

# Auditory Representations of a Graphical User Interface for a better Human-Computer Interaction

György Wersényi<sup>1</sup>,

<sup>1</sup> Széchenyi István University, Department of Telecommunications, Egyetem t. 1,  
H-9026 Győr, Hungary  
wersenyi@sze.hu

**Abstract.** As part of a project to improve human computer interaction mostly for blind users, a survey with 50 blind and 100 sighted users included a questionnaire about their user habits during everyday use of personal computers. Based on their answers, the most important functions and applications were selected and results of the two groups were compared. Special user habits and needs of blind users are described. The second part of the investigation included collecting of auditory representations (auditory icons, spearcons etc.), mapping with visual information and evaluation with the target groups. Furthermore, a new design method for auditory events and class was introduced, called “auditory emoticons”. These use non-verbal human voice samples to represent additional emotional content. Blind and sighted users evaluated different auditory representations for the selected events, including spearcons for different languages. Auditory icons using environmental, familiar sounds as well emoticons are received very well, whilst spearcons seem to be redundant except menu navigation for blind users.

**Keywords:** auditory icon, earcons, blind users, spearcons, GUIB.

## 1 Introduction

Creating Graphical User Interfaces (GUIs) is the most efficient way to establish human-computer interaction. Sighted people benefit from easy access, iconic representation, 2D spatial distribution of information and other properties of graphical objects such as colors, sizes etc. The first user interfaces were text-based, command line operation systems with limited capabilities. Later, hierarchical tree-structures were utilized mostly in menu navigation, since they enable a clear overview of parent-child relations, and causality. Such interfaces are still in use in simple mobile devices, cell phones etc. For the most efficient work the GUIs proved to be the best solution. Nowadays almost all operation systems offer a graphical surface and even command line programs can be accessed by such an interface. Some GUIs also include sounds but in a limited way as an extension to the visual content or for feedback only.

However, the blind community and the visually disabled do not benefit from a GUI. Access to personal computers became more and more difficult for them as the GUIs took over the former command line and hierarchical structures [1]. Although there is a need for transforming graphical information to auditory information for blind users, most so-called “auditory displays” are audio-only interfaces creating a virtual sound-scape, where users have to orientate, navigate and act. These virtual audio displays (VADs) have limited quality, spatial resolution and allow reduced accessibility.

As a result, blind users often use only textual representation of a screen. These text-to-speech (TTS) applications or screen-readers nowadays offer good synthesised speech quality, but they are language-dependent and only optimal for reading textual information. Some programs, such as the most frequently used Job Access With Speech (JAWS) [2] or the Window-Eyes [3] also read icon names and buttons. The user moves the cursor with the mouse or navigates with the TAB-button over the icons and across the screen and information will be read about objects that he crosses. Unfortunately sometimes confusion is created when the objects are read phonetically. A TTS system can not follow a GUI, it is more disadvantageous than helpful in translating graphical information into textual. Tactile translations have encountered many of the same difficulties in representing graphical information in a tactile way [4, 5].

The overriding goal is to create an audio environment where blind users have the same or almost the same accessibility as sighted colleagues do. To achieve this, the most important considerations are the following:

- accessibility and recognition: blind users have to be able to use the interface, recognize items, programs, identify and access them. Some issues to be resolved are: what are the objects, what is the name/type, where are they, what attributes do they have?
- iconic representation: short, easily identifiable sounds, that can be filtered, spatially distributed etc. They have to be interruptable even if they are short.
- safe manipulation: safe orientation and direct manipulation with auditory feedback.

Screen readers and command line interfaces do not currently offer these possibilities. Some stumbling blocks have been

- In contrast to graphics, auditory signals cannot be presented constantly.
- It is hard with auditory displays to get an overview of the full screen and users have to use their short-time memory to remember the content of the screen. Concurrent sound sources are hard to discriminate and/or long term listening to synthesised speech can be demanding (synthesised speech overload).
- Blank spaces of the background (without sound) can lead to disorientation.
- Other graphical information can also be relevant: relatively bigger buttons, font sizes, different colors or blinking may indicate relative importance that is hard to translate into auditory events.
- Grouping of information: the spatial allocation of similar functions and buttons is also hard to map to an auditory interface.

- The static spatial representation of a GUI seems to be the most difficult to transfer and the cognitive requirements for a blind user are quite demanding. Hierarchical structures are easily abstracted but they represent discrete values (menu items). Sonification of continuous data, such as auditory graphs is also in interest [6, 7].

The most critical issue is here navigation: good overall performance by using an auditory display is strongly related to good and fast navigation skills. Navigation without the mouse is preferred by blind users. Keyboard short-cuts and extended presentation of auditory events (spatial distribution, filtering etc.) are useful to expert users. Spatial models are maybe preferable as opposed to hierarchical structures, but both seem to be a good approach to increase accessibility. Learning rates are also an important consideration, because everybody needs time to learn to use an auditory interface.

It is impossible to transfer all the information in a GUI to an auditory interface, so we have to deal with some kind of an “optimal” procedure: the most important information should be transferred, and details and further information (for trained and expert users) can extend the basic auditory events. The goal is that blind users can work with computers, create and handle basic text-oriented applications, documents, e-mails, and also browse the internet for information. They have to be able to save, open, copy, delete, print files. After basic formatting and file managing then sighted users may do any remaining formatting.

## 1.1 Some Previous Results

Earlier investigations tried to establish different auditory interfaces and environments for the visually impaired as early as the 1990s. The *SonicFinder* [8] was an Apple program which tried to integrate auditory icons into the operating system for file handling, but it was not made commercially available primarily because of memory usage considerations. Mynatt and colleagues presented a transformed hierarchical graphical interface, utilizing auditory icons, tactile extension and a simplified structure for navigation in the *Mercator* project [4]. The hierarchical structure was thought to best to capture the underlying structure of a GUI. The project focused on text-oriented applications such as word-processors, mailing programs but neglected graphical applications, drawing programs etc. The TTS module was also included. A basic set of sounds were presented the users as seen in Table 1.

Furthermore, they used filtering and frequency manipulations to portray screen events, e.g. appearing of pop-up windows, selecting items or the number of objects. These were mostly chosen intuitively and were sometimes not very helpful at all, because some sounds are ambiguous (closing a pop-up window can have the same sound as “close” or even some speech feedback) or the related events are not really important (pop-up blocking reduces pop-ups to a minimum). A more general problem is that there are no standards or defined ways to use the simplest modifications in volume, pitch, timbre or spectral content of an auditory event. For instance, the sound of paper shuffling in *Mercator* represented “switching between applications” but this sound is clearly not good in Windows, where a similar sound is mapped with the recycle bin. Different operating systems may require different sound-sets, but the

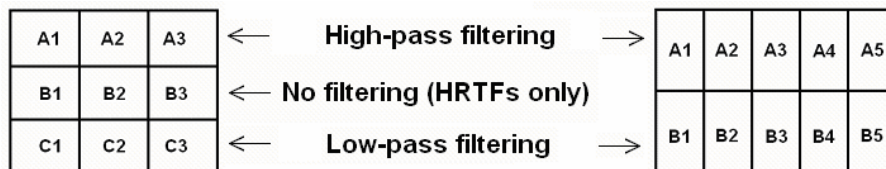
overriding concern is to find the most important applications, functions and events of the screen that have to be represented by auditory events.

Involving sighted people in this quest is desirable both for comparison with blind users, and because it can be advantageous for the sighted user as well: they can examine the efficiency of transition from GUI to auditory interface and finally, they could also benefit from auditory feedback during work.

**Table 1.** Auditory icons introduced by Mynatt for the Mercator [4, 9].

Interface Object	Sound
Editable text area	Typewriter, multiple keystrokes
Read-only text area	Printer printing out a line
Push button	Keypress (ca-chunk)
Toggle button	Pull chain light switch
Radio button	Pa pop sound
Check box	One pop sound
Window	Tapping on glass (two taps)
Container	Opening a door
Popup dialog	Spring compressed then extended
Application	Musical sound

Later, the *GUIB project* (Graphical User Interface for Blind persons) tried a multimodal interface, using tactile keyboards (Braille) and spatial distributed sound, first with loudspeaker playback on the so-called sound-screen, then using headphone playback and virtual simulation [5, 10, 11, 12]. In this project the Beachtron soundcard was used with real-time filtering of the Head-Related Transfer Functions (HRTFs) to create a spatial virtual audio display. A special 2D surface was simulated in front of the listener instead of the usual “around the head” concept. This should create a better mapping of a rectangle computer screen and increase in navigation accuracy with the mouse as well. Listening tests were carried out first with sighted and later with blind users using HRTF filtering, broadband noise stimuli and headphone playback. The results showed an increased rate of headphone errors such as in-the-head localization and front-back confusions, and the vertical localization was almost a complete failure. A follow-up study used additional high-pass and low-pass filtering to bias correct judgements in vertical localization (Fig. 1.) and achieved about 90% of correct rates [13, 14].



**Fig. 1.** A possible scheme for increasing vertical localization judgments. Input signals can be filtered by HPF and LPF filters before or after the HRTF filtering.

Simulation of small head-movements without any additional hardware also seemed very useful in reducing of errors [15, 16]. Spatial distributed auditory events can be used in a special window-arrangement in different resolutions according to the users' experience and routine. In addition, distance information can be used for overlapping windows or other parameters.

In [4] it was reported that blind users have positive response to the project, but they were skeptical about hierarchical navigation schemes. A spatial one seems to be better, primarily for blind people who lost their vision later in life. Users who were born blind have more difficulties in understanding some spatial aspects of the display, but tactile extensions can be helpful to understand spatial distribution and forms.

## 2 Auditory Representations

What kind of auditory events can be used in an auditory display? Different basic sound types have different considerations:

- Speech is sometimes too slow, language-dependent, and synthesised speech overload can happen. A TTS is necessary for textual information but not optimal for orientation, navigation and manipulation.
- Pure tones are easily confused with each other, are not very pleasant to listen to them and mapping is intuitive that needs more learning time.
- Musical instrumentation is easier to listen to, but also needs learning and abstraction because of the intuitive mapping.
- Auditory icons, earcons, spearcons and auditory emoticons, or structured combination of environmental sounds, music, non-speech audio or even speech can create good iconic representations. Iconic everyday sounds can be more intuitive than musical ones [4].

Auditory icons and earcons were the first introduced by William Gaver, followed by others [17-19]. These sounds are short „icon-like” sound events having a semantic connection to the physical event they represent. Auditory icons are easy to interpret and easy to learn. Users may connect and map the visual event with the sound events from the initial listening. A typical example is the sound of a matrix dot-printer that is intuitively connected with the action of printing. Gaver provided many examples of easily learned auditory icons. Unfortunately, there are other events on a screen that are very hard to represent by auditory icons.

Environmental sounds are very good for auditory icons, because they are easily identifiable, learnable, they have a semantic-nomic connection to (visual) events. There are numerous factors that affect the useability of environmental sounds as auditory icons: a brief overview was provided in [20-22]. Among these are the effects of filtering on various types of environmental sounds. Some sounds are resistant against filtering and some completely lose their typical properties depending on the spectral content. Furthermore, some sounds are only identifiable after a longer period of time and thus it is disadvantageous to use them as auditory icons. Ballas gave a time period of about 200-600 ms for a proper recognition of a sound and as a good start to create an auditory icon [23]. At last but not least, context contributes to

recognition: logical, expected sounds will be recognized better than unexpected [24]. On the other hand, unexpected sounds do not have to be too loud to get attention to. Realistic sounds sometimes are inferior to other but more familiar versions of them. Cartoonification may help, or e.g. a gunshot is much different in the real life as it is in movies [25, 26].

On the other hand, earcons are „meaningless” sounds. The mapping is not obvious, so they are harder to interpret and to learn, and have to be learned together with the event they are linked to. An example: the sounds that we hear during start-up and shut down the computer or during warnings of the operation system are well-known after we hear them several times.

Spearcons have already proved to be useful in menu navigations and in mobile phones because they can be learned and used easier and faster than earcons [27-30]. Spearcons are time-compressed speech samples which are often names, words or simple phrases. The optimal compression ratio, required quality and spectral analysis was made for Hungarian and English language spearcons [31]. For the study described later, the Hungarian and German spearcon databases for our study were created with native speakers.

Furthermore, some sound samples can not be classified into the main three groups mentioned above. Based on the results of a user survey, we will introduce a new group of auditory events called auditory emoticons. Emoticons are widely used in e-mails, chat and messenger programs, forum posts etc. These different smileys and abbreviations (such as brb, rotfl, imho) are used so often that users suggested that they be represented with auditory events as well.

Auditory emoticons are non-speech human voice(s), sometimes extended and combined with other sounds in the background. They are related to the auditory icons the most, using human non-verbal voice samples with emotional load. Auditory emoticons – just like the visual emoticons - are language independent and they can be interpreted easily, such as the sound of laughter or crying can be used as an auditory emoticon.

All the above auditory events are intended for use in auditory displays both for sighted and blind users as feedback of a process or activation, to find a button, icon, menu item etc.

### **3 Evaluation and Comparison of User Habits**

After many years of research, the Hungarian Institution of Blind Persons is involved in our survey and we have access to blind communities in Germany as well. The first part of the investigation was to find out how blind persons use personal computers nowadays, what their likes and dislikes are, or their needs for a better accessibility. In order to do this we created a questionnaire both for blind people and for people with normal vision. Based on the answers we selected the 30-40 most important and frequently accessed programs and functions. The second part of the project included the selection and evaluation of sound events (auditory icons, earcons or spearcons) representing these functions. Furthermore, user habits of different age groups and

user routines were also evaluated. Details of the survey and some preliminary results of sighted users were presented and described in [31, 32].

The survey included 100 persons with normal vision and 50 visually impaired (from Hungary and Germany). Subjects were categorized based on their user routines on their ages. Eighty-three percent of the sighted subjects were “average” or “above average” users but only forty percent of the blind users were. It is clear that a large number of blind users often restrict themselves to basic computer use.

The average age of sighted users was 27,35 years and 25,67 for blind participants. Subjects had to be at least 18 years of age and they had to have at least basic knowledge of computer use. Usage was ranked on a scale from 1 to 5, detailed in Table 2. Mean rankings above 3,75 correspond to frequent use. On the other hand, mean rates below 3 points are regarded not to be very important. Because some functions appear several times on the questionnaire, these rates were averaged again (e.g. if „print” has an mean value of 3,95 in Word; but only 3,41 in the browser then a mean value of 3,68 will be used).

**Table 2.** Ranking points for applications and services.

Points	
1	Unknown by the user
2	Known, but no use
3	Not important, infrequent use
4	Important, frequent use
5	Very important, everyday use

Mean results are listed in Table 3. Light grey marked fields indicate important and frequently used programs and applications (mean ranking 3,00 – 3,74). Dark grey fields indicate everyday use and higher importance (mean ranking above 3,75 points). At the end of the table some additional ideas and suggestions are listed without rating. Additional applications suggested by sighted users were: Wave editor, remove USB stick, date and time. Blind users mentioned DAISY (playback program for reading audio books), JAWS and other screen-readers. As mentioned above, the frequent use of emoticons (smileys) in e-mails and messenger applications brought up the need to find auditory representations for these as well.

**Table 3.** Averaged points given by the subjects.

<b>Programs/ applications/ functions</b>	<b>sighted</b>	<b>blind</b>
<b>Number of subjects</b>	<b>100</b>	<b>50</b>
	Total avg.	Total avg.
Internet Browser (icon/starting of the program)	4,62	4,67
E-mail client	4,11	4,67
Windows Explorer	3,98	3,25
My Computer	3,58	3,94
Windows/Total Commander	2,75	1,88
Acrobat (Reader)	4,26	3,33
Recycle Bin	4,41	3,67
Word (Word processor)	4,53	4,56
Excel	3,81	2,94
Power Point	3,14	2,47
Notepad/ WordPad	2,6	2,61
FrontPage (HTML Editor)	2,09	2,24
CD/DVD Burning	3,84	3,94
Music/Movie Player	4,09	4,17
Compressors (RAR, ZIP etc.)	3,41	3,22
Command Line, Command Prompt	2,84	1,83
Printer handling and preferences	3,87	2,64
Image Viewer	3,98	1,62
Downloads (Torrent clients, DC++, GetRight)	2,7	2,06
Virus/Spam Filters	4,29	4,39
MSN/Windows Messenger	3,29	3,17
Skype	2,91	3,39
ICQ	2,33	1,78
Chat	2,06	1,5
Paint	3,05	1,67

Calculator	2,82	2,61
System Preferen., Control Panel	3,6	3,17
Help(s)	2,55	2,98
Search for files or folders (under Windows)	3,32	3,11
My Documents folder on the Desktop	3,52	4,11
JAWS, Screen- readers	1	4,6
<b>FUNCTIONS</b>		
Home button (Browser)	3,53	3,61
Arrow back (Browser, My Computer)	4,22	4,33
Arrow forward (Browser, My Computer)	4,22	4,33
Arrow up („one folder up”, My Computer)	3,53	3,53
Re-read actual site (Browser)	3,31	3,47
Stop loading (Browser)	2,79	2,98
Enter URL address through the key- board (Browser)	4,02	4,05
Favorites, Bookmarks	3,55	3,88
New register tab (Browser)	3,95	2,94
New window (Browser)	3,78	3,53
Search/find for text on the screen (Browser, Docs)	3,35	3,88
Save/open image and/or location (Browser)	3,59	3,19
Print	3,41	2,9



Cut	4,11	3,88
Paste	4,26	4,41
Copy	4,49	4,56
Move	4,26	4,38
Delete	4,14	4,56
New folder (My Computer)	4,24	4,38
Download mails/ open E-Mail client (Mail)	4,1	4,59
Compose, create new mail (Mail)	4,2	4,67
Reply (Mail)	4,14	4,67
Reply all... (Mail)	2,91	2,98
Forward (Mail)	3,26	4,19
Save mail/drafts (Mail)	3,18	3,47
Send (Mail)	4,22	4,67
Address book (Mail)	3,51	3,72
Attachment (Mail)	3,7	4,06
Open	4,43	4,71
Save	4,28	4,71
Save as...	4,29	4,71
Close	4,27	4,65
Rename	4,14	3,94
Restore from Recycle Bin	3,04	3,44

Empty Recycle Bin	3,74	3,83
New Document	4,29	4,53
Spelling (Docs)	3,47	3,41
Font size (Docs)	3,79	3,41
Format: <b>B/U</b> (Docs)	3,98	3,47
Select, mark, highlight text (Docs)	3,91	4,29
Repeat	3,78	2,94
Undo	3,78	2,94
<b>OTHERS</b>		
Waiting... (hour-glass)		
Start, shut-down, restart computer...		
Resize windows (grow, shrink)		
Frame/border of the screen		
Scrolling		
Menu navigation		
Actual time, system clock		
<b>EMOTICONS</b>		

### 3.1 Blind Users

Blind users have different needs sometimes when using personal computers. We observed that:

- Blind users like the icons, as well as programs that are on the desktop by default, such as My Computer and the My Documents folder. They use these more frequently than sighted users, because sighted can easily access other folders and files deeper in the folder structure as well.
- Programs that use graphical interfaces (e.g. Windows Commander) for ease of access are only helpful for sighted users.
- Image handling, graphical programs, movie applications are only important for sighted users. However, the Windows Media Player is also used by the blind persons, primarily for music playback.
- Select and highlighting of text is very important for the blind, because TTS applications read highlighted areas.
- Blind users do not print often.

- Acrobat is not popular for blind persons, because screen-readers do not handle PDF files properly. Furthermore, lots of web pages are designed with graphical contents (JAVA applications) that are very hard to interpret by screen readers.
- Word is important for both groups, but Excel, Power Point use mainly visual presentation methods, so these latter programs are useful for sighted users.
- For browsing the Internet, sighted users are more likely to use the “new tab” function, while blind persons prefer the “new window” option. It is hard to orientate for them under multiple tabs.
- The need for gaming was mentioned by the blind as a possibility for entertainment (specially- created audio games).

The idea of extensions or replacements of these applications by auditory displays was welcomed by the blind users, however, they suggested not to use too much of them, because this could lead to confusion. Furthermore, they stated speech to be unnecessary on a virtual audio display because screen-readers offer speeded up speech anyway.

Blind users mentioned that JAWS and other screen readers do not offer changing the language “on the fly”; so if it is reading in Hungarian, all the English words are pronounced phonetically. This is very disturbing and makes understanding difficult. However, JAWS offers the possibility to set such words and phrases for a correct pronunciation one by one. An interesting note is that JAWS 9.0 does not offer yet Hungarian language, so Hungarian blind users use the Finnish module although, the reputed relationship between these languages has been questioned lately. Another complaint was that JAWS is expensive while the free version of a Linux-based screen reader has a low quality speech synthesizer.

The best method for a blind person to access applications would be a maximum of a three-layer structure (in menu navigation), alt tags in pictures, and the use of the international W3C standards (World Wide Web Consortium) [33]. Only about 4% of the internet web pages follow these recommendations.

As mentioned before, there is a strong need among blind users for audio-only gaming and entertainment. There are currently some popular text-based adventure games using the command line for navigation and for actions. But there is more need for access to on-line gaming, especially for on-line table and card games, such as Poker, Hearts, Spades or Bridge. This could be realized by speech modules, if the on-line website would tell the player the cards he holds and are on the table.

One of the most popular is the game Shades of Doom, a trial version of which can be downloaded from the internet [34]. In a three dimensional environment, the user guides a character through a research base and shuts down the ill-fated experiment. It features realistic stereo sounds, challenging puzzles and action sequences, original music, on-line help, one-key commands, five difficulty levels, eight completely navigable and explorable levels, the ability to create Braille-ready maps and much more. This game is designed to be completely accessible to blind and visually impaired users, but is compatible with JAWS and Window-Eyes if desired.

On the topic of using environmental sounds in auditory displays for the blind, it should be noted that in one comparative study blind people did not perform better in recognizing environmental sounds than sighted people do: the two groups both

performed at a relatively low level of about 76-78% of correct answers. However, blind subjects can be more critical about how auditory icons should sound [10, 35]. Our current investigation (in preparation) about virtual localization of blind persons also showed that in a virtual environment they may not hear and localize better than sighted people.

#### 4 Evaluation of Auditory Events

After determining the most important functions and applications, a collection of sound samples was developed and evaluated based on the comments and suggestions of blind and sighted users. Below is listed the collection of sounds that was previously selected by the users as the “winning” versions of different sound samples. The rating procedure for Hungarian, German and English spearcons and sound samples is based on an on-line questionnaire with sound playback [36]. Figure 2 shows a screenshot of the website, where users rated a sound sample to be *bad* (3 points), *acceptable* (2 points) or *very good* (1 point). According to the German system, the less points are given, the better the results are. Detailed results and evaluation rates are shown here for the auditory icons only (right column in Table 6). All the sound samples can be downloaded from the Internet in wave or mp3 format [32].

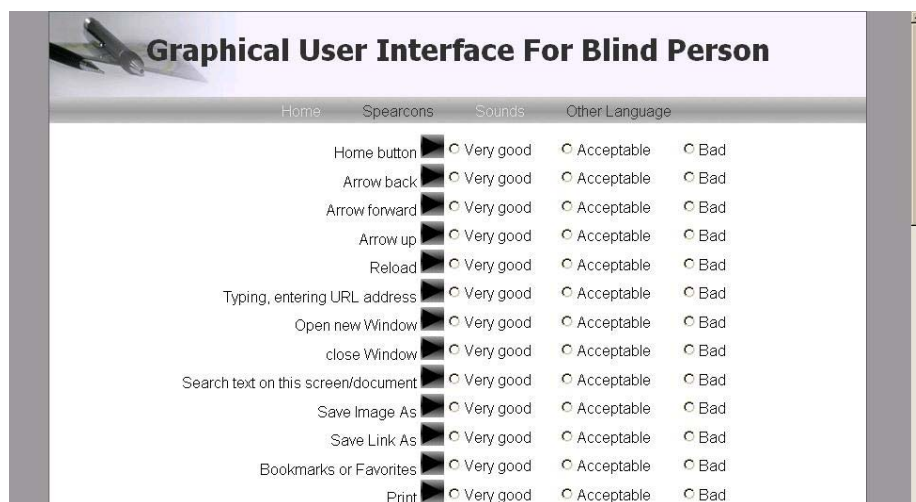


Fig. 2. Screenshot of the website for evaluation.

The approach has been to create sound samples the majority of which have a length of about 1,0-1,5 sec. The actual durations are between 0,6 and 1,7 sec with an mean of 1,11 sec. There are two types of sounds: normal sounds that have to be played back once; and sounds to be repeated (in a loop). The first represents icons, menu items or

short events. Looped signals are supposed to be played back during a longer action or event (e.g. copying, printing).

Sound files were recorded or downloaded from the Internet and were edited by Adobe Audition software in 16 bit, 44100 Hz mono wave format [37, 38]. Editing included simple operations of amplifying, cutting, mixing, fade in/out effects. At the final stage, all samples were normalized to the same loudness level ( $\pm 1$  dB).

A collection of wave data of about 300 files was categorized, selected and evaluated. Subjects were asked to identify the sound (what is it) and judge them by „comfortability” (how pleasing it is to listen to it). Subjects evaluated different sound samples (types) and variations to a given application or event. For example, a sound had to be applied to the action of opening a file. Thus, it had to be determined what „open” sounds like? Possible sound samples include a slide fastener (opening the zipper on a trouser), opening a drawer, opening a beer can or pulling the curtains. We presented different versions of each to insure inclusion of an appropriate representation. In addition, subjects were asked to think about the sound of „close” – a representation in connection with „open”. Therefore, we tried to present reversed versions of opening sounds (simply played back reversed) or using the squeezing sound of a beer can. The reverse playback method can not be applied every time; some samples could sound completely different reversed [39]. Subjects could make suggestions for new sounds as well. If there was no definite winner or no suggested idea at all, a spearcon version was used (e.g. for Acrobat).

The sound files listed in Tables 4-7 (right columns) are included in a ZIP file that can be directly downloaded from <http://vip.tilb.sze.hu/~wersenyi/Sounds.zip>.

#### 4.1 Applications

Table 4 shows the most important applications and programs that have to be represented by an auditory event. These were selected if both blind and sighted users ranked them as “important” or “everyday use” (having a mean ranking of at least 3,00 on the questionnaire), except for the My Documents folder and JAWS because these were only important for blind users.

The sound for internet browsing includes two different versions, both were accepted by the users. It is interesting that the sound sample “search/find” contains a human non-speech part that is very similar in different languages, and is easy to relate to the idea of an “impatient human”. Subjects could relate the intonation to a sentence of “Where is it?” or “Wo ist es?” (in German) or even “Hol van már?” (in Hungarian). It appears a similar intonation is used in different languages to express the feeling during impatient searching. As a result, the same sound will be used in other applications where searching, finding is relevant (Browser, Word, Acrobat etc.). Another idea was a sound of a sniffing dog.

The table does not contain some other noteworthy samples, such as a modified sound for the E-mail client, where the applied sound is extended with a frustrated “oh” in case there is no new mail and a happy “oh” if there is a new mail. Since mail

clients do have some kind of sound in case of existing new mails this sample was not used.

**Table 4.** Collection of the most important programs and applications (MS Windows based). Sound samples can be found under the given names.

Application	Description	Filename
Internet browser (1)	Door opening with keys	Browser1
Internet browser (2)	Knocking and opening a door	Browser2
E-mail client	Bicycle and P.O. box	Mail1
Windows Explorer	Spearcon	S_Explorer
My Computer	Computer start-up beep and fan noise	My Computer
Acrobat	Spearcon	S_Acrobat
Recycle Bin	Pedal of a thin can with the recycle bin sound	Pedal
MS Word	Spearcon	S_Word
CD/DVD burning	Burning flame	Burn
Movie/music player (MS MediaPlayer)	Classic movie projector	Projector
Compressors (ZIP, RAR)	Pressing, extruding machine	Press
Virus/Spam killer	Coughing and “aaahhh”	Cough
MSN Messenger	Spearcon	S_Messenger
Control Panel	Spearcon	S_ControlP
My Documents folder on the desktop	Spearcon	S_MyDocs
Search for files etc.	Seeking and searching with human voice (loop) or dog sniffing	Search_(loop)
JAWS/TTS/Screen Reader appl.	Spearcon, speech	

The events related to the recycle bin also have sound events related to the well-known sound effect of the MS Windows “recycle bin.wav”. This is used if users empty the recycle bin. We used the same sample in a modified way to identify the icon, opening the recycle bin or restore a file from it. The application identification uses the “paper noise” and a thin can pedal together. Restoring a file utilizes the paper noise with human caw. The caw imparts the feeling of a false delete earlier. This thematic grouping was very helpful to identify connected events.

For compressor applications, we used samples of human struggling while squeezing something, e.g. a beer can, but similar sounds appear later in open, close or delete. Similarly, a ringing telephone was suggested for MSN/Windows Messenger, but this sound is used by Skype already. Finally, two different samples for “Help” were selected: a whispering human noise and a desperate “help” speech-sample. Because Help was not selected as a very important function, and furthermore, the first sample was only popular in Hungary (Hungarian PC environments use the term “whispering” instead of “help”, an analog to theatrical prompt-boxes) and the second contains a real English word, these samples were culled from the final listing.

## 4.2 Navigation and Orientation

The sounds in Table 5 were judged to be the most important for navigation and orientation on the screen, primarily for blind persons. Although, blind users do not use the mouse frequently, sometimes it is helpful to know where the cursor is. The movement of the cursor is a looped sound sample indicating that it is actually moving. The actual position or direction of moving could be determined by increasing/decreasing sounds (such as by scrolling) or using HRTF synthesis and directional filtering through headphones [12-14]. This is not implemented yet. Using this sound together with a “ding” by reaching the border of the screen allows a very quick access to the system tray, the start menu, or the system clock which are placed bottom left and right of the screen.

**Table 5.** Collection of important navigation and orientation tasks (MS Windows based). Sound samples can be found under the given names.

Other sounds	Description	Filename
Moving the mouse (cursor)	Some kind of “ding” (loop)	Mouse_(loop)
Waiting for... (sand-glass turning)	Ticking (loop)	Ticking_(loop)
User intervention, pop-up window	Notification sound	Notify
Border of the screen	Some kind of “ding”	Ding (Border)
Scrolling	Increasing and decreasing freq.	
Menu navigation	Spearcons with modifications	
System clock	Speech	S_SystemClock
Start menu	Spearcon, speech	S_StartMenu

In case of menu navigation spearcons have been already shown to have great potential. Modifications to spearcons to represent menu structures and levels can be used, such as different speakers (male, female) or different loudness levels etc. In case of short words, such as Word, Excel, or Cut the use of a spearcon is questionable, since these words are short enough without time-compression in. Users preferred the original recordings instead of the spearcons in such cases. We did not investigate thoroughly what the limit is, but it seems that speech samples with only one syllable and with a length shorter than 0,5 sec. are likely too short to be useful as a spearcon. On the other hand, long words with more vowels become harder to understand after having compressed them into spearcons.

## 4.3 Functions and Events

Table 6 contains the most important and frequently used sub-functions in several applications. The second column indicates where the given function can be found and

some common visual representations (icons) can also be seen. Finally, the last column shows mean values given by blind and sighted users on the homepage by ranking them from 1 to 3 points.

The sounds related to internet browsing have something to do with “home”. Users liked the home button being represented by a doorbell and a barking dog together – something that stereotypically happens when one arrives home. Arrows forward, back and up also have something to do car actions: start-up, reverse or engine RPM boost. Similarly, mailing events have stamping and/or bicycle bell sounds representing a postman’s activity. This kind of thematic grouping is very important in creating auditory events and sound sets. It results in increased learnability and less abstraction is needed. Some of the auditory icons and thematic grouping methods have to be explained but after the users get the idea behind them they use it comfortably. It is recommended to include a short FAQ or user’s manual in a help menu for such sound sets.

Bookmarks/favorites in a browser and the address book/contacts in the e-mail client share the same sound of a book, turning pages and a humming human sound. This is another good example for using a non-speech human voice sample interacting with a common sound and thus creating a better understanding and mapping.

The sound for printing can be used in a long version or looped in case of ongoing printing (in the background this can be more quiet) or as a short sound event to represent the printing icon or command in a menu. The same is true for “copy”: a longer version can be used indicating the progress of the copying action (in the background), and a shorter to indicate the icon or a menu item.











The sound for “paste” is one of the most complex samples. It uses the sound of painting with a brush on a wall, a short sound of a moving paint-bucket and the whistling of the painter creating the image of a painter “pasting” something. This works best for English because in Hungarian and in German a different expression is used and the idea behind this sound has to be explained.

In case of “move” there are two versions: the struggling of a man with a wooden box, and a mixed sound of “cut and paste”: scissors and painting with a brush.

Based on the comments, the action of “saving” has something common with locking or securing, so the sound is a locking sound of a door. As an extension, the same sound is used for “save as” with an additional human “hm?” sound indicating that the securing process needs user interaction: a different file name to enter.





Opening and closing is very important for almost every application. As mentioned earlier, the sounds have to be somehow related to opening and closing something and they have to be in pairs. The most popular version was a zip fly of a trouser to open up and close. The same sound that was recorded for opening was used for closing as well: it is simply reversed playback. The increasing and decreasing frequency should deliver the information. The other sample is opening a beer can and squeezing it in case of closing.

**Table 6.** Collection of the most important actions and functions (MS Windows based). Sound samples can be found under the given names.

Events, Functions	Where?	Visual Representations	Description	Filename	Mean Values
Home button	Internet Browser		Doorbell and dog barking	Homebutton	1,34
Arrow back	Internet Browser, My Computer		Reverse a car with signaling	Backarrow	1,53
Arrow forward	Internet Browser, My Computer		Starting a car	Forwardarrow	2,15
Arrow up	My Computer, Explorer		Car engine RPM increasing	Uparrow	2,68
Re-read, Re-load actual page	Internet Browser		Breaking a car and start-up	Reread	2,31
Typing, entering URL address	Internet Browser		The sound of typing on a keyboard	Keyboard	1,46
Open new /close Browser Window	Internet Browser		Opening and closing sound of a wooden window	Window_open Window_close	1,9
Search/ find text on this screen	Internet Browser, E-mail, Documents		Seeking and searching with human voice (loop)	Search_(loop)	1,87
Save link or image	Internet Browser		Spearcon	S_SaveImageAs S_SaveLinkAs	2,46
Bookmark Favorites	Internet Browser		Turning the pages of a book with human sound	Book	1,99
Printing (action in progress)	Everywhere		Sound of a dot-matrix printer	Print	1,2
Cut	Documents, My Computer, Browser		Cutting with scissors	Cut	1,11
Paste	Documents, My Computer, Browser		Painting with a brush, whistle and can chatter	Paste	2,46



Copy	Documents, My Computer, Browser		Sound of a copy machine	Copy_(loop)	1,57
Move	Documents, My Computer, Browser		Wooden box pushed with human struggling sound or cutting with a scissor and pasting with a brush	Move1 Move2	2,0 2,3
Delete	Documents, My Computer, Browser		Flushing the toilet	Delete	1,32
New Folder...	My Computer		Spearcon	S_New	
New mail, create/compose new message	E-mail		Breathing and stamping	Composemail	2,25
Reply to a mail	E-mail		Breath and stamp (once)	Replymail	2,49
Forward mail	E-mail		Movement of paper on a desk	Forwardmail	2,74
Save mail	E-mail		Sound of save and bicycle bells	SaveMail	2,18
Send mail	E-mail		Bicycle bell and bye-bye sounds	Sendmail	1,99
Address-book	E-mail		Turning the pages with human sound	Book	1,97
Attachment to a mail	E-mail		Stapler	Attach	1,32
Open	Documents, Files		Zip fly up or opening beer can	Zip_up Beer_up	1,22 1,43
Save	Documents, Files		Locking a door with keys	Save	1,72
Save as...	Documents, Files		Locking a door with keys with human "hm?"	Save_as	1,88
Close	Documents, Files		Zip fly down or squeezing beer can	Zip_down Beer_down	1,56 1,82
Rename	Documents, Files		Spearcon	S_Rename	

Restore from the recycle bin	Recycle bin		Original “paper sound” of MS Windows and human caw	Recycleback	2,0
Empty recycle bin	Recycle bin		Original sound of MS Windows (paper sound)	Recycle	1,53
New Document (create)	Documents		Spearcon	S_New	
Text formatting tools	Documents		Spearcons	S_Fontsize S_Formatting S_Bold S_Italic S_Underline S_Spelling	
Mark /select (text)	Documents, Browser, E-mail		Sound of magic marker pen	Mark	1,82

Based on the mean values a total mean value of 1,86 can be calculated (the lower the point the better the sound is). The best values are as low as 1,1-1,5. Only two sounds have worse results than 2,5. This indicates a successfully designed sound set for these functions. A comparison between languages showed only little differences. An explanation phase regarding thematically grouped sounds helped the users to associate the events with the sounds, so this resulted in better ranking points.

#### 4.4 Auditory Emoticons

Table 7 contains the auditory emoticons together with the visual representations. Smileys have the goal of representing emotional content using a few keystrokes and as a result some of them appear to be similar. As smileys try to encapsulate emotions in an easy but limited (graphical) way, the auditory emoticons also try the same using a brief sound. As in real life, some of them express similar feelings. In summary, it can be said that auditory emoticons:

- reflect emotional status of the speaker
- are represented always with human sounds, non-verbal and language independent
- can also contain other sounds, noises etc. for a deeper understanding.

Although there is no scientific evidence that some emotions can be represented better by a female voice than by a male voice, we observed that subjects prefer the female version for smiling, winking, mocking, crying and kissing. Table 7 contains both female and male versions. Users especially welcomed these emoticons.

**Table 7.** Collection of the most important emoticons. Sound samples for female and male versions can be found under the given names.

<b>Auditory Emoticon</b>	<b>Visual Representation</b>	<b>Description</b>	<b>Filename (Female)</b>	<b>Filename (Male)</b>
Smile	☺, :-), :), 😊	chuckle	Smile_f	Smile_m
Laughter	:-D 😂	laughing	Laugh_f	Laugh_m
Wink	:-) 😜	Short “sparkling” sound and chuckle	Wink_f	Wink_m
Mock (tongue out)	:-P 😝	Typical sound of tongue out	Tongue_f	Tongue_m
Surprise	:-o 😲	“oh”	Surprise_f	Surprise_m
Anger	😡, 😠	“grrrrrrrr, uuuhhh”	Anger_f	Anger_m
Perplexed, distracted	:-S, 😕	“hm, aaahhh”	Puzzled_f	Puzzled_m
Shame, “red face”	😳	“iyuu, eh-eh”	Redface_f	Redface_m
Sadness, sorry	☹, :-), :(, 😞	A sad “oh”	Sad_f	Sad_m
Crying, whimper	😭	Crying	Cry_f	Cry_m
Kiss	:-*, 😘, ❤️	Sound of kiss on the cheek	Kiss_f	Kiss_m
Disappointment	:-1, 😞	“oh-hm”	Dis_f	Dis_m

#### 4.5 Presentation Methods

All the auditory representation presented above can be played back in the following ways:

- in a direct mapping between a visual icon or button: the sound can be heard as the cursor/mouse is on the icon/button or it is highlighted, and the auditory event helps first of all the blind users to orientate (to know where they are on the screen).

- during an action in progress, e.g. during copying, deleting, printing etc. in loop.

- after an action is finished and completed as a confirmation sound.

The sounds have to be tested further to find which presentation method is the best for a given action and sound. It is possible that the same sound can be used for both: e.g. first, the sound is played back once as the cursor is on the button “back arrow”, and after clicking, the same sound can be played back as a confirmation that the previous page is displayed.

#### 4.6 Spearcons

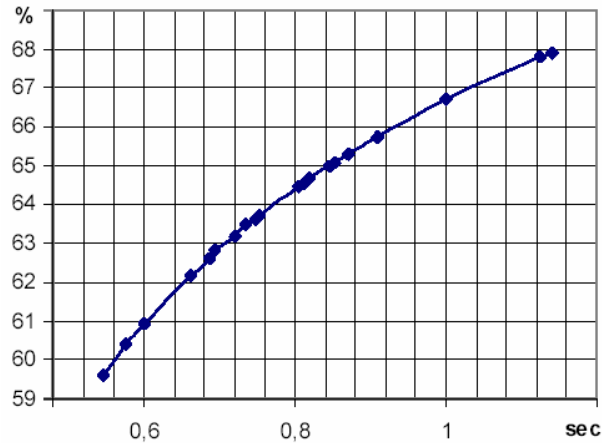
Spearcons, as a version of speeded up speech, were introduced to the Hungarian and German blind and sighted users as well. A MATLAB routine was used to compress original recordings of Hungarian and German words and expressions related to computer usage. Table 8 shows some of the spearcons (here translated in English), duration of original and compressed samples and the compress ratio. Different resolutions of original recordings were tried, from 8 bits to 16 bits and from 8000 Hz to 48000 Hz sampling frequency. Furthermore, the final evaluation regarding the quality of spearcons includes native English speakers and TTS versions as well.

**Table 8.** List of services and features for Hungarian spearcons introduced to blind users. The length and compress ratio is also shown. Original recording was made by a male speaker in 16 bit, 44100 Hz resolution using a Sennheiser ME62 microphone.

Spearcon	Duration (original) [sec]	Duration (compressed) [sec]	Compress ratio [%]
Close	0,87	0,302	65,3
Open	0,812	0,288	64,5
Save	0,687	0,257	62,6
Save as	1,125	0,362	67,8
Search	0,694	0,258	62,8
Copy	0,818	0,289	64,7
Move	0,748	0,272	63,6
Delete	0,661	0,25	62,2
Print	0,752	0,273	63,7
Download	0,853	0,298	65
Stop	0,908	0,311	65,8
Word	0,576	0,228	60,4
Excel	0,599	0,234	60,9
My Computer	0,805	0,286	64,5
Start Menu	0,734	0,268	63,5
Browser	0,845	0,296	65
E-Mail	0,545	0,22	59,6

Spectral evaluation of the spearcons showed that 16-bit resolution and at least 22050 Hz sampling frequency is required. Using 44100 Hz is actually recommended to avoid noisy spearcons [31]: compression has effect on the frequency regions at 4-5 kHz and 16 kHz, so decreasing of the sample frequency or resolution (bit depth) results in a noisy spectrum. A text-to-speech application (SpeakBoard) was also used to save wave files, but listeners preferred original recordings of a human speaker.

The compression ratio is almost linear from 59% to 68% of the duration of the original sample: the longer the sample the higher the compression (Figure 3.). It is always recommended to truncate the samples before compression to remove unnecessary silence at start.



**Fig. 3.** Compression rates (%) as function of duration of the original sample (sec).

For German spearcons we recorded four male native speakers. One set was accent-free, while the other speakers had typical German accents (Saxonian, Bavarian, Frankonian). A current investigation is examining the effects of different accents for German spearcons. All spearcons are made from original recordings in an anechoic chamber using Adobe Audition software and Sennheiser microphones. The Hungarian database was recorded by a native male speaker of 33 years of age. The databases contain 35 words (spearcons) respectively but on the homepage there are 25 for evaluation.

We observed that longer words (having more vowels) are harder to understand after creating the spearcons. Longer sentences (more than 3-4 words) become unintelligible after compression, so this method is not suited for creating spearcons longer than 1-2 words. Although it is not required to understand the spearcon, subjects preferred those they have actually understood. Independent of the fact, whether a spearcon was used or not, all were tested and judged by the subjects. All spearcons were played back in a random order. A spearcon could be identified and classified as follows:

- the subject has understood it the first time,
- the subject could not understand it, and he had a second try,
- if the subject failed twice, the spearcon was revealed (the original recording was shown) and a final try was made.

The evaluation showed that only 12% of the spearcons were recognized on the first try. It was interesting that there was no clear evidence and benefit for using accent-free spearcons: e.g. recognition of the spearcon sometimes was better for the Saxonian version (across all German speakers). Blind persons tend to be better in this task than sighted persons.

In a comparison between German and Hungarian spearcons the German versions got better rankings. Mean value for the 25 spearcons on the homepage was 2,07 for Hungarian language but it was 1,61 for the German versions. We found no clear explanation for this.

Summarized, the best spearcons can be created from good quality recordings of native speakers, who speak relatively slow and articulated. Male speakers are preferred because after compression the speeded up spearcons sound clearer.

## **5 Future work**

Future work includes implementation into various software environments such as JAWS or other Screen Readers that also offer non-speech solutions. The pre-defined samples can be replaced and/or extended with these. In JAWS, words and phrases written on the screen can be replaced by wave files, but actions and events usually can not be mapped with external sound files.

Furthermore, a MS Windows patch or plug-in is planned (in Kernel-level or using the Microsoft Automation or another event logger). This executable file can be downloaded, extracted and installed. It will include a simple graphical user interface with check-boxes for activate and deactivate the sounds and simple environmental settings (e.g. auto start on start-up, default values etc.) and all of the default sound samples, probably in mp3 format.

## **6. Summary**

Fifty blind and hundred users with normal vision participated in a survey in order to determine the most important and frequently used applications, and furthermore, to create and evaluate different auditory representations for them. These auditory events included auditory icons, earcons and spearcons of German and Hungarian language. The German spearcon database contains original recordings of a native speaker and samples with different accents. As a result, a new class of auditory events was introduced: the auditory emoticons. These represent icons or events with emotional content, using non-speech human voices and other sounds (laughter, crying etc). The previously selected applications, programs, function, icons etc. were mapped, grouped thematically and some sound samples were evaluated based on subjective parameters. In this paper the “winning” sound samples were collected and presented. Based on the mean ranking points and informal communications, both target groups liked and welcomed the idea and representation method to extend and/or replace the most important visual elements of a computer screen. This is mostly true for environmental sounds; spearcons are only interesting for blind users in menu navigation tasks, because the screen-reader software offers speeded up speech already. However, becoming an expert user and benefit from all these sounds requires some accommodation and learning time and a guiding explanation or FAQ can ease this process.

## References

1. Boyd, L.H., Boyd, W.L., Vanderheiden, G.C.: The Graphical User Interface: Crisis, Danger and Opportunity. *Journal of Visual Impairment and Blindness*. 496--502 (1990 December)
2. [http://www.freedomscientific.com/fs\\_products/software\\_jaws.asp](http://www.freedomscientific.com/fs_products/software_jaws.asp)
3. <http://www.gwmicro.com/Window-Eyes/>
4. Mynatt, E.D.: Transforming Graphical Interfaces into Auditory Interfaces for Blind Users. *Human-Computer Interaction* 12, 7--45 (1997)
5. Crispian, K., Petrie, H.: Providing Access to GUI's Using Multimedia System – Based on Spatial Audio Representation. *J. Audio Eng. Soc.* 95th Convention Preprint, New York, (1993)
6. Nees, M.A., Walker, B.N.: Encoding and Representation of Information in Auditory Graphs: descriptive reports of listener strategies for understanding data. In: *Proc. of the 14th International Conference on Auditory Display (ICAD 08)*, Paris, 6 pages (2008)
7. Nees, M.A., Walker, B.N.: Listener, Task, and Auditory Graph: Toward a Conceptual Model of Auditory Graph Comprehension. In: *Proc. of the 13th International Conference on Auditory Display (ICAD 07)*, Montreal, pp. 266--273 (2007)
8. Gaver, W.W.: The SonicFinder, a prototype interface that uses auditory icons. *Human Computer Interaction* 4, 67--94 (1989)
9. Mynatt, E.D.: Designing Auditory Icons. In: *Proc. of the International Conference on Auditory Display (ICAD 94)*, Santa Fe, pp. 109--120 (1994)
10. Petrie, H., Morley, S.: The use of non-speech sounds in non-visual interfaces to the MS Windows GUI for blind computer users. In: *Proc. of the International Conference on Auditory Display (ICAD 98)*, Glasgow, 5 pages (1998)
11. Wersényi, Gy.: Localization in a HRTF-based Minimum Audible Angle Listening Test on a 2D Sound Screen for GUIB Applications. *J. Audio Eng. Soc.* 115th Convention Preprint, New York, (2003)
12. Wersényi, Gy.: Localization in a HRTF-based Minimum-Audible-Angle Listening Test for GUIB Applications. *Electronic Journal of Technical Acoustics* 1 (EJTA), <http://www.ejta.org>, 16 pages (2007)
13. Wersényi, Gy.: What Virtual Audio Synthesis Could Do for Visually Disabled Humans in the New Era. AES Convention Paper, presented at the AES Tokyo Regional Convention, Tokyo, Japan, pp. 180--183 (2005)
14. Wersényi, Gy.: Localization in a HRTF-based Virtual Audio Synthesis using additional High-pass and Low-pass Filtering of Sound Sources. *Journal of the Acoust. Science and Technology Japan* 28(4), 244--250 (2007 July)
15. Wersényi, Gy.: Effect of Emulated Head-Tracking for Reducing Localization Errors in Virtual Audio Simulation. *IEEE Transactions on Audio, Speech and Language Processing (ASLP)* 17(2), 247--252 (2009 February)
16. Wersényi, Gy.: Simulation of small head-movements on a virtual audio display using headphone playback and HRTF synthesis. In: *Proc. of the 13th International Conference on Auditory Display (ICAD 07)*, Montreal, pp. 73--78 (2007)
17. Gaver, W.W.: Auditory Icons: using sound in computer interfaces. *Human-Computer Interactions* 2(2), 167--177 (1986)
18. Blattner, M.M., Sumikawa, D.A., Greenberg R.M.: Earcons and Icons: Their structure and common design principles. *Human-Computer Interaction* 4, 11--44 (1989)
19. Gaver, W.W.: Everyday listening and auditory icons. Doctoral thesis, Univ. of California, San Diego (1988)
20. Gygi, B., Shafiro, V.: From signal to substance and back: insights from environmental sound research to auditory display design. In: *Proc. of the 15th International Conference on Auditory Display (ICAD 09)*, Copenhagen, pp. 240--251 (2009)

21. Gygi, B.: Studying environmental sounds the watson way. *The Journal of the Acoustical Society of America* 115(5), 2574 (2004)
22. Gygi, B., Kidd, G.R., Watson, C.S.: Spectral-temporal factors in the identification of environmental sounds. *The Journal of the Acoustical Society of America* 115(3), 1252--1265 (2004)
23. Ballas, J.A.: Common factors in the identification of an assortment of brief everyday sounds. *Journal of Exp. Psychol. Human* 19(2), 250--267 (1993)
24. Gygi, B., Shafiro, V.: The incongruency advantage in elderly versus young normal-hearing listeners. *The Journal of the Acoustical Society of America* 125(4), 2725 (2009)
25. Fernström, M., Brazil, E.: Human-Computer Interaction design based on Interactive Sonification – hearing actions or instruments/agents. In: *Proc. of 2004 Int. Workshop on Interactive Sonification*, Bielefeld Univ (2004)
26. Heller, L.M., Wolf, L.: When Sound Effects Are Better Than The Real Thing. *The Journal of the Acoustical Society of America* 111(5/2), 2339 (2002)
27. Vargas, M.L.M., Anderson, S.: Combining speech and earcons to assist menu navigation. In: *Proc. of the International Conference on Auditory Display (ICAD 03)*, Boston, pp. 38--41 (2003)
28. Walker, B.N., Nance, A., Lindsay, J.: Spearcons: Speech-based earcons improve navigation performance in auditory menus. In: *Proc. of the International Conference on Auditory Display (ICAD 06)*, London, pp. 63--68 (2006)
29. Palladino, D.K., Walker, B.N.: Learning rates for auditory menus enhanced with spearcons versus earcons. In: *Proc. of the 13th International Conference on Auditory Display (ICAD 07)*, Montreal, pp. 274--279 (2007)
30. Dingler, T., Lindsay, J., Walker, B.N.: Learnability of Sound Cues for Environmental Features: Auditory Icons, Earcons, Spearcons, and Speech. In: *Proc. of the 14th International Conference on Auditory Display (ICAD 08)*, Paris, 6 pages (2008)
31. Wersényi, Gy.: Evaluation of user habits for creating auditory representations of different software applications for blind persons. In: *Proc. of the 14th International Conference on Auditory Display (ICAD 08)*, Paris, 5 pages (2008)
32. Wersényi, Gy.: Evaluation of auditory representations for selected applications of a Graphical User Interface. In: *Proc. of the 15th International Conference on Auditory Display (ICAD 09)*, Copenhagen, pp. 41--48 (2009)
33. <http://www.w3.org/>
34. <http://www.independentliving.com/prodinfo.asp?number=CSH1W>
35. Cobb, N.J., Lawrence, D.M., Nelson, N.D.: Report on blind subjects' tactile and auditory recognition for environmental stimuli. *Journal of Percept. Mot. Skills* 48(2), 363--366 (1979)
36. <http://guib.tilb.sze.hu/>
37. <http://www.freesound.org>
38. <http://www.soundsnap.com>
39. Gygi, B., Divenyi, P.L.: Identifiability of time-reversed environmental sounds. In: *Abstracts of the Twenty-seventh Midwinter Research Meeting, Association for Research in Otolaryngology*, 27 (2004)