Available online at www.sciencedirect.com
ScienceDirect
Int. J. Human-Computer Studies 69 (2011) 100–112
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Multiple spatial sounds in hierarchical menu navigation for visually impaired computer users

Jaka Sodnik*, Grega Jakus, Sašo Tomažič

Faculty of Electrical Engineering, University of Ljubljana, Trzaska 25, 1000 Ljubljana, Slovenia

Received 13 January 2010; received in revised form 19 August 2010; accepted 19 October 2010
Communicated by C. Schmandt
Available online 23 October 2010

Abstract

This paper describes a user study on the benefits and drawbacks of simultaneous spatial sounds in auditory interfaces for visually impaired and blind computer users. Two different auditory interfaces in spatial and non-spatial condition were proposed to represent the hierarchical menu structure of a simple word processing application. In the horizontal interface, the sound sources or the menu items were located in the horizontal plane on a virtual ring surrounding the user's head, while the sound sources in the vertical interface were aligned one above the other in front of the user. In the vertical interface, the central pitch of the sound sources at different elevations was changed in order to improve the otherwise relatively low localization performance in the vertical dimension. The interaction with the interfaces was based on a standard computer keyboard for input and a pair of studio headphones for output. Twelve blind or visually impaired test subjects were asked to perform ten different word processing tasks within four experiment conditions. Task completion times, navigation performance, overall satisfaction and cognitive workload were evaluated. The initial hypothesis, i.e. that the spatial auditory interfaces with multiple simultaneous sounds should prove to be faster and more efficient than non-spatial ones, was not confirmed. On the contrary—spatial auditory interfaces proved to be significantly slower due to the high cognitive workload and temporal demand. The majority of users did in fact finish tasks with less navigation and key pressing; however, they required much more time. They reported the spatial auditory interfaces to be hard to use for a longer period of time due to the high temporal and mental demand, especially with regards to the comprehension of multiple simultaneous sounds. The comparison between the horizontal and vertical interface showed no significant differences between the two. It is important to point out that all participants were novice users of the system; therefore it is possible that the overall performance could change with a more extensive use of the interfaces and an increased number of trials or experiments sets. Our interviews with visually impaired and blind computer users showed that they are used to sharing their auditory channel in order to perform multiple simultaneous tasks such as listening to the radio, talking to somebody, using the computer, etc. As the perception of multiple simultaneous sounds requires the entire capacity of the auditory channel and total concentration of the listener, it does therefore not enable such multitasking.

Keywords: Auditory interface; Simultaneous spatial sounds; Visually impaired user; Human–computer interaction; Cognitive workload

1. Introduction

In this day-and-age, the use of computers and other electronic equipment is essential and indispensable. According to Internet World Stats, more than 70% of the population in the developed countries uses computers on a daily basis, either as their primary working tool or just as a communication and entertainment device. Visually impaired and blind computer users are no exception. Normal sighted users use the keyboard, mouse or other pointing devices for input and various types of displays or screens for output. Blind or visually impaired users, on the other hand, are forced to substitute their visual channel with aural and tactile senses. Tactile interfaces have proven to be quite efficient for the users that have been visually impaired from birth or for a longer period of time and grew up with the so-called Braille keyboard. The users that have lost their sight recently mostly perceive the Braille keyboard as very difficult and relatively slow to use. In this paper, we focus on auditory

*Corresponding author. Tel.: +386 1 4768 494; fax: +386 1 4768 266; Mobile: +386 41 561 281.
E-mail addresses: jaka.sodnik@fe.uni-lj.si (J. Sodnik), grega.jakus@fe.uni-lj.si (G. Jakus), saso.tomazic@fe.uni-lj.si (S. Tomažič).

1071-5819/$ - see front matter
doi:10.1016/j.ijhcs.2010.10.004

interfaces which we believe to be easy and natural to use for everybody, and in most cases (depending on the complexity and structure of the individual interface) require less learning time.

1.1. Auditory interfaces

Auditory interfaces are primarily used to complement visual interfaces with the use of different sounds functioning either as alarms or signals of an ongoing background processes. In the case of visually impaired users, sound is used as a primary communication channel between the user and computer, therefore all information has to be presented with sound and audio signals. The auditory interfaces are roughly divided into speech and non-speech interfaces.

The speech interfaces are based on human speech which can be recorded, processed, played or synthesized by a computer. Speech interfaces are easy to use and require no adaptation or learning time due to the naturalness and intelligibility of human speech. Screen readers, the most widely used software by visually impaired users, are based on a speech synthesizer which reads the content of the computer screen using synthesized artificial speech (JAWS).

Non-speech interfaces in most cases comprise of auditory icons (Gaver, 1986) and earcons (Brewster et al., 1993). Auditory icons try to acoustically reproduce an event as realistically as possible. For example, the sound of water flowing from one glass to another can be used to represent the downloading of a file from one computer to another. Earcons, on the other hand, are very abstract and do not entail any semantic relation between an event and the sound used to signal it. Their meaning thus has to be learned a priori, while the meaning of auditory icons can sometimes be learned on the go. Walker et al. (2006) proposed spearcons, a set of audio clips based on a spoken text. In this case, the spoken words or items are sped up until they are no longer comprehensible as speech. Due to their non-intelligibility, their meaning has to be learned, but the learning procedure seems to be much faster than in the case of earcons.

1.2. Spatial audio and spatial auditory interfaces

The term spatial audio or spatial audio signal refers to a sound that originates from an arbitrary spatial position relative to the listener. The mechanism of sound localization in human listeners has been explored by many researchers and the major concepts have been reported in detail. One of the most important findings is a substantial difference in the localization accuracy in azimuth and elevation (Blauert, 1997; Jin et al., 2004). The minimum audible angle (MAA) also differs significantly in the two directions (Sodnik et al., 2004, 2005). The latter can also be interpreted as spatial sound resolution or the minimum audible proximity of two separate sound sources. All these phenomena have to be taken into consideration when applying spatial audio to an auditory interface.

Spatial audio can be effectively delivered through multiple speakers or through headphones. In the case of headphones, audio signals have to be preprocessed or filtered with Head Related Transfer Functions (HRTFs) in order to add the information on the spatial position of an arbitrary sound. HRTFs are filter transfer functions measured separately for each ear for multiple spatial positions relative to the listener (Begault, 1994; Algazi et al., 2001). They can be very accurate if measured individually for each listener or less accurate and more general if measured with dummy-heads (Gardener and Martin, 1994).

The basic idea of spatial auditory interface is that, in addition to the content of the sound signal, the point of its origin can also hold some information for the listener. The meaning or functionality of an auditory icon or other acoustic element can change when its spatial position is changed. Complex structures, such as tables or hierarchical menus, can for example be represented with spatial sounds describing their physical properties and dimensions.

The difference between auditory and visual perceptions can be illustrated with the comparison between parallel and serial communication channels. Sight enables people to simultaneously perceive an extensive amount of information, while hearing is limited almost to one sound at a time in order for it to be perceived clearly. The latter can be improved by spatial separation of more than one sound source. Multiple sounds can be perceived and understood if originating from different spatial positions. This phenomenon has been recorded as the so-called "Cocktail Party Effect", referring to the human ability to filter several simultaneous sounds and to concentrate on one in particular (Arons, 1992; Cohen, 1992; Stifelman, 1994). We believe this is a very promising human characteristic which enables the use of spatial sound in auditory menus in order to increase the information flow between the computer and the user (Hawley et al., 1999; Drullman and Bronkhorst, 2000).

1.3. Auditory representation of hierarchical menu structure

We already pointed out some major differences in human–computer interaction between visually impaired and normal sighted users. Some hardware has been designed specifically for the visually impaired, but there are very few adapted software packages on the market, therefore visually impaired and blind users are forced to work with operating systems and interaction paradigms for normal sighted users such as windows, icons, menus, etc. In the case of visually impaired and blind users, these interfaces are interpreted by a screen reading software giving only basic or no information on the actual structure of interface.

In the present research, we focus on the hierarchical menu structure, which is a major part of any Windows-based application and is therefore often accessed and used also by visually impaired users. Their use of menus is limited to moving up and down the menus and "reading" the items word by word with the use of a screen reader, whereas normal sighted users can navigate through the menus with the aid of