Virtual Localization by Blind Persons

GYÖRGY WERSÉNYI, *AES Member*
(wersenyi@sze.hu)

Széchenyi István University, H-9026, Egyetem tér 1, Győr, Hungary

Localization performance and spatial hearing abilities of blind persons are complex issues. In everyday life we rely on the "fact" that blind people can hear better, without thinking of what "better" means. Localization performance depends on many parameters such as properties of the excitation signal, environmental conditions, individual aspects, and visual influence. Our goal was to create a virtual environment aimed at helping the blind community use personal computers. In developing this environment we were concerned to cover technical and hearing related questions, as well as human factors. At first, this project included sighted subjects and basic properties of the virtual audio system and the applied HRTFs were tested. Subsequently, blind persons have been involved and comparative measurements performed using the same equipment and selected localization tasks. Twenty-eight blind person's localization performances were tested and compared with the results of 40 sighted subjects in a virtual audio environment. Blind subjects tended to be better in detecting movements in the horizontal plane around the head, localizing static frontal audio sources, and orientation in a 2-D virtual audio display. On the other hand, sighted subjects performed better identifying ascending sound sources in the vertical plane and detecting static sources in the back. In-the-head localization error rates and MAA results appeared to be about the same for both groups. The evaluation was also supported by some informal questions.

0 INTRODUCTION

Virtual auditory displays (VADs) are often used in spatial hearing research [1–6]. The term "virtual" reflects that—in contrast to free-field environments—a headphone playback system is used. In free-field listening situations, mostly in anechoic chambers, sound sources are usually played back over a set of loudspeakers and sitting or standing subjects have to solve various localization tasks. To do this, the human hearing system utilizes the interaural time differences (ITD) and the interaural level differences (ILD) between the ear signals. In case of median plane sources, or if ITD and ILD information is not delivering information, the filtering effects of the outer ears (and body) help to identify the sound source location. This set of direction-dependent filters are called Head-Related Transfer Functions (HRTFs) that can also be described in the time-domain as impulse response functions (HRIRs) [1, 7, 8, 9]. HRTFs are defined in a head-related coordinate system, characterized by the angle of incidence φ (azimuth) in the horizontal and δ (elevation) in the median plane.

Virtual audio simulators are usually incorporated with some kind of HRTF filtering. The measurement problems associated with HRTFs, including their spatial resolution or individuality versus dummy-head systems, as well as playback methods from time-domain to frequency-domain representation, are widely discussed in the literature [10–14]. From the literature, it can be concluded that the use of individual or customized HRTFs establish the best possibility for a good localization performance, while dummy-head HRTFs are inferior to human HRTFs [15, 16]. Furthermore, for a correct emulation, the equalization of the playback chain (mostly the transfer function of the headphone) has to be realized [17, 18]. This equalization together with the applied HRTF filtering and the lack of head-movements result in various localization errors that make virtual environments inferior to free-field listening due to such factors as the increased number of localization errors (e.g., front-back reversals), mostly vertical decreased localization performance, in-the-head localization, elevation shifts, etc. [11, 19–21].

Research projects aiming at human-computer interaction for blind people focus mostly on the development and evaluation of VADs. The use of personal computers with sounds in everyday environments usually requires headphone playback to avoid disturbing the aural environment of others. The more sounds a VAD uses the more important this becomes. Computers usually use GUIs (icons, menus, colors, dynamic parameters, spatial distributed items on a two-dimensional screen, etc.) that can only be used by sighted users. Based on former results of others, the GUIB project (Graphical User Interface for Blind Persons) started to develop an audio interface for the blind community to help them use personal computers [21–23]. For a better human-computer interaction the two main questions to be investigated here are: first, how to map the visual-graphical information into sounds; how to represent the functionality of the system into sound by not risking the loss of properties of sound that might represent some aspects actually better than the visual representation. Second, how to play back this information to the users, which includes the question of whether or not to present some kind of spatial information about the sound source.

This paper first gives a short overview about former results, focusing mainly on the GUIB project and auditory displays, followed by an introduction to the measurement procedure used and a comparison of results of blind persons with former results of sighted users. Finally, some common observations based on informal communication with the blind users will be discussed and summarized.

0.1 Auditory Displays for Blind Persons

Visually impaired people have worked with and programmed computer systems since the 1970s. Systems such as the Hal Screen Reader (SuperNova) or the tape-based Versabraille became quite commonly used from the early 1980s. The first investigations that tried to establish different auditory interfaces and environments for the visually impaired for personal computers appeared in the early 1990s.

The concept of an auditory icon first was introduced by Gaver [24, 25] followed by others [26–28]. The SonicFinder [29] was an Apple program that tried to integrate auditory icons into the operating system for file handling. 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 [30]. The hierarchical structure was thought best to capture the underlying structure of a GUI. The project focused on text-oriented applications such as word-processors and mailing programs but neglected graphical applications. It was also reported that blind users had positive response to the project, but that they were skeptical about hierarchical navigation schemes. They concluded that a spatial scheme would be better, primarily for blind people who had lost their vision later in life. Users who were born blind have more difficulty in understanding some spatial aspects of the display, but tactile extensions could be helpful to understand spatial distribution and forms [31].

Currently, besides speech, the main auditory representations of visual objects are earcons, morphocons, auditory icons, and mixed versions of them, such as spearcons, auditory emoticons, etc. [32, 33]. Speech is still for most blind users the most common method of getting information about the screen on every computer platform. Text-To-Speech (TTS) applications and various screen-readers offer good quality speech with adjustable speed and the possibility of integrating external sounds. Auditory icons are more effective than earcons, because they have a deeper semantic mapping between the sound and visual event, while earcons are "meaningless" sounds [27, 34]. Typical auditory icons include familiar everyday sounds such as the sound of a matrix-dot printer or ringing a bell. Typical earcons, include such sounds as warning signals of the operating system or musical notes and usually need interpretation and a longer learning phase. Spearcons are speech-based earcons, specially compressed speech-segments, and words that have been proved to be superior to earcons or speeded-up speech, e.g., in menu navigations [34, 35]. Studies on the usability of spearcons for different languages and evaluation of auditory representations for the most important functions and events of a computer screen together with the introduction of auditory emoticons were recently presented and published [33, 36]. Although, these solutions provide improved accessibility in some cases, they do not replace speech in human-computer interaction.

Blind users contributed to the evaluation of such systems and welcomed the idea and the sound sets that can extend or even replace some parts of the usual TTS interfaces. Nevertheless, the investigations referred to above seldom, if ever, used spatially distributed sounds having directional information.

In the 1990s the GUIB project tried a multimodal interface, using tactile keyboards (Braille) and spatial distributed sounds, first with loudspeaker playback on the so-called "sound-screen," then using headphone playback and virtual simulation [21–23, 37]. In this project the Beachtron soundcard was used with real-time filtering of HRTFs to create a virtual audio environment [38]. In an attempt to create a better mapping of a rectangle computer screen as well as to increase navigation accuracy with the mouse a special two-dimensional surface was simulated in front of the listener, instead of the usual "around the head" concept. Listening tests were carried out first with sighted users using HRTF filtering, settings of ITD information based of head size data, broadband noise stimuli, and headphone playback. The results of these tests showed that increased rates of headphone errors such as in-the-head localization and front-back confusions and vertical localization was almost a complete failure [21, 23].

A follow-up study used additional high-pass and low-pass filtering to bias correct judgments in vertical localization (Fig. 1) and achieved about 90% of correct rates [39, 40]. Emulation of small head-movements without any additional hardware also seemed very useful in reducing errors [20, 41]. In-the-head localization errors were reduced for about one-third of the subjects in case of white noise stimulus, however, front-back error rates were unaffected.

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

Spatially distributed auditory events could be used in special window arrangements in different resolutions according to the user's experience and routine. In addition, distance information could be used for overlapping windows or other parameters [42].

0.2 Localization Performance of the Visually Impaired

Evaluation of the hearing ability of blind persons can include audiometric screening of the hearing threshold (sensitivity), localization tasks in real-life sound fields (mostly anechoic environments), and in augmented virtual realities (VADs).

Audiometric screening of blind subjects revealed no evidence for them having a better sensitivity and lower hearing thresholds indicating no difference at the peripheral evaluation in the hearing system in comparison to sighted subjects [43–45].

Non-virtual experiments showed that improved localization and distance perception can be achieved by training, by learning, and by adapting to the different coloration of signals also for sighted, partly, and late blind subjects [45–52]. Interference between direct and reflected waves creates spectral distortions, coloration of sound (such as comb-filters do) that relate with distance information [53, 54]. Early blind subjects may localize better in the horizontal plane and worse in the vertical plane in contrast to sighted [46, 48, 55–58]. Some blind people are able to determine distance, size, form, or even texture of obstacles based on auditory cues [46, 55, 59, 60]. Improved obstacle sense is mainly due to auditory feedback (reflections) [61, 62]. A research with two early and four late blind subjects, as well as six sighted participants in the anechoic chamber aimed at the ability of obstacle sense [63]. Comb filtered noise was applied to test distance perception. Blind participants performed better; however, differences in the ability of subjects to detect coloration of test signals by the filter was almost the same for both groups, supporting that localization performance is a central process in the brain. Horizontal plane localization was also found to be superior for four congenital blind subjects to four blindfolded sighted using 12 loudspeakers in an anechoic chamber with and without head movements [64]. Although, the number of participants is relatively low, blind persons performed better at determining the sound source direction and at estimating distance. Rotating the head improved only slightly the results for both groups. By comparing behavioral and electrophysiological indices of spatial tuning within central and peripheral auditory space in congenitally blind and normally sighted but blindfolded adults, it was found that blind participants displayed localization abilities that were superior to those of sighted controls, but only when attending to sounds in peripheral auditory space [65]. There is a sharper tuning of early spatial attention mechanisms in the blind subjects.

Real life training of blind persons include different training methods and exercises mostly in order to develop different equipment that improve orientation, safe movement, and avoiding collision with obstacles [53, 61, 62, 66–72]. These Electronic Travel Aids (ETAs) should be easy to use, have light weight during wearing, and not interfere with normal hearing. Virtual reality simulators can also offer training and Orientation and Mobility (O&M) tasks for the blind, including active feedback of auditory and haptic cues, such as beacon signals or vibrations [73–75].

In virtual environments, both spatial cues as well as sound events, have to be reproduced usually over headphones and in case of VADs, after selecting the best sounds for a given event or visual cue. Cobb et al. reported no differences between blind and sighted persons in accuracy for identifying environmental sounds [76]. However, when rating mappings from auditory icons to interface events, blind users gave significantly lower overall ratings of the appropriateness of the mappings [37]. One possible interpretation of this result is that blind users have stronger associations of sounds to physical events, and so to abstract away from the established nomic relationship to a more metaphorical one may be more difficult for them. This might be alleviated with training, but it poses a challenge for the design of auditory interfaces that will be immediately usable for visually-impaired listeners [77, 78]. Another interpretation might be that the auditory icons were not, in fact, very realistic representations of the real life sounds and blind people may have been more sensitive to discrepancies in the mapping. On the other hand, the technique called cartoonification can improve the acceptance of sound events in contrast to realistic mapping (e.g., a gunshot can be better recognized if a representation from a movie is used rather than a recording of a real gunshot). Blind users declared their interest in audio-only games, not only board and card games (chess, poker) but also action games [79]. However these investigations and solutions disregarded the spatial properties of sound sources.

A recent Polish experiment included non individualized HRTFs, individualized HRTFs derived from short-time HRIR measurements, and personalized HRTFs based on the CIPIC database with only nine untrained blind participants [80, 81]. Individual HRIRs were recorded at the blocked ear canal entrance in 5-degree horizontal and 9-degree vertical resolution. Missing directions' HRIRs were interpolated based on the minimal-phase decomposition for reaching a one-degree resolution. The applied headphone response was found to be linear without equalization. White noise sound source was simulated around the head and in the frontal plane. The broadband signal could be localized relatively well: error rates of about 6–14 degrees horizontally and 9–24 degrees vertically were measured depending on spectral content and movement. Localization results showed in-the-head localization and failure of localization in case of non individualized HRTFs. Improvement in localization by blind persons were observed mainly in the horizontal plane in case of broadband stimulus.

On the other hand, other experiments reported localization of blind persons to be significantly worse than sighted participants'. It was suggested that the reason for that could be that early blind persons had no possibility to learn the mapping between auditory events and visual stimuli [64, 82]. One late blind participant had better results than early blind subjects. Again, the number of participants was relatively low, and blind persons had more in-the-head localization errors in case of non individualized HRTFs.

Virtual audio simulators establish audio environments that are considerably different from those encountered in real-life situations. The use of headphones, non-individualized HRTFs, and approximation-based ITD information settings, as well as the lack of head-tracking and reflections during playback can lead to well-known localization errors. Since it is assumed that increased localization performance and source detection ability of blind subjects is based on echolocation and coloration (both rely on the interference and reflection patterns of direct waves), it is also conjectured that attention and training are key factors in developing such abilitities. Nevertheless, loss of important localization cues in virtual environments can lead to increased error rates and inconsistent experimental results, and is also expected to decrease or completely nullify differences between the performance of sighted and visually impaired subjects. The primary goal of this experiment is to confirm similarities and find differences in performance between these groups in selected listening tests within virtual environments based on a larger number of participants.

1 MEASUREMENT SETUP

The measurement setup has been described in details elsewhere [20, 21, 23, 38, 40]. The system is shipped with the circumaural, open-dynamic Sennheiser HD540 headphone and contains a personal computer equipped with a Beachtron DSP sound card that affords easy programming and dynamic settings of source distance and HRTFs. Real-time convolution of the mono input signal and the HRTFs is made in the time-domain. The HRTFs originate from a “good localizer” female in a measurement of Wightman and Kistler [39, 83, 84]. Seventy-two measured HRTFs are available in a form of 75-point minimum-phase-FIR-filter set in 30° spatial resolution (see Appendix). Missing directions are calculated by linear interpolation from the four nearest available measured directions and all subjects used the same HRTF set during the measurement.

The first tests were made with 40 sighted and untrained subjects, all with normal hearing. This included 20 males and 20 females between 21 and 39 years of age (mean 28) [21, 23]. Subjects filled in a questionnaire about personal data (gender, age), computer skills, and headphone user routine. Seven percent wear headphones everyday; 24% often; 59% seldom; and 10% never. The 28 visually impaired participants included 22 males and 6 females. The minimum age was 19, the maximal 64 with a mean value of 34 years. Sixteen subjects were born blind and the remainder lost vision between 6 and 20 years of age, usually degeneratively due to some kind of disease. Regarding headphone usage, 10 subjects reported using them seldom or never, 10 frequently, and 8 daily.

First, a 300 ms white noise burst had to be localized in an absolute localization task. A static sound source was emulated in front and back of the listener to test front-back reversals and in-the-head localization. Following this, a (virtually) moving sound source around the head had to be identified (direction of circling) as well as movements of the sound source up and down in the median plane.

This was followed by a Minimum-Audible-Angle (MAA) discrimination task including two of the 300 ms white noise burst separated with silence. MAA values were determined in case of a moving source left, right, up, and down compared to a static source in the origin in front of the listener. Novelties and general conditions in this MAA-measurement were:

- The use of a 2-D virtual sound screen in the front of the listener. Sources could move only in the horizontal (left and right) and in the median plane (up and down) from the origin in 1° steps. This resulted in the source distances not being constant and the sources not appearing to move around the head.
- Subjects were asked to categorize the sounds into the following groups: “no difference between the sources,” “different sound sources,” and “I’m not sure.” Subjects had the possibility to be uncertain about their sensation: if sound sources seemed to be completely identical or completely different, they selected one of the first two options, in any other cases they were uncertain.
- Burst-pairs had to be discriminated (a) as the second source moves away from the static reference source, then (b) as it moves toward the reference point. We were looking for the nearest point to the reference, from which the subject is able to discriminate the sources with certainty from both directions. If the localization blur can be determined from both direction of movement, a direction-independent localization performance result is obtained.

Finally, a sound source discrimination task was conducted on a 3 × 3 grid in a 2-D VAD. Subjects responded and reported on a questionnaire by answering the dedicated questions of the experimenter.

While today’s techniques allow virtual simulation without any dedicated hardware by using software-only solutions, e.g., the VibeStudio Designer [85], in order to keep all the parameters for a correct comparison of results, we used exactly the same system setup for both groups of subjects, despite the fact that the Beachtron card is not a state-of-the-art solution for virtual audio. All the measurements with sighted and blind participants covered several years from 2003 and instead of upgrading the measurement equipment during this period, so as to obtain an accurate comparison, we decided to use exactly the same setup for all participants. Measurements carried out for comparison with blind participants in 2009 and 2010 were selected based on the experience with the measurement setup and on former results obtained with sighted participants.

Table 1. Comparison of blind and sighted users in a listening test for front-back confusion. Static sound source was simulated in front of the listener. Total and relative numbers are shown for correct answers (left column), correct but uncertain answers (middle) and errors (right). Right-most column shows in-the-head localization rates.

	Front (correct)	Front (unsure)	Back (error)	IHL
Blind (total 28)	12	12	4	22
%	42.9	42.9	14.2	78.6
Sighted (total 40)	5	7	28	32
%	12.5	17.5	70.0	80.0

Table 2. Comparison of blind and sighted users in a listening test for front-back confusion. A static sound source was simulated in the back of the listener. Total and relative numbers are shown for correct answers, correct but uncertain answers and errors (frontal image and no answer).

	Back (correct)	Back (unsure)	Front (error)	No answer (error)
Blind (total 28)	5	7	8	8
%	17.9	24.9	28.6	28.6
Sighted (total 40)	25	13	2	0
%	62.5	32.5	5.0	0

2 RESULTS

This section presents localization results for absolute and MAA measurements of blind participants as well as former results of sighted subjects for comparison. Based on F-tests there was no difference between population variances. Due to relatively large sample sizes Z-tests and T-tests could be used for testing the difference between population proportions. In all cases a significance level 0.05 was used for statistical analysis.

2.1 Localization of a Static Source

The first experiment included absolute localization tasks of a static sound source. The stimulus was a repeated 300 ms white noise burst separated by 400 ms of silence. The main goal was to detect front-back confusion and in-the-head localization rates. The static source was first emulated in the frontal direction ($\delta = \varphi = 0°$). Results were surprisingly good: 24 correct answers (86%). As a control question, subjects were asked whether or not the sound source appeared behind them and half of the subjects were unsure and only guessed, indicating that the results were unreliable (Table 1). The same task using an emulated sound source behind ($\delta = 0°$, $\varphi = 180°$) shows decreased performance, only five reliable and seven good guess answers appeared. Sixteen subjects indicated they thought it was from the front or gave no answer at all (Table 2).

After explaining to the subjects what in-the-head localization means, a large number reported lack of real externalization in case of the frontal source simulation (78.6%). Tables 1–2 show results for blind and sighted users for comparison. While detection of the frontal source was clearly better by blind persons ($Z = \pm1.96$, $Z = 2.845$, $p = 0.0044$), back direction was detected better by sighted users ($Z = \pm1.96$, $Z=-3.649$, $p<0.001$). In-the-head localization rates were almost the same for both groups and quite large ($Z = \pm1.96$, $Z=-0.143$, $p = 0.886$). Subjects reported about in-the-head localization only once because it was not influenced by the direction of the emulated sound source.

Table 3. Comparison of blind and sighted users in a listening test for a moving sound source in the horizontal plane. The sound source was moving around the head to the left. Total and relative numbers are shown for correct answers (left column) and errors.

	Movement to the left (correct)	Movement behind the head only	Other answers
Blind (total 28)	28	0	0
%	100.0	0	0
Sighted (total 40)	20	17	3
%	50.0	42.5	7.5

Table 4. Comparison of blind and sighted users in a listening test for a moving sound source in the median plane. The virtual sound source was rising and falling from the origin in front of the listener. Total and relative numbers are shown for correct answers, correct but uncertain answers and errors in both directions respectively.

	Moving up (correct)	Up (unsure)	Down (error)
Blind (total 28)	5	10	13
%	17.9	35.7	46.4
Sighted (total 40)	34	3	3
%	85.0	7.5	7.5
Blind (total 28)	13	13	2
%	46.4	46.4	7.2
Sighted (total 40)	20	15	5
%	50.0	37.5	12.5

2.2 Localization of a Moving Source

In case of a moving sound source, the task was to detect the source moving around the head in the horizontal plane to the left. This emulation does not use the 2-D rectangle virtual screen concept but the usual around-the-head method. All blind subjects could detect this correctly showing a much better performance than sighted participants did (Table 3). There is an obvious difference between the result for sighted and blind people in for task ($Z = \pm1.96$, $Z = 4.453$, $p<0.001$). Many of the sighted subjects reported the source to be moving only behind the head (in the back hemisphere). Only 50% were able to detect the correct movement. All blind subjects performed this task without error.

The second task for subjects was to detect first the virtual source rising from the origin in the median plane, followed by the movement down from the origin (Table 4). Surprisingly, sighted users performed much better in the case

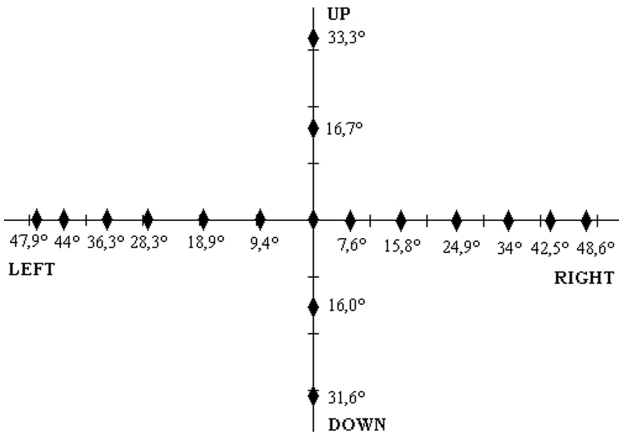

Fig. 2. Mean values as possible source locations for white noise excitation in an MAA measurement with sighted people [14]. The first locations are listed in the right column of Table 5 for comparison.

of rising source than blind users (Z = ±1.96, Z=−5.509, p<0.001). The difference is less and not significant in case of falling virtual sources (Z = ±1.96, Z=−0.290, p = 0.772). For blind users, the falling virtual sound sources were easier to detect than those rising.

Surprisingly, changes of the elevation in the median plane was easily recognized by sighted users. Some authors reported decreased performance from the lower hemisphere. Our results do not support this finding as 92.5% and 87.5% were able to detect the correct direction. However, there was greater uncertainty when the virtual sources were descending.

Among our experimental subjects were eight musicians with absolute pitch but they did not perform these tasks any better by any means. Although all of them reported in-the-head localization, and although front-back and up-down decisions made by them were correct, all were unsure about it.

2.3 Localization in an MAA-Task

Our former measurements with sighted users included an MAA measurement using the same white noise excitation signal. Instead of a single moving source, the task was to detect the difference between two 300 ms bursts separated by 300 ms. The first part of the burst pair was steady (at the origin) as long as the second part was moving away or toward this reference point left, right, up, or down. The task was to detect whether they were similar or different. The MAA was determined where a subject could discriminate the two source directions both when the sounds were moving away *and* toward each other. Our former investigation also used LPF and HPF filtered versions of white noise, as well as several new locations in the 2-D screen [40]. Fig. 2 shows averaged results for sighted users in all directions. For example, if a steady sound source is emulated at the origin and another moving away from and/or toward it on the left side, and the subjects can discriminate them clearly, an MAA of 9.4 degrees on average will result.

Table 5. Comparison of blind and sighted users in an MAA task in front of the listener. A steady sound source in the origin can be discriminated from a moving source by 6.2–16.7 degrees depending on the direction of movement. Maximum, minimum, and mean values are shown for both target groups.

	Blind			Sighted		
Left	18	4	7.7	20	4	9.4
Right	14	4	6.2	14	3	7.6
Up	28	6	14.1	32	8	16.7
Down	30	8	15.2	29	3	16.0
	max	min	mean	max	min	mean

Table 6. Comparison of blind user's results in the same task as shown in Table 5.

	Never/seldom			Often/daily		
Left	16	4	7.8	18	5	7.6
Right	12	4	6.6	14	4	5.9
Up	28	6	14.4	24	8	13.9
Down	30	8	16.2	29	8	14.2
	max	min	mean	max	min	mean

For simplification reasons, we present here white noise and for one MAA location only: for right, left, up, and down movements respectively.

Table 5 shows results for blind and sighted users. Blind users performed somewhat better in the horizontal plane than sighted users, but differences are statistically not significant (T(66) = ±1.996, t = 1.63, p = 0.107 left; t = 1.96, p = 0.054 right; t = 1.98, p = 0.052 up; t = 0.55, p = 0.585 down).

As mentioned, all subjects had to fill out a questionnaire about headphone usage. Thirty-one percent of sighted users and 64% of blind persons were "often/everyday" users. Our former results of sighted users were found to be independent of age and gender, and little improvement in the localization performance was found by subjects using headphones often: mean localization blur in this case was reduced by 0.4°–2.1°. Table 6 shows results for blind persons only. Results were calculated the same way as in Table 5. Little improvement can be seen in the case of subjects using headphones frequently: a mean improvement of about 0.2°–2° was observed for different directions, however, this difference is statistically not significant (T(26) = ±2.379, t = 0.129, p = 0.899 left; t = 0.628, p = 0.535 right; t = 0.219, p = 0.828 up; t = 0.824, p = 0.417 down).

2.4 Localization on a 2-D VAD

The last test included a 2-D virtual audio display simulated in front of the listener in a 3 × 3 grid (Fig. 3). Prior to the listening test, a detailed description was given to the subjects and they reported localization judgments by indicating a field by letter and number. The virtual sound source was emulated in the middle of each field using HRTF filtering, activated one by one in a randomized order. Every subject performed two rounds and, thus, delivered 18 responses.

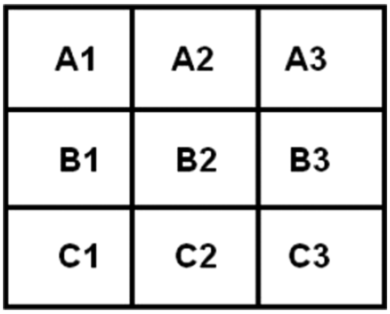

Fig. 3. Spatial resolution of a 2-D virtual audio display in a 3 × 3 grid in front of the listener. Subjects indicate fields by letter and number in the listening test.

Table 7. Results of 28 blind subjects giving 18 answers each in two rounds (504 in total). Columns correspond to active target fields, rows correspond to given answers (see Fig. 3.). Numbers in the diagonal show correct answers.

Active field:	A1	A2	A3	B1	B2	B3	C1	C2	C3
A1 (answer)	38			22			12		
A2		38			14			4	
A3			42			16			8
B1	6			28			8		
B2		12			38			18	
B3			10			30			12
C1	12			6			36		
C2		6			4			34	
C3			4			10			36

The former investigation with sighted subjects using the same setup resulted in a mean score for correct answers of 48.3% [40]. Vertical errors were 47.5% and surprisingly, there was also 4.2% of horizontal errors. In the current investigation, blind users had no horizontal errors at all. Table 7 shows correct answers in the diagonal. Due to the number of subjects, we collected a total of 504 (63.5% correct) and 720 answers respectively (48.3% correct). Blind subjects performed significantly better as they had 15% more correct answers than sighted users did ($Z = \pm 1.96$, $Z = 5.242$, $p<0.001$).

The most common error was to report B1 instead of A1, B3 instead of A3, and C2 instead of B2. All errors were in vertical directions only. Someone was declared to be a "good localizer" if they scored at least 16 correct answers out of 18. We only had two such subjects. More than 11 subjects scored 6. The worst result was 4. The mean value of scores across subjects is 11.5 (64%), whereas for sighted users it was only 8.7 (48.3%).

It was found that users who do not use headphones frequently also had decreased performance in this task. Their mean score was only 10 out of 18 answers, compared with 13 for users who wear headphones often.

Our former test also included a grid of 2 lines and 5 columns to test a decreased vertical and increased horizontal resolution. Sighted users had 50.7% correct answers, 29.3% vertical, and 20.0% horizontal errors. We performed this as an informal test with only two blind subjects. From the 40 given answers only 4 were wrong and all errors were in vertical directions. Although this is 90% correct, it is not a relevant comparison due to the small number of subjects.

3 DISCUSSION

Real-life experiments showed that improved localization and distance perception can be achieved by training and by adapting to the different coloration of signals based on sound reflections [45-54]. Early blind subjects may localize better in the horizontal plane and worse in the vertical plane in contrast to sighted [46, 48, 55-58]. Our results support this observation in the virtual environment, except for back directions. In virtual environments spatial cues have to be reproduced over headphones. An experiment with non-individualized and individualized HRTFs, white noise sound source, and headphone playback resulted in error rates of about 6–14 degrees horizontally and 9–24 degrees vertically [80, 81]. In our case error rates were 4–18 degrees and 6–30 degrees respectively. Localization results showed in-the-head localization and failure of localization in case of non individualized HRTFs. Some improvement in localization by blind persons was observed mainly in the horizontal plane. On the other hand, other experiments reported localization of blind persons to be significantly worse than sighted participants' [64, 82]. At a relatively low number of participants, blind persons had more in-the-head localization errors in case of non individualized HRTFs. We can also support that blind participants have about the same in-the-head localization.

Our results show that some localization tasks can be solved better by blind than by sighted persons. In general, blind users can perform better in virtual localization tasks. The most significant is the 14.7% improvement in the 3 × 3 grid localization task (Table 7). Furthermore, blind users could localize static front sources better than sighted users (Table 1). The movement around the head was perfect for blind persons, but only 50% correct by sighted users (Table 3). This supports former observations that blind subjects have better localization performance in the horizontal plane [46, 48, 55–58, 64, 80, 81].

On the other hand, sighted users performed better if the static source was behind them, mainly due to the fact that most of the sighted users reported sensation only in the back hemisphere (Table 2). The same is reflected in the results that only 50% could determine the movement around the head and 42.5% had an image only in the rear hemisphere (Table 3). The surprising finding was that detection of ascending movements was clearly better than by blind users (Table 4). We can speculate that an increase number of front-back reversals, where frontal images tend to be localized in the back, assists to have more correct answers in case of a sound source actually simulated in the back.

Two parameters, results of the MAA measurement and in-the-head localization rates show almost identical results for both groups (Table 1 and Table 5). Although blind users had better spatial resolution in the discrimination task of about 1–2 degrees horizontally and 1.4% less in-the-head localization reports, these relative differences are too small to be significant.

Some tasks resulted in the same performance results for both groups and so conclusions can be drawn independent of the visual status of the subject. In contrast to static sources if the sound source is moving, the movement helps all subjects to judge ascending/descending directions in the median plane. The same is true for front-back confusion: a movement around the head makes it easier to detect front and back directions. In the two-dimensional 3 × 3 and 5 × 2 tasks sometimes the random placement generator has activated fields next to each other (e.g., A2 after A1 or B3 after C3, etc.). Subjects reported that judging directions was made much easier when there was a reference. If the fields were not neighbors, the task was more difficult. We did not observe any significant difference between genders or between subjects who claimed to have musical background and absolute pitch. Both sighted and visually impaired subjects usually reported not being able to localize vertical sources but only made judgments by the spectral coloration difference because the ascending and descending sources "sound different" (perhaps according to the filtering effect of the HRTFs). Users who often wear headphones performed better in both groups. Our former observation that sources from the left side are typically harder to localize was supported by the blind subjects as well. Results with sighted users showed 2–4° average differences, while the value for the blind users is 1.5°. Sixty-seven percent of sighted subjects had decreased spatial resolution on the left side and only 6% on the right side. For blind persons this was even more: 75% and 7%. We suggest that this may be related to left or right handedness, because most of our subjects were right handed and left handed users had decreased resolution on the right side of the horizontal plane. Only about 10% of our subjects were left handed so this could be something for further investigations. Whether someone was born blind or lost vision later did not influence the results in our case. The best and the worst localizer also were not born blind.

All the participants in this comparative study were untrained and no improvement of localization skills was observed throughout the experiments. Repeated measurements might reveal learning processes and improvements, but in this case, subjects performed the tasks only once. The three-category evaluation allows us to ask participants about the reliability of their answers. Correct but doubtful answers may indicate guessing and lucky judgments rather than real localization. Sighted users also sometimes reported the lack of real localization in the vertical plane and making decisions based on different coloration of the signals only. An increased number of insecure answers may highlight the need for repeated control measurements with a system.

Every significant difference in favor of blind subjects originates in the lower number of front-back reversal rates, which seems to determine successful virtual detection and localization of sound sources. We can speculate that the reason for this is that blind subjects are more accustomed to having sound sources mostly in front of them as they face them often during real-life situations. This may help them to develop the ability of correct detection of virtual frontal sources, even without head-tracking. The only task where sighted subjects significantly out-performed blind subjects was the detection of "upward" movement in the median plane. Further investigations have to be performed in free-field environments to validate these test results and confirm these hypotheses.

At the end of the investigation we had some informal conversations with the blind people. Most of them said that they did not think that they hear better than sighted people but they pay more attention to what they hear. They believed that they develop the ability to discriminate sound sources and to separate them. Not every blind person had developed this skill and sighted users may be able attain the same skills with practice. Blind people think that they obtain a lot of information from reverberations from walls and objects on the street. A corner or crossing can be detected by the loss of the reflections of environmental sounds from the buildings as they walked past them. Also head-movements seem to be very important during this task. Blind people make intensive head-movements as they scan the acoustic field, while sighted people make slow head-movements. Even the surface materials of objects can be detected by sounds. While only about half of the blind participants believed they have a better sensitivity to lower sound pressure levels than sighted people, only some of them had better localization (spatial resolution) abilities.

This investigation is part of an international project in which the goal is to evaluate and compare localization performance of blind and sighted subjects. Besides virtual localization, free-field orientation and navigation tasks are performed to test the effect of acoustic beacon signals on veering during blindfolded walking as well as to examine echolocation during the detection of silent objects [86, 87].

4 FUTURE WORK

Future work includes the implementation of selected auditory icons and emoticons in the computer environment mentioned in the Introduction, as well as the development of an Electronic Travel Aid for the blind community. Both incorporate spatially distributed sounds mostly via headphone playback. In addition, regarding spatial issues, we plan further investigations with blind and sighted people in free-field environments and everyday life situations so as to develop a comparison with non-virtual localization performance.

5 SUMMARY

Virtual localization performance of 28 blind persons was investigated using a 2-D virtual audio display in front of the listener and HRTF synthesis. New results were presented and compared with former results of 40 sighted users with the same equipment and measurement setup. Blind subjects delivered better results on a 3 × 3 grid and in localizing static frontal sources (due to a decreased number of front-back reversals). In the case of moving sources, they were more accurate in determining movements around the head in the horizontal plane. On the other hand, sighted participants performed better during tests in which the task was to listen to ascending movements in the median plane and to identify sound sources in the back. Results of an MAA measurement in front of the listener, the ability to detect descending movements and in-the-head localization rates are almost identical for the two groups. In general, blind subjects performed at least as well as sighted subjects did on this virtual audio display, which indicates that experience and training in real-life environments can lead to better virtual source detection and localization. Other factors, such as gender, age, having a musical background or even absolute pitch, or being born blind had no influence on the results. Evaluation was also supported by informal questions.

6 ACKNOWLEDGMENT

The author would like to thank Gábor Kovács for the help with the statistics and all the participants, especially the elderly and visually impaired. This investigation was supported and funded by the Universitas-Győr Alapítvány in the framework "TÁMOP 4.1.1./A-10/1/KONV-2010-0005."

7 REFERENCES

[1] J. Blauert, *Spatial Hearing* (The MIT Press, MA, 1983).

[2] M. Cohen, and E. Wenzel, "The Design of Multidimensional Sound Interfaces," in W. Barfield, T. A. Furness III (Eds.) *Virtual Environments and Advanced Interface Design*, pp. 291–346 (Oxford Univ. Press, New York, Oxford, 1995).

[3] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia* (Academic Press, London, UK, 1994).

[4] E. H. A. Langendijk, and A. W. Bronkhorst, "Fidelity of Three-Dimensional-Sound Reproduction Using a Virtual Audio Display," *J. Acoust. Soc. Am.*, vol. 107, pp. 528–537 (2000).

[5] E. M. Wenzel, "Localization in Virtual Acoustic Displays," *Presence*, vol. 1, pp. 80–107 (1992).

[6] M. Vorländer, *Auralization: Fundamentals of Acoustics, Modeling, Simulation, Algorithms, and Acoustic Virtual Reality* (Springer, Berlin, 2008).

[7] C. I. Cheng, and G. H. Wakefield, "Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space," *J. Audio Eng. Soc.*, vol. 49, pp. 231–249 (2001 Apr.).

[8] H. Møller, M. F. Sorensen, D. Hammershøi, and C. B. Jensen, "Head-Related Transfer Functions of Human Subjects," *J. Audio Eng. Soc.*, vol. 43, pp. 300–321 (1995 May).

[9] D. Hammershoi, and H. Moller, *Binaural Technique—Basic Methods for Recording, Synthesis, and Reproduction*, in J. Blauert (Ed.) *Communication Acoustics*, pp. 223–254 (Springer Verlag, 2005).

[10] F. Wightman, and D. Kistler, "Measurement and Validation of Human HRTFs for Use in Hearing Research," *Acta Acustica united with Acustica*, vol. 91, pp. 429–439 (2005).

[11] D. R. Begault, E. Wenzel, and M. Anderson, "Direct Comparison of the Impact of Head Tracking Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source," *J. Audio Eng. Soc.*, vol. 49, pp. 904–917 (2001 Oct.).

[12] P. Minnaar, J. Plogsties, and F. Christensen, "Directional Resolution of Head-Related Transfer Functions Required in Binaural Synthesis," *J. Audio Eng. Soc.*, vol. 53, pp. 919–929 (2005 Oct.).

[13] K. A. J. Riederer, "Repeatability Analysis of Head-Related Transfer Function Measurements," presented at the*105th Convention of Audio Engineering Society* (1998 Sept.), convention paper 4846.

[14] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization Using Nonindividualized Head-Related Transfer Functions," *J. Acoust. Soc. Am.*, vol. 94, pp. 111–123 (1993).

[15] P. Minnaar, S. K. Olesen, F. Christensen, and H. Møller, "Localization with Binaural Recordings from Artificial and Human Heads," *J. Audio Eng. Soc.*, vol. 49, pp. 323–336 (2001 May).

[16] H. Møller, M. F. Sorensen, C. B. Jensen, and D. Hammershøi, "Binaural Technique: Do We Need Individual Recordings?" *J. Audio Eng. Soc.*, vol. 44, pp. 451–469 (1996 June).

[17] M Hiipakka, M. Tikander, and M. Karjalainen, "Modeling the External Ear Acoustics for Insert Headphone Usage," *J. Audio Eng. Soc.*, vol. 58, pp. 269–281 (2010 Apr.).

[18] H. Møller, "Fundamentals of binaural technology," *Applied Acoustics*, vol. 36, pp. 171–218 (1992).

[19] P. A. Hill, P. A. Nelson, and O. Kirkeby, "Resolution of Front-Back Confusion in Virtual Acoustic Imaging Systems," *J. Acoust. Soc. Am.*, vol. 108, pp. 2901–2910 (2000).

[20] Gy. Wersényi, "Effect of Emulated Head-Tracking for Reducing Localization Errors in Virtual Audio Simulation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 17, pp. 247–252 (2009 Feb.).

[21] Gy. Wersényi, "Localization in an HRTF-Based Minimum Audible Angle Listening Test on a 2-D Sound Screen for GUIB Applications," presented at the *115th Convention of Audio Engineering Society* (2003 Oct.), convention paper 5902.

[22] K. Crispien, and H. Petrie, "Providing Access to GUI's Using Multimedia System—Based on Spatial Audio Representation," presented at the *95th Convention of Audio Engineering Society* (1993 Oct.), convention paper 3738.

[23] Gy. Wersényi, "Localization in an HRTF-Based Minimum-Audible-Angle Listening Test for GUIB Applications," *Electronic Journal of Technical Acoustics*, vol. 1, http://www.ejta.org, 16 pages (2007).

[24] W. W. Gaver, "Auditory Icons: Using Sound in Computer Interfaces," *Human-Computer Interactions*, vol. 2, pp. 167–177 (1986).

[25] W. W. Gaver, "Everyday Listening and Auditory Icons," *Doctoral thesis* (Univ. of California, San Diego, 1988).

[26] E. D. Mynatt, "Designing Auditory Icons," in *Proc. of the International Conference on Auditory Display ICAD 94* (1994), pp. 109–120.

[27] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, "Earcons and Icons: Their Structure and Common Design Principles," *Human-Computer Interaction*, vol. 4, pp. 11–44 (1989).

[28] L. H. Boyd, W. L. Boyd, and G. C. Vanderheiden, "The Graphical User Interface: Crisis, Danger and Opportunity," *J. Visual Impairment and Blindness*, pp. 496–502 (1990 Dec.).

[29] W. W. Gaver, "The SonicFinder, a Prototype Interface that Uses Auditory Icons," *Human Computer Interaction*, vol. 4, pp. 67–94 (1989).

[30] E. D. Mynatt, "Transforming Graphical Interfaces into Auditory Interfaces for Blind Users," *Human-Computer Interaction*, vol. 12, pp. 7–45 (1997).

[31] F. Grond, T. Drossard, and T. Hermann, "Sonifunction: Experiments with a Function Browser for the Visually Impaired," *inProc. of the International Conference on Auditory Display ICAD 10* (2010), pp. 15–21.

[32] G. Parseihian and B. Katz, "Morphocons: A New Sonification Concept Based on Morphological Earcons," *J. Audio Eng. Soc*., vol. 60, pp. 409–418 (2012 Jun.).

[33] Gy. Wersényi, "Auditory Representations of a Graphical User Interface for a Better Human-Computer Interaction," in S. Ystad et al. (Eds.): *Auditory Display-CMMR/ICAD 2009 post proceedings edition, Lecture Notes in Computer Science (LNCS) 5954*, pp. 80–102 (Springer Verlag, Berlin, 2010).

[34] T. Dingler, J. Lindsay, and B. N. Walker, "Learnabiltiy of Sound Cues for Environmental Features: Auditory Icons, Earcons, Spearcons, and Speech," in *Proc. of the International Conference on Auditory Display ICAD 08* (2008), 6 pages.

[35] D. K. Palladino, and B. N. Walker, "Learning Rates for Auditory Menus Enhanced with Spearcons versusEarcons," in *Proc. of the International Conference on Auditory Display ICAD 07* (2007), pp. 274–279.

[36] Gy. Wersényi, "Evaluation of User Habits for Creating Auditory Representations of Different Software Applications for Blind Persons," in *Proc. of the International Conference on Auditory Display ICAD 08* (2008), 5 pages.

[37] H. Petrie, and S. Morley, "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* (1998), 5 pages.

[38] Crystal River Engineering, Inc., *BEACHTRON – Technical Manual*, Rev.C. (1993).

[39] Gy. Wersényi, "What Virtual Audio Synthesis Could Do for Visually Disabled Humans in the New Era," in *Proc. of the of the AES Tokyo Regional Convention* (2005), pp. 180–183.

[40] Gy. Wersényi, "Localization in an HRTF-Based Virtual Audio Synthesis Using Additional High-Pass and Low-Pass Filtering of Sound Sources," *J. Acoust. Science and Technology Japan*, vol. 28, pp. 244–250 (2007 Apr.).

[41] Gy. Wersényi, "Simulation of Small Head-Movements on a Virtual Audio Display Using Headphone Playback and HRTF Synthesis," in *Proc. of the International Conference on Auditory Display ICAD 07* (2007), pp. 73–78.

[42] A. Martin, C. Jin, and A. van Schaik, "Psychoacoustic Evaluation of Systems for Delivering spatialized Augmented-Reality Audio," *J. Audio Eng. Soc.*, vol. 57, pp. 1016-1027 (2009 Dec.).

[43] M. Bross, and M. Borenstein, "Temporal Auditory Acuity in Blind and Sighted Subjects: A Signal Detection Analysis," *Percept. Mot. Skills*, vol. 55, pp. 963–966 (1982).

[44] I. Starlinger, and W. Niemeyer, "Do the Blind Hear better? Investigations on auditory processing in congenital or early acquired blindness. I. Peripheral functions," *Audiology*, vol. 20, pp. 503–509 (1981 June).

[45] L. Benedetti, and M. Loeb, "A comparison of auditory monitoring performance in blind subjects with that of sighted subjects in light and dark," *Attention, Perception, & Psychophysics*, vol. 11, pp. 10–16 (1972).

[46] C. E. Rice, "Human echo perception," *Science*, vol. 155, pp. 656–664 (1967 Feb.).

[47] W. Niemeyer, and I. Starlinger, "Do the Blind Hear Better? Investigations on Auditory Processing in Congenital or Early Acquired Blindness. II. Central Functions," *Audiology*, vol. 20, pp. 510–515 (1981 June).

[48] C. Muchnik, M. Efrati, E. Nemeth, M. Malin, and M. Hildesheimer, "Central Auditory Skills in Blind and Sighted Subjects," *Scandinavian Audiology*, vol. 20, pp. 19–23 (1991).

[49] L. D. Rosenblum, M. S. Gordon, and L. Jarquin, "Echolocating Distance by Moving and Stationary Listeners," *Ecological Psychology*, vol. 12, pp. 181–206 (2000 Mar.).

[50] J. P. Rauschecker, "Compensatory Plasticity and Sensory Substitution in the Cerebral Cortex," *Trends in Neurosciences*, vol. 18, pp. 36–43 (1995 Jan.).

[51] N. Lessard, M. Pare, F. Lepore, and M. Lassonde, "Early-Blind Human Subjects Localize Sound Sources Better than Sighted Subjects," *Nature*, vol. 395, pp. 278–280 (1998 Sept.).

[52] M. C. Wanet-Defalque, C. Veraart, A. De Volder, R. Metz, C. Michel, G. Dooms, and A. Goffinet, "High Metabolic Activity in the Visual Cortex of Early Blind Human Subjects," *Brain Research*, vol. 446, pp. 369–373 (1988 Apr.).

[53] Y. Seki, T. Ifukube, and Y. Tanaka, "Relation between the Reflected Sound Localization and the Obstacle Sense of the Blind," *J. Acoust. Soc. Jpn.*, vol. 50, pp. 289–295 (1994).

[54] I. G. Bassett, and E. J. Eastmond, "Echolocation: Measurement of Pitch versus Distance for Sounds Reflected from a Flat Surface," *J. Acoust. Soc. Am.*, vol. 36, pp. 911–916 (1964 May).

[55] W. N. Kellogg, "Sonar System of the Blind," *Science*, vol. 137, pp. 399–404 (1962 Aug.).

[56] H-H. Lai, and Y-C. Chen," A Study on the Blind's Sensory Ability," *Int. J. Industrial Ergonomics*, vol. 36, pp. 565–570 (2006).

[57] A. Dufour, O. Despres, and V. Candas, "Enhanced Sensitivity to Echo Cues in Blind Subjects," *Exp. Brain Res.*, vol. 165, pp. 515–519 (2005 July).

[58] M. P. Zwiers, A. J. V. Opstal, and J. R. M. Cruysberg, "A Spatial Hearing Deficit in Early-Blind Humans," *J. of Neuroscience*, vol. 21, pp. 1–5 (2001 May).

[59] C. E. Rice, and S. H. Feinstein, "Sonar System of the Blind: Size Discrimination," *Science*, vol. 148, pp. 1107–1108 (1965 May).

[60] T. Stroffregen, and J. Pittenger, "Human Echolocation as a Basic Form of Perception and Action," *Ecological Psychology*, vol. 7, pp. 181–216 (1995 July).

[61] M. Supa, M. Cotzin, and K. M. Dallenbach, ""Facial Vision": The Perception of Obstacles by the Blind," *Am. J. Psychology*, vol. 57, pp. 133–183 (1944).

[62] M. Cotzin, and K. M. Dallenbach, ""Facial Vision": The Role of Pitch and Loudness in the Perception of Obstacles by the Blind," *Am. J. Psychology*, vol. 63, pp. 485–515 (1950).

[63] T. Miura, T. Muraoka, and T. Ifukube, "Comparison of Obstacle Sense Ability between the Blind and the Sighted: A Basic Psychophysical Study for Designs of Acoustic Assistive Devices," *Acoust. Sci. & Tech.*, vol. 31, pp. 137–147 (2010).

[64] M. Ohuchi, Y. Iwaya, Y. Suzuki, and T. Munekata, "A Comparative Study of Sound Localization Acuity of Congenital Blind and Sighted People," *Acoust. Sci. & Tech.*, vol. 27, pp. 290–293 (2006).

[65] B. Röder, W. Teder-Sälejärvi, A. Sterr, F. Rösler, S. A. Hillyard, and H. J. Neville, "Improved Auditory Spatial Tuning in Blind Humans," *Nature*, vol. 400, pp. 162–166 (1999).

[66] Y. Seki, "Acoustical Design of City for the Visually Handicapped," *J. Acoust. Soc. Jpn.*, vol. 54, pp. 387–392 (1998).

[67] T. Ifukube, T. Sasaki, and C. Peng, "A Blind Mobility Aid Modeled after Echolocation of Bats," *IEEE Trans. Biomed. Eng.*, vol. 38, pp. 461–465 (1991).

[68] N. Pressey, "Mowat Sensor," *Focus*, vol. 11, pp. 35–39 (1977).

[69] L. Kay, "The Design and Evaluation of a Sensory Aid to Enhance Spatial Perception of the Blind," *Electrical Eng. Report 21*, Dept. Elect. Eng., Univ. Canterbury (New Zealand, 1973).

[70] L. Kay, "Auditory Perception of Objects by Blind Persons, Using a Bioacoustic High Resolution Air Sonar," *J. Acoust. Soc. Am.*, vol. 170, pp. 3266–3275 (2000 June).

[71] S. Cardin, D. Thalmann, and F. Vexo, "A Wearable System for Mobility Improvement of Visually Impaired People," *Visual Computer*, vol. 23, pp. 109–118 (2007).

[72] P. Worchel, and K. M. Dallenbach, ""Facial Vision": Perception of Obstacles by the Deaf-Blind," *Am. J. Psychology*, vol. 60, pp. 502–553 (1947).

[73] O. Lahav, D. W. Schloerb, S. Kumar, and M. A. Srinivasan, "BlindAid: A Learning Environment for Enabling People Who Are Blind to Explore and Navigate through Unknown Real Spaces," in *Proc. of Conference on Virtual Rehabilitation* (2008), pp. 193–197.

[74] E. W. Hill, J. J. Rieser, M. M. Hill, M. Hill, J. Halpin, and R. Halpin, "How Persons with Visual Impairments Explore Novel Spaces: Strategies of Good and Poor Performers," *J. Visual Impairment and Blindness*, vol. 87, pp. 295–301 (1993).

[75] O. Lahav, and D. Mioduser, "Construction of Cognitive Maps of Unknown Spaces Using a Multi-Sensory Virtual Environment for People Who Are Blind," *Computers in Human Behavior*, vol. 24, pp. 1139–1155 (2008).

[76] N. J. Cobb, D. M. Lawrence, and N. D. Nelson, "Report on Blind Subjects' Tactile and Auditory Recognition for Environmental Stimuli," *Percept. Mot. Skills*, vol. 48, pp. 363–366 (1979).

[77] B. Gygi, and V. Shafiro, "From Signal to Substance and Back: Insights from Environmental Sound Research to Auditory Display Design," in *Proc. of the International Conference on Auditory Display ICAD 09* (2009), pp. 240–251.

[78] B. Gygi, G. R. Kidd, and C. S. Watson, "Spectral-Temporal Factors in the Identification of Environmental Sounds," *J. Acoust. Soc. Am.*, vol. 115, pp. 1252–1265 (2004).

[79] http://www.audiogames.net/

[80] A. Dobrucki, P. Plaskota, P. Pruchnicki, M. Pec, M. Bujacz, and P. Strumiłło, "Measurement System for Personalized Head-Related Transfer Functions and Its Verification by Virtual Source Localization Trials with Visually Impaired and Sighted Individuals," *J. Audio Eng. Soc.*, vol. 58, pp. 724–738 (2010 Sept.).

[81] M. Bujacz, P. Skulimowski, and P. Strumiłło, "Sonification of 3-D Scenes Using Personalized Spatial Audio to Aid Visually Impaired Persons," in *Proc. of the International Conference on Auditory Display ICAD 11* (2011), 8 pages.

[82] B. J. Cratty, *Movement and Spatial Awareness in Blind Children and Youth* (Ch. C. Thomas Publisher, Springfield, IL, 1971).

[83] S. H. Foster, and E. M. Wenzel, "Virtual Acoustic Environments: The Convolvotron," *Demo system presentation at SIGGRAPH'91, 18th ACM Conference on Computer Graphics and Interactive Techniques* (1991).

[84] F. L. Wightman, and D. J. Kistler, "Headphone Simulation of Free-Field Listening I.-II.," *J. Acoust. Soc. Am.*, vol. 85, pp. 858–878 (1989).

[85] Gy. Wersényi, and H. Fouad, "Listening Tests and Evaluation of Simulated Sound Fields Using VibeStudio Designer" in *Proc. of the International Conference on Auditory Display ICAD 10* (2010), pp. 153–158.

[86] Gy. Wersényi, and J. Répás, "The Influence of Acoustic Stimuli on 'Walking Straight' Navigation by Blindfolded Human Subjects," *Acta Technica Jaurinensis*, vol. 5, no. 1, pp. 1–18 (2012).

[87] Gy. Wersényi, and J. Répás, "Comparison of Localization Performance of Blind and Sighted Subjects on a Virtual Audio Display and in Real-life Environments," presented at the *132nd Convention of Audio Engineering Society* (2012 Apr.), convention paper 8611.

[88] E. M. Wenzel, "What Perception Implies about Implementation of Interactive Virtual Acoustic

Environments," presented at the *101st Convention of Audio Engineering Society* (1996 Nov.), convention paper 4353.

[89] G. Burdea, P. Richard, and P. Coiffet, "Multimodal Virtual Reality: Input-Output Devices, System Integration, and Human Factors," *Int. Journal of Human-Computer Interaction*, vol. 8, pp. 5–24 (1996).

[90] C. Hendrix, and W. Barfield, "The Sense of Presence within Auditory Virtual Environments," *Presence*, vol. 5, pp. 290–301 (1996).

[91] A. T. Mullins,"AudioStreamer: Leveraging the Cocktail Party Effect for Efficient Listening," *Master thesis* (MIT, USA, 1996).

APPENDIX

The Beachtron system was developed by Crystal River Engineering and has served as a low-cost solution for spatial audio [38, 83, 88–89]. It is a high-speed, 16-bit digital-signal processing system capable of producing 3-D sound. The Beachtron is a less expensive variant of the Convolvotron. The system is software compatible with all Crystal River Engineering products and supports the virtual audio protocol. The system uses the HRTFs of a "good localizer" from measurements by Wightman and Kistler [84]. HRTFs are measured by placing microphones in a listener's ears near the eardrums. The listener is seated in an anechoic chamber in the center of a spherical array of 74 loudspeakers arranged at equal intervals. A 4-Hz train of 75 acoustic clicks is played from each of them in turn and the response is recorded and averaged. HRTFs are available in a form of 75-point minimum-phase-FIR-filter set in 30° spatial resolution. To render sounds at specific locations the filters are combined into left and right ear listener-specific filters based on the distance model (simulating atmospheric loss) and individual head diameter data (customization). These filters are downloaded into the memory of the card and output is then convolved with the filters. The filters can be changed as often as every 46 ms (latency). Missing directions are calculated by linear interpolation from the four nearest available measured directions. Real-time convolution of the input signal and the HRTFs is made in the time-domain. One Beachtron card can simultaneously spatialize two sound sources. In addition, it supports external auxiliary inputs. As a result up to eight cards can be used simultaneously. The system is shipped with the circumaural Sennheiser HD540 headphone and can be programmed in C language. The Beachtron maintains a real-time model of the listener's head and head-tracking technology can be applied (e.g., Polhemus). The Convolvotron is a high-speed, digital-signal processing system capable of presenting eight binaural sound sources in a virtual environment. The Acoustetron is a complete integrated 3-D audio workstation for use in high-end VR applications. This system is based on a 15-slot industrialized PC containing sound source and spatialization cards. Complex multisource models (including reflection and Doppler) can be achieved with the modular architecture.

THE AUTHORS

György Wersényi

György Wersényi was born in Győr, Hungary, in 1975. He received the M.Sc. degree in electrical engineering from the Technical University of Budapest, Hungary, in 1998 and the Ph.D. degree from the Brandenburgische Technische Universität (BTU) Cottbus, Germany, in 2002.

He has been a member of the Department of Telecommunications at the Széchenyi István University, Hungary, since 2002. He is an associate professor at the field of telecommunications, audio and video broadcast, and electroacoustics. His current research includes spatial hearing, listening tests, virtual audio displays, HRTF measurement, and dummy-head techniques.

Dr. Wersényi is a member of the Audio Engineering Society, the Hungarian Telecommunication Society, and board member of the International Community for Auditory Display (ICAD).