Correcting Errors in Speech Input During Non-visual Use

Jonggi Hong
Inclusive Design Lab, HCIL
Computer Science Department
University of Maryland
College Park, MD, USA
jhong12@umd.edu

Leah Findlater
Inclusive Design Lab, HCIL
College of Information Studies
University of Maryland
College Park, MD, USA
leahkf@umd.edu

This short paper was presented at the ACM CHI 2017 Workshop on Ubiquitous Text Interaction, May 2017, Denver, Colorado, USA.

© 2017 Authors. Copyright permission was retained by the authors with non-exclusive permission granted to workshop organizers to publish the paper on the workshop website.

Permission to make digital/hard copy of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of publication and its date appear, and notice is given that copying is by permission of the authors. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission.

Abstract

While speech input has improved dramatically in the past few years, reviewing and editing the dictated text during non-visual use is a known challenge. This position paper describes this problem and outlines ongoing and future work plans to address it.

Author Keywords

Speech input; non-visual context; visually impaired users; text entry; error correction.

ACM Classification Keywords

H.5.2. User Interfaces: User-centered design, Input devices and strategies, Interaction styles.

Research Problem and Use Scenario

Speech recognition has improved considerably in the past few years with deep learning advances [18], to the point where it is now faster and more accurate than using a mobile touchscreen keyboard [22]. The importance of speech input is also increasing along with the popularity of wearable and smart devices such as Google Home and Amazon Echo that have small or non-existent visual displays and may need to support at-a-distance interaction.

Unfortunately, reviewing and editing dictated text is a bottleneck for speech input [15]. Despite advances in speech recognition accuracy, it is still hard to completely remove recognition errors due to issues such as the ambiguity of words (*e.g.*, homophones or pronouns) and background noise [14,24]. Using speech itself to correct errors can sometimes have cascading side effects [15] and users may instead prefer to use manual touchscreen input for correction [22].

Speech is particularly useful for *non-visual contexts* such as for blind or visually impaired users, dictation while driving or walking, or at-a-distance interaction with smart devices. The downside, however, is that non-visual use makes correcting errors even more time consuming because reviewing and editing errors needs to be done by listening to the text-to-speech output. Azenkot and Lee [1], for example, showed that blind users use speech input at higher rates than sighted users but also that blind users spend 80% of their speech input time correcting errors [27]. The study also highlighted the difficulty of even identifying that speech recognition errors exist when reviewing the dictated text using a screen reader (i.e., audio output). One example provided is "lost my sight" versus "lost my site", which both sound similar. This problem likely also extends to sighted users in eyes-free contexts.

While we are interested in non-visual use scenarios in general, our specific focus is to support blind and visually impaired users who use screen readers (i.e., audio-based interaction). Though Azenkot and Lee [1] presented text reviewing as a challenge in using speech input by blind users, they did not provide a quantitative analysis of this challenge, nor did they propose mechanisms to address it. These are our goals.

Our Background

Our research team has experience with mobile text entry and with mobile and wearable interaction for blind and visually impaired users.

Touchscreen Text Entry

Hong et al. developed, SplitBoard, a smartwatch QWERTY keyboard where only the left or right half of the keyboard is shown at once, thus increasing the size of individual keys [11]. Users can switch between the two halves with a swipe gesture and select a key by tapping. In a comparison against other keyboards, including QWERTY and ZoomBoard, across different screen sizes, and with varied activity levels (*i.e.*, with/without walking), SplitBoard was faster than ZoomBoard and more accurate than QWERTY [12].

We have also proposed and studied adaptive touchscreen keyboards [6,7,10]. This work has included characterizing ten-finger touchscreen input patterns [7], building and evaluating a keyboard that adapts to each user's unique typing patterns [6], and addressing the difficulty of walking and typing on a smartphone by modeling and compensating for the effects of a user's physical steps on input [10].

Accessible Mobile and Wearable Interaction

We have also conducted several studies on mobile and wearable interaction for blind and visually impaired users. Ye et al. [27], for example, surveyed 215 visually impaired and sighted participants and interviewed 10 participants with visual impairments to explore the challenges and potential accessibility opportunities of mobile and wearable devices. Although the focus was on higher-level themes such as social interaction and privacy, the study confirmed that blind users make use of speech input more often than sighted users. The following situations were commonly cited for not wanting to use speech input: noisy environments, privacy concerns, and quiet but public environments such as churches or libraries.

As another example of our work in this space, Oh et al. [20] proposed and evaluated automated techniques to help novice blind touchscreen users learn touchscreen gestures. The techniques included both automatically generated verbal instructions and gesture sonification, that is, creating sounds to represent the speed and shape characteristics of a gesture. For sonification, changes in pitch were found to be the most distinguishable for communicating a gesture's shape and movement direction. Related to speech input correction, we are exploring whether these and other sonification attributes are useful for audibly highlighting possible errors.

Related Work

Editing Text Using Visual Output

Editing text is known to be a significant challenge when using speech input interfaces. A study by Karat et al. showed that the editing process requires 66% of a user's time when using speech input even with a visual interface [15], although this number may be lower with modern speech recognition engines.

Prior studies have developed interfaces for editing text from speech input. Some of them use multimodal input, combining speech input with other types of input methods such as a touchscreen gestures and a keyboard [8,9,23]. One approach is to combine speech, drawing, and handwriting to correct errors [5]. This approach makes it easier for users to select and correct errors than using speech only. But, it is not suitable in a non-visual context because it requires that users draw gestures at specific locations on the screen (i.e., over the text they would like to correct). Another approach is to suggest a set of alternative words when the user selects a word to correct; these alternatives are suggested based on their similarity in oral pronunciation [13,16,19,26]. However, though the list of alternative words may contain the correct word, navigating that list while using audio-only output (*e.g.*, with a screenreader) is likely to be much less efficient than scanning it visually.

Unimodal speech input has also been investigated for error correction. For example, Choi et al. [4] used speech input as a way to correct errors as well as enter new text. Error correction using only speech input would be appropriate for non-visual contexts, but, as mentioned earlier, unimodal correction suffers from cascading side effects [15]—even when the user has the luxury of visually viewing the results of their speech input and correction efforts.

Non-Visual Text Input

Commercial screenreaders for blind and visually impaired users allow for manual text entry on touchscreen keyboards. For example, the VoiceOver screenreader on Apple iOS devices reads each key aloud to enable blind people to use the standard QWERTY keyboard.

Many research examples also exist for non-visual text entry, although the focus has been on entering rather than correcting text. Non-QWERTY keyboards based on Braille have been designed and evaluated for blind and visually impaired users [2,17,21] . Touchscreen gestures have also been used to select keys without the need for precise visual targeting [3]. However, these keyboards require the users to repeatedly use a touchscreen gestures to explore and edit the text at the level of individual characters. Graffiti is another approach where users could enter text non-visually by drawing characters on a touchscreen [25]; the Apple Watch provides a similar approach to complement speech dictation. This character-based gestural entry, however, is likely to be much more time consuming that speech dictation.

Ongoing and Future Work

In ongoing work, we are studying how well users can identify dictation errors based on text-to-speech output. As already mentioned, Azenkot and Lee [1] found that it is difficult for blind users to identify errors that sound like the intended text (e.g., “site” vs. “cite”). But, how often do such errors occur and what factors affect a user’s ability to detect them? Two factors we are exploring are whether the user is a native English speaker or not (assuming English speech dictation) and the level background noise. Our preliminary findings suggest, for example, that the word error rate is higher with non-native speakers than native speakers. We are also designing and evaluating mechanisms to improve speech-based error identification and correction during non-visual use. Attending the CHI workshop at this formative stage in our research will be useful for helping guide our efforts.

References

1. Shiri Azenkot and Nicole B Lee. 2013. Exploring the use of speech input by blind people on mobile devices. *Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility*, ACM, Article No. 11.

2. Shiri Azenkot, Jacob O. Wobbrock, Sanjana Prasain, and Richard E. Ladner. 2012. Input finger detection for nonvisual touch screen text entry in Perkinput. *Proceedings of Graphics Interface (GI ’12)* d: 121–129.

3. Matthew N Bonner, Jeremy T Brudvik, Gregory D Abowd, and W Keith Edwards. 2010. No-look notes: accessible eyes-free multi-touch text entry. *International Conference on Pervasive Computing*, 409–426.

4. Junhwi Choi, Kyungduk Kim, Sungjin Lee, et al. 2012. Seamless error correction interface for voice word processor. *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, 4973–4976.

5. W Feng. 1994. Using handwriting and gesture recognition to correct speech recognition errors. *Urbana* 51: 61801.

6. Leah Findlater and Jacob Wobbrock. 2012. Personalized input: improving ten-finger touchscreen typing through automatic adaptation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 815–824.

7. Leah Findlater, Jacob O Wobbrock, and Daniel Wigdor. 2011. Typing on Flat Glass: Examining Ten-finger Expert Typing Patterns on Touch Surfaces. *Proceedings of the SIGCHI Conference on*

Human Factors in Computing Systems, ACM, 2453–2462.

8. Arnout R H Fischer, Kathleen J Price, and Andrew Sears. 2005. Speech-based text entry for mobile handheld devices: an analysis of efficacy and error correction techniques for server-based solutions. *International Journal of Human-Computer Interaction* 19, 3: 279–304.

9. Kazuki Fujiwara. 2016. Error Correction of Speech Recognition by Custom Phonetic Alphabet Input for Ultra-Small Devices. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 104–109.

10. Mayank Goel, Leah Findlater, and Jacob Wobbrock. 2012. WalkType: using accelerometer data to accomodate situational impairments in mobile touch screen text entry. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2687–2696.

11. Jonggi Hong, Seongkook Heo, Poika Isokoski, and Geehyuk Lee. 2015. SplitBoard: A simple split soft keyboard for wristwatch-sized touch screens. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 1233–1236.

12. Jonggi Hong, Seongkook Heo, Poika Isokoski, and Geehyuk Lee. 2016. Comparison of three QWERTY keyboards for a smartwatch. *Interacting with Computers*: iww003.

13. David Huggins-Daines and Alexander I Rudnicky. 2008. Interactive asr error correction for touchscreen devices. *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Demo Session*, 17–19.

14. Hui Jiang. 2005. Confidence measures for speech recognition: A survey. *Speech communication* 45, 4: 455–470.

15. Clare-Marie Karat, Christine Halverson, Daniel Horn, and John Karat. 1999. Patterns of entry and correction in large vocabulary continuous speech recognition systems. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 568–575.

16. Yuan Liang, Koji Iwano, and Koichi Shinoda. 2014. Simple gesture-based error correction interface for smartphone speech recognition. *INTERSPEECH*, 1194–1198.

17. Sergio Mascetti, Cristian Bernareggi, and Matteo Belotti. 2011. TypeInBraille: a braille-based typing application for touchscreen devices. *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*, 295–296.

18. Jordan Novet. Google has slashed its speech recognition word error rate by more than 30% since 2012. https://cdn.ampproject.org/c/venturebeat.com/2017/.

19. Jun Ogata and Masataka Goto. 2005. Speech repair: quick error correction just by using

selection operation for speech input interfaces. *INTERSPEECH*, 133–136.

20. Uran Oh, Shaun K. Kane, and Leah Findlater. 2013. Follow that sound: using sonification and corrective verbal feedback to teach touchscreen gestures. *Proceedings of the ACM SIGACCESS International Conference on Computers and Accessibility (ASSETS 2013)*, ACM Press, To appear.

21. João Oliveira, Tiago Guerreiro, Hugo Nicolau, Joaquim Jorge, and Daniel Gonçalves. 2011. BrailleType: unleashing braille over touch screen mobile phones. *IFIP Conference on Human-Computer Interaction*, 100–107.

22. Sherry Ruan, Jacob O Wobbrock, Kenny Liou, Andrew Ng, and James Landay. 2016. Speech Is 3x Faster than Typing for English and Mandarin Text Entry on Mobile Devices. *arXiv preprint arXiv:1608.07323*.

23. Koichi Shinoda, Yasushi Watanabe, Kenji Iwata, Yuan Liang, Ryuta Nakagawa, and Sadaoki Furui. 2011. Semi-synchronous speech and pen input for mobile user interfaces. *Speech Communication* 53, 3: 283–291.

24. Balwant A Sonkamble and Sulochana Sonkamble. 2008. SPEECH RECOGNITION USING THE TEMPLATE APPROACH. *Advances in Computer Vision and Information Technology*: 43.

25. Hussain Tinwala and I Scott MacKenzie. 2009. Eyes-free text entry on a touchscreen phone. *Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference*, 83–88.

26. Lijuan Wang, Tao Hu, Peng Liu, and Frank K Soong. 2008. Efficient handwriting correction of speech recognition errors with template constrained posterior (TCP). *INTERSPEECH*, 2659–2662.

27. Hanlu Ye, Meethu Malu, Uran Oh, and Leah Findlater. 2014. Current and future mobile and wearable device use by people with visual impairments. *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI ’14*: 3123–3132.