Buletinul Ştiinţific al Universităţii "Politehnica" din Timişoara

Seria ELECTRONICĂ şi TELECOMUNICAŢII
TRANSACTIONS on ELECTRONICS and COMMUNICATIONS

Tom 52(66), Fascicola 2, 2007

A HRTF Interface for Visually Impaired People

Laviniu Ţepelea[1], Virgil Tiponuţ[2]

Abstract - For visually impaired people the hear sense replace the seeing sense. To help people moving in an unstructured environment in the real world we try to make a 3D virtual audio reality. The mostly used method are HRTF (Head Related Transfer Functions), a set of audio filters. A simulation model in Matlab demonstrates the possibility to tell visually impaired people where is the obstacles. We use the simulation to make a man-machine interface, which is part of a big project to guide these people to walk and work in the real world.

Keywords: visually impaired people, 3d virtual audio, binaural, HRTF.

I. INTRODUCTION

There are many blind persons around the world. In U.S.A. they are estimated to 11.4 millions of impaired people [1] and in Japan are 307 thousands (65 thousands completely blinds) [2]. Therefore a system guiding them in the real world is very useful.

The goal of this paper is to demonstrate how a system with a HRTF module can guide blind or visually impaired people.

The human hearing system has remarkable abilities in identifying sound source positions in 3D space. Although this process is often aided by visual sense, knowledge and other sensory input, and allows directional positioning in space [3], [4].

In absence of seeing sense, the human hearing is not enough in guiding because of obstacles which not makes sounds. In that case we try to make a 3D audio environment to replace the images.

The interest in 3D-sound techniques has increased recently. Also, there is an increasing level of interaction between the user and the virtual environment [5], [6].

For many years, interest in 3D binaural systems has been in the area of scientific research, simulation and entertainment. [7] Binaural recordings are intended to be reproduced through headphones, and can give the listener a very realistic sense of sound sources being located in the space outside of the head [8].

To answer how 3D audio systems work, it is useful to start by considering how humans can localize sounds using only the ears. A sound generated in space creates a sound wave that propagates to the ears of the listener. When the sound is to the left of the listener, the sound reaches the left ear before the right ear, and thus the right ear signal is delayed with respect to the left ear signal. In addition, the right ear signal will be attenuated because of *shadowing* by the head. Both ear signals are also affected by the torso, head, and in particular, the *pinna* (external ear). The various folds in the pinna modify the frequency content of the signals, reinforcing some frequencies and attenuating others, in a manner that depends on the direction of the incident sound. Thus an ear acts like a complicated tone control that is direction dependent. We unconsciously use the time delay, amplitude difference, and tonal information at each ear to determine the location of the sound. These indicators are called sound localization *cues* [9].

Interaural Time Difference (ITD) and the Interaural Level Difference (ILD) are known to provide primary cues for localization in the horizontal plane [10].

When the sound waves encounter the outer ear flap (the pinna), they in fact interact with complex convoluted folds and cavities of the ear, as we can see in figure 1 [11].

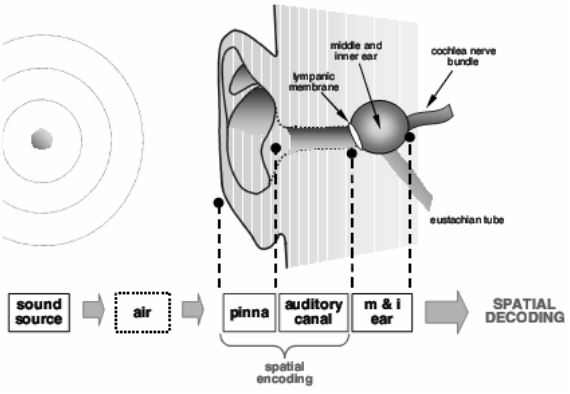

Figure 1 – Acoustic transmission pathway

The main central cavity in the pinna, knows as the *concha*, makes a major contribution at around 5 KHz to the total effective resonance of the outer ear,

[1] University of Oradea, Str. Universităţii 1, 410087, Oradea, Romania, Electronic Department Tel: +40-259-408735, Fax: +40-259-432789, e-mail: ltepelea@uoradea.ro

[2] Politehnica University from Timişoara, Applied Electronic Department, B-dul Vasile Pârvan, Nr. 2, 300223 Timişoara, Romania, e-mail: virgil.tiponut@etc.upt.ro

boosting the incoming signals with approximately 10 to 12 dB at this frequency. The concerted actions of the various acoustic effects create a measurable set of characteristics known as the *Head-Related-Transfer-Function* (HRTF), which comprises three elements: (a) a near-ear response, (b) a far ear response, (c) an inter-aural time delay. To illustrate these effects, characteristics of horizontal-plane HRTF are shown in figure 2 for an azimuth angle of -50^0.

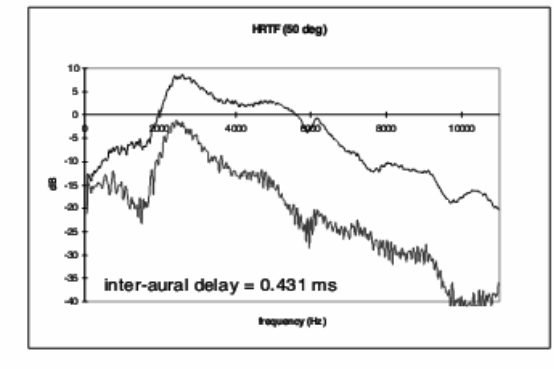

Figure 2 – Typical HRTF characteristics

The three characteristic elements of an HRTF can be synthesized electronically, and then delivered to a listener's ears via headphones or loudspeakers [11].
Therefore, in addition to the primary 3D sound cues (ITD, ILD), there are several additional secondary cues that contribute to our localization capability, such as shoulder, torso, pinna reflections, local room reflections and psychological cues. Shoulder contributes with reflections in certain positions. They provide a strong reflection from lateral sources, with a short path-length of around 10 cm between direct sound and reflections. The effects are most important for side-positioned sources, especially for height effects, where the shoulders tend to mask sources which are below 30^0 depressions [12]. The pinna is responsible to the position of source in vertical plane, which is named *elevation,* because he modifies the frequencies relative to the position of source above or below the head plane. All these effects can be described by a complex frequency response function called the Head Related Transfer Function (HRTF). The corresponding impulse response is called Head Related Impulse Response (HRIR) [10]. In engineering terms, these propagation effects can be captured by two transfer functions, H_L and H_R, which specify the relation between the sound pressure of the source and the sound pressure at the left and right ear drums of the listener. These so-called Head Related Transfer Functions are acoustic filters that vary both with frequency and with the azimuth θ, elevation φ and range r to the source, like in figure 3 [13].
Because of complexity of reflection process there is more difficult to realize a simulation model to be accurately and short in calculus, therefore the HRIR was empirically determined and from here we got the results, HRTF filters.
Finally, there are before mentioned psychological cues present in everyday man life, which work together with the audio cues to build an *image* of the world around us. For example, if you hear the sound of a helicopter flying, you expect it to be up in the air and not downwards. If a dog were to bark nearby, you would expect it to be downwards [12]. These psychological cues are useful, but in the case of native blinds people, the absence of these cues are replaced with the necessary few hours of training with the guiding system.

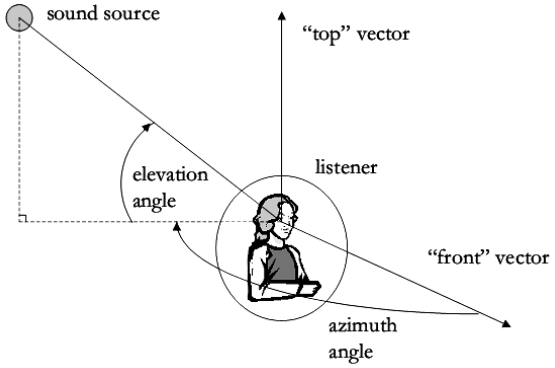

Figure 3 – Azimuth and elevation specify source direction

II. HEAD RELATED TRANSFER FUNCTIONS

The HRTF varies significantly between different individuals due to differences in the sizes and shapes of different anatomical parts like the pinna, head shoulders and torso. Applications in the creation of virtual auditory displays require individual HRTF's for perceptual fidelity [10].
There are two sets of measurements made by two different teams: HRTF from MIT Media Lab. and from CIPIC Interface Laboratory. Measurements from MIT Media Lab. was made using a mannequin named KEMAR dummy-head (Knowles Electronic Manikin for Acoustic Research) [14]. The measurements were made in MIT's anechoic chamber. The KEMAR is an anthropomorphic manikin whose dimensions were designed to equal those of a median human. The pinna used was modelled from human pinna. In total, 710 measurements were made at different locations around the KEMAR. When synthesizing a location that is not in the measured set, HRTF's from four adjacent locations are interpolated [9]. When obtaining HRTF measurements for use in an auditory virtual reality, it is clearly impractical to measure the HRTF for each possible position relative to the listener and therefore the HRTF's are measured for a discrete subset of positions only. However, as previously described, the HRTF corresponding to each unique position is itself unique. As a result, a virtual auditory environment must be able to handle the synthesizing of a sound source which corresponds to a *non-sampled* HRTF positions which will arise when considering a moving sound source, moving listener or a stationary sound source at a position finer than the *grid* of sampled positions. The simplest interpolation technique is linear interpolation whereby the desired HRTF is obtained by taking a linear average of the neighbouring HRTF's. This technique

results in HRTF's which are acoustically different when compared to the actual measured HRTF of the desired target location. However, localization accuracy is not affected by linear interpolation of nonindividualized HRTF's even with a large interval separating the sampled HRTF measurements. Various other interpolation techniques can also be used, such as the more complex spline interpolation techniques, used in various other fields, including computer graphics. Regardless the interpolation technique actually used, some method must be used to handle the fact that it is clearly impractical to measure and store HRTF responses for each location in space relative to the listener [16].

As previously mentioned, the pinna of each person differ with respect to size, shape and general make-up, leading to differences in the filtering of the sound source spectrum, particularly at higher frequencies. The higher frequencies are attenuated by a greater amount when the sound source is to the rear of listener as opposed to the front of the listener. In fact, in the 5kHz to 10kHz frequency range, the HRTF's of individuals can differ by as much as 28dB. This high frequency filtering is an important cue to source elevation perception and in resolving any front-back ambiguities.

Despite the benefits which may be offered to a listener through the use of individualized HRTF's, the process of actually collecting a set of an individuals HRTF's is an extremely difficult, time consuming, tedious and delicate process requiring the use of special equipment and environments, such as an anechoic chamber. It is therefore very impractical to use individualized HRTF's and as a result, nonindividualized HRTF's are used instead [16].

The full set of HRTF from MIT, used worldwide, consists of functions, measured at 72 different azimuths and at 14 different elevations. Functions cover the whole azimuth (0^0 – 360^0) and elevation range from -40^0 to 90^0. Directions in the azimuth are 50 apart, but the division in elevation is not uniform. Each function consists of 512 samples with a sample frequency of 44.1 kHz. [15] Utilizing the symmetry of the head, the KEMAR was setup with two different pinnas. The left pinna in KEMAR was a normal pinna, while the right pinna was a slightly larger one.

Those measurement realised by CIPIC Laboratory consists of 45 individual HRTF datasets obtained from 43 different human subjects (27 men and 16 women) and a KEMAR mannequin (with two different pinna models). For each subject, a total of 1250 measurements were taken at each ear, 25 different azimuths and 50 different elevations [16].

III. MATLAB IMPLEMENTATION OF 3D AUDIO ENVIRONMENT

For this simulation we have used the HRTF database from MIT Laboratory and an implementation realized by Kaibo Liu and Syed Hassan [14]. There are two types of data sets that MIT have made publicly available. The *compact data set* is a set of impulse responses which has been preprocessed to compensate for recording equipment response and other factors, and are ready to be used directly. The other set of data, called *full data set* is what they actually recorded when they were generating the data. We preferred to use the full data set for various reasons. The full data set has 512 taps for the FIR filter instead of 128 taps in the compact data set. We hope this will generate a better 3D sound. Another advantage is that we could use data for two different pinna set instead of a single pinna pair. Finally, using the full data set allowed us to realise the difference that is caused by not compensating for the recorded equipment's response and this effect is implemented in the *diffused field* option in the GUI.

Like we mentioned before MIT made the measurements with two sets of pinna. Therefore, we implemented in the GUI those sets of measurements. The pinna named pinna set 1 is a normal pinna while that named pinna set 2 is that larger one.

The microphones in the KEMAR's ears, when record sound, also picked up the ear canal resonance of the manikin's ears. When these HRTF's are used to generate sound, the listener will hear the KEMAR ear canal resonance in addition to his own ear canal resonance. Besides that, as mentioned earlier, the full data set contains the recording system's response too. A possible way to eliminate the measurement system's response, as well as effect of ear canal resonance is to normalize the measurements with respect to an average across all directions (called diffuse-field equalization). Since neither the measurement system response nor the ear canal response change as a function of sound direction, they will be factored out of the data. To find the diffuse filed data, the magnitude squared responses of all responses is averaged, which results in power average across all directions. We have provided user the choice of using the diffuse field data. If he chooses to use it, we use the pre-computed value of inverse diffuse field to process the sound before playing. In general, the sounds synthesized using diffuse-field data can be localized better. The purpose to provide the option there is to allow the users to evaluate its effect themselves.

The graphical user interface takes the azimuth and elevation value, or range of values, which the user wants to generate the sound for, and when sound from a particular direction is to be played, the main program loads the corresponding HRTF data, and filters the sound using that data. The sound is played using the headphones. To generate sound from a particular direction, user can just click on that particular direction related to the head on the virtual room, like in figure 4. A dot is displayed to visually indicate the chosen direction, and the corresponding azimuth is displayed in the text box below. By default, white noise is used to simulate 3D effects because it has all the frequencies and hence can adjust

the HRTF effects in the best way. We can use also other sounds, like *glass, gun, helicopter or* any other.

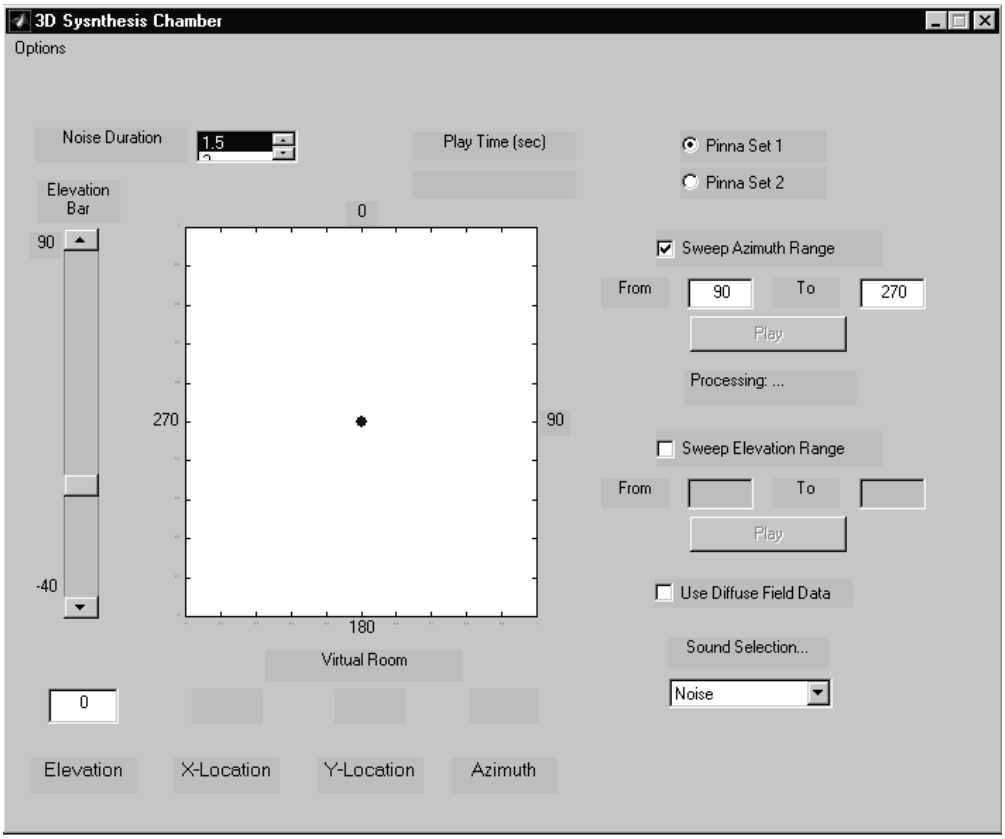

Figure 4 – Simulation in Matlab of 3D virtual environment

IV. CONCLUSIONS

Overall the sound generated by the program is satisfying. But because of nonindividualized HRTF used, there is front-back confusion. A front-back confusion results when the listener receives the sound to be in the front when it should be in back, and vice-versa. Elevation is also a little difficult to *guess*, which is common in 3D audio systems. Now we try to get better synthetic 3D audio environment to make a man-machine interface for an effective guidance system with the use of local navigation with sensors and global navigation with GPS-GPRS module for the visually impaired or blind people.

REFERENCES

[1] M. La Plante and D. Carlson, "Disability in the United States: Relevance and Causes", U. S. Department of Education, National Institute of Disability and Rehabilitation Research, Washington D. C., 2000.

[2] H. Mori and S. Kotani, "Robotic travel aid for the blind: HARUNOBU-6", Proc. 2nd Euro. Conf. Disability, Virtual Reality & Assoc. Tech., Skövde, Sweden, ECDVRAT and university of Reading, pp. 193-202, UK, 1998.

[3] N. Rober, S. Andres, M. Masuch, "HRTF simulations through acoustic raytracing", Department of Simulation and Graphics, School of Computing Science, Otto-von-Guericke-University Magdeburg, Germany.

[4] Durand R. Begault, 3D Sound - For Virtual Reality and Multimedia, NASA Ames Research Center, 2000.

[5] Fabio P. Freeland , Luiz Wagner P. Biscainho, Paulo Sergio R. Diniz, "Efficient HRTF interpolation in 3d moving sound", Universidade Federal do Rio de Janeiro, Brasil

[6] C. Phillip Brown, Richard O. Duda, "A Structural Model for Binaural Sound Synthesis", Fellow, IEEE Transactions On Speech and Audio Processing, Vol. 6, No. 5, September 1998.

[7] C. Jin, T. Tan, A. Kan, D. Lin, A. van Schaik, K. Smith, M. McGinity, "Real-time, Head-tracked 3D Audio with Unlimited Simultaneous Sounds", Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display, Limerick, Ireland, July 6-9, 2005.

[8] David A. Burgess, "Real-Time Audio Spatialization with Inexpensive Hardware", Graphics Visualization and Usability Center - Multimedia Group, Georgia Institute of Technology, Atlanta, Georgia.

[9] William G. Gardner, Ph.D., "3D Audio and Acoustic Environment Modeling", Wave Arts, Inc., Arlington, March 15, 1999.

[10] Vikas C. Raykar , Ramani Duraiswami, Larry Davis, B. Yegnanarayana, "Extracting significant features from the HRTF", Proceedings of the 2003 International Conference on Auditory Display, Boston, MA, USA, July 6-9, 2003.

[11] Alastair Sibbald, "Virtual audio for headphones", Sensaura Ltd., 2000.

[12] Alastair Sibbald, "Digital Ear Technology", Sensaura Ltd., 2001.

[13] Richard O. Duda, "Modeling Head Related Transfer Functions", Preprint for the Twenty-Seventh Asilomar Conference on Signals, Systems & Computers, October 31-November 3, 1993.

[14] Kaibo Liu, Syed Hassan, "Matlab Implementation of 3D Synthetic Environment", Electrical and Computer Engineering Computer Lab.

[15] Rudolf Susnik, Jaka Sodnik, Anton Umek and Saso Tomazic, "Spatial sound generation using HRTF created by the use of recursive filters", EUROCON 2003 Ljubljana, Slovenia.

[16] Bill Kapralos, "Auditory Perception and Virtual Environments", Ph.D. Qualification Exam Document, Department of Computer Science, York University, North York, Ontario, Canada, January 29, 2003.