Designing Mobile Phones using Silent Speech Input and Auditory Feedback

Kamer Ali Yuksel
Computer Vision and Pattern Analysis Laboratory (VPALAB), Sabanci University, Orhanli - Tuzla, 34956 Istanbul
kamer@sabanciuniv.edu

Sinan Buyukbas
Faculty of Arts and Social Sciences, Sabanci University, Orhanli - Tuzla, 34956 Istanbul
sinanbuyukbas@sabanciuniv.edu

Serdar Hasan Adali
Computer Graphics Laboratory, Sabanci University, Orhanli - Tuzla, 34956 Istanbul
serdaradali@sabanciuniv.edu

MobileHCI 2011, Aug 30–Sept 2, 2011, Stockholm, Sweden.
ACM 978-1-4503-0541-9/11/08-09.

Abstract
In this work, we have propose a novel design for a basic mobile phone, which is focused on the essence of mobile communication and connectivity, based on a silent speech interface and auditory feedback. This assistive interface takes the advantages of voice control systems while discarding its disadvantages such as the background noise, privacy and social acceptance. The proposed device utilizes low-cost and commercially available hardware components. Thus, it would be affordable and accessible by majority of users including disabled, elderly and illiterate people.

Keywords
Mobile phone, silent speech interface, vibration sensors

ACM Classification Keywords
H5.2 [Information interfaces and presentation]: User Interfaces. Input devices and strategies; B 4.2 Input Output devices.

Introduction
Mobile phones are essential for everyday lives of most of the people and new features are continuously being added to them since their release, mainly due to their competitive market. Most of their features are not indispensable for the mobile connectivity and not being used by the majority of their users. Due to mainly the accessibility of them, voice call and text messaging are

still core features of a mobile phone among others including MMS, e-mail, Internet, music, gaming and photography [2]. Although extra features are useful for some people, they tend to increase the complexity and cost of mobile devices and their network infrastructure. Thus, they do not represent the actual preference of the majority of users despite their increasing market share of 14% [2]. In this work, we have presented a novel design for a basic mobile phone considering only the fundamental communication features, in order to make it more accessible by the majority of users both in terms of cost and interaction complexity. Thus, target users of the proposed device are people who prefer to use only core functionalities (e.g. voice call and SMS) through an affordable, accessible, small device based on a hands-free and auditory interface. This actually consists of a broad group of population including especially elderly, disabled or illiterate people.

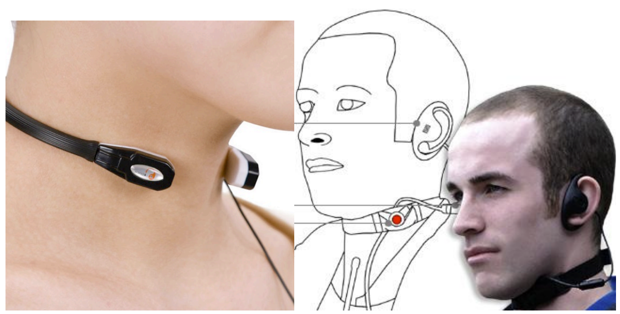

Figure 1. Conceptual designs for the proposed wearable mobile device using SSI together with auditory feedback

The proposed device is in the form of a headset and takes input through a silent speech interface (SSI) and gives auditory feedback through an earphone (Fig. 1) [1]. Proposed ways of interaction are preferred as they are more natural, intuitive and accessible than their alternatives that exist on current mobile devices [5, 6]. Moreover, they require less space, resources for operation and enable low-cost and miniaturized hardware to be designed. Finally, they are hands-free and they do not require line of sight; thus, do not depend on motor or visual skills of users.

Methodology

Speech control is generally considered being the most intuitive and natural way of interacting with devices [6] and the idea of interpreting the silent speech was popularized since the Kubrick's "2001 – A Space Odyssey" science-fiction film and auditory feedback is sufficient for most of the core features [5]. A SSI basically identifies phonemes that an individual pronounces using non-auditory sources of information without actually using the sound of their vocalization. This allows speech processing for synthesizing the actual speech or recognition of different commands even in the absence of an intelligible acoustic signal such as while non-audibly whispering [1]. SSI systems preserve the advantages of conventional speech-based voice control interfaces and they can also be utilized by speech-handicapped, operates in highly noisy environments and suitable when silence-required, mainly due to the privacy and social acceptance. In other words, commands can be softly whispered without concerning the existence and implications of the background noise or the other people around. Furthermore, they can easily be utilized to distinguish commands given during a phone call. In this work, we only consider three alternative non-invasive technologies available to build SSI systems [1]. The first technology is a throat microphone, which captures speech signals from both sides of the Adam's apple and has been used by fighter pilots since many decades to communicate within highly noisy environments. The second is in-ear microphone, which is inserted into the ear canal and has similar functionalities. Final technology is non-audible murmur (NAM) microphone, which consists of consists of a silicon conductor and a condenser microphone and placed on the neck below ear. All of these technologies are resistant to ambient background noise because they are based on body conductive signals. Also, They are cheap and commercially available in compact forms. Moreover, it is shown on literature that they are able to successfully recognize non-audible whisper with high accuracies [3, 4]. The in-ear microphone or NAM microphone can be utilized to have a more compact device due to the proximity of their targeted area to the ear. The throat microphone may be utilized as a separate module as well for visual design purposes.

Interaction

The device may use one of above-mentioned hardware to recognize single-word commands (e.g. "Answer", "Reject") or commands that may be followed by parameters such as "Call Elliot" or "Call 555-1234. Commands and parameters can be accurately recognized because they consist of a small set of keywords, numbers and additional terms such as the names in the address book of the user. When the voice call is initiated, the user may still use SSI to give commands, as they are distinguishable from the audible speech captured by the same interface. For example, the user would be able to terminate the call or use other features by silently speaking during the conversation. Also, users may also give commands for tasks that are not related to the communication such as checking the signal strength or battery level through the auditory interface. Moreover, textual contents can also be dictated to the device and their transcription can be automatically generated for sending text messages. The proposed device, which is essentially designed for communication purposes, would still support other functionalities such as playing music from a specific album, artist or playlist. Finally, it can be also used for more complex interactions through external peripherals or embedded extensions. For instance, a mini-projector or camera can be mounted to the device when visual feedback or photography is preferred or motion sensors (e.g. accelerometer, gyroscope) can be embedded to the throat microphone itself in order to perform gestures when necessary.

References

[1] Denby, B., Schultz, T., Honda, K., Hueber, T., Gilbert, J. M. and Brumberg, J. S. Silent speech interfaces. Speech Communication 2010, vol. 52, no. 4 270–287.

[2] Falaki, H., Mahajan, R., Kandula, S., Lymberopoulos, D., Govindan, R. and Estrin, D. Diversity in smartphone usage. *In Proc. MobiSys 2010,* ACM Press (2010), 179–194.

[3] Jou, S. C., Schultz, T. and Waibel A. Adaptation for soft whisper recognition using a throat microphone. *In Proc. ICSLP 2004*, 1493-1496.

[4] Jou S. C., Schultz, T. and Waibel, A. Whispery speech recognition using adapted articulatory features *In Proc. ICASSP 2005*, vol. 5, 1009

[5] Li, K. A., Baudisch, P. Hinckley, K. Blindsight: eyes-free access to mob ile phones. *In Proc. CHI 2008,* ACM Press (2008), 1389–1398.

[6] Starner, T. E., The role of speech input in wearable computing. IEEE Pervasive Computing 2002, 89–93.