PAN ARCHIVES OF ACOUSTICS
Vol. 40, No. 4, pp. 561–567 (2015)

DOI: 10.1515/aoa-2015-0055

Virtual Sound Localization by Blind People

Larisa DUNAI, Ismael LENGUA, Guillermo PERIS-FAJARNÉS, Fernando BRUSOLA

Universitat Politècnica de València
Camino de Vera s/n, Valencia 46022, Spain; e-mail: {ladu, islenlen, gperis, fbrusola}@upv.es

(*received March 13, 2015; accepted October 9, 2015*)

The paper demonstrates that blind people localize sounds more accurately than sighted people by using monaural and/or binaural cues.

In the experiment, blind people participated in two tests; the first one took place in the laboratory and the second one in the real environment under different noise conditions. A simple click sound was employed and processed with non-individual head related transfer functions. The sounds were delivered by a system with a maximum azimuth of 32° to the left side and 32° to the right side of the participant's head at a distance ranging from 0.3 m up to 5 m.

The present paper describes the experimental methods and results of virtual sound localization by blind people through the use of a simple electronic travel aid based on an infrared laser pulse and the time of flight distance measurement principle. The lack of vision is often compensated by other perceptual abilities, such as the tactile or hearing ability.

The results show that blind people easily perceive and localize binaural sounds and assimilate them with sounds from the environment.

Keywords: virtual sounds, localization, distance, azimuth, blind people.
PACS no. 43.66.Qp, 43.66.Yw

1. Introduction

The sound source localization ability is one of the most important factors for human survival. This ability helps humans perceive and detect danger, localize any object from the environment and determine the differences among climatic conditions: wind, snow, rain etc.. Sound source localization is the "law or rule by which the location of an auditory event is related to a specific attribute or attributes of a sound event" (Blauert, 1997). Sound source localization is influenced by acoustical cues such as the interaural time difference (ITD), interaural level difference (ILD), torso and pinnae (Brungart, Rabinowitz, 1999; Fitzpatrick *et al.*, 2000). Torso and pinnae are two main factors that influence the ITD. Depending on the human head position, the sound waves reach the ear at different times, which influences the precision of the direction of sound source perception and localization. Mendelson *et al.* demonstrated that sound influences visual epistemic behaviour in humans (Mendelson *et al.*, 1976). However, simultaneous perception often causes misrepresentation of the information contained in the audio and visual stimuli (Kunka, Kostek, 2012).

This experiment analyzes virtual sound source localization through headphones on blind people by using artificial vision, which transforms the surroundings into acoustic sounds. As previously mentioned, a lack of vision is often compensated by hearing and/or another perceptual ability (JASA, 8 May 2012; Wonder and Charless, 2005). Over 285 million people are blind or partially sighted in the world, and 39 million people are completely blind. They represent 0.7% of the global population, and 90% of the blind people live in developing countries (WBU, 2012). They learn to make use of the sounds, tactile feelings, temperature etc to help them in their habitual life. After the Second World War, with sensor invention, many technologies have been introduced to the blind community to satisfy the basic desire of mobility and communication. More than 45 electronic travel aids (ETA) for blind people have been developed during the last century (Dunai *et al.*, 2013; Dakopoulos *et al.*, 2012). These aids are based on ultrasounds, vibration, Braille, acoustics, Global Position Systems and

synthetic speech and are implemented into the white sticks, mobile phones, or small computers and glasses; none of these technologies have been proven to be reliable enough for communication, navigation, control, or recognition of products and people. The development of ETA devices for blind people requires studies of all psychoacoustic cues and the behaviour of the electronics on the sound source generation and delivery, in addition to the behaviour of the blind people with sound source localization and echolocation (Thaler, 2013).

The sound source localization has been studied from a pure psychoacoustic experimental point of view in anechoic and semi anechoic chambers by studying the threshold of the interaural time difference in humans (Hartmann *et al.*, 2013; Brugera *et al.*, 2013). For the sound source localization in real environments, where the sounds are continuously moving, the effect of noises and esters is more complex due to influencing psychological factors (Dunai *et al.*, 2010). These data help in understanding how humans localize sounds and determine the thresholds in ITD by using sine tones and complex sounds.

2. Equipment design

Humans use a wide range of information for navigation such as depth, azimuth, and the inter click time interval. Before building the system, it is necessary to define the aspects of the visual scene. The visual scene represents the most important features for navigation and object identification because it includes the presence of the objects and their positions in space. The auditory system, which is capable of combining information by classes of cues and by frequencies to synthesize a unitary spatial image, plays a crucial role for navigation. Moreover, the auditory system solves a difficult problem when localizing sounds, mainly when there is more than one sound source (Takahashi, Keller, 1994).

The main drawback of the existing systems is the complexity of the computational algorithm and the high cost of the necessary resources. Regarding the speech navigational systems, they give precise information, although they cannot provide real-time information. In addition, they can be confused with human speech.

To solve all of these disadvantages, a cognitive system, which could provide access to the spatial information surrounding the participant via a sensory system and audio interface, was developed (Mora *et al.*, 2006).

The system is composed of a three dimensional complementary metal oxide semiconductor laser system, a FPGA (field programmable gate array) and a stereo headset (see Fig. 1). A disadvantage with the laser is that its long time contact with the human eyes causes vision disorder. Due to safety concerns, the system was not permitted for commercialization.

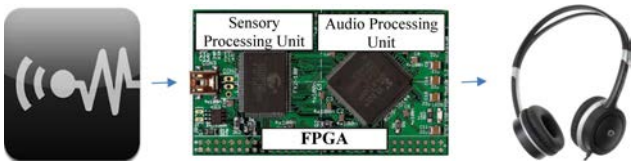


Fig. 1. Schematic representation of the system.

The system is based on a computational algorithm that assigns a two dimensional position to any object falling in the overlapping region of the sensor field of vision. From each of the 64 pixels of the sensor, a light beam is emitted. In this way, the exact distance between the sensor and the object for a specific azimuth angle is obtained. The computational algorithm attributes a virtual random sequence of short binaural sounds to each image pixel according to the obtained distance. Those sounds are previously generated in an anechoic chamber and convolved with a click sound (the methodology is described in Subsec. 2.2). Each one of the 64 array of sounds is reproduced with a delay of 8 ms. Each sensor corresponds to one directional sound starting from 32° on the left up to 32° on the right side of the human head. In this way, the blind people are able to perceive the whole frontal image through binaural sounds. It is important to remark that an important innovation of the proposed system is that two-dimensional environment information is acquired and represented by two-dimensional sounds.

2.1. Sensory system

It is desirable for the visual information input unit to be small and lightweight because these devices will be mounted on the participant's head. The sensory system with all of the optical components, analogue and digital electronics and laser were assembled on a pair of glasses as shown in Fig. 2. The maximum distance reached by the sensor is 5 m at 64° in azimuth. The patented measurement principle is based on a time of flight measurement of pulse-modulated laser light utilizing a high-speed photosensitive CMOS sensor and infrared laser pulse illumination. The analogue signals of several laser pulses are averaged on the chip to reduce the required laser power and also to increase the measurement accuracy. A fully solid state micro system is embedded on the FPGA (Field Programmable Gate Array). The advantage of using these sensors is providing an exact distance for both the horizontal and frontal planes. In addition, they reduce the necessary processing time for the calculation. The information from the CMOS sensor is used in the audio representation module when decisions are made for generating the appropriate sound map. We have used this data input device for accuracy and response time.

Fig. 2. System components. The laser was assembled on a pair of glasses.

2.2. Audio interface

The sensory modules provide two main types of data on the participant frontal scene: one – the location and direction of the objects and second – the set of coordinates where a horizontal plane passing at eye level cuts the surface of the existing object. The audio interface is able to synthesize the set of sounds to be delivered in real time by means of a convolution operation between every spatial filter provided by the sensor system. The audio interface presents audio information to the participant, representing a limited area of the subject frontal scene. This area consists of a plane that is horizontal to the participant's head, located at the height of the ears. The used sounds are very short and impulsive sounds, a type of click that is processed for this system. The information is transmitted to the participant via headphones.

A click sound of 2048 samples at a sampling rate of 44.100 Hz and 47 ms was used in the experiment. The generated sound is convolved with previously measured non-individual head related transfer functions (64 HRTFs in azimuth at each 0.96° and 16 levels in distances from 0.5 m to 5 m).

A maximum length binary sequence (MLBS) has been used as a sound source to measure the HRTFs in an anechoic chamber and Kemar dummy head (Dunai *et al.*, 2011).

The HRTF measurement is based on the calculation of the impulse response for both ears by using the filter in frequency domain. For a sound signal x1(n) reproduced by the speakerphone, the registered response can be calculated as:

$$Y_1 = X_1 LFM, \tag{1}$$

where X_1 is the representation of the sound x1(n) in the transformed frequency domain, L is the transfer function of the speakerphone and all of the reproduction equipment, F is the transfer function of the space between the speakerphone and human ear, and M is the transfer function of the microphone from the human ear and all recording equipment.

To calculate the filter, it was necessary to generate a x2(n) reproduced by headphones such as the registered response of Y_2 equal to Y_1. Y_2 can be calculated:

$$Y_2 = X_2 HM, \tag{2}$$

where H is the transfer function of the headphone and all of the reproduction equipment.

Given that Y_1 is equal to Y_2, we obtain:

$$X_1 LFM = X_2 HM. \tag{3}$$

Solving for X_2 as the registered sound by headphones from Eq. (3) we obtain:

$$X_2 = \frac{X_1 LFM}{H}. \tag{4}$$

From Eq. (4), the filter T measured for the impulse responses for each ear can be calculated as:

$$T = LF/H. \tag{5}$$

The T filter represented by Eq. (5) is measured for one unique speakerphone for one sound position and one ear. However, it is necessary to measure the transfer functions for both ears simultaneously. Therefore, the transfer functions from the speakerphone to the microphone for both ears for different spatial positions are given by:

$$\begin{aligned} Y_1/X_1 &= LFM\ X_2, \\ Y_2 &= X_2 HM. \end{aligned} \tag{6}$$

If we multiply Y_1/X_1 by the inverse of Y_2/X_2, we obtain the digital filter T:

$$T = \frac{LF}{H}. \tag{7}$$

The impulse response is obtained by circular cross-correlation between the MLBS from the system input and the response at the output signal $y(n)$:

$$h(n) = \Omega_{sy}(n) = s(n) * \Phi y(n), \tag{8}$$

where Φ represents the circular cross correlation, and $s(n)$ represents the response of the excitation of the MLBS.

A fast Hadamard transform (FHT) was used to reduce the computational time of the impulse response $h(n)$:

$$h(n) = \frac{1}{(L+1)s[0]} P_2 \langle S_2\{H_{L+1}[S_1(P_1 y(n))]\}\rangle, \tag{9}$$

where P represents the permutation matrices, S represents the redimension matrices, and $y\ HL+1$ is the Hadamard matrices of L+1 degree.

In other words, the binaural cues influence the generation of the virtual sounds. The azimuth displacement of the virtual sound source was obtained by the interaural phase shift of the left and right signals in terms of time difference between the left and the right ear. Each virtual sound corresponds to one pixel from the sensor module and covers an area of 60°.

3. Test description and characteristics

Four participants took part in the test; two of them were blind, one (A) was a partially blind participant who was experienced in testing various types of electronic way-finding technologies, and the fourth had normal vision. The second blind participant (B) was a young blind man who lost his vision in 2003. The third one (C) was blind since birth. All of the participants had a normal hearing ability, demonstrating correct perception of the virtual sounds. The four different participants took part in the experiment to study the audio perception and the sound localization thresholds and to compare the behaviour of the participants with different types of blindness and that of the normal vision participant.

The performance of virtual sound localization by using stereovision in the real environment was evaluated in the two experiment groups: an indoor navigation experiment and an outdoor navigation experiment. The indoor navigation experiment took place in a large hall that was 15 m in length and 10 m in width. The outdoor navigation experiment took place in an open park. In both experiment groups, the test participants were subjected to environmental noises and could be distracted easily. The time interval between each set of virtual sounds was 153 ms. The virtual sound source localization is based on subjective and objective data, the data that were extracted from the subject during the test. The subjective data describe the information collected from the participants regarding the object position as indicated by hand and voice, which included the sound source position in the real word and the distance between the subject and the sound source. The objective data are based on the time recorded from the start to sound localization and the deviation of the original sound source and the pointed direction.

The study consisted of two stages: (1) familiarization with the system functionality and virtual sounds and (2) navigation in the real indoor and outdoor environments. At the beginning of the first phase, the subject received a short and concise explanation about the features of the system and how to manage it. A series of tasks were included regarding sensor acquisition, audio feedback and volume.

One of the objectives of the tests was the "externalization" of the sound source. The participants should perceive the sounds as coming from outside – from the objects themselves – rather than being in the ear. In this way, the recognition of the location of the objects and the perception of their height, width and distance can be achieved.

In the second stage, the exploration and the recognition of the objects, where the participant walked from a certain point towards the point where the object was placed, were studied. During the walking task, the participants conducted a series of exercises consisting of overcoming several soft objects with different dimensions that were arranged in a specific way. Some different situations were taken into account: a single column (Fig. 3a), a free passage between two columns (Fig. 3b), a wide wall (Fig. 3c), a single column in the front of a wall (Fig. 3d), and finally, an outdoor area (Fig. 3e).

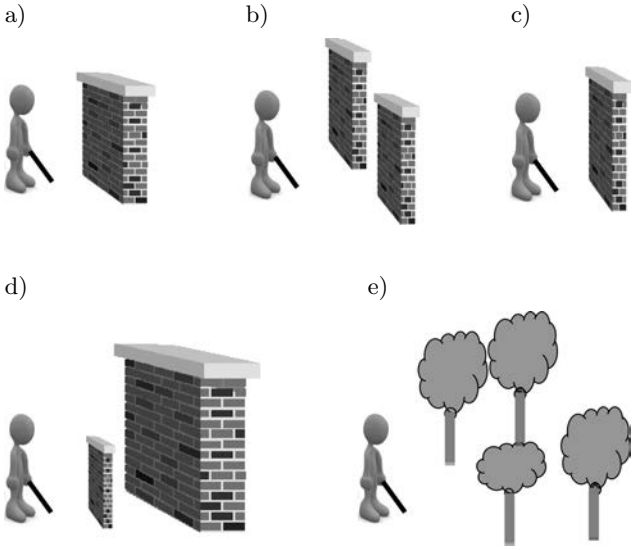


Fig. 3. Experimental scenarios: a) single column detection; b) two column detection and pass through them; c) a wide wall; d) a column detection in front of a wide wall and e) an outdoor experiment.

In all of the situations, the participants start the experimental test at a distance of 3 m from the object, faced in a direction such that the distance between the wall and the participant is greater than 5 m. Thus, neither the system will detect any object nor the subject will hear any sound from the system. The participant starts looking around to explore the environment. Whenever he detects an object in his direction of view, he receives external beeping sounds through stereo headphones. In the case that the participant approaches the column (object), the sounds increase in intensity. The intensity level of the spatial sound is inversely proportional to the distance. At the same time, the participant must comment on what he hears and how he perceives the sounds using his own imagination. In the case that there is more than one object (situations of two columns Fig. 3b, a column situated in the front of a wall, as shown in Fig. 3d, and in the outdoor environment, as shown in Fig. 3e), the participant listened to some differences in the intensity level of the sounds depending on the distance of each object. For the situation of the column situated in the front of a wall, the participant will hear an intense sound sequence coming from the column and a secondary sound, as a background, coming from the wall.

In every experiment, the participant is requested to correctly indicate the edges with his arms outstretched and to gauge the distance and width of the gap (Fig. 4).

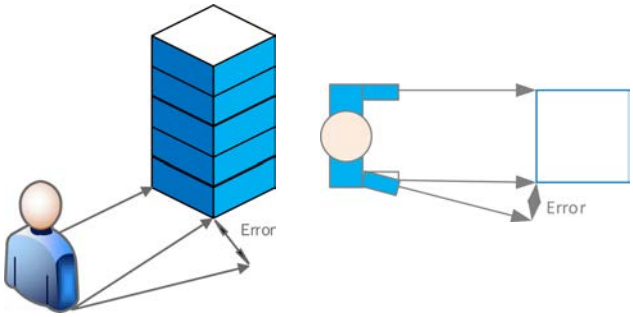


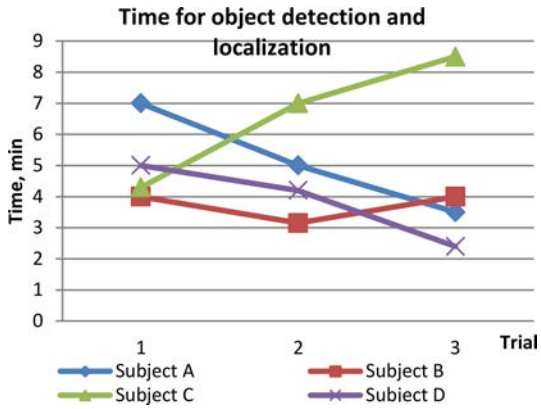
Fig. 4. Object volume and direction recognition representation.

4. Results and discussion

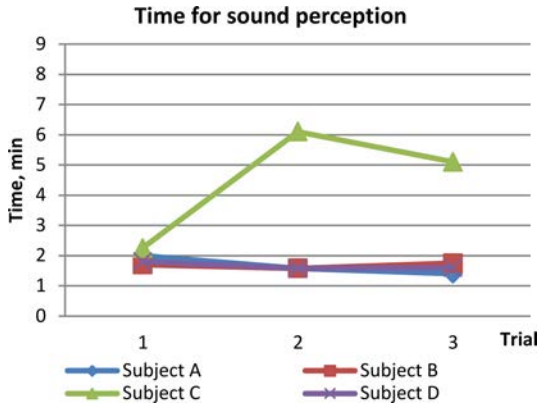
This section describes the results obtained with virtual sounds reproduced by a portable Toshiba computer, where the sound sources were selected by using a stereo vision system and real environment objects.

The data were collected in the five aforementioned scenarios from four subjects. These data are shown in Fig. 5. Figure 5a indicates the times in which the participants perceived the sounds for different trials (1, 2 and 3). Figure 5b shows the average times required by the participant for the detection and location of the objects in trials 1 using a white cane and only the developed device in trials 2 and 3.

a)



b)



c)

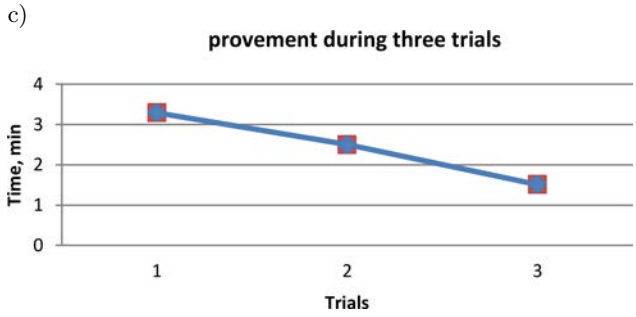


Fig. 5. Results of preliminary tests on auditory localization: a) time for sound perception; b) time for object detection and location, sound localization and object detection evolution after the first three trials, and c) the averaged sound source perception and object detection after three more trials for all blind participants.

The figure indicates that the average time for sound source perception was quite similar for all of the participants in each trial, with the exception of subject C who had difficulties in learning the system. The tests indicate that the error in distance perception is ±40 cm, which indicates that the subjects were wrong by only one or two steps. Regarding the height and width, the error rate was ±10 cm for a distance of 3 m between the participant and the object. The azimuth or the direction deviation error was measured offline by using image processing and tracking. The distance

error was determined by the subjective data collected from the participants.

As it can be observed in Fig. 5, the familiarization with the system functionality requires a short time (only a few minutes). Due to the sound frame rate, which is quite slow, walking with a normal rhythm was difficult. There were individual differences in performances; some participants had little difficulty with the guidance mode. In Fig. 5, it can be observed that subject C, who was completely blind since birth, had some difficulties when externalizing the sounds and perceiving the object localization. For this participant, additional time in the trainings was necessary. In general, the participants performed very well in these trials. After some trials, the participants were able to perceive the sound origin and to localize the objects in a few seconds. The average time for the sound perception was 2.32 min and for object detection 4.86 min. The total average time for completing these exercises was 3.59 minutes. The results obtained from Fig. 5a and 5b are relatively long; however, this time could be improved with training classes, as shown in Fig. 5c, where we can perceive almost linear fitting.

After the indoor trial sessions, additional tests were developed in an outdoor environment. After a training day, the participant was able to find the way outside quite easily. The exercise was certainly complex because to go outside the participant should pass through the door of the laboratory and through a corridor that has many corners to finally get outside. The blind participant was able to detect and to gauge the size of the obstacles such as a bicycle or a car (Fig. 6). The average walking time represents the average time for all of the participants over three different days: the first test outdoors was on day 1, and the tests on the next day (2) and after a week (3) required using only the device. The grey color represents the repetition of the same route with the same complexity. From Fig. 6, it can be observed that the blind participants easily improved their time scores for sound localization, decision making and navigation tasks. Regarding the number of hits, the same number occurred on the first two days; when the experiment was repeated after one week, they made more errors in decision making due to the large time between trials, which was shorter on the first day.

From the experiment, it was observed that the blind participants perceived sounds during all of the trials; they could externalize them and make a decision as to where the obstacle was and that they had to avoid it. In the outdoor trial, participant C had great difficulties in sound perception and sound externalization. He was unable to understand the system functionality; he could not assimilate that the binaural sounds that he heard through the headphones represented a sound from the environment and that it represented an object that he had to avoid. Participant C was not excluded

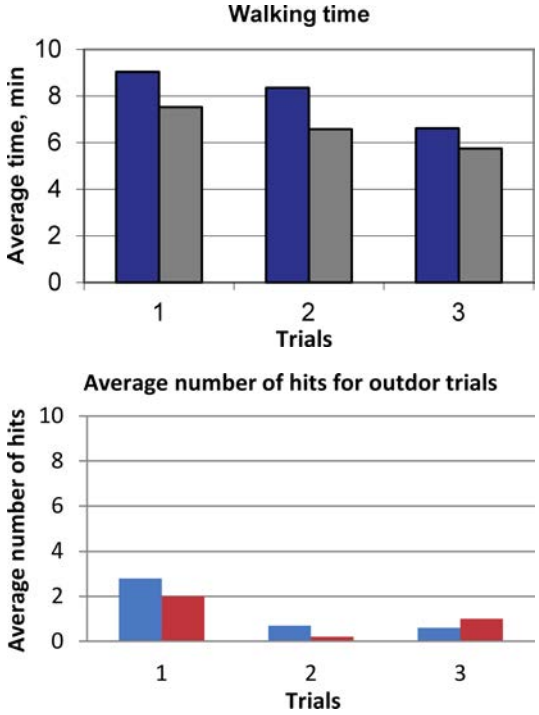


Fig. 6. Results of the preliminary tests for outdoor trials. In the left plot, the average walking time (blue) is represented for the three routes, and the grey color represents the repeated trials with the same complexity. In the right plot, the average number of hits represent the average number of hits for the same trials.

from the experiment because he represented a good case to analyze the training methods for blind people with hearing difficulties.

Another situation was based on detecting the presence of other persons standing in the front or moving. In this situation, the participant was able to gauge the distances at which other people were located. The results suggest that with the system used, it is possible to perceive the presence, position and dimensions of the detected object. This results indicates that the virtual sounds are a promising solution as a part of the participant interface of a blind navigation system. Due to the 64 lasers working continuously in the frontal horizontal plane, all objects that cross the lasers are detected and transformed into binaural sounds. As the 64 binaural sounds were in the azimuth, blind participants had more information and precision regarding the object surface horizontally. All of the participants demonstrated the ability of directional sound source localization with the experiment. They perfectly determined the left, the right and the center sounds. They also perceived the sound source movement from left to right and were able to perceive and localize more than one sound source simultaneously at different spatial positions with an interclick interval of 8 ms between sounds in the azimuth and an interclick interval of 5 ms

in the sound level. Performances were better than that of previously investigated models. The virtual acoustical sounds have the additional advantage of consuming less time than conventional speech or simple sine tones of a scanned image.

The presented data augment and expand the previous studies, which demonstrated the utility of direct perceptual cues for navigation. It will be important for future work to compare the guidance models in more complex environments, such as train stations, supermarkets, and stairs. Another topic for further research is the sound source localization and sound cues characteristics.

The presented findings may have good applications in guiding navigation for the blind people. An advantage of the system is that it could be integrated with other navigational systems, such as GPS and other visual interfaces.

One important result of the acoustic interface is that headphones do not exclude real sound appearing from outside.

5. Conclusions

The results show that blind people are able to perceive and localize virtual sounds through headphones. In addition, they were able to externalize perceived sounds through the headphones and interpret them from the real environment after a short practice. Due to the importance of time for blind people, the results are presented as the dependence of spent time for perceiving and localizing sounds by indicating the direction and the place from where the sounds come. The experiment demonstrated that the blind participants have a more developed hearing ability than sighted people and that sound source localization is one of the important factors for the survival of the blind people.

References

1. Blauert J. (1997), *Spatial Hearing: The Psychophysics of Human Sound Localization*, Revised edn, The MI Press, Cambridge, MA, USA.
2. Brugera A., Duani L., Hartmann W.M. (2013), *Human interaural time difference thresholds for sine tones: The high-frequency limit*, J. Acoust. Soc. Am., **133**, 5, 2839–2855.
3. Brungart D.S., Rabinowitz W.M. (1999), *Auditory localization of nearby sources. Head-related transfer functions*, J. Acoust. Soc. Am., **106**, 3, 1465–1479.
4. Dakopoulos D., Bourbakis N.G. (2010), *Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey*, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, **40**, 1, 25–35.
5. Dunai L., Peris-Fajarnés G., Lluna E., Defez B. (2013), *Sensory Navigation Device for Blind People*, The Journal of Navigation, **66**, 349–362.
6. Dunai L., Peris-Fajarnés G., Defez Garcia B., Santiago Praderas V., Dunai I. (2010), *The influence of the Inter-Click Intervalo n Moving Sound Source Localization for Navigation Systems*, Acoustical Physics, **56**, 3, 384–353.
7. Dunai L., Peris-Fajarnés G., Magal T., Defez Garcia B., Santiago Praderas V., Dunai I. (2011), *Virtual moving source localization through headphones*, InTech, 269–282.
8. Fitzpatrick D.C., Kuwada S., Batra R. (2000), *Neural Sensitivity to Interaural Time Differences: Beyond the Jeffress Model*, The Journal of Neuroscience, February 15, **20**, 4, 1605–1615.
9. Hartmann W.M., Dunai L., Qu T. (2013), *Interaural Time Difference Thresholds as a Function of Frequency*, Advances in Experimental Medicine Basic Aspects of Hearing, Physiology and Perception, **787**, 239–246.
10. Kunka B., Kostek B. (2012), *Objectivization of Audio-Visual Correlation Analysis*, Archives of Acoustics, **37**, 1, 63–72.
11. Mendelson Morton J., Haith Marshall M. (1976), *The Relation between Audition and Vision in the Human Newborn*, With Commentary by James J. Gibson; with reply by the authors; with Further Note by James J. Gibson, Monographs of the Society for Research in Child Developments, 41(4, Serial No. 167).
12. Mora J.L.G., Rodriguez-Hernandez A.F., Martin F., Castellano M.A. (2006), *Seeing the world by hearing: Virtual Acoustic Space (VAS) a new space perception system for blind people*, Proceedings of the 2nd Information and Communication Technologies Conference, ICTTA'06, IEEE pp. 837–842.
13. Takahashi T.T., Keller C.H. (1994), *Representation of multiple sound sources in the owl's auditory space map*, The Journal of Neuroscience, **14**, 8, 4780–4793.
14. Thaler L. (2013), *Echolocation may have real-life advantages for blind people: an analysis of survey data*, Front Physiol., **4**, 98.
15. Wonder S., Charles R. (2005), *Loss of sight and enhanced hearing: a neural picture*, Plos Biol, **3**, 2, e48.
16. WBU [World Blind Union] (2012), Visual impairment and blindness. Retrieved March 3th, 2014. Media center, http://www.who.int/mediacentre/factsheets/fs282/en/.
17. ASA [Acoustical Society of America] (2012), 'Blindness' may rapidly enhance other senses. ScienceDaily. 8 May 2012. ScienceDaily, www.sciencedaily.com/releases/2012/05/120508152002.htm.