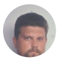
Efficient Algorithm for Extracting Essential Head Related Impulse Response Data for Acoustic Virtual Reality Development



David-George Cristea
Technion - Israel Institute of Technology

Ivan Bogdanov
Polytechnic University of Timisoara

Efficient Algorithm for Extracting Essential Head Related Impulse Response Data for Acoustic Virtual Reality Development

ZOLTAN HARASZY, DAVID-GEORGE CRISTEA, SEBASTIAN MICUT, IVAN BOGDANOV
Department of Applied Electronics
POLITEHNICA University of Timisoara
Bvd. Vasile Parvan 2, 300223 Timisoara
ROMANIA
zoltan.haraszy@etc.upt.ro, george.cristea@etc.upt.ro, sebastian.micut@etc.upt.ro, ivan.bogdanov@etc.upt.ro
http://www.ea.etc.upt.ro/index.php/en/personal/cadre-didactice/haraszy-zoltan

Abstract: - The new Acoustic Virtual Reality (AVR) concept is often used as a man-machine interface in electronic travel aid (ETA), that help blind and visually impaired individuals to navigate in real outdoor environments. According to this concept, the presence of obstacles in the surrounding environment and the path to the desired target will be signalized to the blind subject by burst of sounds, whose virtual source position suggests the position of the real obstacles and the direction of movement, respectively. The practical implementation of the AVR concept requires the so-called Head Related Transfer Functions (HRTFs) to be known in every point of the 3D space and for each involved individual. In the present paper, we describe an efficient algorithm for extracting essential Head Related Impulse Response (HRIR) data, we apply it for one person's HRIRs from the Listen Ircam HRTF database. To verify its applicability, after describing the experimental setup, listening tests are also conducted and results are compared with the results of listening tests without using the proposed algorithm. Finally, conclusions and future research ideas are also presented.

Key-Words: - head related transfer functions, artificial neural networks, man-machine interface, acoustic virtual reality, visually impaired, localization experiment.

1 Introduction

In the last decade, many research efforts have been devoted to develop electronic travel aid (ETA) with new features, used by blind and visually impaired persons to navigate in real world environments [1]-[10]. These devices, based on sensors and signal processing, improve the mobility of the blind users (in terms of safety and speed), in unexplored or dynamically changing environment.

In spite of these achievements, the traditional tools (white can and guiding dogs [1]) still remain the most used travel aid by the blind people's community. The main drawbacks of existing assistive tools are their limited abilities to detect obstacles in front of the subject and the level of technical knowledge required to operate these tools.

The success of an ETA device is highly influenced by the man-machine interface used between the blind person and the travel aid. It has been proved [6] that AVR concept can be successfully used in order to implement such a man-machine interface.

The man-machine interface, based on the AVR and implemented in our previous work [11], has been suggested by the high sensitivity and accuracy of the human hearing [12]. This human-computer interface must be implemented in such a way to be accepted by the community of the blind people.

According to the AVR concept, the presence of obstacles in the surrounding environment and the path to the desired target will be signalized to the blind subject by burst of sounds, whose virtual source position suggests the position of the real obstacles and the direction of the movement. Actually, the visual reality is substituted by an appropriate acoustic virtual reality.

The rest of the paper is organized as follows. In Section 2, the proposed algorithm is explained and its advantages are shown. Section 3 describes a concrete case where the proposed algorithm is applied to a chosen subject, the obtained result is used to train the neural networks. In Section 4, the measurement scenario is presented and the obtained results are compared to the results obtained in [13].

The last section is devoted to conclusions and future research plans.

2 The proposed algorithm

As we mentioned before in this paper, the purpose of the algorithm is to extract the essential part of HRIR data from the already measured HRIR data of an existing database. To be more precise, for example, applying the algorithm to a person's HRIRs from the Listen Ircam database, where each compensated HRIR has a total length of 512 samples, we are able to obtain HRIR data with a length around 60-150 samples, depending on the samples' values and on the parameters of the algorithm.

The necessity of such an algorithm came to our mind after we succeeded to implement an appropriate AVR on an evaluation board [14]. In order to design and implement such an AVR, we must take the structure, weight and bias values of an already trained ANN (one ANN for each ear as described in Section 4 of this paper or, similarly, in [13]) and store it somehow on the evaluation board. On our board, we had two options:

- the first was to store them in the internal memory of the board, which obviously reduces the execution time when generating the desired acoustic signals using the HRIR calculated locally on the board;
- the second was to store them on a memory card, which is a much slower option, because in this case when generating the desired acoustic signals for headphone listening using the locally calculated HRIRs each weight and bias value must be read from the memory card (instead of the internal memory) which introduces a significant delay in execution, prolonging the whole cycle of acoustic signal generation.

If the dimensions of the two ANNs, one for the left side HRIRs and one for the right side HRIR, are too big, it is possible that one cannot implement them (cannot store the weight and bias values for both ANNs) using the first option and must use the second, but if the HRIRs are shortened enough using our algorithm, it could be possible to use the first option with all of its advantages.

The algorithm has two main parameters: GetInRule and GetOutRule. These two parameters determine how long the filtered HRIRs should be. Please note that the filtered HRIRs are the HRIRs after having applied the algorithm on them. Fig. 1 and Fig. 2 give an example of how an original HRIR (shown in Fig. 1) is filtered by the algorithm. As can be seen, the algorithm basically cuts out the impulse from the HRIR between two particular samples (shown in Fig. 2).

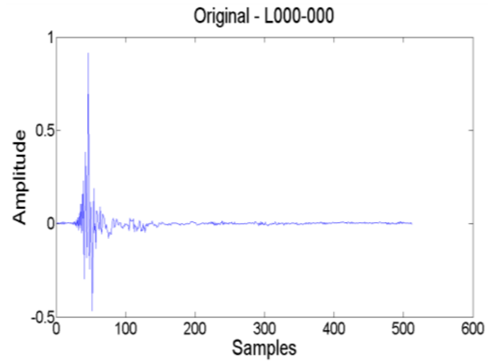


Fig. 1 – Original HRIR data from database.

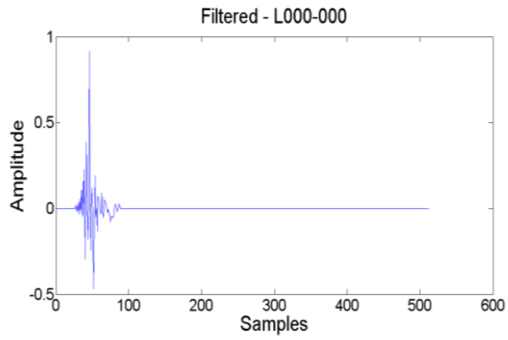


Fig. 2 – Filtered HRIR data, which is the result of our algorithm.

The algorithm is based on calculating the starting point (from where to cut the HRIR) and the ending point (until where to cut the HRIR) of the relevant HRIR data in order to extract it from the existing HRIR data, which most commonly has a length of 512 samples with a sampling rate of 44100 Hz. In order to this, we take as reference three HRIRs:

- the first reference is the HRIR from the left of a subject (azimuth -90 or 270 degrees, elevation 0 degrees), this should give us the starting point for the left side HRIRs (where the sound arrives faster) and the ending point for the right side HRIRs;
- the second reference is the HRIR from the front of a subject (azimuth 0 degrees, elevation 0 degrees), this was considered for error checking;
- the third reference is the HRIR from the right of a subject (azimuth 90 degrees, elevation 0 degrees), this should give us the starting point for the right side HRIRs (where the sound arrives faster) and the ending point for the left side HRIRs.

Please note that all angles are given with respect to the vertical-polar coordinate system, which is the most commonly used coordinate system in the literature.

After obtaining the starting and ending points for the three chosen references for each side (left and

ISBN: 978-1-61804-023-7

right), we determined the starting point for the whole left side by taking the minimum from the three reference starting points and, for the ending point, we took the maximum from the three reference ending points. We applied the same method for the right side. So, once we had which samples to retain for both sides, we filtered all the HRIRs to obtain the relevant HRIR data of interest.

These filtered HRIR were used to train the two ANNs, which then were used in our listening tests to determine the applicability of our algorithm. In the next section, we explain our methodology for subject selection and the structure of the two ANNs.

3 Subject selection and the ANN structure

In our experiments, we used the Listen HRTF Database, which is a public HRIR database, for the training of the ANNs. The database includes measurement data of 49 test subjects. For each subject there are 187 pairs of HRIRs, each of these HRIRs corresponding to a particular point (uniquely defined by a particular azimuth-elevation couple) in the 3D space.

From all these 49 subjects one single subject was chosen based on the anthropometric similarities with our test person. The chosen subject from the Listen HRTF Database was the subject known as IRC_1031. The compared anthropometric measurements were presented in Table 1. We used the same anthropometric measurements like in [15]. The training of the ANNs was conducted using the data set corresponding to the selected subject.

Extent	Measurement	IRC_1031		Subject	
		L	R	L	R
x_1	Head width	153		162	
x_3	Head depth	216		225	
x_{12}	Shoulder width	530		550	
d_1	Cavum concha height	20	20	18	19.3
d_3	Cavum concha width	17	17	15.4	15.4
d_4	Fossa height	22	23	21.2	16.7
d_5	Pinna height	62	66	66.5	64.7
d_6	Pinna width	28	31	26.9	28.8

Table 1 – Selected anthropometric measurements in [15]. Specified values are expressed in millimeters.

Two ANNs are necessary in order to generate a pair of HRIRs (one ANN for each ear). Each of the two ANNs is a simple multilayer perceptron feed-forward backpropagation network. The structure of the ANNs, shown in Fig. 3, is similar with the one described in [16], but will be different for the left and right ear of the subject, opposed to the cases met in [13] and [16]. The network consists of three parts: an input layer of source nodes (2 inputs), a hidden layer (100 neurons) and an output layer (L, respectively R neurons, depending on the ANN used for the left or right side HRIRs). The number of neurons from the hidden layer was determined experimentally and is considered optimal for our application. The two network inputs will be the azimuth and the elevation of the desired virtual sound source. As a result, each of the two ANNs gives us a set of L (left side), resp. R (right side) values corresponding to the HRIRs for the desired virtual sound source.

After applying the proposed algorithm to the chosen subject's 187 pairs of HRIRs, we can set the number of output neurons of the two ANNs, which are trained using this set of HRIRs. These are 146 for the left side ANN (left network structure is 2-100-146), respectively 149 for the right side (right network structure is 2-100-149).

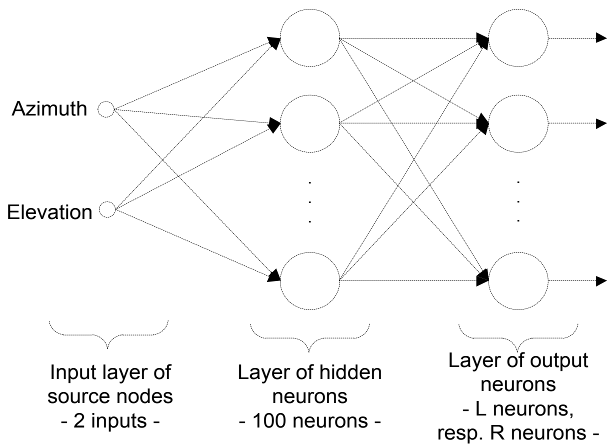


Fig. 3 – The architecture of the proposed ANNs.

As a result of applying our algorithm the length of the HRIRs was reduced to 146 samples for the left side and to 149 samples for the right side from 512 samples (original HRIR sample length in most public HRTF/HRIR databases). This reduction of the number of HRIRs samples can be very useful when implementing the whole AVR concept on an evaluation board, as explained in Section 2.

The purpose of the present paper is to evaluate the performances of the chain composed of the used network structures and the proposed algorithm through listening/localization experiments. The experimental setup, presented in Section 4, proposes to determine if the practically obtained sound really offers the appropriate perception of the desired virtual sound source in the auditory space of an arbitrary subject.

4 Experimental setup and results

The experimental setup used in our experiments is presented in Fig. 4.

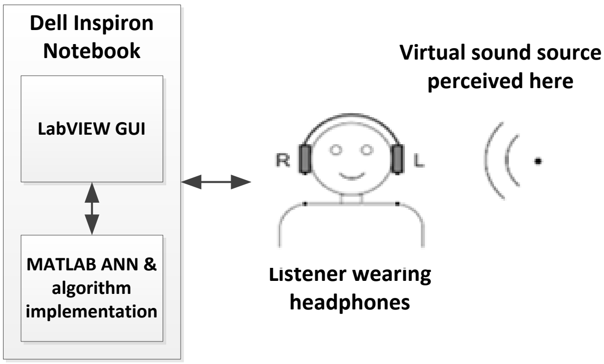


Fig. 4 – The experimental setup.

Our experiments were conducted on a Dell Inspiron 1520 notebook running Windows 7 Ultimate 32-bit operating system with the following hardware configuration: Intel Core 2 Duo 2.00 GHz CPU, 2GB DDR2 RAM, 160 GB HDD. We used Sennheiser HD 448 headphones for headphone listening of the generated or stimulus acoustic signals.

The software environment is an improved version of the graphical user interface (GUI), presented in [13] and was developed using National Instruments' LabVIEW. Fig. 5 shows a snapshot of the running GUI. The basic idea was to design a basic version of the AVR concept, already presented in [13], [16]. One can obtain a simple AVR, which means obtaining the left and right acoustic signals intended for headphone listening, whose virtual sound source is a certain point (specified by an azimuth-elevation couple) in the 3D space. The acoustic signal for each ear is obtained by convoluting a locally generated sound (for example: chirp, pink noise or white noise, from the implemented LabVIEW GUI) with the corresponding HRIR samples according to the desired point, where an obstacle should be signalized. These HRIR samples can be obtained from a public database or can be calculated by the user using its own method. An example of such a method is shortly described in Section 3 and in [13].

In our case the HRIR samples, for a particular point for the left or right ear, are obtained by simulating an already trained ANN. In order to make this possible the LabVIEW GUI had to be interfaced with MATLAB using LabVIEW's MathScript RT Module. Basically a MATLAB window is opened every time the GUI is launched in LabVIEW in order to be able to provide the necessary HRIR samples for convolution by simulating the already trained ANNs for both ears. The inputs values (azimuth-values), as presented in Section 3, are coming from the LabVIEW GUI. The neural network objects are saved on the hard drive in .mat format.

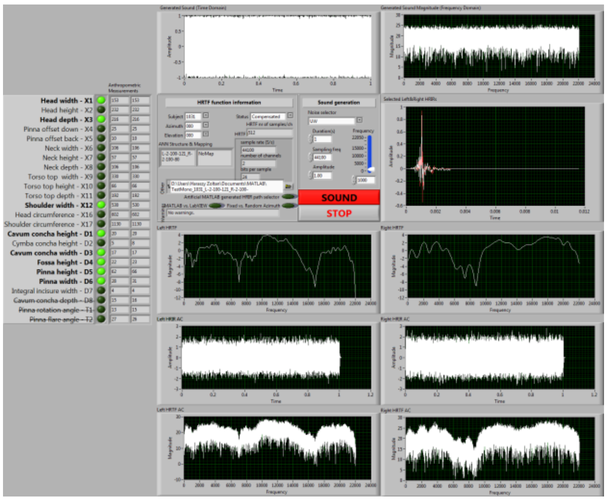


Fig. 5 – A snapshot of the running GUI.

As can be seen in Fig. 4, our proposed algorithm and the neural networks are implemented in MathWorks' MATLAB environment using the Neural Network Toolbox.

The selected transfer functions and network training function, used in our experiments, were the same as in [13]. The number of training epochs used is 100000 for both ANNs and was determined experimentally.

In order to be able to validate the proposed algorithm from Section 2 of the current paper, we used the same measurement scenario as in our previous experiment [13]. The listening tests were conducted in our Bioinspired Systems' Laboratory from the Department of Applied Electronics, in Politehnica University of Timisoara.

A short description follows. The test subject was placed on a rotating chair, with a field compass placed horizontally between its legs for simple angle measurement. A loudspeaker was placed straight ahead of the test subject at the same height with the subject's head. Expressed in azimuth-elevation coordinates the loudspeaker position is at (0˚;0˚). The distance between the loudspeaker and the subject's head was fixed at 1,5 meters. The loudspeaker served as a reference sound source in this localization experiment. The reference signal coming from the loudspeaker was a burst of uniform white noise of 1 s with 0,2 s silence and was repeated throughout the localization experiment.

Acoustic signals were played through the headphones to the subject. The stimulus signals were generated using the GUI from LabVIEW. The

ISBN: 978-1-61804-023-7

subject was told to rotate the chair without rotating its head and modifying the relative height from the ground until the virtual sound source of the acoustic signals (stimulus signals) corresponds with the physically present sound source position (reference signal). An observation is necessary: the subject was not allowed to rotate its head and, also, not allowed to modify the relative height measured from the ground, in order to do all the measurements in the horizontal and frontal plane (where all the measurements took place). When the two sources coincide, the angular displacement (azimuth) is read from the field compass. The chosen acoustic signal was a burst of uniform white noise of 1 s with 0,3 s silence and was played repeatedly until the test subject gave an estimation of the virtual sound source azimuth in its auditory space by remaining in the same position with the rotating chair. The total number of randomly generated stimulus signals was 140, compared to 54 in our previous experiment [13]. In the next phase, the virtual sound source position and the actually used azimuth-elevation pair is compared (by calculating the absolute difference between the used and measured azimuth values) to determine the accuracy of the sound localization, this way the performance of the whole system. In order to better classify the gravity of this error, we divided the 0º-90º interval into 15º intervals, as can be seen in Fig. 7.

The obtained results of the conducted listening experiments are represented in Fig. 6. In frame of this experiment the stimulus signals were only generated in the frontal plane and at 0˚ elevation (coming from the loudspeaker placed straight ahead of the subject). For azimuth, this means angles between – 90˚ and 90˚.

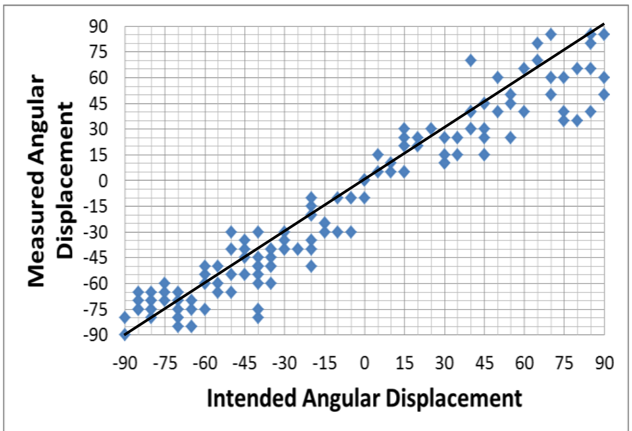


Fig. 6 – Obtained localization results.

Fig. 7 gives us an overview of the obtained absolute errors. One can see that most of the errors that appeared during the measurements were located in the first error interval [0,15) degrees – 64,29%. Little over 90% of the listening test resulted in errors smaller than 30 degrees.

After comparing the results from Fig. 7 with the ones from H. Hu et al. [15], we can state that the localization performances presented in the current paper are better than in [15], if we compare them with the results obtained using non-individual HRTFs. But, when we consider the results obtained using individual HRTFs, the localization using our method is less accurate.

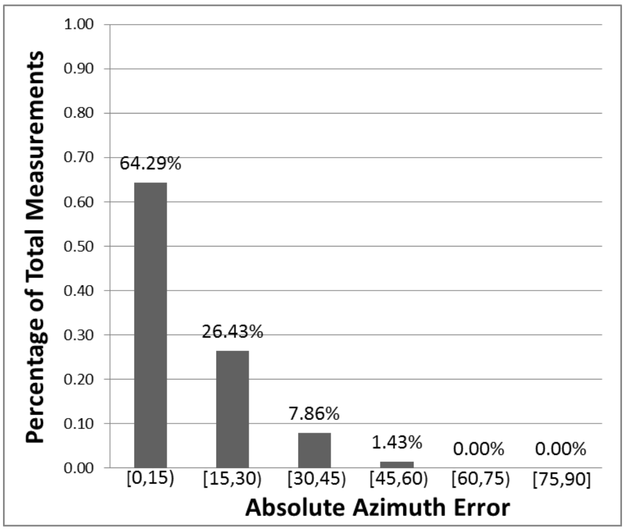


Fig.7 – Absolute azimuth errors classification with respect to the total number of listening test.

Also, after comparing the results from the current paper with the ones from our previous paper [13], one can observe that, first of all the number of conducted listening tests were significantly greater (140 vs. 54). When we look at the listening results from Table 2 in [13] and compare them to the ones in Fig. 7, we can state that the current results are better, which is an interesting observation.

6 Conclusions and future research

Based on the results of the current research (as shown in Fig. 5 and Table 2), it can be stated the proposed algorithm can be applied to extract relevant HRIR data from one person's HRIRs. The proposed algorithm can be applied if there is at least one subject included in the Listen HRTF or any other available database, whose anthropometric measurements are close to the test person's selected anthropometric measurements. The algorithm has its advantages (+) and drawbacks (-):

- the necessary ANN structures are simpler (fewer network outputs after applying the algorithm on a data set); (+)
- the time necessary for ANN training is shortened; (+)

ISBN: 978-1-61804-023-7

- when implementing the whole AVR chain on an evaluation board, the network structure is the main factor, when it comes to decide where to store (for example, internal memory or external memory card) the weight and bias values. This point of view can be translated into slower or much faster execution time for each acoustic signal generation cycle; (+)
- it is possible to lose precision in localizing the obstacles, which are signalized using the AVR concept, although results from Section 5 do not confirm this statement. (-).

As future research plans, the authors would like to mention the optimization of the presented solution. Also, the authors intend to create their own HRIR database for testing purposes.

Acknowledgements

This work was partially supported by the strategic grant POSDRU 6/1.5/S/13, Project ID6998 (2008), co-financed by the European Social Fund – Investing in People, within the Sectorial Operational Program Human Resources Development 2007-13.
This work was partially supported by CNCSIS – UEFISCDI PNII – IDEI Grant Contract No. 599/19.01.2009.

References:

[1] A. Helal, S, Moore, B. Ramachandran-Drishti, An Integrated Navigation System for Visually Impaired and Disabled, *International Symposium on Wearable Computers (ISWC),* 2001, pp. 149-156.

[2] V. Kulyukin, C. Gharpure, J. Nicholson, S. Pavithran, RFID in Robot-Assisted Indoor Navigation for the Visual Impaired, *IEEE/RSJ Intern. Conference on Intelligent Robots and Systems,* Sendai, Japan (IROS), 2004.

[3] I. Ulrich, J. Borenstein, The GuideCane – Applying Mobile Robot Technologies to Assist Visually Impaired, *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans,* Vol. 31, No. 2, 2001, pp. 131-136.

[4] S. Soval, I. Ulrich, J. Borenstein, Robotics-based Obstacle Avoidance Systems for Blind and Visually Impaired, *IEEE Robotics Magazine,* Vol. 10, No. 1, 2003, pp. 9-20.

[5] H. Shim, J. Lee, E. Lee, A Study on the Sound-Imaging Algorithm of Obstacles Information for the Visually Impaired, *The 2002 Intern. Conf. on Circuits/Systems, Computers and Communications (ITC-CSCC)*, 2002, pp. 29-31.

[6] V. Tiponut, A. Gacsadi, L. Tepelea, C. Lar, I. Gavrilut, Integrated Environment for Assisted Movement of Visually Impaired, *Proceedings of the 15th International Workshop on Robotics in Alpe-Adria-Danube Region, (RAAD 2006),* ISBN: 9637154-48-5, 2006, pp. 234-239.

[7] V. Tiponut, S. Ionel, C. Caleanu, I. Lie, Improved Version of an Integrated Environment for Assisted Movement of Visually Impaired, *Proceedings of the 11th WSEAS International Conference on SYSTEMS*, ISBN: 978-960-8457-90-4, ISSN: 1790-5117, 2007, pp. 87-91.

[8] R. Z. Shi, T. K. Horiuchi, A Neuromorphic VLSI Model of Bat Interaural Level Difference Processing for Azimuthal Echolocation, *IEEE Trans. Circuits and Systems,* Vol. 54, 2007, pp. 74-88.

[9] J. Reijniers, H. Peremans, Biometric Sonar System Performing Spectrum-based Localization, *IEEE Trans. on Robotics,* Vol. 23, No. 6, 2007, pp. 11511159.

[10] N. Bourbakis, Sensing Surrounding 3-D Space for Navigation of the Blind, *IEEE Engineering in Medicine and Biology Magazine,* Vol. 27, No. 1, 2008, pp. 49-55.

[11] V. Tiponut, Z. Haraszy, D. Ianchis, I. Lie, Acoustic Virtual Reality Performing Man-machine Interfacing of the Blind, *Proceedings of 12th WSEAS International Conference on SYSTEMS*, ISBN: 978-960-6766-83-1, ISSN: 1790-2769, 2008, pp. 345-349.

[12] J. Blauert, *Spatial Hearing – The Psychophysics of Human Sound Localization, Revised Edition*, The MIT Press, 1997.

[13] Z. Haraszy, D.-G. Cristea, V. Tiponut, T. Slavici. Improved Head Related Transfer Function Generation and Testing for Acoustic Virtual Reality Development, *14th WSEAS International Conference on SYSTEMS*, ISBN: 978-960-474-199-1, ISSN: 1792-4235, 2010.

[14] Z. Haraszy, Hardware and Software Implementation of a simple Acoustic Virtual Reality, Technical Report, 2011.

[15] H. Hu, L. Zhou, H. Ma, Z. Wu, HRTF personalization based on artificial neural network in individual virtual auditory space, *Applied Acoustics*, 2008; 69(2), pp.163–172.

[16] Z. Haraszy, D. Ianchis, V. Tiponut, Generation of the Head Related Transfer Functions Using Artificial Neural Networks, *13th WSEAS International Conference on CIRCUITS*, ISBN: 978-960-474-096-3, ISSN: 1790-5117, 2009.