

Chapter XVIII

Mouth and Ear Interaction

Dirk Schnelle

Technische Universität Darmstadt, Germany

ABSTRACT

This chapter gives an overview of the challenges that have to be mastered while working with audio. The vision of ubiquitous computing involves challenges for the future workplace. Tasks to be performed by workers are becoming more and more complex, which leads to an ever-increasing need to deliver information to workers. This can be information from a manual or instructions on how to proceed with the current task. Workers typically have their hands busy and use of a mouse or a keyboard will force them to stop working. Mouth and ear interaction can be performed without focusing attention on the device. But audio is also a medium that is not easy to handle. This chapter provides an understanding of why audio-based interfaces are difficult to handle and you will also be provided with some pointers as to how these challenges can be mastered to improve the quality of applications involving mouth & ear interaction.

MOTIVATION

Ubiquitous computing allows the delivery of information to the worker while she is engaged in her task. The mobile worker typically wears a small computer and the information is transmitted from a server to this wearable computer. Since workers typically have their hands busy while performing their tasks, the use of hands and eyes devices, that is, mouse and keyboard, will force them to interrupt working. Use of the acoustic channel does not have this limitation. Audio can

be used to deliver information to the worker and to enable her to interact with the system using voice commands. In addition, the acoustic channel is still functional under extreme cases, for example, darkness and limited freedom of movement. A headset requires only limited space. Moreover the acoustic channel can be used in addition to other modalities, like conventional screens and buttons.

A drawback of voice is that it becomes unusable in noisy environments. Noise coming from the environment may cause the speech recognizer to

detect unwanted commands and makes it harder for the user to listen to the system's output. In this case, graphical interfaces are preferable.

Nevertheless, in many cases the use of audio has some advantages to be used exclusively or in addition to other interfaces in smart environments. Melvyn Hunt (1992) stated that: "Reports and theses in speech recognition often begin with a cliché, namely that speech is the most natural way for human beings to communicate with each other."

This statement of Hunt is still true and the question arises why we see so little use of voice interfaces in today's human-computer interaction. Speech is certainly one of the most important means of communication for humans. It has been used from the beginning of mankind to transfer customs and knowledge from one generation to the next. Although new communication technologies have been developed throughout history, speech remains a very efficient way of communication. But is this also true for the computer as the counterpart?

Hunt's statement expresses implicitly the common expectation that a computer can be treated similarly to a human, leading to the vision of the patient listening homunculus in a computer's shape: as a result, less pretentious but more realistic and helpful applications were of low interest for academic research. This vision requires that computers have the active and passive communicative competence of a human, a theory that is supported by Schukat-Talamazzini (1995). It is also a message of Shneiderman (1986), who states that although the recognition rate is increasingly accurate for dictation systems, adoption outside the disabled user community has been slow compared to visual interfaces. In the past years, the area of natural language processing made a lot of progress in research towards this vision, but we are still a big step behind. Amy Neustein (2001) states that this "...does not seem far fetched when we consider how the field of linguistics, with its

wide spectrum of methods to study interactive speech, provides the building blocks for spoken language systems that simulate human dialog."

But she has to admit that this progress is only a necessary first step toward an open collaborative relationship between computational linguists and conversation analysts. As a consequence, this leads to more interactive conversational interfaces. But we will need more of this collaboration. Today the use of the computer as a patient listening homunculus remains a vision. Otherwise the big advantage of audio, being more natural, would have displaced the ubiquitous graphical user interface. In fact, banking companies, who offer both Web-based customer self-service and telephony-based self-service, are observing a trend in favor of the graphically oriented medium.

Nevertheless, the use of voice-based user interfaces seem to be promising as an alternative to graphical interfaces, or to be used in addition in the shape of multimodal interfaces. Imagine a warehouse picker who needs to pick different items from different shelves. The worker's hands are busy, calling for the use of voice interaction, and the high mobility of the worker makes carrying additional equipment impractical.

This chapter is organized as follows: the section "Dialog Strategies" describes the main differences between voice-based interfaces and graphical interfaces and introduces the four main concepts for voice-based interaction. In order to use voice as an input medium the definition of grammars is required. Grammars are a fundamental concept to define the valid input of a speech recognizer. The section "Domain Specific Development" introduces current approaches to using them in more than a single context. The section "Mouth & Ear interaction in Smart Environments" describes the concept of a browser for smart environments with a focus on mouth & ear interaction. The concept is similar to a Web browser, using a minimal command-set for interaction. Besides the main problems of browsing, the development of mouth

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/mouth-ear-interaction/21778

Related Content

Design and Development of Member Platform of Public Platform for Service Outsourcing Association

Ran Gao and Liu Ye (2018). *International Journal of Advanced Pervasive and Ubiquitous Computing* (pp. 1-22). www.irma-international.org/article/design-and-development-of-member-platform-of-public-platform-for-service-outsourcing-association/199773

Telecentre-Based Community Wireless Networks: Empowering Rural Community in Uganda

Dorothy Okello and Julius Butime (2013). *Social and Economic Effects of Community Wireless Networks and Infrastructures* (pp. 160-177). www.irma-international.org/chapter/telecentre-based-community-wireless-networks/74452

The Information Construction of Wind Farm Based on SIS System

Yao Wan-Ye and Yin Shi (2013). *Global Applications of Pervasive and Ubiquitous Computing* (pp. 127-134). www.irma-international.org/chapter/information-construction-wind-farm-based/72937

Impact of Computer and Video Games on the Development of Children

Kalawati Malik (2009). *Risk Assessment and Management in Pervasive Computing: Operational, Legal, Ethical, and Financial Perspectives* (pp. 343-351). www.irma-international.org/chapter/impact-computer-video-games-development/28465

Smart Home Communication System

Tianze Li (2017). *International Journal of Advanced Pervasive and Ubiquitous Computing* (pp. 1-15). www.irma-international.org/article/smart-home-communication-system/182524