Int. J. Human-Computer Studies 73 (2015) 1–11

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Enhanced auditory feedback for Korean touch screen keyboards ☆

Yongki Park [a], Hoon Heo [a], Kyogu Lee [a,b,*]

[a] Music and Audio Research Group, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Republic of Korea
[b] Center for Interactive Media Art and Technology, Advanced Institutes of Convergence Technology, Suwon, Republic of Korea

ARTICLE INFO

Article history:
Received 30 October 2013
Received in revised form
11 June 2014
Accepted 3 August 2014
Communicated by C. Schmandt
Available online 9 August 2014

Keywords:
Multimodal interaction
Auditory feedback
Text entry
Touch screen device
Mobile

ABSTRACT

With the increasing popularity of touch screen mobile devices, improving the usability and the user experience while inputting text on these devices is becoming increasingly important. Most conventional touch screen keyboards on mobile devices rely heavily on visual feedback, while auditory feedback seldom includes any useful information about what is being inputted by the user. The auditory feedback usually simply replicates the sounds produced by a physical keyboard. This paper describes the development of an enhanced auditory feedback mechanism for a Korean touch screen keyboard called the enhanced auditory feedback (EAF) mechanism. EAF has subtle phonetic auditory feedback generated using the acoustic phonetic features of human speech. While typing with EAF, users can acquire non-invasive auditory clues about the keys pressed. In this work, we compare conventional auditory feedback for a touch screen keyboard used in touch screen mobile devices with that of EAF and explore the possibility of using enhanced auditory feedback for touch screen keyboards.

1. Introduction

Mobile, handheld touch screen devices are becoming increasingly common. Smartphones and tablet PCs support an ever-expanding array of activities. Using touch screen devices, users can access e-mails, write text messages, and input numerous kinds of information. As the range of applications for mobile devices grows, mobile text input is becoming equally important. To support the expansion of text input activities, mobile devices employ various text input methods, such as handwriting recognition, voice dictation, and a variety of touch screen keyboards with various layouts.

Among these methods, the QWERTY touch screen keyboard is the most popular method for text input. However, conventional touch screen keyboards rely heavily on visual feedback. The auditory feedback from conventional touch screen keyboards only replicates the sounds made by a physical keyboard, it seldom includes any useful information. To support more general-purpose computing tasks that rely heavily on efficient text entry such as word processing, spreadsheets, and even email, touch screen platforms need to provide effective mechanisms for entering text that are on par with those of traditional physical keyboards. Thus, our work is focused on exploring whether the use of enhanced auditory feedback on a touch screen device can provide users with support for efficient text entry. We believe that this is an important question to answer in order for touch screen input to be used in mainstream computing platforms that support the majority of computing tasks, including desktop-style tasks.

This paper presents the enhanced auditory feedback (EAF) mechanism for Korean touch screen keyboards, which is generated using the acoustic phonetic features of human speech. Typing with EAF, users obtain non-invasive and *private* auditory clues about the keys he/she has pressed. Using EAF only the user who is expecting specific auditory feedback can discern the provided auditory clue as he/she is pressing the specific keys. EAF was designed to provide minimum amount of auditory clue about the pressed keys to insure one's privacy. Moreover, mimicking the sounds of human speech helps the users recognize with more ease the different keys they are pressing.

The remainder of this paper is organized as follows. Section 2 discusses the development of auditory feedback, touch screen technology, and related text entry research. Section 3 gives the definition of acoustic phonetic feature and describes how EAF is generated using the feature. Sections 4 and 5 describe the structure of the Korean language and its associated QWERTY keyboard layout, respectively. Section 6 outlines the experimental framework used to evaluate the performance of EAF. Section 7 presents the results of user studies conducted. Finally, Section 8 presents our conclusions and outlines our plans for future work.

2. Background

2.1. Auditory feedback

The concept of using sounds that people are used to hearing naturally in everyday life was first proposed by Gaver. Gaver's

☆This paper has been recommended for acceptance by C. Schmandt.

* Corresponding author. Tel.: +82 31 888 9139; fax: +82 031 888 9148.
E-mail address: kglee@snu.ac.kr (K. Lee).

http://dx.doi.org/10.1016/j.ijhcs.2014.08.002
1071-5819/

concept uses natural everyday sounds to represent events and objects in a computer interface. Auditory icons constitute brief auditory feedback that represents objects, functions, and actions. A user's prior knowledge and natural auditory association with sound sources are used to recognize auditory icons. The Apple SonicFinder used these so-called auditory icons for the first time in its interface, and they are still in use today.

Auditory cues are divided into three categories on the basis of their abstraction level: iconic, metaphorical, and symbolic. While iconic representations of auditory cues try to acoustically reproduce an event as realistically as possible, an analogy between an event and an associated sound is established in metaphorical auditory cues. For many objects, functions and actions in a computer interface, where there is no clear iconic representation, earcons can be an effective sonification.

According to Brewster et al. (1992), earcons are non-verbal audio messages "composed of motives, which are short, rhythmic sequences of pitches with variable intensity, timbre, and register." The earcons have the highest level of abstraction; there are no semantic relations between an event and a sound in an earcon, but rather an arbitrary audio signal for event representation. Moreover, Brewster et al. (1993) states that earcons can be designed to represent a single item and its position in a hierarchical structure, either in audio-only interfaces, or multimodal interfaces. Pirhonen and Holguin (2002) showed that earcons can successfully improve the usability of multimodal interfaces for mobile use.

Simple auditory icons that represent an event or object by playing a typical sound (e.g., using the rattling sound of a trash bin to represent the deletion of a file) can easily be interpreted by almost any listener. However, developing highly compatible auditory icons for more abstract events (e.g., changing the display profile of a mobile phone) is difficult. In addition, a user's ability to recognize the auditory icon without training is reduced. For earcons, the meaning needs to be learned a priori and is not transferable to another earcon. Jones and Furner (1989) and Lucas (1989) showed that there is no significant difference in efficiency between simple auditory icons and earcons. Moreover, according to Walker et al. (2006), their efficiency does not vary when used in combination with spoken menu items for locating different items in a hierarchical structure.

The above-mentioned study introduced an auditory cue called spearcons. To generate Spearcon, text of a menu item is converted to speech and then the resulting audio clip is speeded up until it is no longer a comprehensible speech. According to Walker et al. (2006), the combination of spearcons with a spoken menu text gives it only a slight advantage over menus that are spoken only, but a more significant advantage over earcons. Hearcons, three-dimensional (3D) abstract auditory objects positioned in an auditory interaction realm, have subsequently been created to support the navigation of web pages and hierarchical menus. They constantly emit sound and can be manipulated with the use of a pointing device.

As has already been proven, events within a hierarchical structure can be successfully represented by sound events. Sound event can represent information about the meaning of an event as well as its position in a hierarchy. Sound event clearly has limited amount of information it can represent, but it is clear that by enriching the way in which the event is represented through sound can facilitate the inclusion of additional information.

2.2. Touch screen and text entry

Touch screen devices have been available for decades; however, only since the recent emergence and subsequent ubiquity of smartphones and tablet PCs has urgent effort been made to understand the strengths and limitations of touch screen technology. Touch screen interaction was introduced in 1970 by Johnson and Fryberger (1972); however, only after Boie's multi-touch screen was demonstrated in 1984 and the research done by Buxton (1986) on two-handed input in 1986 was touch screen devices deemed relevant for use in modern society. In the late 1980s, commercial products featuring touch screen interface, such as the Apple Newton and the IBM Simon, began appearing.

A number of research papers on the technology and interaction techniques of tablet and touch screen computing were published in the 1990s. The DigitalDesk-a "Computer Augmented Environment," which allowed interaction via the pointing of a finger, was introduced by Wellner (1993). Kurtenbach et al. (1997) presented the results of research conducted on a GUI design for touch screen devices that was geared towards increasing the quality of the input. Leganchuk et al. (1998) presented in their work the manual and cognitive benefits of two-handed input applied to traditionally one-handed computer tasks.

It was the 2000s, however, that saw a major growth in development of multi-touch technology, and the period when multi-touch actually moved beyond the laboratory and into the hands of consumers. Introduction of major commercial product using the multi-touch technology came in 2007. Microsoft announced the Microsoft Surface which is an interactive tabletop computing platform. In that same year, Iphone was introduced by Apple, changing the perception and the way users interact with multi-touch device.

The research being presented in this paper focuses on using new auditory feed back to improve text entry using multi-touch devices. This work is built upon efforts that originally considered text entry on single-touch and mobile platforms. MacKenzie and Soukoreff (2002b) introduced models for developing and studying text entry on mobile devices. MacKenzie and Soukoreff (2002a) also studied techniques for evaluating such systems. Soukoreff and MacKenzie (2004) refined their previous works to give better measurement of error rates. Clarkson et al. (2005) presented a longitudinal study of mini-QWERTY keyboard use, while Wobbrock and Myers (2006) introduced a method for the analysis of character-level errors in input streams for unconstrained text entry evaluations, and new metrics for refining text entry methods.

Even though a number of researchers have conducted studies geared towards improving touch screen keyboards, most of these studies explore the fundamental problems of text input on small handheld, mobile devices. Screen sizes, different touch screen keyboard layouts, and new multimodal feedback methods were used as factors to improve touch screen keyboards in these studies. The relationship between the size of a touch screen-based touch screen keyboard and data entry rates was investigated by Sears (1993).

In their study they found that users made and corrected fewer errors when entering text on larger keyboards, data entry rates were higher on larger keyboards, and participants preferred larger keyboards. MacKenzie and Zhang (1999) developed a new touch screen keyboard layout called OPTI and demonstrated that after about 4 h of practice, users entry rates are higher with the OPT1 layout than with a conventional QWERTY touch screen keyboard layout.

Non-speech auditory feedback has proven effective in improving interaction on mobile devices. In studies conducted by Brewster and Tasker (2003), Holland and Morse (2001), Sawhney and Schmandt (2000), and Pirhonen and Holguin (2002), information has been successfully presented via auditory channels to users while allowing users to visually focus on navigating through their physical environments. We were also heavily influenced by the study of multi-touch virtual keyboard by Varcholik et al. (2012).