VIII Congresso Ibero-americano de Acústica
acústica 2012
1-3 outubro
Évora - Portugal

VIRTUAL SOUND LOCALIZATION BY BLIND PEOPLE

Larisa Dunai, Ismael Lengua, Guillermo Peris-Fajarnés, Beatriz Defez

Camino de Vera s/n, Valencia, 46022, Spain Universitat Politècnica de València
ladu@upv.es, ilengua@upv.es, gperis@upv.es, bdefez@upv.es

Resumen
El presente artículo describe los métodos experimentales y los resultados obtenidos en la localización de sonidos virtuales con personas ciegas. La falta de visión esta casi siempre compensada por otras habilidades, así como la táctil, olfato y auditiva. Se ha demostrado que las personas ciegas, tienen una habilidad de localizar los sonidos que las personas con visión, utilizando las características auditivas monaurales y binauraless.
En el experimento, las personas ciegas han completado dos ejercicios: uno en laboratorio y el segundo en el entorno abierto, bajo distintas condiciones de ruido. Se ha utilizado un sonido llamado "click", cual ha sido procesado y convolucionado con las Funciones de Transferencia a la Cabeza HRTFs.
El espacio de sonificación es desde los 30º a la izquierda a 30º a la derecha en una distancia desde 0.3m hasta 5m.
Los resultados han demostrado que las personas ciegas perciben y localizan los sonidos binaurales y los asimilan como si procedieran desde el entorno.

Palabras-clave: sonidos virtuales, localización, distancias, azimut, personas ciegas.

Abstract
The present paper describes the experimental methods and results of virtual sounds localization by blind people. The lack of vision is often compensated by other perceptual abilities, as tactile or hearing ability. It is demonstrated that blind people localize sounds more accurately than sighted people, by using monaural and/or binaural cues.
In the experiment, blind users were supposed on two experiments, one took place in the laboratory and the second one in the real environment under different noise conditions. A simple click sound has been used and processed with non individual Head Related Transfer Functions.
The sounds has been delivered by a system with a maximum azimuth of 30º at the left side and 30º at the right side of the user head for a distance from 0,3m up to 5m.
The results show that blind people easily perceive and localize binaural sounds and assimilate them with the sounds proceeding from the environment.

Keywords: virtual sounds, localization, distance, azimuth, blind people.

PACS no. 43.66.Qp, 43.66.Yw

1 Introducción

Sound source localization ability is one of the most important factor for human survival. This hability helps humans to perceive and detect dangers and make the difference between changes of climatic

conditions, 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” [1]. Sound source localization is influenced by acoustical cues such as the Interaural Time Difference (ITD), Interaural Level Difference (ILD), torso and pinnae [2]. For virtual sound sources localization an important role plays the Head Related Transfer Functions (HRTF), which represent the transfer characteristics of the sound source in a free field to the listener’s external ears [1].
The present paper describe the methods and the reults of the virtual sound sources localization by blind people by usin a non-individual Head Related Transfer Functions.
In the second chapter the equipment used for sound localization is briefly described. the equipment is composed by a sensory device which measures the distance from the device to the environment objects. The measured distances are transformed into virtual sounds and delivered through headphones. Also the methods of generation and convolution of the acoustical sounds is described.
In the third chapter the experimental procedures are described.
In the fourth chapter the results and discutions over the method are presented.

2 Equipment design

Humans use a wide range of information for navigation such as depth, azimuth etc. Due to this fact, it is not easy to decide which information can be the best acoustically interpreted. Therefore, before building the system is extremely necessary to define the aspects of the visual scene which represent the most important features for navigation and object identification (presence of the objects and its position in space). The auditory system, which is capable of combining information by classes of cues and by frequencies in order to synthesize a unitary spatial image, plays a crucial role for the navigation. Indeed, the auditory system solves a difficult problem when localizing sounds, mainly when there is more than one sound source [3].

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.

In order to solve all these disadvantages, a cognitive system, which would provide access to the spatial information surrounding the user via sensory system and audio interface, was developed [4]. Figure 1 shows the general scheme of the developed system. That figure shows schematically the processing steps for converting the sensor data input into an acoustic stimulus.

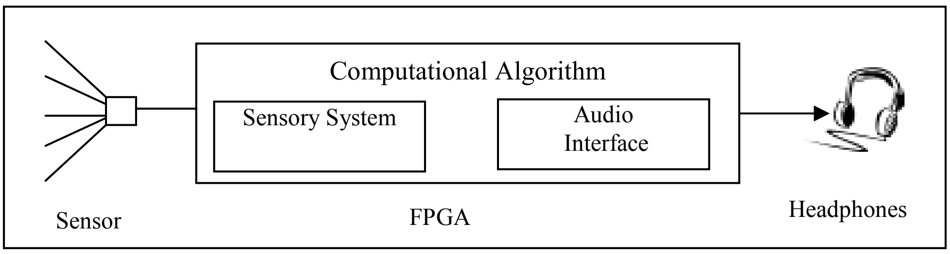
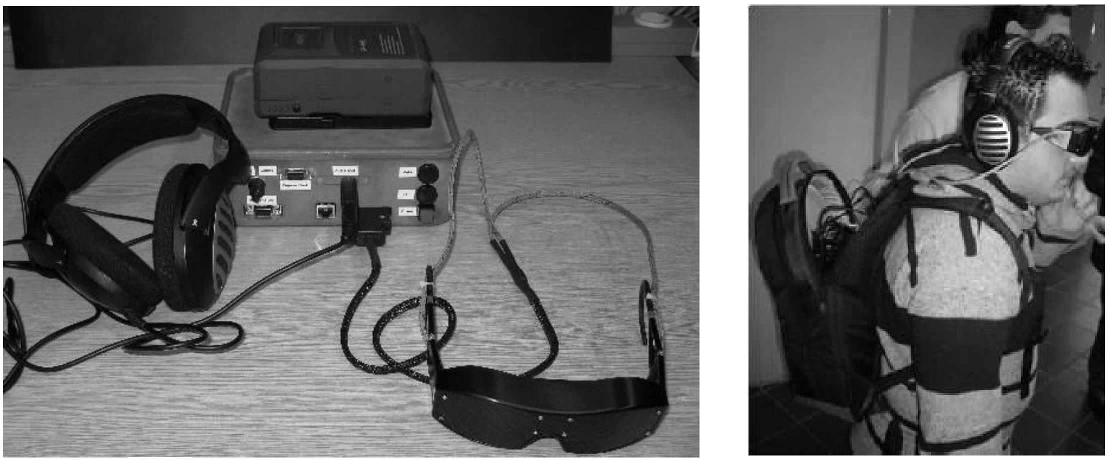


Figure 1 – Schematic representation of the System.

The system is based on a computational algorithm which 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 beaming 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 to each pixel, according to the obtained distance, a virtual random sequence of short 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 user head. The sensory system with all the optical components, analogue and digital electronics and laser were assembled on a pair of glasses, as shown in Figure 2a. The maximum distance reached by the sensor is 5 m at 64º in azimuth. The patented measurement principle is based on Time of Flight measurement of pulse-modulated laser light utilising a high-speed photosensitive CMOS sensor and infrared laser pulse illumination. The analogue signals of several laser pulses are averaged on chip to reduce the required laser power and also to increase measurement accuracy. A fully solid state micro system is embedded on FPGA. The advantage of using these sensors is to provide an exact distance both in a horizontal and in a frontal plane. In addition, they reduce the necessary processing time for the calculation. The information from CMOS sensor is used in the audio representation module when decisions are made for generating the appropriate sound map.



a) b)

Figure 2 – Overview and system composition: a) Components of the operational system b) Full device picture.

2.2 Audio Interface

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

A click sound of 2048 samples at sampling rate of 44.100 Hz and 47 ms has been used on the experiment. The generated sound is convolved with previously measured non-individual Head Related Transfer Functions (64 HRTF's in azimuth at each 0.96º and 16 levels in distances from 0.5m to 5m).

3 Tests description

Several tests were developed in order to verify the different features of the system. These tests were carried out on four users; three of them were blind, whereas the fourth had normal vision. One of blind users (A) was a partially blind user, who was experienced in testing various types of electronic way-finding technologies. The second blind user (B) was a young blind man who lost his vision in 2003. The third one (C) was blind since birth. All participants had normal hearing abilities demonstrating correct perception of the virtual sounds.

The tests were carried out on the basis of a learning protocol. The indoor navigation environment was a large hall with 15 m length and 10m width. The outdoor navigation environment was a park. The time interval between each set of virtual sounds was 153ms.

The study consisted of two stages: (1) familiarization with the system functionality and virtual sounds and (2) navigation in the real indoor and outdoor environment. 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.

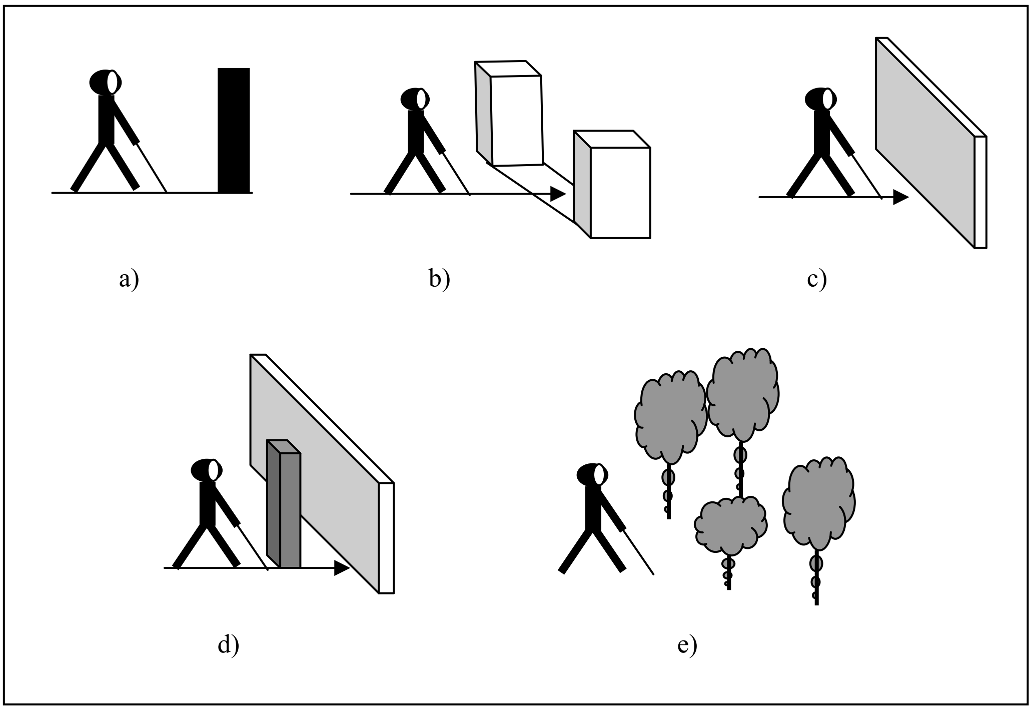


Figure 3 – Experimental scenarios: a) Single column detection; b) Two columns detection and pass through them; c) A wide wall; d) A column detection in front of a wide wall and e) Outdoor experiment.

One of the objectives of the tests was the “externalisation” of the sound source. The users should perceive the sounds like 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 (Figure 4).

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

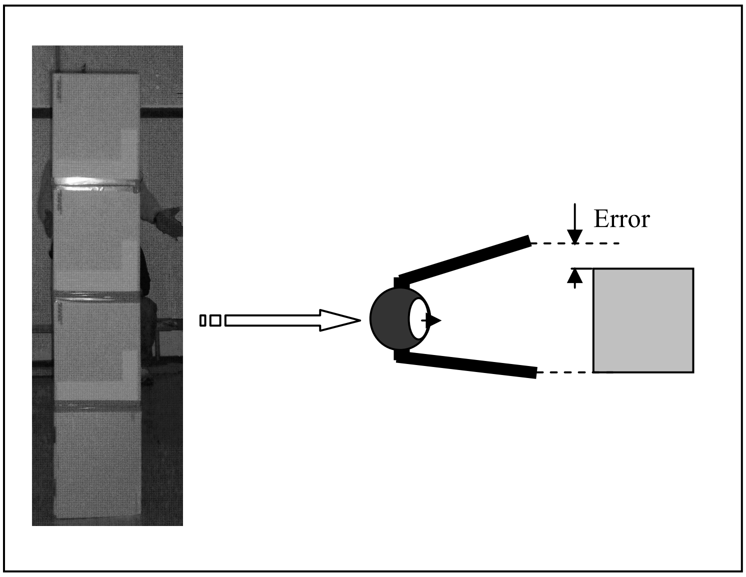


Figure 4 – Distance and volume recognition representation.

In all situations, the subject starts the experimental test at distance of 3 meters far from the object, faced in a direction such that the distance between the wall and the user is longer than 5 m. Thus, neither the system will detect any object nor the subject will hear any sound from the system. The user starts looking around in order to explore the environment. Whenever he detects an object in his direction of view, he receives an external beeping sound through stereo headphones. In the case that the subject gets near the column (object), the sounds increase in intensity. The intensity of the spatial sound is inversely proportional to the distance. At the same time, subject must comment what he listens and how he perceives the sounds using his own imagination.

In the case that there are more than one object (situations of two columns Figure 3 b), a column situated in the front of a wall Figure 3 d) and in the outdoor environment Figure 3 e)), the subject listened some differences in the intensity of the sounds depending on the distance of each object. For the situation of the column situated in the front of a wall the subject will listen an intense sound sequence coming from the column and a secondary sound, as a background, coming from the wall.

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

4 Results and discutions

The data were collected in the five aforementioned scenarios from four subjects (three blind and one subject with normal vision). These data are shown in Figure 5. Figure 5 (a) indicates the times in which the users perceived the sounds for different trials (1, 2 and 3). Figure 5 (b) shows the average times required by the user for the detection and location of the objects in trials 1, 2 and 3. The figure indicates that the average time for sound source perception was quite similar for all users in each trial, with exception of subject C who had difficulties in the learning of the system. The tests indicate that the error in distance perception is +/- 40cm, which means that the subjects were wrong with only one or two steps. Regarding to the height and width the error rate was +/-10 cm.

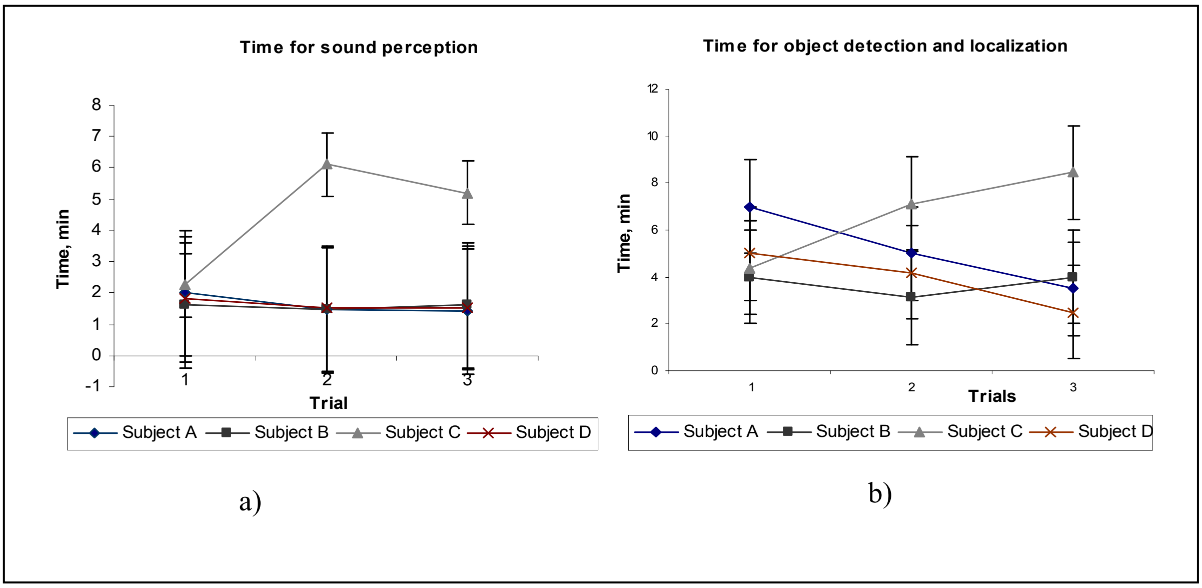


Figure 5. Results of preliminary tests on auditory localization; (a) Time for sound perception, (b) Time for object detection and location.

As it can be seen in Figure 5, the familiarization with the system functionality requires a short time (only few minutes). Due to the sound frame rate which is quite slow, to walk with a normal rhythm was difficult. There were individual differences in performances: some participants had little difficulties with the guidance mode. In Figure 5 it can be seen that the subject C, who is total blind since birth, has some difficulties when externalizing the sounds and perceiving the object localization. For this user, additional time in the trainings was necessary. In general, participants performed very well on these trials. After some trials, the users were able in few seconds to perceive the sound origin and to localize the objects. The average time for the sound perception was 2.32 min and for the object detection 4.86 min. The total average time for completing these exercises was 3.59 minutes.

After the indoor trial sessions, additional tests were developed in outdoor environment. After a training day, the user was able to find the way outside quite easily. The exercise was certainly complex, since in order to go outside the user should pass through the door of the laboratory, through a corridor with many corners and finally to arrive outside. The subject was able to detect and gauge the size of obstacles such as a bicycle or a car, though she was not able to identify them.

Another situation was based on detecting the presence of other persons standing in the front or moving. In this situation, the user was able to gauge the distance at which other people were. The results suggest that with the used system it is possible to perceive the presence, the position and the dimensions of the detected object. This is indicating that the virtual sounds are a promising solution as a part of the user interface of a blind navigation system. Performances were better than other models investigated before. The virtual acoustical sounds have the additional advantage of consuming shorter time than conventional speech.

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

The present findings may have good applications in guiding navigation for the blind people. An advantage of the system is that 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 shows that all blind people are able to perceive and localize virtual sounds through headphones. Also they were able to externalize the perceived through headphones sounds and interpret them as providing from the real environment after a short practice. Due to the importance of the time for blind people the results are presented in dependence of spent time for perceiving and localization of the sounds and showing the dirrection and the place where the sounds come from. The experiment demonstrate again that blind useers have more developed hearing abilities than visual people and sound source localization is one of the important factor for blind people survival.

References

[1] Blauert J. Spatial Hearing: The Psychophysics of Human Sound Localization. *Revised edn, The MI Press*, Cambridge, MA, USA, 1997.

[2] Brungart D.S.; Rabinowitz W. M. Auditory localization of nearby sources. Head-related transfer functions, *J. Acoust. Soc.Am.* 106(3), 1465-1479 1999.

[3] Takahashi T. T.; Keller C. H. Representation of multiple sound sources in the owl's auditory space map. *The Journal of Neuroscience*, 14(8), pp. 4780-4793, 1994.

[4] Mora J.L. G.; Rodriguez-Hernandez A.F.; Martin F., Castellano M.A. 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, 2006.