14

Sonification of 3D Scenes in an Electronic Travel Aid for the Blind

Michal Bujacz, Michal Pec, Piotr Skulimowski,
Pawel Strumillo and Andrzej Materka
Institute of Electronics, Technical University of Lodz
Poland

1. Introduction

Sight, hearing and touch are the sensory modalities that play a dominating role in spatial perception in humans, i.e. the ability to recognize the geometrical structure of the surrounding environment, awareness of self-location in surrounding space and determining in terms of depth and directions the location of nearby objects. Information streams from these senses are continuously integrated and processed in the brain, so that a cognitive representation of the 3D environment can be accurately built whether stationary or in movement. Each of the three senses uses different cues for exploring the environment and features a different perception range (Hall, 1966). Touch provides information on the so called near space (termed also haptic space), whereas vision and hearing are capable of yielding percepts representing objects or events in the so called far space.

Spatial orientation in terms of locating scene elements is the key capability allowing humans to interact with the surrounding environment, e.g. reaching objects, avoiding obstacles, wayfinding (Gollage, 1999) and determining own location with respect to the environment.

An important aspect of locating objects in 3D space is the integration of percepts coming from different senses. Understanding distance to objects (depth perception) has been possible by concurrent binocular seeing and touching experience of near space objects (Millar, 1994). For locating and recognition of far space objects, vision and hearing cooperate in order to determine distance, bearings and the type of objects. The field of view of vision is limited to the space in front of the observer whereas hearing is ominidirectional and sound sources can be located even if occluded by other objects.

Correct reproduction of sensory stimuli is important in virtual reality systems in which 3D vision based technologies are predominantly employed for creating immersive artificial environments. Many applications can greatly benefit from building acoustic 3D spaces (e.g. operators of complex control panels, in-field communication of combating soldiers or firemen). If such spaces are appropriately synthesized, perception capacity and immersion in the environment can be considerably enhanced (Castro, 2006). It has been also evidenced that if spatial instead of monophonic sounds are applied, the reaction time to acoustic stimuli becomes shorter and the listener is less prone to fatigue (Moore, 2004). Because of the enriched acoustic experience such devices offer (e.g. spaciousness and interactivity) they are frequently termed auditory display systems. Recently, such systems gain also in importance

in electronic travel aids (ETA) for the blind. The sensory substitution concept is employed in these devices for auditory navigation of the user (Hersh, 2008), (Strumillo, 2006).
The presented studies focus on the problem of 3D space representation by means of auditory cues for the purposes of aiding the mobility of visually impaired persons. The task is simplified by extensive image processing and scene segmentation calculations, allowing sonification to focus on translating geometrical features and spatial location of major 3D scene elements (e.g. potential obstacles or walls). The number of simultaneously generated sound streams can be limited to the perceptive capacity of a human. Seminal trials were conducted in which different space sonification scenarios were tested with the participation of both sighted and visually impaired volunteers.
The chapter is organized as follows. In Section 2 an overview of ETAs employing auditory displays is given, along with the introduction of the system developed at the Technical University of Lodz. Section 3 discusses the processing steps employed in the prototype ETA - scene reconstruction, segmentation and sonification. Section 4 goes into more detail about sonification and the developed sound coding scheme. Section 5 presents the HRTF measurement system constructed for the project, as well as the results of localization tests with virtual sound sources in 3D space. Section 6 moves on to simulations of the ETA prototype and tests of the developed sonification scheme in virtual reality.

2. Review of sonic outputs of electronic travel aids

One of the first reported electronic travel aids for the visually impaired was built by a Polish scientist Kazimierz Noiszewski in 1897 (Capp, 2000). Dubbed "the artificial eye" - the device used photosensitive Selenium cells and a buzzer to convert light to sounds of strength proportional to the registered brightness. Since Noiszewki's pioneering work there have been many attempts undertaken to use hearing as a sensory substitute for the lost vision, with significant leaps made in the 1960s using ultrasonic sensors and in the 1990s using modern computer technology
Depending on the amount of conveyed information, modern ETAs can be divided into two main groups:

- obstacle detectors, that use laser or ultrasound sensors, but offer limited information
- environmental imagers, which convert scene images or 3D data into rich but complicated sound patterns.

The disadvantage of the simple obstacle detectors is that they are less useful for understanding space, while the main shortcoming of more complex environmental imaging systems is the requirement of a large degree of focus of the user and prolonged training. All sonic devices must also take care not to overly burden the sense of hearing (Bregman, 1999). Consequently, no single ETA has been widely accepted by the community of the visually impaired; however, a few experienced relative success and will be discussed in this section.
The most basic obstacle detectors are built on the concept of the white cane, which remains the basic mobility aid for the blind and can be regarded as an extension of the sense of touch. Such devices as UltraCane (Hoyle, 2003) or Teletact (Damaschini et al., 2005) use ultrasound or laser sensors correspondingly and simple auditory or vibration output for further extension of the cane's reach. These devices provide extra head level protection that is not offered by a standard white cane. Each of these systems uses some form of energy emitted into the environment. The reflected signals are analyzed and if an obstacle is present in the nearby space a simple alert sounds.

More complex obstacle detectors that use ultrasonic waves to scan the scene and convert the reflected signal into sounds are the Sonic Pathfinder (Heyes, 1984) and the KASPA (Kay's Advanced Spatial Perception Aid) system (Kay, 1974). The head-mounted Sonic Pathfinder uses three sonar beams and generates a simple sound code, comprising of frequencies proportional to distances to obstacles. The KASPA system offers better resolution by making use of frequency-modulated (FM) signals for echo location. The ultrasonic transducers are mounted on a standard white cane. KASPA communicates object distance by pitch, but also enhances the scene perception as the timbre of the sounds depends on the texture of scanned objects.
The flaw of ultrasonic devices is that the emitted beams diverge with distance and their angular precision of locating obstacles deteriorates. On the other hand, the laser based devices can be interfered by strong ambient light.

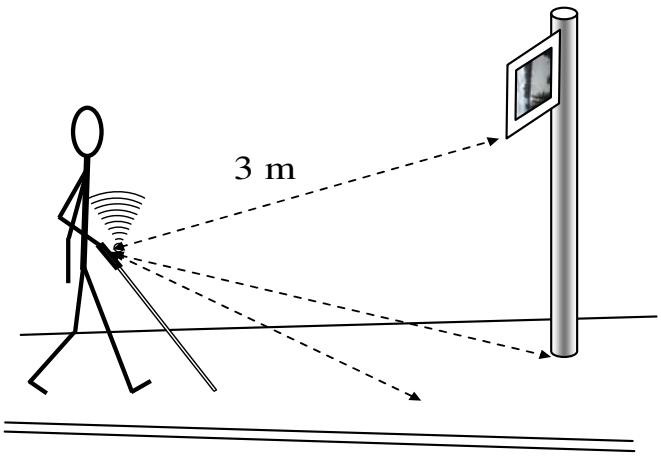


Fig. 1. An example of an electronic obstacle detector in action; laser or ultrasound beams emitted from cane handle are reflected form obstacles and communicated to the user by audio or tactile cues, e.g. providing head level protection that is not offered by a standard white cane

A number of current projects aim at building ETA's that feature functionalities of environment imagers that convert images (e.g. stereoscopic images) of a scene into a scheme of sound codes (auditory icons) or tactile patterns reflecting scene obstacles. Because of the much lower perception capacity of the human auditory (or tactile) system in comparison to visual perception, some form of image analysis or data pre-selection must be applied before the auditory code can be synthesized and presented to the blind user. One such system built in the Nederland, the vOICe, scans the image of a scene in short one-second cycles (Meijer, 1992). Image contents in the columns of pixels that move from left to right are converted into sound patterns. Each column of pixels is used as the spectrum for the synthesized sound, with higher positioned pixels corresponding to higher frequencies and the pixel brightness determining the magnitude of the frequency component. The resulting auditory code is very rich in information, but far from intuitive, thus the users require on average a few months of training before independent navigation trials can be commenced. Note also that the obstacles are not necessarily represented by bright image regions.

Another interesting system was designed and built in the University of La Laguna, Spain (Gonzales-Mora, 1999), under the name of Espacio Acustico Virtual (EAV). This ETA combines stereovision and Head Related Transfer Function (HRTF) technologies. Miniature cameras are mounted into glasses that are attached to headphones of special construction. Stereovision enables 3D reconstruction of a scene that is represented by a collection of equally spaced points. Each such point is characterised by its distance and direction as seen from the ETA. Such a collection of points are sources of acoustic generators which periodically and simultaneously generate short sharp sounds. Locations of points in the environment are determined by a pair of HRTF filters and their distance is reflected by the loudness and phase shift. The shortcoming of the system is the acoustic overload the listener is confronted to. The project has reached the phase of a prototype, but has not been commercialized yet.
The indicated environmental imaging systems, however, do not attempt to pre-process images (or select scene objects) to match the sensory bandwidth mismatch between the human sight and the sense of hearing.

3. 3D scene sonification concept

The ETA system under development at the Technical University of Lodz, Poland, utilizes stereovision and HRTF technologies to provide real-time conversion of video into sound streams. The predominant assumption made in the construction of this system is to limit the amount of auditory information that is presented to a blind user to the most useful minimum. A block diagram of the auditory space representation system is shown in Fig. 2 and the processing steps are described in subsections below. The three main modules of the constructed ETA system perform the following tasks:

- acquisition of stereovision image sequences and their processing for real time 3D scene reconstruction and segmentation,
- coding and synthesis of sound streams associated with selected scene objects,
- filtering of the sound output through individualized HRTFs (Head Related Transfer Functions) for spatial sound illusion via headphones.

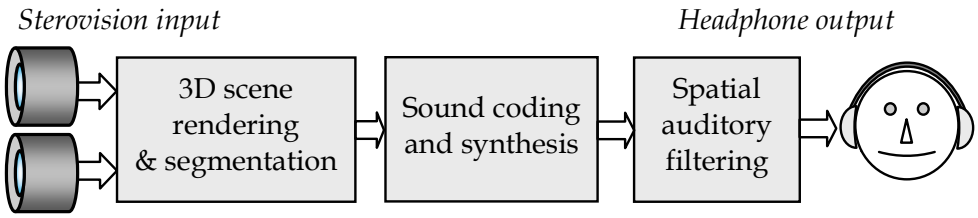


Fig. 2. Schematic of the auditory space representation system

3.1 Scene reconstruction

Scene reconstruction is the determination of 3D coordinates of points belonging to the scene. Usually, a reconstructed scene is presented in the form of a 2.5D image - i.e. a 2D array of depth values.

A number of computer vision solutions were considered for use in the developed ETA, but stereovision was chosen due to its inexpensiveness, passiveness and the experience the authors have had with the technology.
Stereovision reconstruction consists of comparison of images of the same scene as viewed by two cameras from different locations in space. A point in 3D space with coordinates *(X; Y; Z)* will appear in two different image coordinates when viewed by two cameras $p_L(x_0; y_0)$ and $p_R(x; y)$. If the camera axes are parallel and the images have been rectified and corrected for geometric distortions (Brown, 2003), the disparity can be calculated as the difference between the horizontal coordinates of the point in the left and right image:

$$d(p_L; p_R) = x_0 - x \tag{1}$$

Disparity can directly be used to find the depth of the reconstructed point:

$$Z = Bf/d, \tag{2}$$

where B – distance between cameras' optical axes, f – focal length of cameras.
Disparity is calculated for each point for which a corresponding point can be found in the second camera's image, which is determined using correlation of a small window of pixels (default: 5x5). One of the major problems of stereovision is that the maximum of correlation is often difficult to determine, as smooth surfaces will result in large highly correlated regions. For such regions depth estimation is marked inconclusive. Only points for which the correlation maximum was clearly found are considered correctly reconstructed.
Basic stereovision reconstruction was later improved on by camera ego-motion estimation. Estimating the movement vectors of the cameras allowed to better predict positions of correlating pixel groups, improving framerates, as well as leading to sub-pixel depth accuracy thanks to interpolation of the positions of scene elements. A more detailed description of how the processing of the stereovision data is carried out is given in (Skulimowski, 2007 & 2008).

Fig. 3. Stereovision scene reconstruction - the left and right images and the result of the reconstruction; Grey areas represent pixel groups for which proper depth estimation was not possible

3.2 Scene segmentation

After successful 3D reconstruction, the vision module performs segmentation by implementing original algorithms for building a 3D model of the scene (Skulimowski, 2009). The model consists of planes and other objects interpreted as scene obstacles. This type of model is justified by noting that most man-made environments assume such spatial geometry, e.g. streets and buildings' walls, corridors, halls, rooms etc.

An iterative algorithm detects surfaces basing on the depth map. The map is overlaid with a mesh of triangles, with vertices in points of well determined depth. Each triangle's normal vector and plane equation is calculated, and those with similar coefficients to their neighbours are defined as belonging to a common plane. Iteratively, planes are combined if they have similar normal vectors. After a number of iterations, all remaining points which formed groups too small to qualify as planes are grouped with theirs spatial neighbours and marked as obstacles. Clouds of points classified as obstacles are then analyzed to estimate the obstacle's size, shape and orientation. An example of the segmentation procedure and its output is shown in Fig. 4. Note that the plane representing the floor surface in cancelled out and is not meant for sonification.

Fig. 4. The scene segmentation process: original depth map in pseudo-color (left), detection of planes overlaid on the original image (center), and grouping of objects which remain after removal of segmented surfaces (right)

3.3 Scene sonification

The output of the vision module is further converted into a data stream that encodes the size, distance, and angular direction of the obstacles, as well as the equation coefficients of planes present in the scene. This stream of data is used for controlling the sound synthesis module that generates the auditory streams, each representing a unique scene element. Different sonification schemes are assigned to the selected obstacles and planes.

To create a spatialized perceptive illusion of sound streams attached to scene elements (obstacles and planes), the sound presentation module processes the sound streams using listeners' individual HRTFs. The final output of the system are spatially filtered stereophonic sounds arriving in the listener's headphones.

3.3.1 Development of the sound coding scheme

The term "sound code" is used to describe the method of representing the attributes of a real object with parameters of a virtual audio source, so that the sound carries information useful for a visually impaired traveller. A number of trials where volunteers judged various sound codes (Bujacz, 2006) led to the eventual design of a sonar-like sonification method called "depth scanning".

The first sound coding concept did not utilize scene segmentation and was based on direct depth information. The reconstructed scene was automatically swept with a simulated narrow-beam sensor (Bujacz, 2006) and the sound output was given in the form of MIDI synthesized musical tones corresponding to the distance to the nearest obstacle covered by the beam. A screenshot of the program and the horizontal scanning concept is shown in Fig. 5. The simulated ETA was tested with participation of 10 sighted volunteers. Results showed

that the sound coding method was very inefficient; however, it allowed accurate scene recognition after just 3-4 hours of practice. Participants were able to draw the shape of a room observed through the code after 1-3 minutes of observation, as well as slowly navigate simple labyrinths.

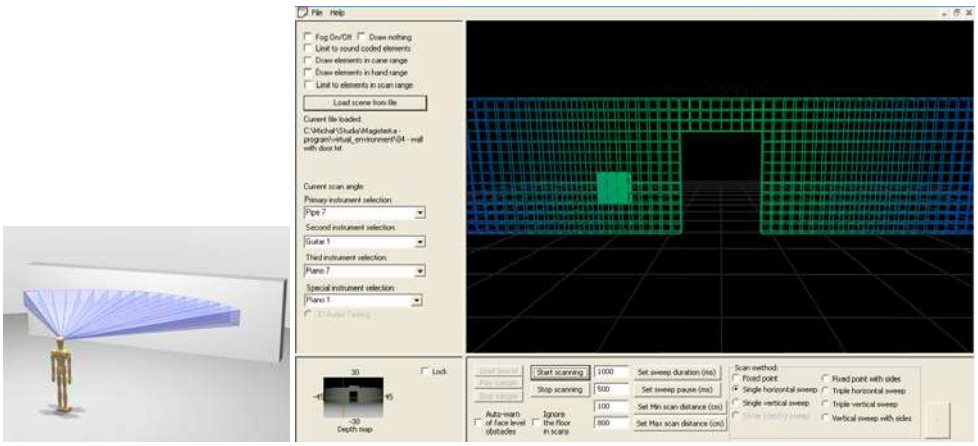


Fig. 5. "Musical range finder" scene sonification concept; the scene is periodically scanned horizontally producing sounds of tone inversely proportional to distance (left); Simulation software (right), note the graphics window highlights the currently sonified scene section

After successful trials of the scene segmentation algorithm, a more efficient sound code was developed to utilize the knowledge of a parameterized scene description (Strumillo, 2006). In addition to MIDI sounds, the simulation programs used formant-based sound synthesis in an attempt to recreate vowel sounds to add an extra dimension to the sound code (Pełczyński, 2006). The formant-based synthesis was also used in further described studies on sound localization.

A wide selection of sound codes was prepared, so that feedback from visually impaired testers could be obtained. The test participants were tasked with identifying the primary sound characteristics they easily and quickly recognized. The majority of testers decided that the ability to quickly interpret information about nearest obstacles and the scene layout was of most importance. Pure musical tones were deemed more pleasant to continuous frequency changes and caused less dissonance during simultaneous playback of multiple sources. Synthesized musical instruments were preferred over artificial sounds, and timbre manipulation through formant filtering was judged too difficult to easily interpret. Methods of informing of multiple scene elements were also tested. Simultaneous playback of all sounds with periods proportional to the distance was deemed overly noisy and difficult to interpret. However, the very instinctive solution of increasing the period of sounds with proximity to an object was noted as having potential for special warnings, such as alerting about very near face-level obstacles. Scanning the scene for obstacles horizontally, vertically or distance-wise was also considered, and the last method proved most useful for travel.

Bregman's theories about the functioning of auditory perception (Bregman, 1990), especially the concept of auditory streams, were of big importance during the sound code design. A sound stream is usually a single sound source in a specific location; however, multiple sounds played in unison or close succession integrate into a single stream. This can be a

desired effect, e.g. in music; however, the idea behind the sound code was to achieve an opposite effect. Each scene element was to be perceived as a separate auditory event. For this reason the sound streams were separated temporally (default 0.2 s) and spectrally by pitch and tone.
The final version of the sound code was prepared after the surveys with potential blind users and aimed at being most clear and instinctive to interpret. Distant objects were encoded with more quiet sounds, which were also played back later during the scanning cycle, i.e. object distance was coded with sound amplitude and time delay from the start of a scanning cycle. Larger objects were assigned longer and lower sounds thus encoding object size with tone and duration. HRTF filtering was used to give the sound sources the illusion of originating from scene elements and it will be discussed in more detail in the next section.
In the "depth scanning" sound code, the sources are presented in order of their proximity to the observer. A good way to illustrate this coding concept is to picture a virtual scanning plane that moves through the scene and releases sound sources as it intersects various scene elements. The scene sonification method is illustrated in Fig. 6.

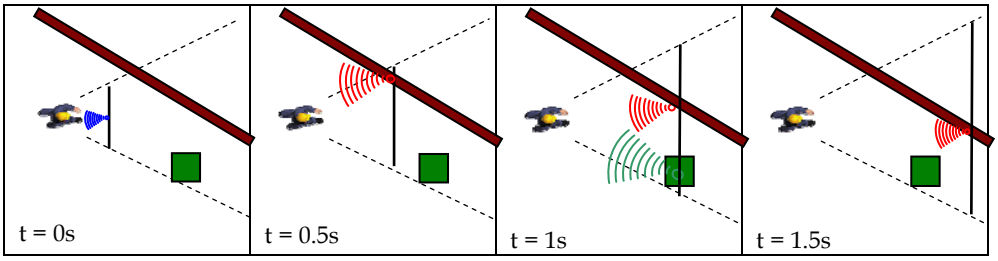


Fig. 6. One cycle of the depth scanning sound code concept. A virtual scanning plane (solid vertical line) moves away from the observer through a scene with one wall (red) and one obstacle (green). The detection range is marked with dashed lines and the sound sources which are released as the scanning surface intersects scene elements are shown with icons whose size is proportional to their loudness. After a cycle completes and a short period of silence (default 0.5s) the scanning starts again.

The number of sound sources presented during a scanning cycle can be limited and is set to 4 nearest elements as default. The scanning can be relatively accurate to a range of 15 m; however, a range of only 5 m was chosen as only the nearest scene elements were deemed important for safe travel. The listener can adjust the speed of the scanning. The default scanning period is 2 s. A number of listeners sped up the scanning to 1.5 s periods; however, most preferred to slow it down to 2.5 - 3 s during training.
Different types of scene elements were assigned sounds of different musical instruments. The prototype only features recognition of two classes of elements - walls and generic obstacles; however, the sound code assumes 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.
The sound code utilizes audio files pre-generated with a Microsoft General MIDI synthesizer and modulated with 5% noise (14 dB SNR). They are stored in collections of 5s long wave files of full tones from the diatonic scale (octaves 2 to 4) referred to as banks. The range of pitches in a bank can be selected independently for each scene element type. The default setting for obstacles spans from 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 Program 1), while the bass end of a calliope synthesizer (General MIDI Program 83) was used for walls. The instruments were chosen during preliminary surveys with blind testers; however, it was the wish of most trial participants to be able to assign their own choice of sounds if possible (Bujacz, 2005).

4. Measurement and verification of HRTFs

In parallel to the ETA project, research on head related transfer functions (HRTFs) was carried out. The HRTFs were measured in an anechoic chamber using special equipment designed for efficiency of data collection (Fig. 7). Data was collected in the full azimuth range (θ = 0° to 360°) with a 5° step and a broad elevation range (φ=-45° to 90°) with a 9° step. Twelve sighted and eight visually impaired volunteers took part in the HRTF measurements. The measurement and processing procedure performed in an anechoic chamber enables recording of the head related impulse responses (HRIRs) to the sounds produced by all speakers, while the listener sits in a revolving chair with the microphones placed at the entrances to his ear canals. The measured HRIRs were converted into a format supported by the NASA SoundLAB environment (Miller, 2001) used for later trails directly or as libraries for custom written software (Pec, 2008).

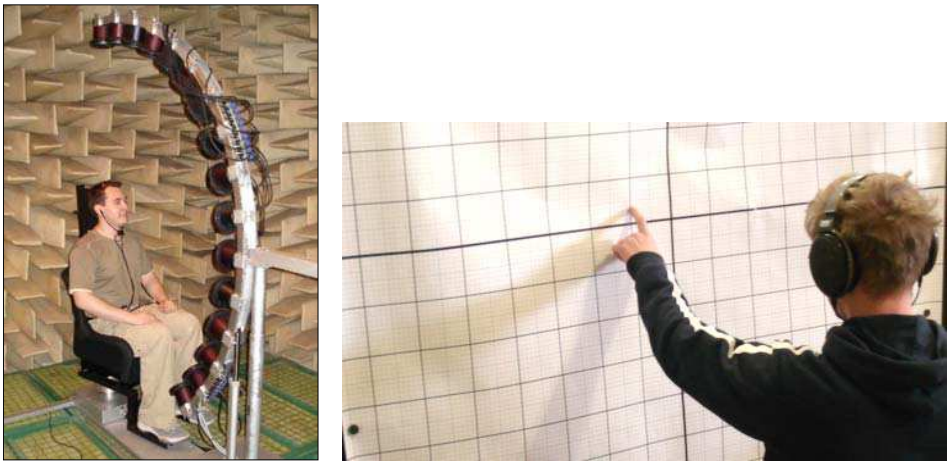


Fig. 7. The HRTF measurement setup in an anechoic chamber (left) and sound localization trials (right)

Most volunteers who took part in the measurements also participated in externalization and localization trials, which served to verify the correctness of data collection and the usefulness of personalized HRTFs. In the first trials, out of 15 volunteers, 12 achieved full sound externalization and 2 gained limited externalization after short training. Different static and moving virtual sources located in the frontal space hemisphere (to avoid front-back confusions) were presented to the volunteers using their personalized and generic HRTFs. Volunteers were to point to the location of perceived sounds on a grid in front of them (Fig. 7). The results of the localization trials with different virtual sound sources for both sighted and blind volunteers are shown in Figs. 8 - 11.

The blind volunteers localized the spatialized sound sources with larger errors (12.5°deg) than sighted individuals (8° on average). The better results for sighted participants are likely due to better trained localization skills thanks to visual feedback training throughout their entire lives (Zahorik 2001). Although the results are not presented in the charts, congenitally blind volunteers made more errors than those who lost sight at a later age. The visually impaired volunteers stressed the need for early-age rehabilitation and training to improve sound localization skills. Despite errors being larger than expected from similar studies (Wersenyi, 2007), the concept of using HRTFs for sonifying the obstacles was proven to be worth considering. Wideband sounds are mandatory for accurate localization, and moving sources are localized with a slightly higher precision than static ones.

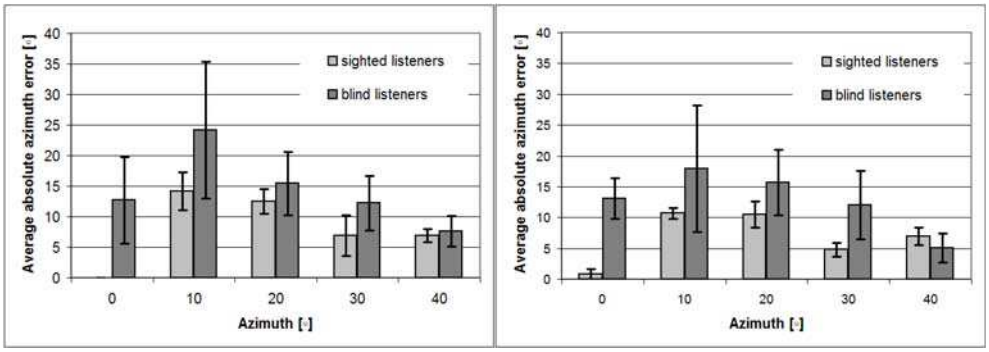


Fig. 8. Comparison of azimuth localization errors for sighted and blind listeners: (left) narrow-band formant synthesized vowel "a" sound, (right) wideband chirp; error bars represent standard deviation among all test participants

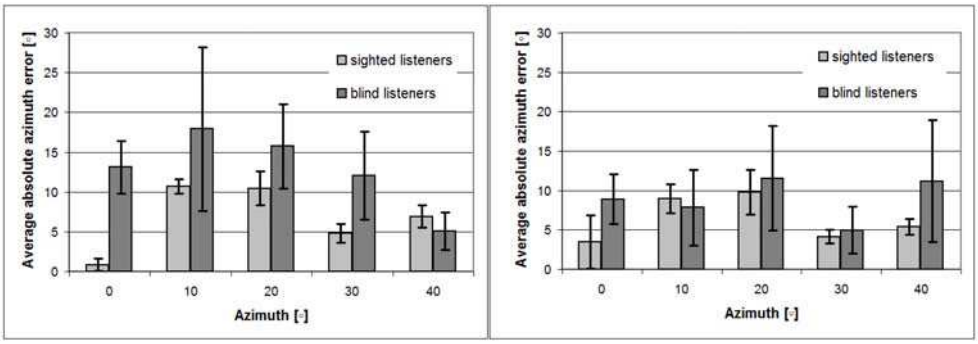


Fig. 9. Comparison of azimuth localization errors for sighted and blind listeners: (left) static wideband sounds (right) oscillating wideband sounds; error bars represent standard deviation among all test participants

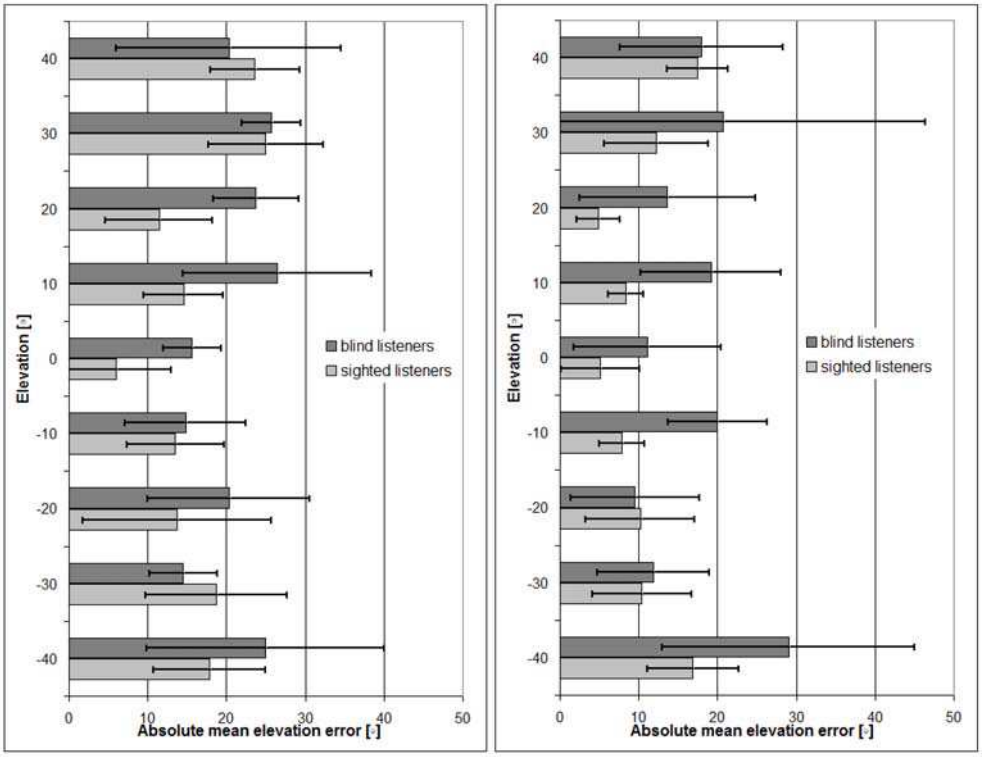


Fig. 10. Comparison of elevation localization errors for sighted and blind listeners: (left) narrow-band formant synthesized vowel "a" sound, (right) wideband chirp; error bars represent standard deviation among all test participants

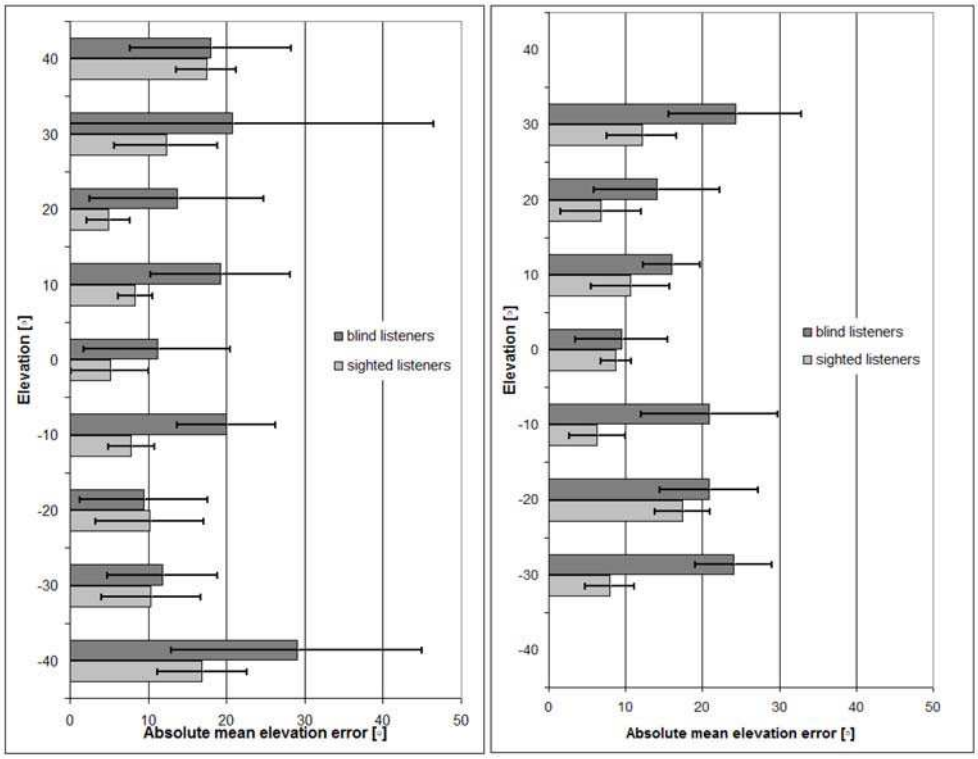


Fig. 11. Comparison of elevation localization errors for sighted and blind listeners: (left) static wideband sounds (right) oscillating wideband sounds; error bars represent standard deviation among all test participants

6. Virtual reality trials

After static trials with virtual sources, we tested the sonification concept in conditions allowing for head movements relative to the sounds, as active perception plays an important role in sound localization (Kato, 2003). A VirtualResearch V6 head mounted display with an InterSense InterTrax 3DOF tracker were used for the trials. Because the tracker only detected rotational movements and previous tests showed that performance in trials requiring translational movements in a VR environment were too dependant on previous experience with 3D environments, such as games (Bujacz, 2005), the VR tests were designed in a way to avoid the necessity of any translational movements.

During the VR trials 10 volunteers, aged 22 to 49, tried to describe simple sonified scenes, such as those presented in Fig. 12 or localize virtual sound sources in discrete positions. Obstacles could be located in three positions along each axis: horizontally – left/center/right, vertically – low/eye-level/high, and depth-wise – near/middle/far. Walls could be positioned in five different ways: parallel to the camera axis on the left or right, at a 45°angle on the left or right, and perpendicularly to the camera axis in front of the observer. Errors of a single position in any direction are referred to as "small", while of two or more positions as "large".

The participants of the trials were all sighted and their personalized HRTF characteristics were collected in the study discussed in Section 5. All participants were trained with the use of the sound code for 15-30 mins by observing scenes visually after hearing them sonified. During the actual trials the first batch of 10 scenes was also considered training and not taken into account for the calculation of final results. The participants continued to learn throughout the trials, as after providing an answer based on the sound code, they were allowed to verify it visually.

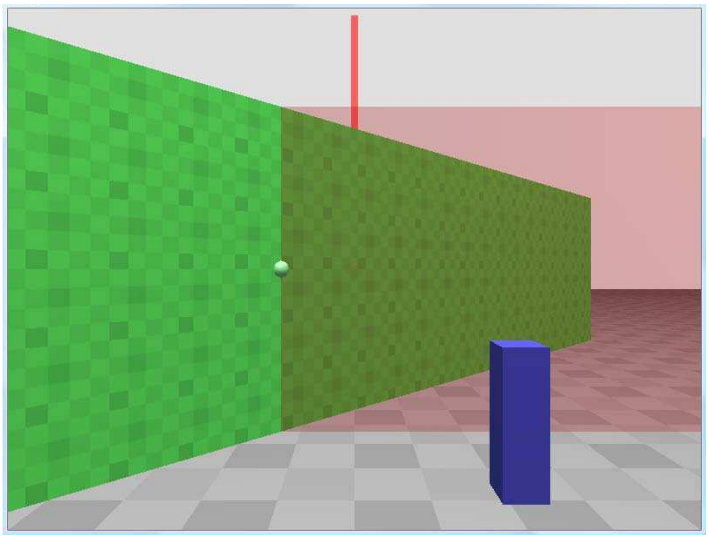


Fig. 12. A sample scene used in the VR sound localization and scene recognition trials; during trials only the floor and zero azimuth marker on the horizon were visible, while participants identified the position of invisible scene elements basing on the sounds heard

The VR trials consisted of three stages, each containing 80 randomized scenes – 40 filtered with personalized HRTFs and 40 with HRTFs of a default individual. In the first stage, the volunteers were tasked with localizing a single object which could appear in one of 27 positions (3x3x3). The sound code was set to fast 1 s cycles and the participants, even though not timed, were asked to make their guesses quickly and instinctively. The second stage consisted of describing simple scenes with a single small obstacle and a wall. This time the sound code was slowed down to 2.5 s cycles. The final third stage consisted of describing scenes with a single wall and two obstacles of different size.
The trial results averaged for the 10 participants, along with standard deviation markings, are shown in Figs. 13 to 16. Personalized HRTFs give a clear advantage in sound localization, especially in terms of large errors; however, small errors are still frequent.
Personalized HRTFs clearly provide better localization accuracy than non-personalized ones; however, errors are still very frequent, especially in locating the vertical position of sources. The main advantage of the personalized HRTFs is visible in the significantly decreased number of large errors, i.e. up-down confusions. A surprising result is the lessened 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.

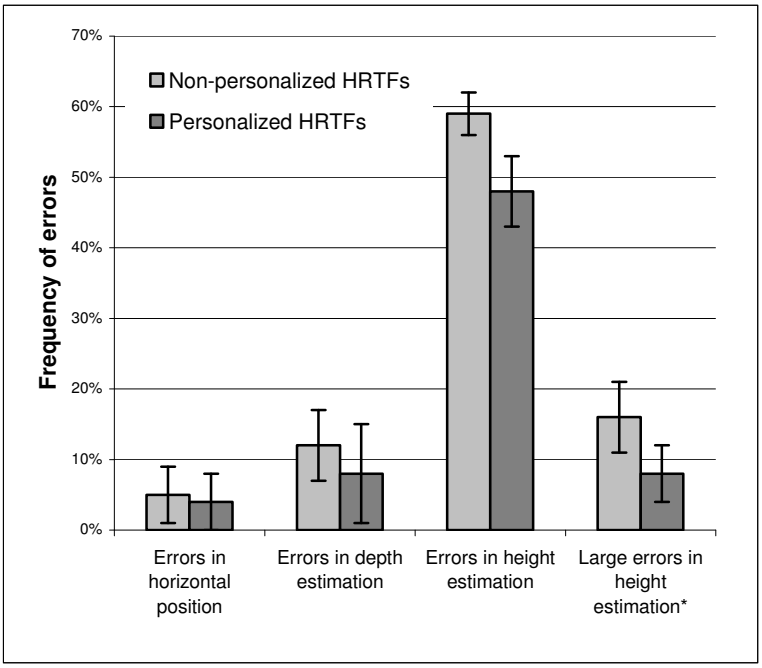


Fig. 13. Percentage of errors in single source localization trials; *) large errors mean a mistake of two discrete positions, which only occurred in cases of up-down confusions

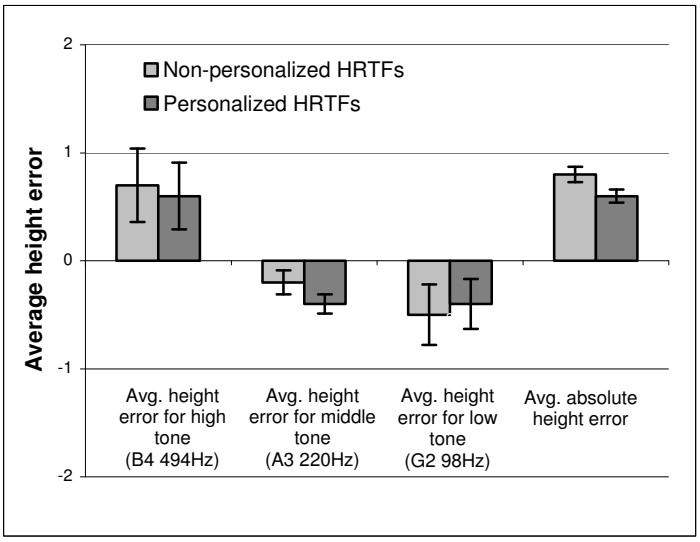


Fig. 14. Directionality of height errors. An error value of +1.0 means a source was localized one level higher than it should be, -1.0 one level lower. Errors were in the range of -2.0 to +2.0, though up-down confusions were not frequent

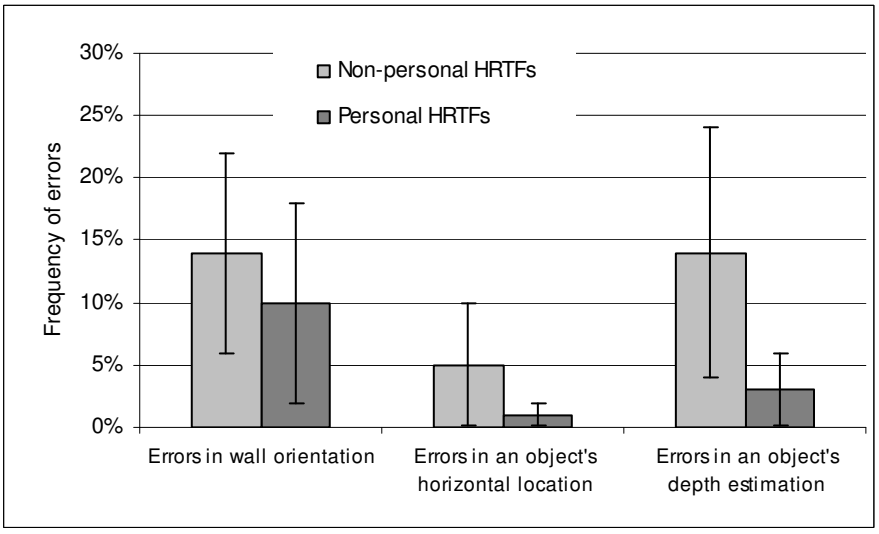


Fig. 15. Percentage of errors during the localization of one wall and one obstacle

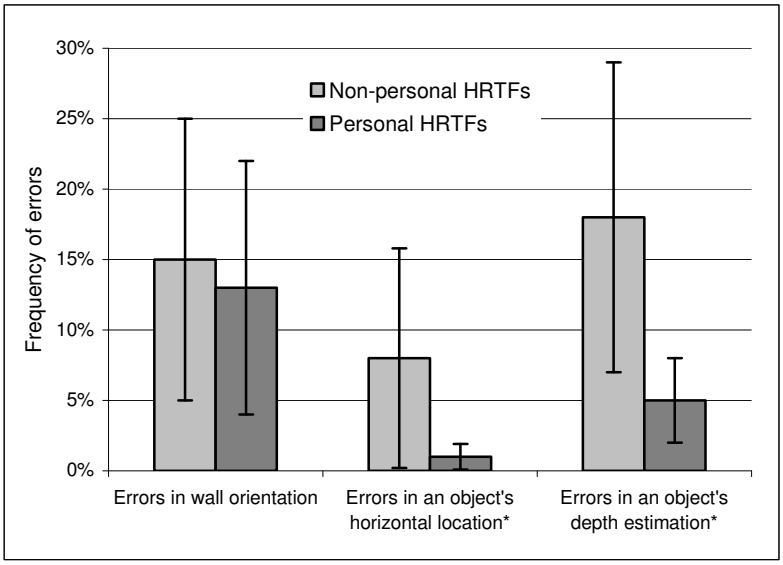


Fig. 16. Percentage of errors during the localization of one wall and two obstacles; the results are averaged for the two obstacles, as errors in decoding the relative position (both horizontal and depth) of two objects never occurred and identification of the larger and smaller object was always correct

Errors in scene recognition were unfortunately frequent. Especially the vertical localization of obstacles leaves a lot to wish for; however, the proposed sound code is effective enough to move on to trials in real scenes once the hardware of the ETA prototype is fully functional.

Results presented in Fig. 14 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.

7. Conclusions and future work

In this chapter we have outlined main results of the research devoted to sensory substitution systems in which spatial location and geometric features of 3D scene objects are converted into spatialized sound icons in order to offer an auditory scene representation system for the visually impaired.

The conclusions from the presented studies are that the use of personalized HRTFs improves externalization and localization of virtual sources, although the improvement is small compared to non-personal HRTFs. The proposed sound encoding concept is instinctive and easy to learn, as well as efficient at warning of nearby obstacles and orientation in simple scenes. An original scene sonification concept was proposed in which scene obstacles important for user safety are segmented out from the scene images, assigned unique sound icons and selected for auditory display. The cyclic depth scanning method and sequential sonifying of obstacles concur with Bregmans' theory of sound streams, i.e. a

low number of selected sound streams are presented only so that the user can easily track them while in movement.
Further research is needed to judge the usefulness of the prototype when users need to focus on the actual task of walking and navigating in real environments. Real-world trials with a portable prototype and visually impaired participants are in preparation.
Results of the presented work can be of use in virtual reality systems in which immersion in virtual world can be further improved by supporting 3D imaging of objects with 3D auditory sensation of the surrounding acoustic scenes.

8. Acknowledgements

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 third author is a scholarship holder of the project entitled "Innovative education [...]" supported by the European Social Fund.

9. References

Benjamin, J., Malvern, J., (1973), "The new C-5 laser cane for the blind." In: *Proc. Carnahan Conf. on Electronic Prosthetics*, Univ. of Kentucky.

Bourbakis, N. (2008). Sensing Surrounding 3-D Space for Navigation of the Blind, *IEEE Engineering in Medicine and Biology Magazine*, Jan/Febr. 2008, 49–55

Bregman S. (1990). *Auditory Scene Analysis: the Perceptual Organization of Sound,* A Bradford Book, The MIT Press, Cambridge, Massachusetts

Brown, M.Z., Burschka, D. & Hager, G.D. (2003). "Advances in computational stereo", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 8, 993–1008

Bujacz, M. & Strumillo, P. (2006): Stereophonic Representation of Virtual 3D Scenes – a Simulated Mobilty Aid For the Blind, *XI Symposium AES: New Trends In Audio And Video*, 157–162

Capp, M., Picton, P., (2000) The optophone: an electronic blind aid, *Engineering Science and Education Journal*, June 2000, 137-143

Castro-Toledo, D.; Magal, T.; Morillas, S. & Peris-Fajarnés, G. (2006). 3D Environment Representation through Acoustic Images. Auditory Learning in Multimedia systems, *Current Developments in Technology-Assisted Education*, 735–740

Damaschini, R.; Legras, R.; Leroux, R. & Farcy, R. (2005). Electronic Travel Aid for the Blind people, in *Assistive Technology: from Virtuality to Reality*, Pruski, A. & Knops, H. (Eds.), 251–260

Dobrucki, A., Plaskota, P., Pruchnicki, P., Pec, M., Bujacz, M., Strumillo, P., (2010). Measurement System of Personalized Head Related Transfer Functions and Its Verification by Virtual Source Localization Trials with Visually Impaired and Sighted Individuals, Journal of Audio Engineering Society, vol. 58, no. 9, pp. 724–738.

Gollage, R. G. (Ed.) (1999). *Wayfinding behaviour: cognitive mapping and other spatial processes*, John Hopkins University Press, Baltimore , USA

Gonzalez-Mora, J. L., Rodriguez-Hernandez, A., Rodriguez-Ramos, L. F., Diaz-Saco, L., Sosa, N., (1999). *Engineering Applications of Bio-Inspired Artificial Neural Networks.*

Springer Berlin/Heidelberg, Ch. Development of a new space perception system for blind people, based on the creation of a virtual acoustic space, 321-330.

Hall, E. T. (1966). *The Hidden Dimension*, Doubleday, Garden City, N.Y.

Hersh, M. A. & Johnson, M. A. (Eds.) (2008). *Assistive Technology for Visually Impaired and Blind People*, Springer–Verlag, London Limited

Heyes, D. A., (1984). The sonic pathfinder: A new electronic travel aid. *Journal of Visual Impairment and Blindness* 77, 200-202.

Hoyle, B. S. (2003). The Batcane – mobility aid for the vision impaired and the blind, *IEE Symposium on Assistive Technology*, 18–22

Kato, M., Uematsu, H., Kashino, M., Hirahara, T., (2003) "The effect of head motion on the accuracy of sound localization", *Acoustical Science and Technology*, Vol. 24, No.5, 315-317.

Kay, L., (1964). An ultrasonic sensing probe as a mobility aid for the blind. *Ultrasonics* April-June.

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

Meijer, P., (1992). An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering* 39, 112-121.

Moore, B. C. J. (2004). *An introduction to the psychology of hearing*, Elsevier Academic Press, London, UK

Millar, S. (1994), *Understanding & representing space*, Clarendon Press, Oxford.

Miller, J., (2001), SLAB: A software-based real-time virtual acoustic environment rendering system. In: *Proceedings of the 2001 International Conference on Auditory Display*, Espoo, Finland.

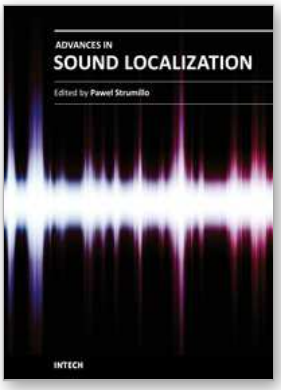
Pelczynski, P., Strumillo, P., Bujacz, M., Formant-based speech synthesis in auditory presentation of 3D scene elements to the blind, ACOUSTICS High Tatras 2006 - *33rd International Acoustical Conference - EAA Symposium*, Štrbské Pleso, Slovakia, October 4th - 6th, 2006, 346-349.

Skulimowski, P., Bujacz, M., Strumillo, P., (2009). *Image Processing & Communications Challenges.* Academy Publishing House EXIT, Warsaw, Ch. Detection and Parameter Estimation of Objects in a 3D Scene. 308-316

Skulimowski, P. & Strumillo, P. (2007). Obstacle localization in 3D scenes from stereoscopic sequences. *Proc. of the 15th European Signal Processing Conference (EUSIPCO 2007)*, September 3-7, Poznań, Poland, 2095–2099

Skulimowski, P. & Strumillo, P. (2008). Refinement of depth from stereo camera ego-motion parameters, *Electronics Letters*, vol. 44, no. 12, 729–730

Strumillo, P.; Pelczynski, P.; Bujacz, M. & Pec, M. (2006). Space perception by means of acoustic images: an electronic travel aid for the blind, *ACOUSTICS High Tatras 06 - 33rd International Acoustical Conference – EAA Symposium*, Štrbské Pleso, Slovakia, October 4th – 6th, 2006, 296–299



Advances in Sound Localization

Edited by Dr. Pawel Strumillo

ISBN 978-953-307-224-1
Hard cover, 590 pages
Publisher InTech
Published online 11, April, 2011
Published in print edition April, 2011

Sound source localization is an important research field that has attracted researchers' efforts from many technical and biomedical sciences. Sound source localization (SSL) is defined as the determination of the direction from a receiver, but also includes the distance from it. Because of the wave nature of sound propagation, phenomena such as refraction, diffraction, diffusion, reflection, reverberation and interference occur. The wide spectrum of sound frequencies that range from infrasounds through acoustic sounds to ultrasounds, also introduces difficulties, as different spectrum components have different penetration properties through the medium. Consequently, SSL is a complex computation problem and development of robust sound localization techniques calls for different approaches, including multisensor schemes, null-steering beamforming and time-difference arrival techniques. The book offers a rich source of valuable material on advances on SSL techniques and their applications that should appeal to researches representing diverse engineering and scientific disciplines.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Michal Bujacz, Michal Pec, Piotr Skulimowski, Pawel Strumillo and Andrzej Materka (2011). Sonification of 3D Scenes in an Electronic Travel Aid for the Blind, Advances in Sound Localization, Dr. Pawel Strumillo (Ed.), ISBN: 978-953-307-224-1, InTech, Available from: http://www.intechopen.com/books/advances-in-sound-localization/sonification-of-3d-scenes-in-an-electronic-travel-aid-for-the-blind

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821