SONIFICATION OF 3D SCENES USING PERSONALIZED SPATIAL AUDIO TO AID VISUALLY IMPAIRED PERSONS

Michał Bujacz, Piotr Skulimowski, Paweł Strumiłło

Institute of Electronics, Technical University of Lodz
211/215 Wólczańska Str. B-9 building
PL 90-924 Łódź Poland
bujaczm@p.lodz.pl, piotr.skulimowski@p.lodz.pl pawel.strumillo@p.lodz.pl

ABSTRACT

The presented research concerns the development of a sonification algorithm for representation of 3D scenes for use in an electronic travel aid (ETA) for visually impaired persons. The proposed sonar-like algorithm utilizes segmented 3D scene images, personalized spatial audio and musical sound patterns. The use of segmented and parametrically described 3D scenes allowed to overcome the large sensory mismatch between visual and auditory perception. Utilization of individually measured head related transfer functions (HRTFs), enabled the application of illusions of virtual sound sources. The selection of sounds used was based on surveys with blind volunteers.

A number of sonification schemes, dubbed sound codes, were proposed, assigning sound parameters to segmented object parameters. The sonification algorithm was tested in virtual reality using software simulation along with studies of virtual sound source localization accuracy. Afterwards, trials in controlled real environments were made using a portable ETA prototype, with participation of both blind and sighted volunteers. Successful trials demonstrated that it is possible to quickly learn and efficiently use the proposed sonification algorithm to aid spatial orientation and obstacle avoidance.

1. INTRODUCTION

The authors' research currently focuses on technology aiding the blind, especially the design of an electronic travel aid (ETA), dubbed Naviton, that would assist the visually impaired in independent travel.

The process of the ETA development begun with questionnaires conducted among 20 blind volunteers, who were asked to indicate the main difficulties in their everyday activities and opinions of currently available technologies. The survey results were very similar to those reported by Hersh [1]. The following barriers were pointed out by the majority of the interviewed visually impaired persons:

1. Being unable to travel safely and independently
2. Limited navigation and orientation
3. Inability to access visual or textual information in the environment (e.g. road signs, bus numbers)

The blind volunteers were also asked to comment on their use and opinions of currently available ETAs. Their most common complaint was that devices they used or attempted to use, utilized overly unnatural or complex sonification schemes for conveying information about the environment, requiring large perceptual effort and training. Another problem reported by the blind community were the poor ergonomics of many electronic aids. Use of some ETAs requires additional devices that need to be attached to the body or clothing. Handling of such devices is another design aspect of ETAs that has not found considerable attention. Despite the dynamic development of electronic and telecommunication technologies, it seems that no single ETA device has been widely accepted and used by the blind community. A simple mechanical device, i.e. the long cane, is still the main and frequently the only aid used by the visually impaired. This encourages research on developing better solutions to support the blind in safe and independent travel.

The paper is organized as follows. First, a short review of a number of most known ETA solutions is provided. The list does not include all the available devices, especially it excludes those with non-auditory output; however, it lists the solutions that most inspired the authors. The next section describes the main assumptions of the developed ETA prototype and the sonification algorithm it utilizes. Because spatial audio is an important aspect of the proposed solution, a section is devoted to the system for fast measurement of personalized head related transfer functions (HRTFs) designed by the authors. Section 5 presents a number of software-based trials conducted with both blind and sighted volunteers during the ETA development. The trials concerned: virtual sound source localization accuracy, development of the sound code, and simulation of the sonification algorithm in a virtual reality environment. Section 6 describes a study of an early portable prototype in controlled real world environments. The last section summarizes the paper and lists future goals of the authors.

2. SONIFICATION SOLUTIONS IN ELECTRONIC TRAVEL AIDS

The history of ETAs reaches as far back as the XIX century, with the construction of the first selenium-cell based photosensitive optophone [2]; however, it was not until the advent of semiconductor electronics that functional and wide-spread devices were developed.

There are numerous ways to categorize ETAs. They can be intended as primary or secondary (complementing a long cane or a guide dog) travel aids. They can be limited to simple obstacle detection, or attempt to present more information to aid orientation and navigation. ETAs can use vibration, sound or electric impulses as output, and sonar, laser, or video inputs. The amount of data processing is also a categorizing factor – ETAs can translate input data to output directly, or attempt to process it, e.g. by detecting peaks, differentiating distance or even recognizing shapes.

The systems the authors wish to review are those focusing on obstacle and environment imaging through auditory means. Although tactile or even electric solutions exist, it is clear that sound can convey much more information. The presented systems are reviewed roughly in order of increasing complexity.

The Kay Sonic Torch [3], developed in the 1960s is the first modern ETA that found a widespread user base. It was widely tested by blind youths as a potential primary aid, a replacement for the long cane. The Kay Sonic Torch was a handheld device that used a narrow-beam ultrasonic sensor (40-80 kHz), and produced a mono auditory output in the hearing range by multiplying the reflected and transmitted signal. The output was a repeating wide-band sound, with pitch corresponding to range and timbre corresponding to the texture of the reflecting surface.

The same sonification principle was later used in the head mounted stereo version of the device named the Sonicguide [4]. The signals from its two ultrasonic receivers were delivered to each ear independently, allowing a natural interaural amplitude difference and delay. The Sonicguide was less successful than it's predecessor, so recently a miniaturized cane-mountable modern version of the sonic torch called the Bat 'K' Sonar-Cane [Kay, 2006] has been released.

The Teletact [5] is a very good example of a modern torch-like sonified ETA. It utilizes a triangulating laser telemeter, which is a much more precise sensor than ultrasonic sonar, and offers a choice of auditory or tactile output. For sonification, the distance information is converted into one of 28 notes of inversely proportional pitch, i.e. the smaller the distance to an obstacle, the higher pitched the sound. A similar solution was proposed by the first author in his M.Sc. thesis [6].

The Sonic Pathfinder [7] is one of the earliest ETAs that utilized input processing algorithms to intelligently limit the presented information. Input from three ultrasonic receivers is analyzed and a stereophonic output informs the user of the nearest obstacle (peak in the sonar ranging input), with a musical note corresponding to the distance. Additionally, the system varies the output depending on the estimated speed of travel, the number and position of the detected peaks and the motion of the obstacles. The Pathfinder can also aid in shoreline following (e.g. moving along a curb or wall), in which case the changing pitch of notes signifies straying from the path.

The vOICe [8] is a computer based sensory substitution system, which sonifies 2D grayscale images. The sonification algorithm uses three pixel parameters: horizontal position, vertical position and brightness to synthesize a complex audio output. The images are coded in horizontal sweeps. The vertical coordinates of pixels are assigned different sound frequencies and the amplitude of the frequency component corresponds to a given pixel's brightness. This coding method creates a short one second audio signal, with numerous frequency components and a time-varying spectrum. Although the sonification algorithm is very difficult to learn, it has been demonstrated that prolonged use and correct training allows a blind user's brain to adapt itself to the new perception method [9]. The use of 2D images as input greatly limits the device as a mobility aid, so the use of stereovision is being considered by the system's designers.

Another sonification solution that should be mentioned is the Spanish prototype ETA called EAV - Espacio Acustico Virtual (Virtual Acoustic Space) [10]. The system uses stereoscopic cameras to create a low resolution (16x16x16) 3D stereo-pixel map of the environment in front of the user. Each occupied stereopixel becomes a virtual sound source filtered with the user's HRTFs. All of the virtual sources produce a sharp sound (Dirac delta function) in unison; however, since they are spatialized, they arrive in the ears at varying times and with different amplitude spectra. Theoretically, clouds of virtual sources create an illusion of whole objects producing sounds.

Generally, ETA systems can be divided into two extremes. At one end we have systems which provide an overabundance of data (KASPA, vOICe, EAV) and require the user to learn to filter out the most useful information in the audio messages. On the other, we have the ETAs that deliver limited information that requires little focus to interpret (Bat Cane, Teletact). There are some devices which attempt to select and present only the most useful information (Sonic Pathfinder) and the ETA in development by the authors is intended to be such a solution.

3. ETA PROTOTYPE SOUND CODE CONCEPT

One of the main decisions made during ETA design is the amount and type of information to be delivered to the blind user. Information can range from a binary alert about the presence of an obstacle in a senor's path, to complex sound patterns representing distance to obstacles, 2D or 3D scene images.

The authors' main goal was to develop a device that would deliver as much useful information as possible, but in a very simple, easily understandable way. To make it possible, it is necessary to use a rich input and various computer algorithms to reduce the amount of data at the output to a necessary minimum.

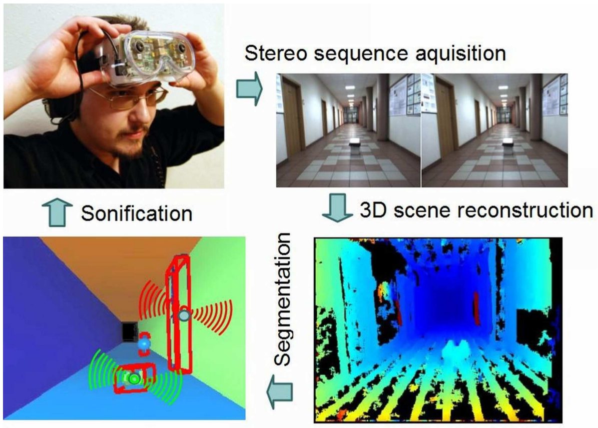


Figure 1: The electronic travel aid concept.

The ETA solution proposed by the authors utilizes stereovision input, and a spatial sound output. Stereovision was chosen as the scene acquisition method because of its inexpensiveness, passiveness and the authors' image processing experience.

The reconstructed 3D scene undergoes segmentation in order to prepare a parametric scene model for sonification. The concept is illustrated on Figure 1. The segmentation algorithm consists of two main steps: firstly, all major planes are detected and subtracted from the scene, and secondly, remaining scene elements are approximated by cuboid objects. Plane detection is an iterative process in which surfaces defined by triangles formed by points in the depth map are merged with similar neighbors. After the threshold of similarity is not met by any triangles, the remaining points are grouped into clouds, and the centers of mass and dimensions of the objects are calculated. Details of the segmentation process can be found in [11].

The default model used for input of the sonification algorithm was limited to four surfaces, described by plane equations, and four objects, described by position, size and rotation parameters.

Simulation trials and surveys with 10 blind testers served to shape the final version of the sonification algorithm. The main concept was to translate the parameters of segmented scene elements into parameters of sounds, in a way that was simple, instinctively understandable and pleasant to the ear. The chosen assignment was to represent the distance to an object with pitch and amplitude of the assigned sound. Sound duration was made proportional to the size of an object. Using spatial audio, the virtual sound sources are localized to originate from the directions of the scene elements. Different categories of elements can be assigned different sound/instrument types. The sounds are played back in order of proximity to the observer.

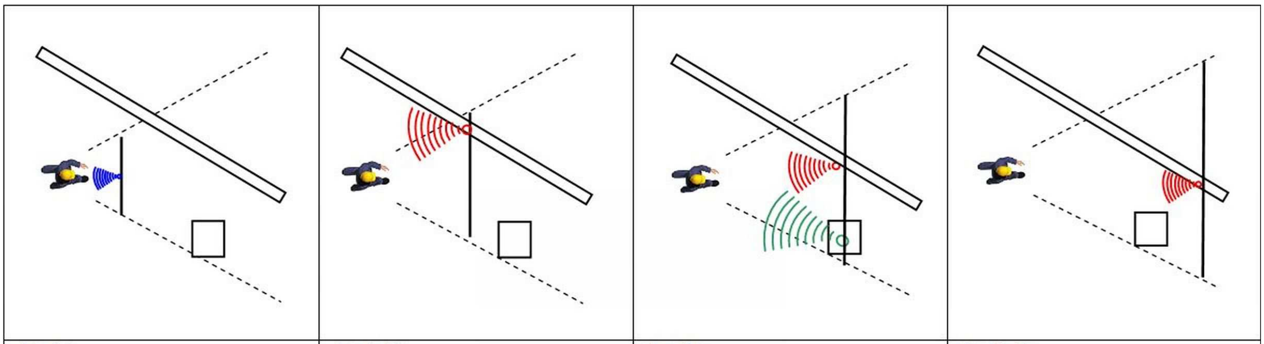
The sonification process, dubbed "depth-scanning", can be illustrated by a virtual scanning plane as illustrated in Figure 2. As the plane moves away from the observer, it releases the playback of the virtual sources assigned to scene elements. When sonifying walls, the sound continues as long as the plane intersects the wall and the virtual source moves along its surface at eye-level. The default scanning period has been set to 2 s, with 0.5 s of silence between consecutive scans.

The sound code utilized audio files pre-generated with a Microsoft General MIDI synthesizer and modulated with 5% noise (14 dB SNR) to ensure a wide spectrum suitable for HRTF spatial filtering. They were stored in banks of 5 s long wave files of full tones from the diatonic scale (octaves 2 through 4). A simple synthesizer program was written to pre-generate sound banks of various instruments.

The range of pitches in a bank can be selected independently for each scene element type. The default setting for obstacles spanned from the full musical tone G2 (98 Hz) to B4 (493 Hz) and for walls from G3 (196 Hz) to G4 (392 Hz). The default instrument for obstacles is a synthesized piano sound (General MIDI P.1), while the bass end of a calliope synthesizer (General MIDI P. 83) was used for walls.

The sound selection was influenced by the concepts of stream segregation and integration as proposed by Bregman [12], as the authors' intent was to make individual virtual sources easily segregated from the entirety of the device's stream. The instruments were chosen to differ spectrally in both pitch and tone, and a minimum delay between the onsets of two sounds was set to 0.2 s. The final selection of sounds was also influenced by preferences of surveyed blind testers during trials presented in Section 5.

To allow efficient modular design of the ETA prototype software, the sonification algorithm was ported to the Mircrosoft DirectShow (DS) multimedia framework technology. Modular architecture is useful when a group of programmers separately develops different aspects of the project, as it was in this case. DirectShow's advantage is its multimedia oriented design, allowing for fast exchange of buffers of various data types, in our case first stereovision images, depth maps and parametric scene descriptions. Details of the implementation were presented by Szypiński [28].



t = 0s	t = 0.5s	t = 1s	t = 1.5s
- start of scan signalled with a brief reference sound - the virtual scanning plane starts moving away from the observer	- the source assigned to the wall starts playing as the scanning plane intersects it	- the wall's sound source moves along with the scanning plane - the obstacle's sound source plays briefly	- the wall's source dies out at the end of the scan - after a short pause the scanning cycle repeats

Figure 2: The depth scanning concept presented on a simple scene with a wall and a single obstacle. The virtual scanning plane (thick vertical line) moves away from the observer releasing sound sources assigned to scene elements.

4. PERSONALIZED SPATIAL AUDIO

A key part of the developed sonification system was the utilization of externalized spatial virtual sources, i.e. providing the ETA users with the illusion of sounds originating from specific locations in space. The effect was achieved by filtering sounds via appropriate head related transfer functions (HRTFs), which were measured individually for each user.

Using HRTF-based spatial audio has a number of advantages - aside from allowing to localize virtual sources in all of 3D space, not just in the median plane, the sounds can also be localized more precisely [13] and are less tiring to a listener during prolonged exposure [14]. Spatial audio has been successfully used to aid the blind in navigation [15] or in orientation training [16].

HRTF sets can be generalized, either mathematically [17] or by measuring them for a dummy, such as KEMAR [18]; however, their magnitude spectra can differ between two persons by as much as 20 dB for various frequencies [13] and there are several advantages to using individual personalized HRTFs [19][20]. The author's previous studies showed that the use of HRTFs measured individually for each listener results in more likely externalization and improved sound localization when compared to non-personalized spatial audio [21][22].

A system for fast HRTF measurement was designed and constructed with the aid of the University of Wroclaw [22]. The system's main goal is efficient measurement of a large set of HRTFs for all directions in space surrounding a listener.

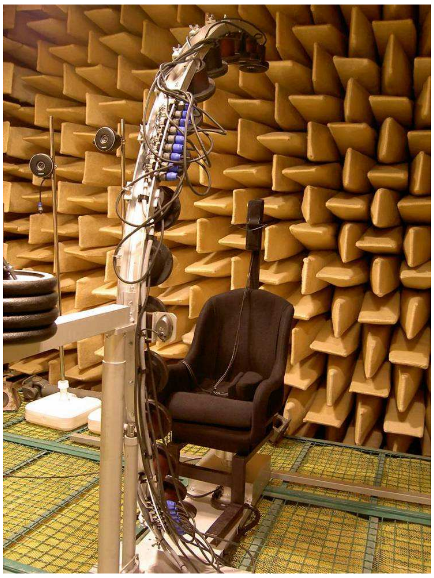


Figure 3: The HRTF measuring equipment in an anechoic chamber.

The measurement equipment consists of an arc of speakers mounted next to a chair on a stepper motor as seen in Figure 3. The device is placed inside a small anechoic chamber. For each listener, the equipment is configured so that the listener's head is exactly in the center of speaker arc and on the axis of rotation of the chair. The speakers play short bursts of wide spectrum sounds, such as noise or chirps. These sounds are recorded by microphones placed inside the entrances to the listener's ear canals. The speakers on the arch play in quick succession, after which the automated chair rotates by a preset angle and the speakers play again in order to sample all directions around the listener. Data is collected in the full azimuth range, $\theta = 0°$ to 360°, and a broad elevation range, $\varphi = -54°$ to 90°. The maximum possible angular resolution in azimuth is 1°, but a default step of 5° was chosen as a balance between precision and time of measurement. The vertical angular resolution is limited by the size and number of the speakers and is set to 9°.

Two 10-15 min measurement runs are performed: first, a reference run without a listener present, but with microphones positioned on thin wires and the equipment set up to the listener's height, and second, the actual measurement run with microphones at the entrances to the listener's ear canals. This procedure allows to normalize the measured impulse responses in real time using the reference responses, thus removing the influence of the equipment characteristics on the gathered data.

5. SIMULATION TRIALS

In order to verify the individually measured HRTFs and the proposed sonification algorithm, a number of trials were performed using various simulation software. The NASA SoundLAB software [23] was used directly for the localization trials of virtual sources, and later as a library component for the virtual environment programs prepared by the authors.

5.1. Source localization

Ten blind and ten sighted volunteers took part in verification trials of their personalized HRTFs. The simple experiment consisted of listening to a virtual sound source played through stereo headphones and pointing to its perceived origin on a grid at arm's length. Various aspects of virtual source localization were checked, such as the type of sounds used, noise level mixed into sounds, source motion and the disability of the listeners. The types of sounds compared were narrowband sounds (synthesized vowels), wideband chirp sounds and white noise. Moving sounds were oscillated 20° horizontally, vertically or diagonally.

Despite the relatively small number of participants, a number of observations were made. The blind volunteers localized the virtual sound sources with larger errors than sighted individuals and elevation errors were much larger than errors in azimuth. Wideband sources were localized with most accuracy. Blind volunteers' average error in azimuth was 8° and in elevation it was 14°, while for sighted blindfolded volunteers the values were 6° and 12° respectively. Wideband sounds allowed for most accurate localization, with average errors 7° and 13° in azimuth and elevation respectively. Moving sources were localized with a slightly higher precision than static ones, which confirms the theory that changes in spectrum over time aid in localization [13]. Localization was also more precise near the center of vision. The plots of the averaged trial results for static sources are shown in Figure 4.

Overall the errors were larger than what is suggested in literature [24] and there were large deviations in the results (for each listener individually and between all listeners). The reason for this might be that the method of assessment – placing a marker on a 2D board at arm's reach while blind or blindfolded was in itself a source of errors. A reference trial with real sources placed near the board (a coin being scratched) resulted in average errors of 6° in elevation and 4° in azimuth for all volunteers. More details about the conducted localization trials can be found in [21] and [22].

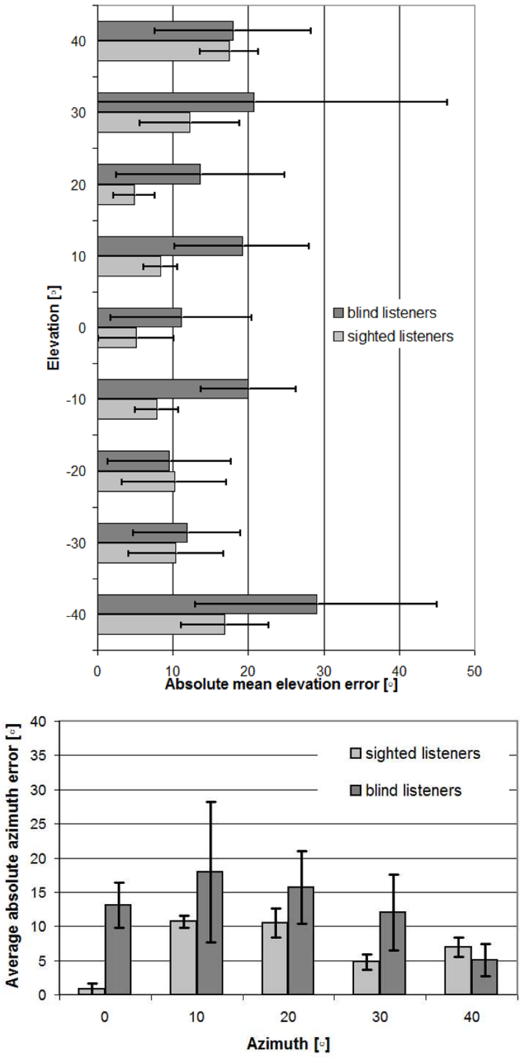


Figure 4: Average errors for localization of static wideband virtual sources.

5.2. Sound code design

The natural sound code assignment and the ”depth-scan” presentation method used in the final prototype were a consequence of a number of earlier studies. Before the possibility of performing tests in real-world environments, most trials took place in virtual ones, with simulation software that allowed control over various sonification parameters.

A code generator program was prepared, that allowed synthesis of various sound patterns basing on the parameters of a number of adjustable scene elements (large cuboid objects). For example, sound amplitude or pitch could be correlated to an object's size or the object's distance from the observer. The program was presented to 10 blind volunteers to gather feedback on the proposed sonification algorithms. The participants were tasked with identifying primary sound characteristics they easily and quickly recognized. The blind listeners also gave their opinions as to what information about obstacle parameters was most vital to them. Basing on those answers the participants were asked to propose their own sonification schemes, which were quickly synthesized using the software and evaluated. Finally, the sonification schemes proposed by the authors were evaluated. The assignment of obstacle to sound parameters most preferred by the visually impaired volunteers is presented in Figure 5.

The method of presenting sounds that was most instinctive to most participants was presenting them in order of proximity. Even participants who themselves suggested schemes containing multiple simultaneously playing sources, such as each source being repeated with a frequency related to proximity, abandoned the concept after listening to the synthesized code they proposed.

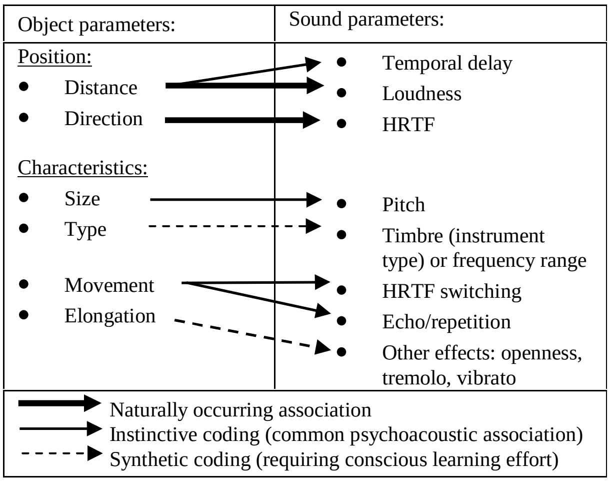


Figure 5: Average errors for localization of static wideband virtual sources.

5.3. Virtual reality trials

A simulation program, which tested the proposed sonification algorithm in a virtual reality environment, was prepared next. The application also allowed virtual source localization trials in conditions where the observer could move his head.

A VirtualResearch V8 head mounted display with an InterSense InterTrax 3DOF tracker were used to allow experimentation with spatial audio and active head movements. The program displayed a virtual 3D environment to the user and allowed to observe it through auditory codes. The user could move around the virtual environment using a keyboard and a mouse or a VR helmet. Sound codes were composed from pre-synthesized wave file collections and filtered through personalized HRTFs to provide a spatial audio illusion.

The sonification method based on a depth-scanning surface that was parallel to the viewing plane and moving cyclically away from the observer. As the scanning plane intersected scene elements, virtual sound sources were created that corresponded to the position and type of object.

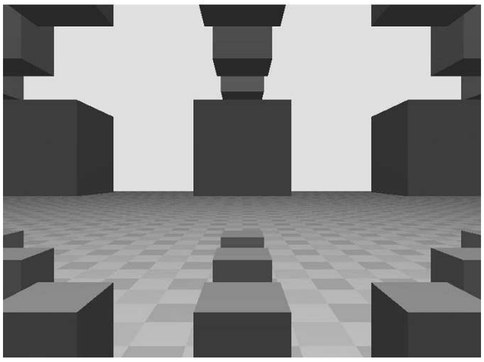


Figure 6: Discrete source positions in VR localization trial.

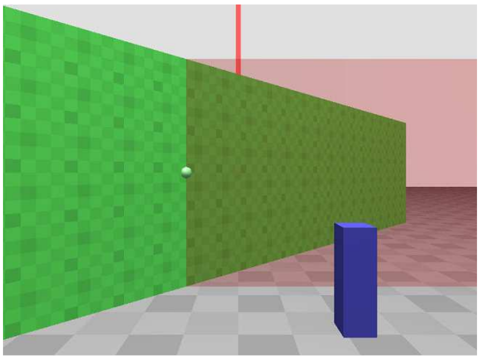


Figure 7: Sample scene from virtual reality trials. The transparent surface represents the scanning plane.

Different types of scene elements were assigned sounds of different musical instruments. The VR trials only featured recognition of two classes of elements - walls and generic obstacles; however, the sound code assumed future expansion of categories to include elements such as moving obstacles, with possible recognition of human silhouettes, and various surface discontinuities, i.e. curbs, drop-offs, stairs and doors. Each of those classes could be assigned a separated bank of sounds.

During the VR trials 10 sighted volunteers, aged 22 to 49, tried to localize virtual sound sources in discrete positions (Figure 6) and describe simple sonified scenes, consisting of a wall and a number of obstacles (Figure 7). All trials were static - i.e. they did not require translation movement from the observers, just rotational movements of the head in the VR helmet. Such limitation was introduced because when virtual environments are navigated by the blind, they have no clear imagination of the implementation of movements by a joystick [25] while for sighted volunteers, previous trials have shown a very high dependence of movement efficiency on previous experience with virtual 3D environments, e.g. from computer games [6].

All subjects were able to very quickly learn the principle of "depth scanning" and describe scenes presented with the auditory code with much accuracy. Errors in scene recognition were infrequent, aside from the very random vertical localization of obstacles, which however was in agreement with literature [18] and the author's previous studies [22]. The proposed sound code was concluded to be effective enough to move on to trials in real scenes with a portable ETA prototype. Another conclusion is that even though personalized HRTFs clearly provide better localization accuracy than non-personalized ones, the difference is not large and the perception of vertical position of virtual sources leaves a lot to wish for.

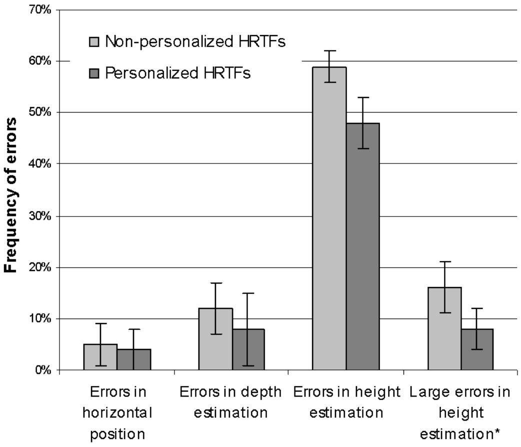


Figure 8: Percentage of errors in single virtual source localization trials (a large height error is an up-down confusion).

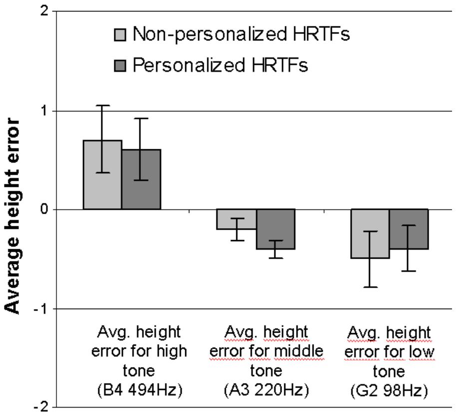


Figure 9. Directionality of height errors. An error value of +1.0 means a source was localized one level higher than it should be (from the positions on Figure 6), -1.0 means one level lower.

Results presented in Figure 9 show a strong psychoacoustic correlation of pitch with perceived height. When in the first stage of the tests closer scene elements were encoded with high pitched sounds, they were nearly always perceived higher than their HRTFs actually placed them and were the main source of the large errors. The phenomena is known as the Pratt effect and originates from early association of high tones with "higher" physical location [27].

The main advantage of the personalized HRTFs is visible in the significantly decreased number of large errors, i.e. up-down confusions. An interesting result is the fact that the use of individualized HRTFs lessened the frequency of errors in all other aspects of the sound code, not only those directly related to sound localization. The cause of those improvements might be better externalization or the fact that the sounds seem more natural when filtered through personal HRTFs [26].

6. REAL WORLD TRIALS

The latest progress of the research was the design and construction of a portable ETA prototype. The purpose of the prototype was to provide real-time conversion of video input into a sonified output. The video source was the commercially available Point Grey Bumblebee2 Firewire stereovision module in a custom head-mount. A mid-range laptop PC performed the necessary data processing and could be carried in a special backpack (however, it was more comfortable to perform most tests on a 10m cable tether). The laptop allowed 3D scene reconstruction and segmentation with approximately 10 fps.

The pilot study was to test obstacle avoidance and orientation in real world scenes. For the safety of the trial participants, the tests were conducted in a controlled environment. Colored cardboard boxes (40x40x80cm) were used as obstacles, and their texturing guaranteed very accurate stereovision reconstruction and smooth operation of the segmentation algorithms.

The trial participants were 5 blind and 5 blindfolded volunteers for whom personalized HRTFs were previously collected. Each group consisted of three men and two women, aged 28-50. The volunteers tested the efficiency of obstacle detection, avoidance and orientation in a safe environment. Trial times, collisions and scene reconstructions served as objective results. Subjective feedback was also collected through surveys during and after the trials. Volunteers were given only 5 minutes of instruction and were allowed short familiarization with the prototype. The tests were designed in such a way as to provide simultaneous training by growing in complexity.

The first test consisted of pointing out the location of a single obstacle, then walking to it and stopping at arm’s reach from it. The average time to localize an obstacle was 4.1 s (with default scanning rate of 2s) and it was done with 96% accuracy in azimuth and 91% accuracy in distance. Blind participants were slightly slower at this task (4.2 s) than blindfolded volunteers (3.9 s).

The second test was similar, but involved two obstacles and the participants were tasked with walking to the further one. The volunteers averaged 5.4 s on this task, with 96% accuracy in both azimuth and distance, and 5.6 s and 5.2 s averages for blind and blindfolded participants respectively.

The final tests involved walking to a playing radio speaker located 5 m away while avoiding obstacles and choosing the least difficult path. All volunteers correctly avoided collisions with obstacles, though minor brush-ups occurred in 7% and 10% of cases, for scenes with two and four obstacles respectively. The most common cause of the minor collisions were attempts to walk between closely spaced (60cm) obstacles. This might have indicated errors in orientation; however, in 89% of cases the volunteers claimed they clearly perceived both obstacles and consciously decided to walk between them.

Figure 10. Prototype ETA used by a blind volunteer in an obstacle avoidance and orientation trial.

An important observation from the prototype trial was that although personalized HRTFs were used for all participants, one bind and two sighted volunteer reported they did not experience a clear externalization illusion. They explained that they were able to interpret the sonified information based on stereo cues alone. Interestingly, the remaining seven of ten participants commented on the immersive spatial sound illusion and how the sounds allowed to form a mental image of the scenes. This demonstrates the subjectiveness of the studies connected to spatial 3D audio.

7. CONCLUSIONS AND FUTURE WORK

An original approach to sonification of the environment for the visually impaired is presented. The proposed system processes stereovision images of 3D scenes to segment out key elements for auditory presentation. The depth scanning sonification concept that employs HRTFs and assignment of unique sound icons to selected scene objects was experimentally verified in two trials. In the first virtual reality trial (conducted with the use of VR helmet) 10 blindfolded participants reported their auditory perceptions of the sonified virtual 3D scenes. These trials, despite revealing large errors in estimating height of the obstacles, demonstrated that the participants were capable of grasping general spatial structure of the sonified environments and accurately estimate scene layouts. In the second set of trial, 5 blind and 5 blindfolded volunteers navigated real world scenes with cardboard obstacles using a wearable prototype. Stereovision scene image sequences were converted in real-time into sound codes according to the introduced depth scanning sonification method. Both groups of trial participants performed with better than 90% accuracy in estimating direction and depth of the obstacles. All trial participants expressed enthusiasm as to the future of the project. The blind and sighted volunteers performed with very similar error rates. The main differences were that blindfolded participants localized obstacles slightly faster in static trials, while the visually impaired travelers walked the obstacle courses much faster, likely due to their own experience and greater confidence.

The authors’ current work focuses on the construction of a small portable prototype and trials in real world outdoor settings are scheduled for the near future. Further trials will also examine the consequences of increased familiarity with the sonification algorithm on the efficiency and safety of travel.

8. ACKNOWLEDGMENT

This work has been supported by the Ministry of Science and Higher Education of Poland research grant no. N N516 370536 in years 2009-2010 and grant no. N R02 008310 in years 2010-2013. The authors would like to thank members of the Polish Union of the Blind for volunteering in the trials.

9. REFERENCES

[1] Hersh, M. & Johnson, M. (Eds.) *Assistive Technology for Visually Impaired and Blind People* Springer, London, 2008

[2] Capp, M. & Picton, P. "The optophone: an electronic blind aid" *Eng. Science and Education J*, 9(3), pp.137-143, 2000

[3] Kay, L. "An ultrasonic sensing probe as a mobility aid for the blind" *Ultrasonics*, 2, p. 53, 1964

[4] Kay, L. "A sonar aid to enhance spatial perception of the blind : Engineering design and evaluation" *Radio and Electronic Engineer*, 44, pp. 605-627, 1974

[5] Farcy, R. & Bellik, Y. "Locomotion assistance for the blind" *Applied Optics, Universal Access and Assistive Technology*, 2002

[6] Bujacz, M. Stereophonic representation of a virtual 3D scenes - a simulated mobility aid for the blind, M.Sc. Thesis, Technical University of Lodz, 2005

[7] Heyes, D. A. "The Sonic Pathfinder: A New Electronic Travel Aid" *Journal of Visual Impairment and Blindness*, 77, pp. 200-202, 1984

[8] Meijer, P. "An Experimental System for Auditory Image Representations" *IEEE Transactions on Biomedical Engineering*, 39, pp. 112-121, 1992

[9] Merabet, L.; Battelli, L.; Obretenova, S.; Maguire, S.; Meijer, P., Pascual-Leone, A. "Functional recruitment of visual cortex for sound encoded object identification in the blind" *Neuroreport*, 20(2), pp. 132-138, 2009

[10] González-Mora, J. L.; Rodríguez-Hernández, A.; Rodríguez-Ramos, L. F.; Díaz-Saco, L. & Sosa, N. "Development of a new space perception system for blind people, based on the creation of a virtual acoustic space" *Eng. Applications of Bio-Inspired Artificial Neural Networks*, Springer Berlin / Heidelberg, pp. 321-330, 1999

[11] Skulimowski, P.; Bujacz, M. & Strumiłło, P. "Detection and Parameter Estimation of Objects in a 3D Scene Image" *Processing & Communications Challenges*, Academy Publishing House EXIT, Warsaw, 1, 308-316, 2009

[12] Bregman, A. *Auditory Scene Analysis: the Perceptual Organization of Sound* The MIT Press, 1990

[13] Wightman, F. L. & Kistler, D. J. "Factors Affecting the Relative Salience of Sound Localization" *Binaural and Spatial Hearing in Real and Virtual Environments* Cues Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey, pp. 1-24, 1997

[14] Bronkhorst, A. W.; Veltman, J. & van Breda, L. "Application of a Three-Dimensional Auditory Display in a Flight Task" *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 38, pp. 23-33, 1996

[15] Loomis, J.; Golledge, R. & Klatzky, R. "Navigation System for the Blind: Auditory Display Modes and Guidance" *Presence: Teleoperators and Virtual Environments*, 7(2), pp. 193-203, 1998

[16] Seki, Y. & Sato, T. "A Training System of Orientation and Mobility for Blind People using Acoustic Virtual Reality" *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 99, 1, 2010

[17] Dobrucki, A. & Plaskota, P. "Computational modelling of head-related transfer function." *Archives of Acoustics*, 32(3), pp. 707-728, 2007

[18] Gardner, M. & Gardner, R. "Problem of localization in the median plane: effect of pinnae cavity occlusion" *The J of the Acoustical Society of America*, 53, pp. 400-408, 1973

[19] Moller, H.; Sorensen, M.; Jensen, C. & Hammershoi, D. "Binaural Technique: Do We Need Individual Recordings?" *J. of Audio Engineering Society*, 44(6), pp. 451-469, 1996

[20] Cheng, C. I. & Wakefield, G. H. "Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space" *J. of Acoustic Society of America*, 49, pp. 231-249, 2001

[21] Pec, M.; Bujacz, M. & Strumiłło, P. "Individual HRTF Measurements for Accurate Obstacle Sonification in an Electronic Travel Aid for the Blind" *Proc. International Conference on Signals and Electronic Systems*, 2008

[22] Dobrucki, A.; Plaskota, P.; Pruchnicki, P.; Pec, M.; Bujacz, M. & Strumillo, P. "Measurement System for Personalized Head-Related Transfer Functions and Its Verification by Virtual Source Localization Trials with Visually Impaired and Sighted Individuals" *Journal of Audio Engineering Society*, 58(9), pp. 724-738, 2010

[23] Miller, J. "SLAB: A Software-Based Real-Time Virtual Acoustic Environment Rendering System" *Proc.of the 2001 Int.Conf. on Auditory Display*, Espoo, Finland, 2001

[24] Wersenyi, G. "Simulation of small head-movements on a virtual audio display using headphone playback and HRTF synthesis" *Proc. the 13th International Conference on Auditory Display*, Montréal, Canada, 2007

[25] Frauenberger, C. & Noisterning, M. "3D audio interfaces for the blind" *Proc. of the 2003 International Conference on Auditory Display*, Boston, MA, USA, 2003

[26] Valentino, G.; Rudolph, P.; Datta, P. & Stuart, M. "BACIS-binaural audio communications and intercom system" *Proceedings of the IEEE National Aerospace and Electronics Conference*, 2000

[27] Talbot, M. & Cowan, B. "Trajectory capture in frontal plane geometry for visually impaired" *Proc. of the 12th Int. Conf. on Auditory Display*, London, UK, 2006

[28] Szczypiński, P.; Pełczyński, P.; Szajerman, D. & Strumiłło, P. "Implementation of Computer Vision Algorithms in DirectShow Technology" *Advances in Soft Computing*, 84, pp. 31-38, 2010