A User Study of Auditory Versus Visual Interfaces for Desktop Computer

Sodnik, Jaka and Tomažič, Sašo

Abstract— *This paper describes a user study on the interaction with auditory interfaces for desktop computers intended to be used by blind computer users. Two new auditory interfaces are proposed, and their effectiveness and efficiency are compared to a standard visual interface. All three interfaces represent a hierarchical menu structure of a simple word processing application. The two auditory interfaces consist of spatialized auditory cues implemented in two different spatial configurations. The AH auditory menu has a ring shaped horizontal configuration of the sound sources, whereas the AV auditory menu has a vertical configuration of the sound sources. The use of spatial sounds enables the simultaneous playback of multiple sounds (i.e. menu commands), which increase the information flow between the user and the application. A pair of headphones is used for the playback of the auditory interfaces, while the 17" computer screen is used for the visual menu. The interaction with the interface is preformed by the QUERTY computer keyboard. The task completion times, interaction efficiency, perceived workload and overall user satisfaction are evaluated. Sixteen test subjects who reported no visual deficiencies were asked to perform five different tasks with two auditory interfaces and a classic visual GUI. The main goal of the experiment was to compare all three interfaces and select the more efficient and better evaluated of the two auditory ones. The experiment proved that both auditory interfaces are effective to use in the interactions with desktop computers; however, they are not faster than the visual interface. The AV auditory interface proved to be more efficient and better evaluated by the users and will therefore be used in further experiments with visually challenged computer users.*

Index Terms— *Human-computer interaction, spatial audio, desktop computer, blind users.*

1. INTRODUCTION

Nowadays there are over a billion computer users worldwide and the number is still increasing rapidly. Visually impaired people constitute a significant number of computer users which use the computers especially for various word processing applications and web browsing. The majority of existing computer applications for blind computer users is based solely on simple voice synthesizers which read the content of the documents word by word [1][2]. They are sometimes combined with tactile keyboards. We believe all those systems suffer from slowness and low information flow due to the narrow human auditory channel.

We developed a special auditory interface for controlling the desktop application which could also be used by blind users. In this paper we focus primarily on the hierarchical auditory menus for the interface that can be controlled via the computer keyboard. The interface is based on multiple spatial sounds played simultaneously, which enables high information flow between the user and the computer and therefore improves the user interaction with the computer.

The main goal of our user study was to evaluate the effectiveness of such an acoustic interface in the interaction with an application. Three different interfaces - two auditory and one visual - were compared, all of them based on the same hierarchical menu structure. The GUI in the visual interface resembles the interfaces of the typical word processing application such as MS Word or Open Office Writer. The menu items in the two acoustic interfaces are presented with spatial sounds coming from different pre-fixed positions. All three interfaces were evaluated with an extensive user study, resulting in sets of objective measurements and subjective user comments. The main objective of the experiment was to select one of the two proposed acoustic interfaces to be used in the future development of auditory software for blind computer users.

2. RELATED WORK

Several researchers focused on spatial auditory interfaces used for browsing or manipulating data. Many different solutions were proposed, each of them with its specific benefits and limitations.

Schmandt described a virtual acoustic environment used for browsing collections of related audio files [3]. The user had the possibility of traveling up and down the virtual hallway by head motion, passing various rooms on the left

and right side and selecting and playing audio clips.

A surrounding three-dimensional audio interface for various user applications was proposed by Frauenberger and Noistering [4]. The binaural audio rendering implemented directional hearing and room acoustics via headphones. The users could move around freely using a joystick. The system was capable of producing a virtual audio reality that functioned as some sort of a user interface for people with a visual impairment.

A prototype of such an auditory interface was then evaluated with common interactions like tasks with menus, text input and dialogs [5]. The authors concluded that such an auditory interface could be used effectively for presenting various interface components and that there is no significant difference in the user performance between normal sighted and visually impaired users. They also pointed out the importance of the quality of spatial sounds. The real world applications such as MS Explorer also proved to be manageable by auditory display [6].

An example of an auditory interface for a communication device in a vehicle was proposed and evaluated by Sodnik et al. [8]. The interface was based on speech commands positioned in space and manipulated through a custom interaction device.

Mynatt proposed a general methodology for transforming graphical interfaces into non-visual interfaces [7]. Various salient components of graphical interfaces were transformed into auditory interfaces by using the so-called auditory icons. These types of interfaces were designed primarily to meet the needs of blind computer users.

3. AUDITORY INTERFACE DESIGN

In our study, we designed a simple word processing application that is be controlled by the proposed auditory interfaces. The hierarchical menu structure of the application is based on MS Word menu structure. The auditory interface enables all the operations needed for handling documents. The menu commands are represented with sound sources (recorded voice commands). At each level of the menu, all available options are presented to the user. After the user selects an option, a new level of the menu is loaded. The options at the basic or main level of the menu are File, Edit, Format, Tools and Help. If File is selected, the options on the next level are New, Open, Save, Print, Page Setup and Exit. When the user selects an option in the basic level and moves to the next level, a gentle music starts playing in the background. Different music is assigned to each branch of the menu (File, Edit, Format, etc.) in order to identify the selected branch. The sound sources at various levels of the menu are organized in two different ways, thus constituting two auditory interfaces.

A. AH interface

In the AH interface, the individual sound sources are placed on a virtual horizontal ring that surrounds the user's head [8]. The spatial angle between two individual sources changes with the number of sources constituting the ring. For example, if there are six simultaneous sources, the angle between the sources is 360° / 6 = 60°, if there are four sources, the angle is 360° / 4 = 90°, etc. The selected source is always directly in front of the user.
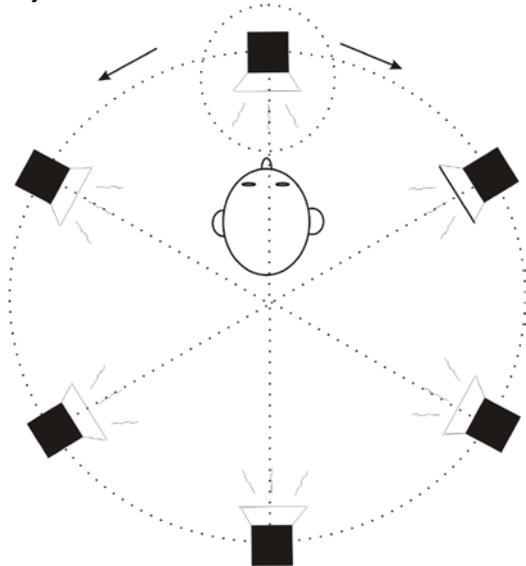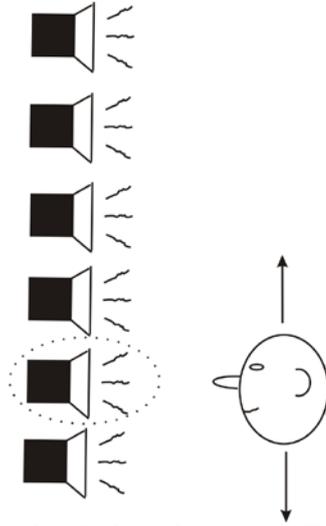

Fig. 1. The basic principle of AH menu. The individual sound sources are placed on a virtual horizontal ring that surrounds the user's head.

The ring can be turned in any direction (left or right) in order to select the desired menu option. The user is not located directly in the centre of the virtual ring, but slightly closer to the front. As the front item is the closest to the user, it is therefore also the loudest one.

B. AV interface

In the AV interface, the individual sound sources of a specific level are aligned vertically in front of the user.

The vertical alignment resembles the typical menus used in graphical interfaces. The user can browse between the options by moving up and down. Again, the option or the command directly in front of the user is the selected one. Important features of the AV interface are the pegs at both ends of the menu. When the top or bottom of the menu is reached, the user has to go back in the opposite direction, while in the AH menu the horizontal circle can rotate indefinitely.

Fig. 2. The basic principle of AV menu. The individual sound sources of a specific level are aligned vertically in front of the user.

In t he A V i nterface, t he c entral pi tch of the background music changes according to the current position of the user. If the user is located at the bottom of t he c urrent m enu, t he pi tch i s very low, and i f he i s l ocated at t he t op of t he menu, the pitch is high.

In the A V m enu, t hree s patial s ound s ources are played simultaneously, r epresenting t he closest t hree m enu opt ions. T he us er c an therefore always hear the selected, the previous and the following option.

C. Interaction

In both interfaces, the interaction is preformed with the use of a QUERTY computer keyboard. In the AH interface, the "LEFT" and " RIGHT" k eys are used for turning the c ircle i n t he des ired direction, whereas i n the AV i nterface, t he "UP" and "DOWN" k eys are used for moving the user up and dow n in the m enu. The "Q" k ey is used for confirming the s elected m enu opt ion and moving to the next level. The "W" key is used for canceling the option and moving one step back in the menu. The "SPACE" key is used to replay the current c hoice w ithout m oving t he us er up or down in t he A V or t urning t he c ircle i n A H interface.

The i nteraction i n t he v isual i nterface is performed with a c omputer mouse. The menu is shown on a 17" computer screen.

D. Sound reproduction

The s ound c ommands w ere r ecorded in a professional r ecording s tudio by a female speaker. T he s ound c lips w ere then s ped t o double speed (similar t o " spearcons" pr oposed by Walker [9]) in order to increase the navigation speed. The auditory interfaces were developed in Java programming language with OpenAL sound positioning library. T he latter w as used to create spatial sounds at v arious positions by using Creative X-Fi CMSS-3D technology on X-Fi

Extreme Gamer sound card. Sennheiser HD 270 headphones w ith a v ery good at tenuation of ambient noise (from -10 dB to -15 dB) were used for the playback of the auditory menus.

4. USER STUDY

A t otal of 16 t est s ubjects participated in the experiment. T hey al l r eported normal sight and hearing. T hey w ere as ked t o per form 5 di fferent tasks:

1. C hange l anguage of t he doc ument t o US English.
2. Change font size to 16.
3. Sort table in the document by field 2.
4. Change font color to blue.
5. G o t o H elp and ac tivate t he pr oduct by phone call.

All test subjects performed the tasks with three different interfaces: a visual interface (V), the AH interface and the AV interface – 15 tasks per test subject. The audi tory i nterfaces (AH and A V) were used w ithout any v isual s upport (in bot h cases the computer screen was blank).

A 5 -minute br eak w as as signed af ter eac h interface in order for the test subjects to answer some ques tions on t he s ubjective ev aluation of the interface and to c omplete t he Q uestionnaire for User I nterface S atisfaction (i.e. t he Q UIS test).

In order to eliminate t he l earning ef fects, t wo groups of 8 t est s ubjects w ere f ormed. E ach group performed the tasks with the interfaces in a different order:

1. group: V, AH, AV.
2. group: V, AV, AH.

Since the same menu structure was used with all t hree i nterfaces, t he v isual i nterface (V) w as used in the experiment primarily to introduce the menu structure to the test subjects. That is why it was used as the first interface in both groups.

The t est s ubjects w ere as ked t o complete the tasks as fast as possible. The duration times and the navigation performance were logged automatically.

5. RESULTS

In the five tasks preformed by 16 test subjects five variables were evaluated:

- task completion times
- interaction efficiency
- QUIS test
- overall ranking of the interfaces
- subjective comments on the experiment and interfaces

The m ain r esults w ith c orresponding interpretations are s ummarized i n t he f ollowing subchapters. T he anal ysis of variance (ANOVA) is us ed t o es tablish t he significant di fferences i n the results.

A. Task completion times

The task completion time was measured automatically for all three interfaces. The timer started right after the task was:

- read to the user by the voice synthesis system in the auditory interfaces or
- shown on the screen in the visual interfaces.

The timer stopped when the task was completed. The average task completions times in seconds are shown in Figure 3.

A significant difference between the three interfaces could be confirmed for tasks 1, 2 and 3, with the visual interface always being significantly faster than the two auditory interfaces:

$F_{Task1}(2, 21) = 11.367$, $p < 0.01$;
$F_{Task2}(2, 21) = 3.464$, $p = 0.05$;
$F_{Task3}(2, 21) = 6.523$, $p < 0.01$.

No significant difference could be noted for tasks 4 and 5.
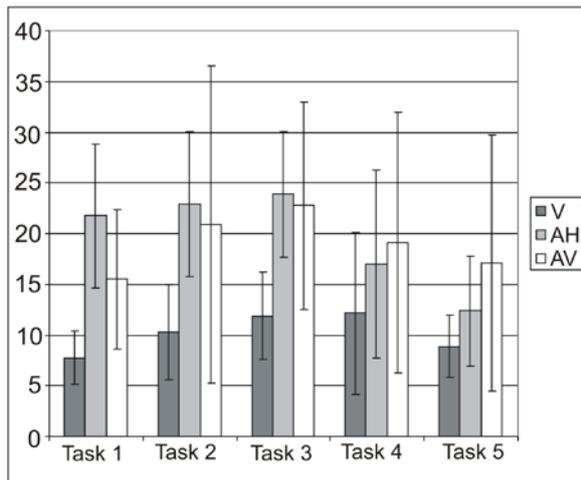

Fig. 3.   The average task completion times for all five tasks shows a significant difference between the three interfaces for tasks 1,2 and 3.

The shortest task completions times were measured for the visual interface. The confidence intervals for the visual interface were small as well, which signifies small variations of the results. On the other hand, no significant difference could be found between the two auditory interfaces. The latter can be confirmed also by comparing the average completions times of all tasks:

TABLE I
THE AVERAGE TASK COMPLETION TIMES FOR THE THREE
INTERFACES SHOW NO SIGNIFICANT DIFFERENCE BETWEEN THE TWO
AUDITORY INTERFACES

Interface	Time / s
V	10.2
AH	19.6
AV	19.1

We expected the visual interface to be the fastest one, since the majority of test subjects were already skilled in using MS Word GUI for editing documents. We believe that the use of a computer mouse in the visual interface and the use of a keyboard in the auditory interfaces also contributed to the established difference.

B. Interaction efficiency

The interaction efficiency was measured by counting the movements of the user in the menu (i.e. how many times the user pressed the "LEFT" and "RIGHT" or "UP" and "DOWN" keys) and the selections of various menu items (i.e. how many times the user pressed the selecting buttons). Both parameters were measured automatically by the application itself. The results were then compared with the optimal interaction procedure necessary to complete a specific task.

The interaction efficiency was measured only for the auditory interfaces. Figure 4 shows the average efficiency (the sum of movements and selections) per task:
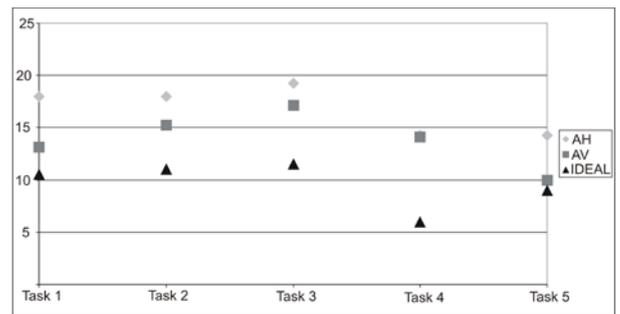

Fig. 4.   Menu navigation performance shows higher effectiveness of AV interface.

The AV interface proved to be more effective and easier to use since less navigation was required for completing the tasks.

The interaction in both cases proved to be far from ideal. We believe the interaction could be improved significantly with continuous practice and a better knowledge of the menu structure.

C. QUIS test

The QUIS test was designed to assess the users' subjective satisfaction with specific aspects of the human-computer interface. In our experiment, we intended to measure the reaction of the users to the software used in the experiment. We asked the users to rank the interface on a scale of 1 to 5 (1 being entirely false and 5 being entirely true) based on the following statements about each individual interface:

1. the interface was more wonderful than terrible (Q1)
2. the interface was more easy than difficult (Q2)

3. the interface was m ore s atisfying t han frustrating (Q3)
4. the interface w as m ore adequat e t han inadequate (Q4)
5. the interface was more stimulating than dull (Q5)
6. the i nterface w as m ore f lexible t han r igid (Q6)
7. it was easy to learn how t o oper ate t he system (Q7)
8. it was eas y to explore new features by trial and error (Q8)
9. it was eas y t o r emember nam es and us e commands (Q9).

The significant difference between the interfaces c an be es tablished f or t he ques tions Q1, Q4, Q5, Q6 and Q7:

$F_{Q1}(2, 21) = 12.014$, $p < 0.01$;
$F_{Q4}(2, 21) = 7.000$, $p < 0.01$;
$F_{Q5}(2, 21) = 5.509$, $p < 0.05$;
$F_{Q6}(2, 21) = 5.619$, $p < 0.05$;
$F_{Q7}(2, 21) = 7.768$, $p < 0.01$.

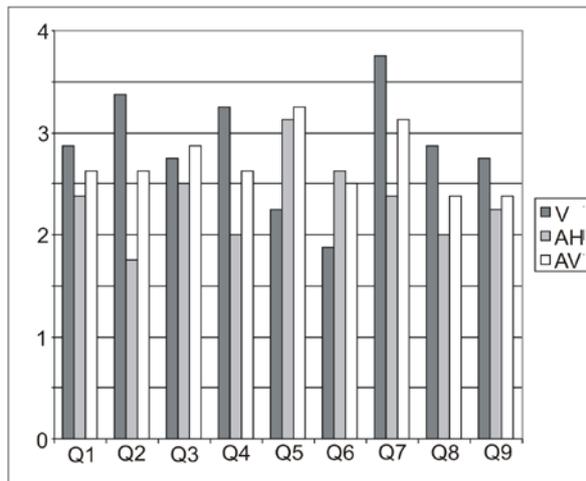Figure 5 s hows t he av erage QUIS scores of the individual interfaces:

Fig. 5.   QUIS scores for the three interfaces show significant differences between the interfaces for questions Q1, Q4, Q5 Q6 and Q7.

The visual i nterface got s ignificantly hi gher scores in questions Q 1, Q 4 and Q 7. T he l atter signifies t hat the test s ubjects f ound t he v isual interface much easier to use than the two auditory i nterfaces. T hey al so des cribed i t as more adequate and s tated it was easier to learn the commands and their positions in the menu.

The auditory interfaces scored significantly more points in questions Q 5 and Q 6. The users described t he audi tory i nterfaces as m ore stimulating and more flexible than the visual one.

No s ignificant di fference c ould be established in questions Q3, Q 8 and Q 9. I t m eans t hat al l three interfaces s eemed t o be equal ly s atisfying and equally difficult when exploring new features by t rial and er ror and al so for remembering the

names and use of commands.
We believe the users found the auditory interfaces more s timulating s ince a new and interesting p rototype (the auditory i nterface) was compared to a well known and widely used GUI.

D. *Overall ranking and subjective comments*

We also asked the users t o r ank t he t hree interfaces w ith m arks f rom 1 t o 3 (1 bei ng t he worst interface and 3 bei ng the best one). Table 2 shows the average marks for the interfaces:

TABLE 2
THE AVERAGE MARKS FOR THE THREE INTERFACES SHOW THE SELECTION OF THE AV INTERFACE AS THE BETTER OF THE TWO AUDITORY INTERFACES.

Interface	Mark
V	2.94
AH	1.25
AV	1.81

The us ers s elected t he AV i nterface as t he better of the two auditory interfaces. We collected some s ubjective c omments j ustifying t heir decision:
• the vertical arrangement of t he s ources i n the AV interface seems m ore nat ural as i t resembles the GUI menus (the vertical columns);
• the peg at both ends of t he AV menu increases the overall orientation in the menu since t he us ers al ways k now w hen t hey reach the top or the bottom of the individual menu; i n t he A H m enu, t he v irtual circle cand be rotated indefinitely without a peg at each end of the menu.
Other comments about the auditory interfaces:
• the background music which intends to identify the individual submenu is more disturbing than helpful;
• the s patial s ound per ception and understanding of t he m ultiple simultaneous sources is better in the AH menu.

6. *DISCUSSION*

The main goal of this study was the evaluation of an ac oustic i nterface as a s ubstitute f or t he traditional v isual i nterface (V) f or a simple word processing app lication. T he t hree m ain v ariables measured in the experiment were task completion time, navigation performance and overall us er s atisfaction. O ur m ain goal was to compare the two auditory interfaces (AH and AV) and select one of them for our future research.

Both audi tory i nterfaces pr oved to be significantly slower than the visual interface, which w as an ex pected outcome. Visual perception is m uch f aster t han t he auditory one. In addition, the resolution and the capacity of the visual channel are much better than those of the auditory c hannel. O n t he ot her hand, w e c ould

not confirm any significant difference in the task completion times between the two auditory interfaces.

The second parameter we evaluated in the study was the navigation performance of the users when solving tasks. By navigation performance we mean the overall orientation of the test subjects in the menu, which can be measured by following the movements of the user in the menu and the movements of the menu itself (turns, movements, clicks). The average navigation performances of the two auditory interfaces were measured and then compared to the ideal task completion navigation. According to the results, the AV interface enabled better navigation than the AH interface. We find the result somewhat surprising since we expected the AH interface to more efficient. It has already been established in [10] that the horizontally distributed spatial sounds can be localized much more efficiently and faster than the vertically distributed sound sources. It appears that the test subjects concentrated primarily on just one selected sound source and not so much on the two additional simultaneous sources.

The comparison of the subjective evaluations also shows a preference of the test subjects for the AV interface. Most of the users confirmed the AV interface to be more efficient and user friendly. The latter is confirmed also by a significantly higher overall rank.

Taking into consideration the results discussed above, the AV interfaces proved to be more efficient and preferred by the majority of the users.

7. CONCLUSION

With the designed interfaces we intend to improve the use of various desktop applications for the visually impaired users. We explored the use of spatial sound in order to increase the auditory information flow between the computer and the user. A simple interaction based on only two keys for controlling the position in the menu and two additional keys for confirming or canceling various options should effectively replace the mouse navigation in graphical interfaces.

The most important results of the evaluation show that both proposed auditory interfaces can be used effectively for manipulating documents in word processing applications, but one of them is slightly more efficient and also preferred by the users. The latter will therefore be included in our future research. We intend to develop a 3D auditory computer desktop to be used by blind computer users. The auditory desktop will support free space navigation and selection of various applications, similar to the movements of the mouse in the graphical interface. Spatial

sound will play an important role for user orientation and accurate perception of the desktop structure.

REFERENCES

[1] American Foundation for the Blind, Word Processing, http://www.afb.org/ProdBrowseTaskResults.asp?TaskID=425&SpecID=26

[2] TSBVI, Adaptive Technology, http://www.tsbvi.edu/technology/overview.htm

[3] C. Schmandt, "Audio Hallway: a Virtual Acoustic Environment for Browsing," ACM Symposium on User Interface Software and Technology, 1998, pp. 163-170.

[4] C. Frauenberger, M. Noistering, "3D Audio Interfaces for the Blind," Proc. of the Int. Conf. on Auditory Display, Boston, MA, USA, 2003, pp. 281-283.

[5] C. Frauenberger, V. Putz, R. Holdrich, "Spatial Auditory Displays: A Study on the Use of Virtual Audio Environment as Interfaces for User with Visual Disabilities," Proc. of the 7th Int. Conf. on Digital Audio Effects (DAFx'04), Naples, Italy, 2004, pp. 384-389.

[6] C. Frauenberger, V. Putz, R. Holdrich, T. Stockman, "Interaction Patterns for Auditory User Interfaces," ICAD Proc., 2005, pp. 154-160.

[7] E. D. Mynatt, "Transforming Graphical Interfaces into Auditory Interfaces," Doctoral dissertation, Georgia Institute of Technology, 1995.

[8] J. Sodnik, C. Dicke, S. Tomažič and M. Billinghurst, "A user study of auditory versus visual interfaces for use while driving," Int. Journal of Human-Comput. Stud., vol. 66, no. 5, 2008, pp. 318-332.

[9] B. N. Walker, A. Nance and J. Lindsay, "Spearcons: Speech-Based Earcons Improve Navigation Performance in Auditory Menus," Proc. of the Int. Conf. on Auditory Display (ICAD 2006), London, England, 2006, pp. 63-68.

[10] J. Sodnik, R. Sušnik, M. Štular, S. Tomažič, "Spatial sound resolution of an interpolated HRIR library," Applied Acoustics, vol. 66, no. 11. pp. 2005, pp. 1219-1234.G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in Plastics, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.