tiny QWERTY Soft Keyboards. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*: 669–678. <http://doi.org/10.1145/2702123.2702388>
8. I Scott MacKenzie and Kumiko Tanaka-Ishii. 2015. *Text entry systems*. <http://doi.org/10.1016/B978-012373591-1/50001-2>
9. Is MacKenzie, Rw Soukoreff, and Joanna Helga. 2011. 1 thumb, 4 buttons, 20 words per minute: design and evaluation of H4-writer. *Proceedings of the 24th annual ...*: 471–480. <http://doi.org/10.1145/2047196.2047258>
10. S Oney, Harrison, a Ogan, and J Wiese. 2013. ZoomBoard: a diminutive qwerty soft keyboard using iterative zooming for ultra-small devices. *Proceedings of CHI 2013, C*: 2799–2802. <http://doi.org/10.1145/2470654.2481387>
11. Srinath Sridhar, Anna Maria Feit, Christian Theobalt, and Antti Oulasvirta. 2015. Investigating the Dexterity of Multi-Finger Input for Mid-Air Text Entry. *Proceedings of the ACM CHI'15 Conference on Human Factors in Computing Systems 1*: 3643–3652. <http://doi.org/10.1145/2702123.2702136>
12. Keith Vertanen. 2016. Inviscid Text Entry and Beyond. *CHI Extended Abstracts on Human Factors in Computing Systems*: 3469–3476. <http://doi.org/10.1145/2851581.2856472>