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 a mouse and choose between multiple menu items simultaneously. This is the issue we wanted to address and improve in our research. With the use of simultaneous spatial sounds, multiple menu items can be presented simultaneously also in the auditory domain. We chose two different spatial configurations of the sources—vertical and horizontal. In the vertical configuration, the menu items – i.e. the sound sources – were located one above the other, while in the horizontal configuration all the sound sources were located in the same horizontal plane around the user's head.

### *1.4. Related work*

The question of representing hierarchical menus with spatial sounds has been addressed before. In this section, we present some significant work closely related to our research.

Audio Windows, proposed by Cohen and Ludwig (1991), is one of the earliest attempts in this field. It consisted of a 3D auditory display based on spatial sound which used gestural input for data manipulation. The users wore headphones and data glove for pointing to specific areas of the interface and choosing between various elements.

Input based on hand gestures was used also by Crispien et al. (1996) and Savadis et al., (1996), complemented also by a 3D pointing device and speech recognition. The interface itself was based on a hierarchical menu structure represented with spatial audio cues surrounding the user's head. The speech and non-speech auditory items were positioned on a virtual ring which could be rotated in any direction.

The idea of a speaker rotating around the user's head was utilized by Kobayashi and Schmandt (1997). Their work was a further development of the AudioStreamer (Schmandt and Mullins, 1995) and it was used for browsing through audio recordings. The spatial positions of multiple simultaneous sound sources represented different temporal positions within the audio recordings. The interaction with the system was based on a touchpad. The system enabled the creation of a new speaker (to rewind or fast forward) or switching to a different, already existing speaker.

Brewster et al. (2003) proposed a similar auditory interface for a mobile device. Spatialized auditory cues were localized in the horizontal plane around or in front of the user's head. Hand and head gestures were used as input for selecting the items in the menu and triggering various events on the mobile device while on the go.

An extensive work in this area was performed by Frauenberger et al., (2004). They proposed several auditory interfaces based on spatial sound intended for visually impaired and blind users. They presented a hierarchical menu structure and text input fields in a virtual auditory room (20 m × 20 m × 7 m), where the individual menu items were arranged in a semicircle in front of the user. The individual items were presented with descriptive auditory icons. The interaction was based on head gestures and keyboard input. They concluded that spatial sound can be effectively used for human–computer interaction, as they proved that there was no significant difference between the performance of visually impaired and normal sighted users.

The authors established that auditory interfaces should be designed without having visual concepts in mind (2005, 2006). They introduced a set of mode-independent interaction patterns along with their transformation in the auditory domain. The auditory interface for MS Explorer was chosen as an example in order to demonstrate some problems with the representation of sound.

A different approach to the aural presentation of a graphical interface was proposed by Jagdish and Gupta (2008). The prototype called SonicGrid was intended to help visually impaired and blind users to navigate in GUI-based environments in a non-linear spatial presentation. It was based on an interactive grid layer on the screen giving non-speech sound feedback. The horizontal direction of auditory elements was coded as a stereo signal which changed volume between the left and the right speaker. The mechanism of changing the pitch was used for the vertical dimension. Auditory elements at lower physical positions were presented with low-pitch signals and elements at higher positions with high-pitch signals. The authors reported the system to be effective but also requiring a certain amount of learning time before actual use.

In our previous research, we proposed a prototype of a spatial auditory interface for controlling a communication device in a car (Sodnik et al., 2008). A group of test subjects was observed while operating a driving simulator and performing different tasks with the communication device. Two auditory interfaces (a spatial and a non-spatial one) were compared to a standard visual interface based on a small display attached to the dashboard. The research showed that the use of an auditory interface while driving is significantly safer than the use of a visual interface due to less distraction of the driver. On the other hand, auditory interfaces proved to be slower. The comparison of the spatial and non-spatial interfaces showed no significant difference between the two.

The main goal of the present research was an effective auditory presentation of the hierarchical menu structures of a typical MS Windows application. We focused on visually impaired and blind computer users, since they deal with auditory interfaces on a regular basis. In our prototype, MS Word menu structure was transformed into an auditory domain with the aid of several different auditory interfaces. Individual menu items were presented with speech-based sounds in different spatial configurations. Taking into consideration the most important Frauenberger's conclusions of (2005,2006), we proposed two different interface concepts that were completely independent of the existing visual interfaces. The visual representation of MS Windows-based menu structures consists of a horizontal and vertical alignment of elements. The elements in the main menu are usually presented horizontally, requiring the user to move left and right in order to choose the desired option, while in the submenus they are aligned vertically. In our auditory interfaces, we focused either on the horizontal or the vertical dimension and aligned all menus and submenus in the same way in order to have the same interaction pattern at all levels.

Both interfaces could be used with or without spatial audio. In the spatial auditory interfaces, up to three spatial sound sources or menu commands were presented simultaneously in order to increase the auditory information flow between the computer and the user. The latter has already been reported and suggested as effective (Hawley et al., 1999; Drullman and Bronkhorst, 2000).

Our first prototype of a similar interface – without simultaneous spatial sounds – was tested with a group of normal sighted users and compared to a visual interface with the same menu structure (Sodnik and Tomazic 2009, 2010). As expected, the visual interface outperformed the auditory interfaces in all evaluated variables. The vertical auditory interface was chosen to be the better one of the two auditory interfaces based on the subjective opinion of the users. However, there was no statistically significant difference between the two auditory interfaces. The most important findings and user feedback of the research were taken into consideration when building the auditory interfaces for the present evaluation study with visually impaired and blind test users.

### *1.5. Main research contribution*

All the authors mentioned in the survey presented different types of spatial auditory interfaces and evaluated their efficiency with user studies. Most of them reported the usefulness of spatial sound in such interfaces without proving their advantage over non-spatial auditory interfaces. Most of them did not use simultaneous audio sources; however, some suggested the method as a possible improvement of their work. Our main research goal was to compare similar auditory interface with and without spatial sound in a controlled environment. We observed the measurable effects of adding multiple spatial sounds, for example in task completion times and navigation performance, as well as in the users' experience with different interfaces and the cognitive workload. The latter was measured with a special set of preset questions and the general comments and suggestions of users. We concentrated on visually impaired and blind test subjects. Some of them also participated as system design advisers and most of them as system evaluators.

Our main research questions were:

1. Can spatial auditory interfaces offer an important improvement to non-spatial interfaces in terms of task completion times and navigation performance within the menu structure?
2. What is the cognitive workload when dealing with spatial auditory interfaces (with multiple simultaneous choices) compared to non-spatial ones?
3. Are there any significant differences between horizontal and vertical spatial auditory interfaces?

Our work is different from previous research in several ways:

1. we concentrated solely on visually impaired and blind computer users due to their substantial experience with auditory interfaces;
2. we compared the same auditory interfaces in spatial and non-spatial conditions;
3. we compared the use of one choice versus multiple simultaneous choices in auditory interfaces;
4. we also focused on the cognitive workload of the test subjects;
5. we compared the vertical and horizontal configuration of the sound sources.

Our hypothesis was that spatial sound can increase the information flow in computer–human interaction by representing multiple simultaneous menu choices. We expected the test subjects to be able to complete their tasks faster and with less interaction when using spatial auditory interfaces. We also expected visually impaired and blind users to positively evaluate the new interfaces and perceive them as an interesting improvement in their everyday interaction with the computer.

## 2. Auditory interface design

Our auditory interface represents a hierarchical menu structure. It is an interface for a simple word processing application, such as MS Word or Wordpad. It enables all operations for creating and editing text documents. A small portion of the structure is presented in Fig. 1.

Individual menu items are presented with sounds—i.e. prerecorded spoken commands. The commands are then played with double speed, which in turn makes them similar to spearcons (Palladino and Walker, 2008), only somewhat slower and still fully intelligible without previous learning. At each level of the menu, the user can select between various available menu items. For example, there are six different items in the main menu: File, Edit, Format, Tools, Table and Help. After selecting one item, a submenu with new options is loaded. Gentle background music is assigned to each main branch of the menu. The music starts playing when the user leaves the main menu and enters one of the submenus. The central pitch of the background music changes according to the depth of the user's position in the menu. The pitch lowers

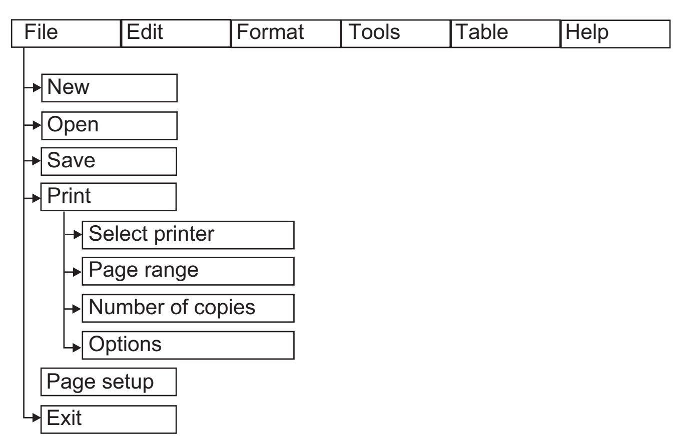


Fig. 1. The hierarchical menu structure was a simplified version of the MS Word menu structure. The structure was defined with the external XML file and prerecorded voice files.

when the user moves one level down into the submenu and rises when he or she moves back up. The changing pitch is intended to give feedback on the current position within the menu.

The sound commands of the individual menu or submenu are organized in two different spatial configurations—horizontal and vertical. Each configuration can be used with or without the spatial effect, thus forming four different auditory conditions.

### 2.1. Horizontal interfaces—H1 and H3

In the horizontal interface, the sound sources or menu commands are placed on the horizontal plane of a virtual ring surrounding the user's head. The radius of the ring is constant, while the spatial angle between the sources changes with their number. For example, if there are six simultaneous sources, the angle between the sources is $360^\circ/6 = 60^\circ$; if there are five sources, the angle is $360^\circ/5 = 72^\circ$, etc. The ring can be turned left or right, thus enabling browsing between individual items in the menu. The angle of a single act of turning is equal to the angle between the sources. In this way, the arrangement of the sound sources remains the same, when the sources shift to the selected direction.

The major difference between the non-spatial H1 and the spatial H3 interface is in the number of sources or commands presented simultaneously. In H1, only one command is played at a time and the user is thus unaware of the entire arrangement of the menus. He or she moves from item to item and selects the desired one. In H3, a maximum of three closest commands are presented almost simultaneously, enabling the user to select any of them. The maximum number of three spatial sources was chosen based on the results of our previous user studies with simultaneous spatial sounds (Sodnik et al., 2008; Sodnik and Tomazic, 2009, 2010). Two sound sources are played in the cases when only two menu items are available.

The initial idea was to present multiple sound sources simultaneously at different spatial positions. Due to the use of the same type of sound signal – i.e. voice command – in all cases, the spatial differentiation of multiple sources does not enable a clear perception and comprehension. Therefore, a short pause of approx. 200 ms is applied between the playbacks of each individual source. For example, the playing sequence of three simultaneous sources is:

$$\text{central source} - 200\text{ ms of silence} - \text{left source} - 200\text{ ms of silence} - \text{right source}$$

With the use of a short pause, we achieved a clear perception and comprehension of the speech sources. The length of the pause (i.e. 200 ms) was determined experimentally.

In the paper, we use the term ''simultaneous sources'' for describing multiple spatial sources played sequentially, but presented simultaneously, describing currently available menu options.

The principles of H1 and H3 are illustrated in Fig. 2.

As demonstrated in Figs. 1 and 2, there is a difference in the absolute angles of the sources relative to the user. In case of H1, the selected item is always in front of the user. The latter is not the case in H3, since up to three items can be selected at any time.

### 2.2. Vertical interfaces—V1 and V3

In the vertical interface, menu commands are aligned vertically in front of the user. The user can move up and down and select the desired command. In the introduction, we mentioned a substantial difference in the localization accuracy in azimuth and elevation. In order to increase the relatively poor elevation perception, an attempt of artificial coding of elevation is used (Susnik et al., 2008). Since up to three vertical sound sources are presented simultaneously, the central source is played with normal pitch, the pitch of the upper source is elevated by 15% and the pitch of the lower source is decreased by 15%. As reported by Susnik et al. (2008), the manipulation of pitch seems to be the most natural way of emphasizing the differences in elevation.

There is no circularity in the vertical interfaces. When the top or the bottom is reached, the user has to start moving in the opposite direction to continue browsing between the items in the menu. The major difference between menus V1 and V3 is in the number of commands presented simultaneously and the number of commands that are available to the user. The upper limit of simultaneous commands in V3 is three. As in the case of H3, a pause of 200 ms between individual playbacks of simultaneous sources is applied in V3 as well.

The main difference between the horizontal (H) and vertical (V) conditions is in the basic arrangement of menu commands and also in the circularity. In the horizontal menu, a user can move in any direction at will, while in the vertical menu there is a peg at both ends of the individual menu or submenu (Fig. 3).

### 2.3. Interaction and navigation

The interaction is based on a standard "QUERTZ" computer keyboard, which is the most common input device used by normal sighted users and blind users as well. The navigation in all four interfaces is based on arrow keys, with "left" and "right" arrows used for rotating the virtual ring in the horizontal menus, and "up" and "down" arrows used for moving the user up and down in the vertical menu.

In H1, the "down arrow" or "W" key are used for confirming the selected menu item and "up arrow" is used for moving one step back in the menu tree—to the previous menu. In V1, "right arrow" or "A" key are used for confirming the selection and "left arrow" is used for moving back in the menu tree.

In H3, "Q", "W" and "E" keys are used for selecting one of the three available options. In the case of just two available items, only "Q" and "E" are used. As in H1, "up arrow" is used for moving one step back in the menu tree.

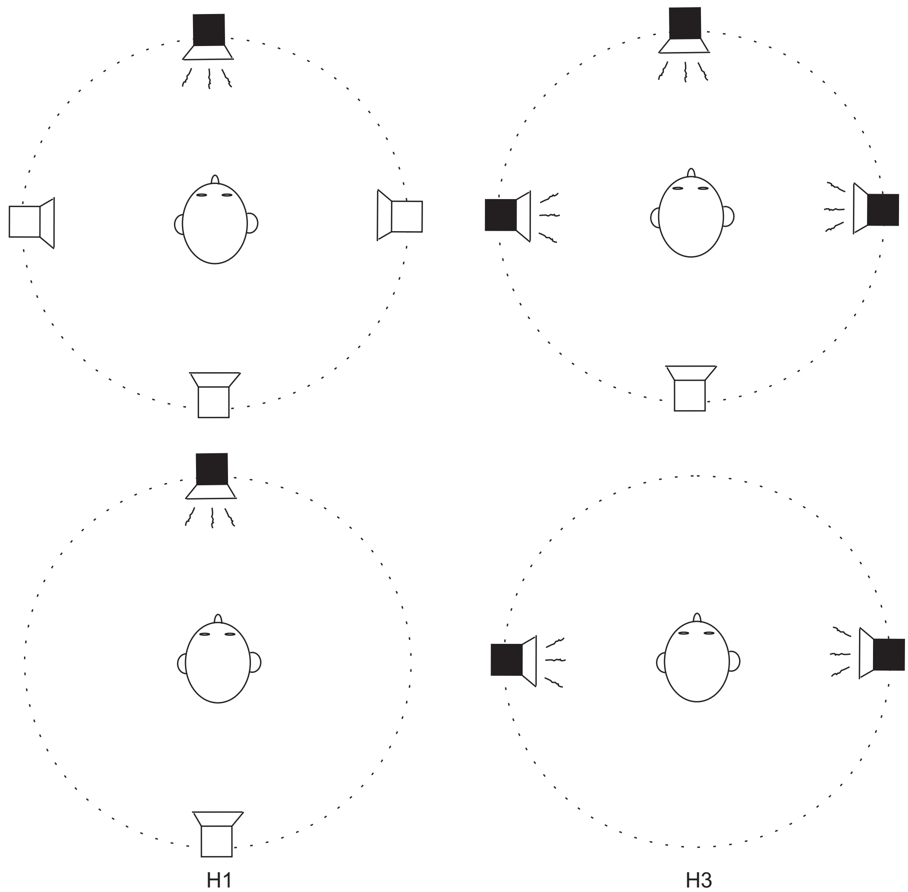


Fig. 2. In the horizontal interfaces, the virtual sound sources were distributed equally around the user's head. The virtual circle could be rotated in any direction. In H1, only one sound source was active at a time, while in H3 up to three sources were presented simultaneously. Fig. 2 shows examples of H1 and H3 auditory menus with four available items (upper pair), and H1 and H3 with two available auditory menus (lower pair).

In V3, "Q", "A" and "Y" are used for selecting the desired option and "left arrow" is used for moving back in the menu tree.

In the non-simultaneous (H1 and V1) conditions, the users preferred to use the arrow keys ("up" and "right") for confirming their menu selections. This simplified the navigation as only one hand was required for the interaction, which the users were familiar with from typical MS Windows applications.

The "Escape" key is used for starting a new task and "Space" key is used for repeating the instructions for each task.

### 2.4. Sound reproduction

The entire system was developed in Java programming language. The generation and reproduction of spatial sound was based on JOAL library (JOAL), a Java wrapper of OpenAL (OpenAL) library. The latter enables the generation of multiple sound sources at various spatial positions relative to the listener. The listener's coordinates as well as the coordinates of the sources are defined in a Cartesian coordinate system and can be dynamically changed or updated at any time. The OpenAL library uses direct hardware support for 3D sound if available. We ran the evaluation tests on Fujitsu Siemens Lifebook (Series E) with Realtek High Definition Audio soundcard which has no hardware support for 3D sound.

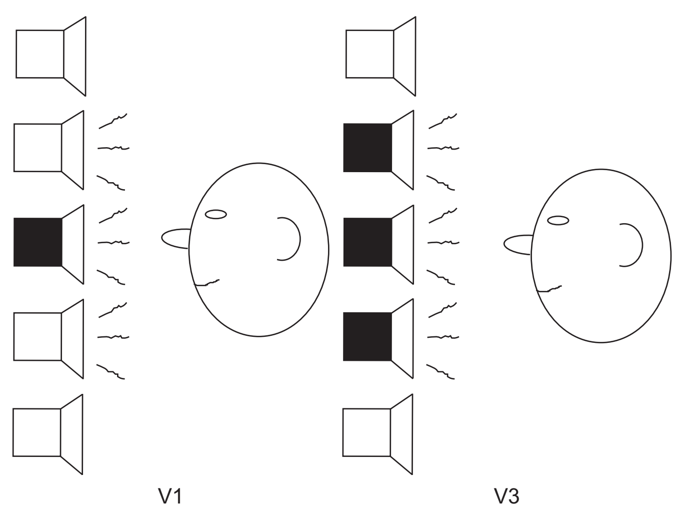


Fig. 3. In the vertical interfaces, the virtual sound sources were aligned vertically in front of the user. The listener could move up and down from top to bottom. In V1, only one sound source was active at a time, while in V3 up to three sources were presented simultaneously. Fig. 3 shows examples of V1 and V3 auditory menus with five available items.

Sennheiser HD 270 studio headphones were used for playback. Due to their very good attenuation of ambient noise (from −10 dB to −15 dB), no special quiet room was required for evaluation studies.

The sound commands in the auditory menus were recorded in a professional recording studio. As already mentioned, the recordings were sped up to double speed before being built into the menus.

A Java based FreeTTS (FreeTTS) speech synthesizer was used for reading the instructions of the tasks to the test subjects.

## 3. User study

Our user study was performed in order to evaluate the efficiency of multiple spatial versus non-spatial sounds for hierarchical menu navigation in a Windows-based environment. The test subjects performed a set of tasks in four different auditory conditions, two spatial ones and two non-spatial ones. The effect of different arrangements of the sound sources in spatial auditory interfaces was also observed.

### 3.1. Test subjects

A total of 12 (4 blind and 8 visually impaired) volunteers participated in our studies. All of them had at least average computer usage skills. Two highly proficient computer users were involved in preliminary studies and development stage of the interfaces (one of them is a computer science teacher for visually impaired and the other is the head of informatics in a large company).

The average age of the test subjects was 33.4 years. Five test subjects were born with visual impairment or lost their vision in their childhood, whereas the others lost their vision recently, from 4 to 10 years ago. All test subjects reported to have normal hearing. None of them has ever participated in similar studies before.

### 3.2. Tasks and experiment conditions

We primarily observed how the users performed different tasks with Windows-like hierarchical menus. The application used in the experiment was a simplified version of the MS Word application. The menu structure enabled all basic word processing functions with corresponding feedback, but without any actual editing or manipulation of text.

Four sets (TS1, TS2, TS3 and TS4) of ten tasks were chosen in order to compare four different interfaces (H1, H3, V1 and V3). Here is an example of the task set TS1:

1. Open new blank document.
2. Change language to French.
3. Change font style to bold.
4. Go to help and activate product by email.
5. Change background color to red.
6. Delete one cell in the table.
7. Undo last command.
8. Set right margin of the document to 2 cm.
9. Change document protection to read-only.
10. Customize toolbars to formatting.

The number of ten tasks within one set was chosen based on preliminary studies performed with normal sighted users. The completion time for the entire set was approximately 5 min, including the short brakes between individual tasks. A certain degradation of performance was noticed when increasing the number of tasks to 15 or even 20. We believe the latter was caused by the drop of concentration when performing the test continually for more than 5 min.

The tasks in the four sets were chosen based on the fact that they were of approximately the same difficulty and that they required the same number of operations to finish the individual task. The instructions for each task were read to the test subject by a speech synthesizer and could be repeated at any time if necessary. An additional set of five test tasks (TTS) was used as a warm-up exercise in order for the test subjects to get acquainted with each new interface.

Four different auditory interfaces formed four different experiment conditions: H1, H3, V1 and V3. H3 and H1 were horizontal auditory menus with and without the addition of spatial sounds. V1 and V3 were vertical auditory menus.

### 3.3. Experiment procedure

All test subjects were asked several basic questions about their age, computer skills, visual impairment and possible hearing disabilities. They were also informed that they can quit the evaluation study at any time and decline any further tests.

At the beginning, the test subjects were given a short explanation of the hierarchical menu structure and the approximate location of various commands. They reported the memorization of the structure to be very easy due to its resemblance to the general menu structure of the MS Word application. All test subjects reported to have at least moderate knowledge of MS Word due to the fact that all of them use it on a daily basis.

Before each new experiment condition (i.e. a different interface), the test subjects were given approx. 3 min to try out the interface for themselves. This, in turn, was followed by the test tasks (TTS) in order to practice the task solving procedure.

Each official task started when the "Escape" key was pressed by the test subject. The instructions were first read clearly and slowly by a synthesized voice. They could be repeated at any time by pressing the "Space" key. The successful conclusion of the task was announced by the same synthesized voice saying "Task completed".

In order to eliminate the learning effects between the different interfaces, four groups of three participants were formed. Each group performed the task sets (TS1, TS2, TS3 and TS4) with the conditions (H1, H3, V1 and V3) in a different sequential order:

1. group: V1, V3, H1, H3;
2. group: H1, H3, V1, V3;
3. group: V3, V1, H3, H1;
4. group: H3, H1, V3, V1.

After the test with the individual interface, the test subjects were asked to evaluate it by answering a set of Questionnaire for User Interface Satisfaction (QUIS) (QUIS, 2006) questions. The users were also interviewed in order to get their personal evaluation of the experiment. Some questions were based on NASA TLX workload test (The Task Load Index) (Hart and Wickens, 1990) in order to evaluate the cognitive workload of the test subjects when performing the tasks.

The duration times of the tasks and user navigation performance were logged automatically by the application. The time measurement started after the "Escape" key was pressed and finished when the "Task completed" message was read to the user. By navigation performance we mean the total number of actions performed by the user or his or her movements within the menu structure.

The measurements collected were the following:

1. task completion time;
2. navigation performance;
3. QUIS test;
4. additional questions and comments on cognitive workload.

## 4. Results

### 4.1. Task completion time

The task completion time was measured automatically by the application. The measurement started when the "Escape" key was pressed by the test subject and ended when the task was actually completed. At the beginning, the instructions on each individual task were read by a synthesized voice. They could be repeated at any time by pressing the "Space" key.

The average task completion time was calculated for each individual interface or experiment condition. Fig. 4 shows the average times with the corresponding confidence intervals based on standard deviation.

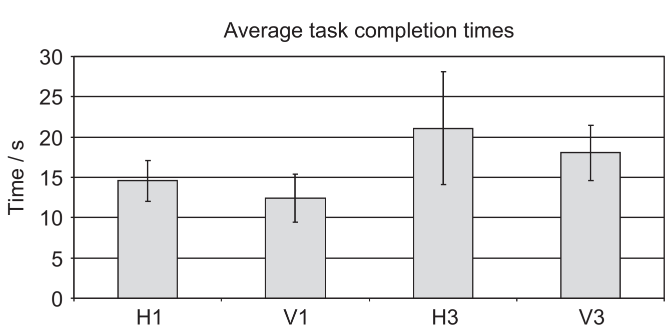


Fig. 4. The average task completion times of all tasks with four interfaces (H1—non-spatial horizontal interface; H3—spatial horizontal interface; V1—non-spatial vertical interface; V3—spatial vertical interface).

As can be seen in Fig. 4, non-spatial sound interfaces (H1 and V1) were faster than interfaces with spatial sound (H3 and V3). A significant difference between the four interfaces was confirmed also by the one-way within groups ANOVA test, which yielded the following result: $F(3,396) = 26.600$, $p < 0.001$. A post-hoc Bonferroni test with a .05 limit on familywise error rate confirmed that task completion times for the H1 and V1 interfaces were significantly lower than task completion times for the H3 and V3 interfaces.

The significance values were:

- H1 vs. H3: $p < 0.001$;
- H1 vs. V3: $p = 0.008$;
- V1 vs. H3: $p < 0.01$;
- V1 vs. V3: $p < 0.01$.

The test, however, showed no significant difference between H1 and V1 ($p = 0.168$), but showed a significant difference between H3 and V3 ($p = 0.027$), with V3 being faster.

The results reported here do not confirm our hypothesis about interactions with spatial audio interfaces being faster. It seems that spatial audio slowed down the user interaction, although the information flow between the computer and the user was undoubtedly higher due to multiple simultaneous sounds or menu commands.

### 4.2. Navigation performance

In the horizontal interface, the user was required to turn the virtual ring left and right, while in the vertical interface, the user had to move up or down the menu in order to select the desired command. Besides the difference in the alignment of the commands, the interfaces also differed in the fact that the vertical ones had a peg at the end of the submenus.

The navigation within an individual menu was logged automatically. Each rotation of the menu (in the case of H1 and H3) or each vertical movement (in the case of V1 and V3) as well as command selection or cancellation was logged automatically.

The log files demonstrate different task solving strategies employed by the test subjects. For example, some users used the same interaction pattern in spatial audio and non-spatial audio interfaces. In H3 and V3, there were up to three simultaneously available menu items and any of them could be selected at any time. Despite this fact, some users concentrated solely on the central source and moved within the menu several times until the desired source was in the center, which resulted in a longer and unnecessary navigation.

The final sums of all the test subjects' actions (i.e. selections + movements) were calculated. Fig. 5 shows the average sums for all four interfaces.

The results demonstrate that spatial auditory interfaces (H3 and V3) required less interaction for completing the tasks. The ANOVA test confirmed a significant difference between the four conditions: $F(3,396) = 12.869$, $p < 0.001$. A post-hoc

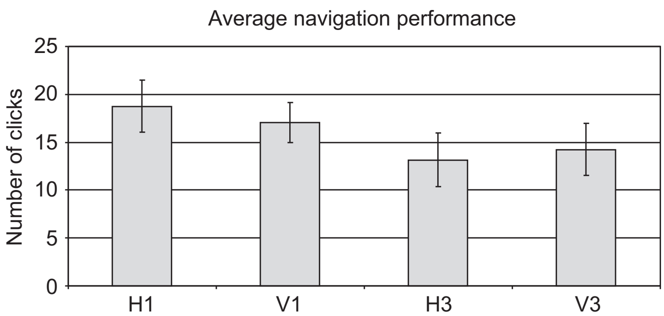


Fig. 5. The average navigation performance (the sum of key presses) of all tasks with the four interfaces (H1—non-spatial horizontal interface; H3—spatial horizontal interface; V1—non-spatial vertical interface; V3—spatial vertical interface).

Bonferroni confirmed a significant difference between the spatial (H3 and V3) and non-spatial (H1 and V1) interfaces.

The significance values were:

- H1 vs. H3: $p < 0.001$;
- H1 vs. V3: $p < 0.001$;
- V1 vs. H3: $p = 0.01$;
- V1 vs. V3: $p < 0.035$.

There was no significant difference between H3 and V3 ($p = 1.000$) or between H1 and V1 ($p = 0.548$).

These results confirm our hypothesis about a more efficient interaction with spatial auditory interfaces.

### 4.3. QUIS test

With the QUIS test, we intended to measure the users' subjective evaluation of the interfaces with a set of questions on the overall satisfaction of the test subjects with an individual interface, system capabilities and interaction procedures. The goal was to get a quantitative evaluation of the system based on a series of subjective answers. The users were asked to evaluate a set of statements with a mark from one to five.

The statements on the overall reaction to the interfaces were the following:

The interface was more:

1. terrible (1) than wonderful (5);
2. difficult (1) than easy (5);
3. frustrating (1) than satisfying (5);
4. inadequate (1) than adequate (5);
5. dull (1) than stimulating (5);
6. rigid (1) than flexible (5).

The statements on the learning abilities were the following:

It was easy or difficult to:

7. learn to operate the system (1–5);
8. explore new features by trial and error (1–5);
9. remember names and use commands (1–5);
10. perform tasks straightforwardly (1–5).

The statements on the system capabilities:

11. the system speed was: two slow (1) or fast enough (5);
12. system reliability was: unreliable (1) or reliable (5);
13. the system was designed for all levels of users: never (1) or always (5).

Most test subjects were uncertain about the meaning of question 6 and thus provided us with no answer. The answers to this question were therefore excluded from further analyses.

The average scores for all questions are listed in Table 1.

The majority of users preferred non-spatial auditory interfaces to spatial auditory ones. Both non-spatial interfaces significantly outscored the spatial ones: $F(3,476) = 50.881$, $p < 0.001$. The significance values for post-hoc tests were:

- H1 vs. H3: $p < 0.001$;
- H1 vs. V3: $p < 0.001$;
- V1 vs. H3: $p < 0.01$;
- V1 vs. V3: $p < 0.01$.

There was no significant difference between the two spatial interfaces (H3 vs. V3: $p = 0.659$) or between the two non-spatial ones (H1 vs. V1: $p = 0.051$).

The spatial auditory interfaces, however, outscored the non-spatial ones in two areas (questions 5 and 11).

The users found both spatial auditory interfaces significantly more stimulating and less dull than the non-spatial ones: $F(3,36) = 6.435$, $p = 0.001$. Table 2 shows the average scores for question 5:

The answers to question 11 demonstrated that the spatial interface H3 was significantly faster than the other three interfaces: $F(3,36) = 4.109$, $p = 0.013$ (Table 3).

The final results of QUIS test do not support the use of simultaneous commands in auditory interfaces. The majority of test subjects found such interfaces interesting and fun to use, but not on a regular basis and not for Windows-based tasks. It is surprising that most users found the spatial auditory interface H3 faster, although the tasks were completed significantly faster with non-spatial interfaces.

Table 1
The average scores of QUIS test for all questions on a scale from 1 to 5 (with 5 being the best possible mark).

| H1 | V1 | H3 | V3 |
|---|---|---|---|
| 3.60 | 3.96 | 2.63 | 2.83 |

Table 2
The average scores of QUIS test for the question whether the individual interfaces was more dull (1) than stimulating (5).

| H1 | H3 | V1 | V3 |
|---|---|---|---|
| 2.6 | 3.8 | 2.8 | 4.0 |

Table 3
The average scores of QUIS test for the question whether the individual interfaces was slow (1) or fast enough (5).

| H1 | H3 | V1 | V3 |
|---|---|---|---|
| 2.8 | 4.1 | 2.9 | 3.0 |

It seems that the users had the impression of being faster due to less navigation and key pressing.

### 4.4. User subjective comments

We were interested in how the users felt when solving tasks with different interfaces. Each user was asked a set of questions in order to obtain some descriptive and informal answers. Our main goal was to get their feedback on the cognitive workload in terms of mental, physical and temporal demand, and also on the general satisfaction with the interfaces. Some examples of the questions were:

- How much mental and perceptual activity was required for selecting the desired option, moving within the menu, etc.?
- How much physical activity was required?
- How hard did you have to work (mentally and physically) to accomplish your level of performance?
- What do you think about the background music in the submenus?
- etc.

#### 4.4.1. Non-spatial auditory interfaces

No specific comments were made about non-spatial interfaces. The test subjects reported their resemblance to the typical Windows-based auditory menus except for the difference in the interaction procedure. In our case, each auditory menu isolated a selection of menu items in one dimension (horizontal or vertical). The test subjects needed a couple of test tasks to adopt the pattern and after that reported no further problems. The cognitive workload was relatively low and not higher than when dealing with any existing auditory interface. The physical activity was also reported as similar to the one they were used to.

They all noticed the difference at the ends of the submenus. The vertical menu ended with a peg, requiring that the user moved back when the bottom or top was reached, while in the horizontal menu the users could move circularly. The majority of users did not mind the difference, while some of them thought the menus with pegs improved the overall orientation. Some suggested the peg to be optional, for example enabled when learning the menu structure and navigation and then disabled for fast task solving.

#### 4.4.2. Spatial auditory interfaces

The majority of users reported that the use of spatial auditory interfaces requires a lot of mental concentration and cognitive workload. They complained about it being very difficult to listen to three simultaneous sources and to perceive and understand their content. This requires absolute focus and no disturbances in order to be successful. The latter was also noticed by the person conducting the tests, since a relatively long reaction time was required by the test subjects in order to select one of the three available menu items after hearing all of them.

Sometimes the high temporal demand caused the test subjects to forget the goal of their tasks, requiring them to play the instructions again. It was the test subjects that suggested a short pause (currently 200 ms) between the simultaneous sound sources in order for them to be distinguishable from one another.

Some of the test subjects concentrated solely on one sound source, usually the center one, and tried to ignore the others. They described the interfaces as too noisy and suggested the option of turning the other sounds off. Some of them suggested some sort of two-mode interfaces with optional spatial support when needed or required.

The test subjects were also asked to compare the horizontal and vertical interfaces. Most of them confirmed a much better localization of sound sources in the horizontal interfaces, but did not report any effect on their overall performance. The coding of elevation with a difference in central pitch seemed to be a very good solution. The test subjects did not report any problems separating and localizing sources at different elevations.

Some test subjects approved of the multi-sounds environment. They described the spatial sounds as an excellent variegation in the otherwise boring and dull auditory interfaces. The difference in central pitch as an additional information on the elevation of individual voice commands in the vertical menus was perceived as a great solution. Some test subjects suggested something similar also for the horizontal menus, for example using three different voices or speakers for three simultaneous sources.

#### 4.4.3. General comments

All test subjects were enthusiastic about the idea of background music functioning as an indicator of submenus and changes in pitch indicating the current depth in the menu. The users reported the music to be an efficient feedback and also great variegation in the auditory menu.

As already mentioned earlier, the voice commands in the menus were played with double speed. The test subjects reported no problems with understanding the meaning; however, some suggested the speed of playback to be changeable.

The test subjects also gave some comments regarding the headphones that were required for the playback of spatial sounds. Headphones in general prevent the users to perceive other sounds from their environment and thus do not enable them to listen to the radio or talk to someone when interacting with the computer.

## 5. Discussion

In our study, we designed and evaluated spatial auditory interfaces for visually impaired and blind computer users, focusing on hierarchical menu navigation as a typical MS Windows based interaction. By spatial auditory interface we mean the use of multiple simultaneous commands at various spatial positions relative to the listener—i.e. the user. A simple application was developed for performing different word processing tasks via an auditory interface. Twelve blind or visually impaired test subjects participated in the user study, in which they performed ten different tasks with four different interfaces. We evaluated the interfaces with regards to task completion times, navigation performance and QUIS test results. Subjective comments and suggestions were also collected.

The main goal of the study was the evaluation of spatial sound in such interfaces. Two spatial auditory interfaces were compared to two identical non-spatial ones. We were interested whether spatial sound can be used effectively in such interfaces and what its major benefits or drawbacks are. We started the study with the hypothesis that spatial sound and the simultaneous presentations of multiple options can speed up the interaction with the menu and enable a much faster completion of various tasks. We also intended to compare two different interaction patterns and two different spatial configurations of the sound sources—horizontal and vertical.

We established a significant difference between the spatial and non-spatial auditory interfaces; however, the results did not confirm our hypothesis. The spatial auditory interfaces proved to be slower, although they enabled the completion of tasks with fewer interactions and key presses. We believe the latter can be explained with a high cognitive workload perceived and reported by the test subjects. Solving tasks with spatial auditory interfaces required absolute concentration of the test subjects. Although up to three sound sources can be identified and understood simultaneously, such listening and comprehension requires high concentration and mental demand. The time required by the brain to process the selection of one of the sources slowed the users down and increased the time between individual actions and interactions with the interface. It is interesting that the users had an impression of finishing the tasks faster with spatial interfaces, which, however, was not really the case.

In non-spatial auditory interfaces, only one menu option or command is available at a certain time and there is always only one possible interaction pattern for the command to be executed. The latter enables the users to establish some sort of routine in solving tasks and to remember the positions of commands and the combinations of keys they need to press in order to reach them. In the spatial auditory interfaces with multiple simultaneous choices and various possible interactions or selection processes, memorizing such navigation routine is much harder and less probable. We believe this fact also contributed significantly to longer task completion times in the case of multiple simultaneous choices. On the other hand, the interaction patterns were shorter and required less navigation within the menus. Based on the studies of interaction patterns and the reaction times of the users, we can establish that the choice of navigation pattern contributed much less to the final performance compared with the simultaneous representation of multiple menu choices and the consequential mental demand on the user.

We did not establish any significant differences between the vertical and horizontal interfaces, although we expected the horizontal configuration to be more efficient and better accepted by the users. Sound sources at different horizontal positions or azimuths can in general be localized much more accurately than sound sources at different elevations. This phenomenon was already explained in the introduction.

We believe the high mental demand was the main reason why the users preferred non-spatial auditory interfaces and gave them higher marks in QUIS test. They found them more wonderful, easier to use, more satisfying and more adequate, but less stimulating. They also thought non-spatial interfaces were easier to learn to operate, explore new features by trial and error, and remember names and use of commands. They said that performing tasks was more straightforward and that the system reliability was higher. In addition, the test subjects thought that non-spatial interfaces are more appropriate for all levels of users. As already mentioned, they also claimed spatial interfaces were faster than non-spatial ones.

Visually impaired and blind people are used to multitasking when interacting with computers. They rely solely on the auditory channel, but they can still listen to music or somebody talking to them while at the same time using an auditory interface. They do not approve of headphones, since they occupy their auditory channel completely and thus prevent them from performing multiple simultaneous tasks. However, headphones are mandatory in spatial auditory interfaces in order to play spatial sounds correctly. Perhaps multi-channel systems could be used as well, but in this case a more or less total silence would be required in order to perceive spatial sounds correctly.

## 6. Conclusion

Visually impaired and blind people are forced to use their auditory channel as the primary communication channel when interacting with computers and other machines. The auditory channel enables much lower bit rates than the visual channel, used by normal sighted computer users. The information in the auditory channel has to be transferred in serial in order to be understood. One possible idea of increasing the auditory information flow is the use of spatial sounds. More sounds can be perceived simultaneously if they originate at different spatial positions relative to the listener.

The use of multiple spatial sounds for the navigation in a hierarchical menu does not seem to be the best solution in auditory interfaces for blind computer users. Spatial auditory interfaces with multiple sounds increase the information flow between the computer and the user, thus enabling shorter navigation within menu structure. On the other hand, spatial auditory interfaces are slower than non-spatial ones, since the perception of multiple simultaneous sounds requires a very high degree of concentration and extremely high cognitive workload. The latter reflects in slower reaction times and increased brain activity.

Spatial sound can be an interesting complement to various non-spatial auditory interfaces. Interaction with such systems can be challenging and fun for some time, but not when they are used constantly and for everyday computer tasks. In such cases, blind computer users want their auditory channel to be occupied as little as possible in order to be able to perform more simultaneous tasks. Spatial sounds should therefore be used in games and entertainment environments, where total concentration and high mental demand are expected and even desired. For everyday computer tasks, such as word processing or mail composition, a hybrid auditory interface might be appropriate, in which the use of simultaneous spatial sounds can be turned on or off.

It is important to point out that a group of 12 blind or visually impaired test subjects represent a relatively small number of evaluators. Although some significant differences between the interfaces were established and supported by the ANOVA statistical tests, we believe the experimental power could be increased by involving more test subjects and increasing the number of test trials. All participants were skilled computer users, but novice users of the proposed auditory interfaces. We believe the continuous use of spatial auditory interfaces or a more extensive training period could improve the navigation performance and shorten the required task completion times.

In the present experiment, only standard and widely available hardware was used. The relatively low quality of spatial sound reproduction and localization accuracy did not prove to play an important role, since the test subjects had no problems localizing sound sources at different spatial positions. Irrespective of this fact, we believe that some improvement in this area could also be made and that a high quality spatial sound could improve the interaction with spatial auditory interfaces. We also mentioned a major drawback of using headphones in our experiment, since they block the auditory channel completely, despite the fact that they at the same time enable the interfaces to be used in noisy environments and with a high degree of privacy. Perhaps non-isolated headphones without attenuation of the surrounding sounds should be considered as well. On the other hand, the use of multiple speakers and surround systems could increase the accuracy of spatial sound localization and make the auditory channel of the listeners available for other sounds as well. In this case, the users would, however, need to work in silent and isolated environments.

## Acknowledgements

This work has been supported by the Slovenian Research Agency within the research project: "Assisting visually impaired people in using the computer with the aid of spatial sound". The authors would like to thank Mr. Jožef Gregorc from the "Inter-municipal society of blind and visually impaired people in Kranj" for his feedback and help with the development of the system and the organization of the user study group.

## References

Algazi, V.R., Duda, R.O., Thompson, D.M., Avendano, C., 2001. The CIPIC HRTF Database. In: Proceedings of the 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics. Mohonk Mountain House, New Paltz, pp. 99–102.

Arons, B., 1992. A review of the cocktail party effect. Journal of the American Voice I/O Society 12 (July), 35–50.

Begault, D.R., 1994. 3-D Sound for Virtual Reality and Multimedia. Academic Press, Cambridge.

Blauert, J., 1997. Spatial Hearing: The Psychophysics of Human Sound Localization, revised ed. MIT Press, Cambridge, USA.

Brewster, S.A., Wright, P.C., Edwards, A.D., 1993. An evaluation of earcons for use in auditory human–computer interfaces. In: Proceedings of the INTERCHI '93 Conference on Human Factors in Computing Systems, Amsterdam, pp. 222–227.

Brewster, S., Lumsden, J., Bell, M., Hall, M., Tasker, S., 2003. Multimodal 'Eyes-Free' interaction techniques for wearable devices. SIGCHI Conference on Human Factors in Computing Systems 5 (1), 473–480.

Cohen, M., Ludwig, L.F., 1991. Multidimensional audio window management. International Journal of Man–Machine Studies 34 (3), 319–336.

Cohen, J., 1992. Monitoring background activities. In: Proceedings of the First International Conference on Auditory Display, Santa Fé, USA, Addison-Wesley, pp. 499–532.

Crispien, K., Fellbaum, K., Savidis, A., Stephanidis, C., 1996. A 3D-auditory environment for hierarchical navigation in non-visual interaction. In: Proceedings of the Third International Conference on Audio Display (ICAD '96), Palo Alto, USA, pp. 18–21.

Drullman, R., Bronkhorst, A.W., 2000. Multichannel speech intelligibility and talker recognition using monaural, binaural and three-dimensional auditory perception. Journal of the Acoustical Society of America 107 (4), 2224–2235.

Frauenberger, C., Putz, V., Holdrich, R., 2004. Spatial auditory displays: a study on the use of virtual audio environments as interfaces for users with visual disabilities. In: Proceedings of the Seventh International Conference on Digital Audio Effects (DAFx'04), Naples, Italy, pp. 384–389.

Frauenberger, C., Putz, V., Holdrich, R., Stockman, T., 2005. Interaction patterns for auditory user interfaces. In: Proceedings of the 11th International Conference on Auditory Display (ICAD06), Limerick, Ireland, pp. 154–161.

Frauenberger, C., Stockman, T., 2006. Patterns in auditory menu design. In: Proceedings of the 12th International Conference on Auditory Display (ICAD06), London, UK, pp. 141–147.

FreeTTS. From: 〈http://freetts.sourceforge.net/docs/index.php〉 (01/2010).

Gardner, B., Martin, K., 1994. HRTF measurements of a KEMAR dummy-head microphone. MIT Media Lab Perceptual Computing—Technical Report #280.

Gaver, W., 1986. Auditory icons: using sound in computer interfaces. Human Computer Interaction 2 (2), 167–177.

Hart, S.G., Wickens, C., 1990. Workload assessment and prediction. In: Booher., H.R. (Ed.), MANPRINT: an approach to systems integration. Van Nostrand Reinhold, New York, pp. 257–296.

Hawley, M.L., Litovsky, R.Y., Colburn, S., 1999. Speech intelligibility and localization in a multi-source environment. Journal of Acoustic Society of America 105 (6), 3436–3448.

Internet World Stats. Usage and population statistics. From: 〈http://www.internetworldstats.com/stats.htm〉 (01/2010).

Jagdish, D., Gupta, M., 2008. Sonic grid: an auditory interface for the visually impaired to navigate GUI-based environments. In: Proceedings of the IUI'08, Maspalomas, Gran Canaria, Spain, pp. 337–340.

JAWS for windows {version 9.0}. Freedom Scientific. From: 〈http://www.freedomscientific.com/products/fs/jaws-product-page.asp〉 (01/2010).

Jin, C., Corderoy, A., Carlile, S., Schaik, A., 2004. Contrasting monoaural and interaural spectral cues for human sound localization. Journal of the Acoustic Society of America 115 (6), 3124–3141.

JOAL. From 〈https://joal.dev.java.net/〉 (01/2010).

Kobayashi, M., Schmandt, C., 1997. Dynamic Soundscape: mapping time to space for audio browsing. Human Factors in Computing Systems. In: Proceedings of the CHI 1997, pp. 194-201.

NASA TLX for Windows. From: 〈http://www.nrl.navy.mil/aic/ide/NASATLX.php〉 (01/2010).

OpenAL. From: 〈http://connect.creativelabs.com/openal/default.aspx〉 (01/2010).

Palladino, D.K., Walker, B., 2008. Navigation efficiency of two dimensional auditory menus using spearcons enhancements. In: Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting, San Antonio, TX, pp. 1262–1266.

Savadis, A., Stephanidis, C., Korte, A., Crispien, K., Fellbaum, K., 1996. A generic direct-manipulation 3D-auditory environment for hierarchical navigation in non-visual interaction, Proceedings of Assets'96. ACM, New York, USA, pp. 117–123.

Schmandt, C., Mullins, A., 1995. AudioStreamer: exploiting simultaneity for listening, Proceedings of the CHI. ACM, pp. 218–219.

Sodnik, J., Sušnik, R., Bobojević, G., Tomažič, S., 2004. Smerna ločljivost navideznih izvorov zvoka pri človeku. Electrotechnical Review 71 (3), 121–127.

Sodnik, J., Sušnik, R., Štular, M., Tomažič, S., 2005. Spatial sound resolution of an interpolated HRIR library. Applied Acoustics. 66 (11), 1219–1234.

Sodnik, J., Dicke, C., Tomazic, S., Billinghurst, M., 2008. A user study of auditory versus visual interfaces for use while driving. International Journal of Human–Computer Studies 66 (5), 318–332.

Sodnik, J., Tomazic, S., 2009. Spatial auditory interface for word processing application. In: Proceedings of the Second International Conferences on Advances in Computer–Human Interactions (ACHI '09), Martinique, France, pp. 271–276.

Sodnik, J., Tomazic, S., 2010. A user study of auditory versus visual interfaces for desktop computer. Transactions on Advanced Research 6 (1), 17–22.

Stqifelman, L.J., 1994. The cocktail party effect in auditory interfaces: a study of simultaneous presentation. MIT Media Laboratory Technical Report.

Susnik, R., Sodnik, J., Tomazic, S., 2008. An elevation coding method for auditory displays. Applied Acoustics 69 (3), 233–241.

Walker, B.N., Nance, A., Lindsay, J., 2006. Spearcons: speech-based earcons improve navigation performance in auditory menus. In: Proceedings of the International Conference on Auditory Display (ICAD 2006), London, England, pp. 63–68.