3D Sound applied to the design of Assisted Navigation Devices for the Visually Impaired

José Francisco Lucio Naranjo, Roberto Tenenbaum, Henry Patricio Paz Arias, Luis Alberto Morales Escobar and Carlos Efraín Iñiguez Jarrín

***Abstract*— This work presents an approach to generate 3D sound by using a set of artificial neural networks (ANNs). The proposed method is capable to reconstruct the Head Related Impulse Responses (HRIRs) by means of spatial interpolation. In order to cover the whole reception auditory space, without increasing the network complexity, a structure of multiple networks (set), each one modeling a specific area was adopted. The three main factors that influence the model accuracy --- the network's architecture, the reception area's aperture angles and the HRIR's time shifts --- are investigated and an optimal setup is presented. The computational effort to process the ANN is shown to be slightly smaller than traditional interpolation methods and all error calculation reached very low levels, validating the method to be used in the design of a 3D sound emitter capable of provide navigation aid for the visually impaired. Two approaches are presented in order to detect obstacles, one which makes use of computational vision techniques and other with laser proximity sensors.**

***Index Terms*—Acoustical Virtual Reality, Auralization, Artificial Neural Networks, HRIR Interpolation, ETA devices.**

***Resumen*— Este trabajo presenta un abordaje para generar sonido 3D utilizando un conjunto de redes neuronales artificiales (RNAs). El método propuesto es capaz de reconstruir la Respuestas Impulsivas Asociadas a Cabeza Humana (HRIRs) mediante interpolación espacial. Con el fin de cubrir todo el espacio de recepción auditivo, sin aumentar la complejidad de la red, fue adoptada una estructura de múltiples redes (conjunto), cada una modelando un área específica. Los tres factores principales que influyen en la exactitud del modelo --- la arquitectura de la red, ángulos de apertura de la zona de recepción y los cambios de tiempo del HRIR --- son investigados y es presentada una configuración óptima. El esfuerzo computacional necesario para procesar la RNA muestra ser menor que métodos tradicionales de interpolación y todos los cálculos de error alcanzan niveles muy bajos, validando el método para ser utilizado en el diseño de un emisor de sonido 3D capaz de proporcionar asistencia en la navegación de discapacitados visuales. Dos enfoques se presentan con el fin de detectar obstáculos, uno que hace uso de técnicas de visión computacional y otro con sensores de proximidad de láser.**

***Palabras clave*— Realidad Virtual Acústica, Aurilización, Redes Neuronales Artificiales, Interpolación de HRIR, Dispositivos ETA.**

I. INTRODUCTION

It is well known that navigation could represent a dangerous activity for the visually impaired. Nonetheless, the recent development of ETA devices (Electronic Travel Aid) provides means to detect obstacles (position, distance or even size). In such a case, the human hearing attribute can be one of the best ways to communicate a save path to walk avoiding obstacles. In this sense, virtual acoustic reality is the most powerful tool to consider for generating a 3D sound emitter.

Nowadays, acoustical simulation encompasses not only the assessment of acoustical parameters, such as levels and reverberation times, but accounts with a powerful tool: The auralization. It consists in generating the sound heard by a subject when immersed into a simulated environment, which can be a simple room, a somewhat complex theater, an industrial plant or even an urban space. This sound must be reproduced to be heard by a human being, in an environment free from sound reflections, i.e, in an anechoic chamber with a cross-talk cancellation system --- which is unavailable for most users --- or through an equalized headphone set.

The sound that arrives at the ears entrance is altered due to diffraction and absorption, among other phenomena [1], which is highly dependent on the wavefront incidence direction. The human head and torso constitute natural acoustic filters from every source position to the ears entrance. Such filters can be modeled as Finite Impulse Response (FIR) systems, known as Head-Related Impulse Responses (HRIRs). They are responsible to confer the 3D sound sensation in virtual

José F. Lucio Naranjo, Departamento de Informática y Ciencias de la Computación, Facultad de Ingeniería de Sistemas, Escuela Politécnica Nacional, Quito, Ecuador; (email: jose.lucio@epn.edu.ec)

Roberto A. Tenenbaum, Laboratório de Instrumentación en Dinâmica, Acústica y Vibraciones – LIDAV, Departamento de Modelagem Computacional, Instituto Politécnico, Universidade del Estado del Rio de Janeiro, Nova Friburgo, Brasil; (email: ratenenbaum@gmail.com)

Luis Morales is currently acting as titular professor at the Escuela Politécnica Nacional of Ecuador, and he is manager of Electronic Instrumentation Laboratory of the Faculty of Electrical and Electronic Engineering from the EPN. (email: luis.moralesec@epn.edu.ec)

Ing. Paz is currently acting as computer science titular professor at the National Polytechnic School (EPN) of Ecuador. (email: henry.paz@epn.edu.ec)

MSc. Carlos Iñiguez is acting as titular professor at Escuela Politécnica Nacional of Ecuador (EPN), and he is currently leading the software development for the research project ¨Simulation of Acoustic Wave Propagation in closed areas¨, at EPN. (email: carlos.iniguez@epn.edu.ec)

environments and to provide directional cues for human ability to distinguish the sound source locations [2].

Considering that auralization is a virtual reality process and that HRIR modeling is an important part of such systems, computational load is always a subject of concern. The most common solution consists in reducing the number of mathematical operations and/or simplifying the models involved in the auralization process. Nevertheless, such approach may cause a significant reduction of the 3D sound sensation and therefore compromise the receiver ability to recognize the sound source's direction.

Although, on real or simulated environments, the sound wavefront may come from anywhere and, even in the most complete HRIR databases, these functions are measured for discrete spatial locations. Such spatial discreteness leads to undesired audible effects, mainly when fast sound source movements are reproduced, producing 'clicks' or subtle changes [3]. In this case, two solutions are available: To round off the sound direction to the closest measured one; or to interpolate the closest functions to achieve a better estimative of such direction characteristics. Several techniques have been presented for HRIR interpolation [4, 5, 6, 7] while other were developed for continuous HRIR [8, 9, 10].

In a previous work, a method based on Artificial Neural Networks (ANNs) was introduced for interpolating HRIRs [11]. In that model an ANN committee is responsible for several reception areas around the listener, where each network models the functions (HRIRs) inside those areas.

The ANN main task is to synthesize an interpolated HRIR, receiving as input a vector with the direction of the desired sound source. This is accomplished thanks to a training procedure, in which several target functions are generated by using the bilinear interpolation technique (as shown in Fig. 1) and repeatedly presented to the network for learning purposes. In this approach, the interpolated (target) function is given by weighted sum of four known HRIR (measured ones):

$$\widehat{\mathrm{HRIR}}(n) = \sum_{i=1}^{4} \alpha_i \mathrm{HRIR}_i(n), \qquad (1)$$

where α_i are the weights for the measured $\mathrm{HRIR}_i(n)$ [12].

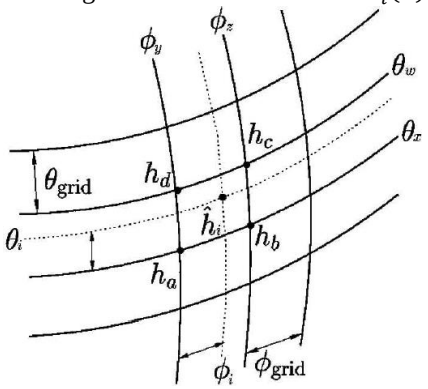


Figure 1: Bilinear interpolation scheme for a given direction coordinate *i*.

It is worth mention that the auralization procedure also involves a convolution product between an interpolated HRIR (corresponding to the sound source position) and an arbitrary anechoic signal. Nevertheless, this interpolation task constitutes an optimization that will facilitate the implementation of a 3D sound emitter, a constitutive part of the electronic travel aid (ETA) device for the visually impaired that is currently being developed.

In this work, the procedure to determine the optimal architecture and the reception area size, is presented. First, the model is presented, then the different steps to setup and to train the ANN to achieve the HRIRs interpolation capability are described. Then, the model performance is evaluated according to different error criteria. Finally, two approaches are presented in order to provide means to identify obstacles in which virtual sound sources will be simulated.

II. MODEL DESCRIPTION

An ANN is consisted by a set of simple processing elements, called artificial neurons, which have mutual influence behavior via a network of excitatory or inhibitory weights [13]. In order to define the optimal values of these synaptical weights (randomly initiated at first), a common supervised learning procedure is used. In such training procedure, which is called *backpropagation* [14], the ANN's weights are adjusted proportionally to their contribution to the observed error between the network output and the desired outputs (targets).

In this case, a *feed-forward multi-layer* ANN [15] was chosen to implement the HRIR interpolator system, due to the fact that the multi-layer feature provides means to optimize the ANN response by adjusting the network architecture (number of layers and number of neurons in each layer).

The network architecture, their number of inputs, outputs and internal layers depend on the data nature of excitation and output. Working as an interpolator, the input is the sound source direction, represented in a spherical coordinate system by two angular variables: The azimuth (ϕ) and the elevation (θ) angles. The desired outputs are the HRIR coefficients (time samples), whose targets are obtained from bilinear interpolation technique, as shown in Fig. 1.

The number of samples originally defined by the measurements conducted by Gardner and Martin [16] was 512 at a sampling rate of 44.1 kHz. However, most part of these samples tends to zero, due to the natural decay of the HRIR functions. Therefore, the most significant information is contained, approximately, in the first 100 samples [17]. Therefore, the number of samples chosen for the output functions was 128, since it facilitates any post signal processing with "power of two"' Fast Fourier Transform (FFT) algorithm.

The error used for synaptical weight optimization during the ANN's training is the mean squared error (MSE) between the target functions and the network output. The network architecture has an input vector with 2 elements (azimuth and elevation), one hidden layer with *L* neurons and the output layer with 128 neurons, as shown in Fig. 2.

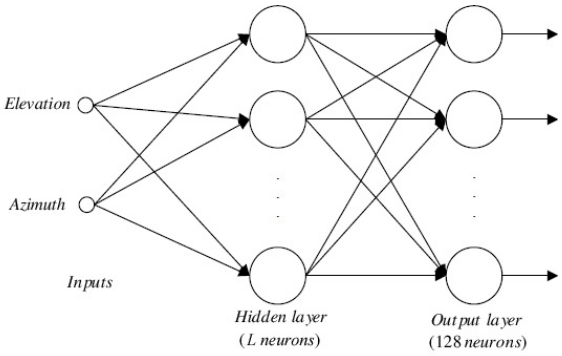


Figure 2: Network architecture for HRIR interpolation.

In order to improve the system performance, the listener surrounding space is split in several areas, as shown in Fig. 3. The subdivision area criteria is based on the tradeoff between network complexity and modeling error. Each region encapsulates several HRIRs, whose similarity depends on the aperture angle, i.e., wider regions present less correlated functions, especially for incidence direction where the sound suffers higher diffraction effects.

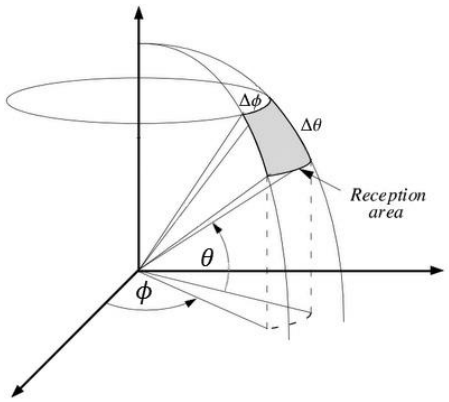


Figure 3: Network operation area limited in azimuth by ϕ and $\phi+\Delta\phi$ and in elevation by θ and $\theta+\Delta\theta$.

A similar research [18] uses just one network to cover a wide reception area. Although a complex architecture was used, the errors still were not small enough for an auralization processing. Therefore, it was assumed that smaller areas will facilitate the training procedure (producing smaller errors) and guaranteeing simple network architectures (single hidden layer with a few neurons for fast processing speed) due to the similarity of the functions involved. The numerical results will show if this assumption is correct.

Since each target function comes from an interpolation procedure whose input can be any arbitrary direction, it is possible to generate as many input/target pairs as needed. In order to cover the entire reception area, a grid distribution was applied to establish the position of the training parameters. The positions of the validation parameters were randomly chosen. All this taking care not to produce a pair where exist a measured HRIR, as shown in Fig. 4. The measured functions will be used later in order to test the accuracy of the RNA model.

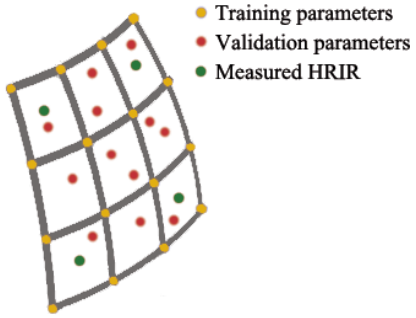


Figure 4: Input/target pairs distribution inside the reception area.

III. CRITERIA FOR PERFORMANCE EVALUATION

The model performance, including training error and target-output (mis)matching, is governed mainly by the following factors:

a. **Number of neurons in the hidden layer**: It defines the specificity in which a network will respond to training. An over-dimensioned number of neurons, besides demanding more computational effort to be processed, may lead the network to learn unimportant details of the training parameters, losing its generalization capacity for new parameters (overfitting). On the other hand, too few neurons in the hidden layer may decrease the network capacity to learn and, therefore, lose the output resolution.

b. **Reception area aperture**: When working with areas, the similarity of the HRIR inside such areas may contribute or degradate the network performance, according to the spectral characteristics of HRTFs, i.e., magnitude and phase, since the training is done in the time domain. Inside small areas there are only a few measured functions, which cause high similarity in the target functions, since they are derived from interpolation of very close locations. Besides, closer functions (HRIRs) present very small delays and highly correlated spectrum. On the other hand, wider areas, such as 90°, 180° or even 360° of azimuth, will include targets with very low similarity, both in spectrum and delays. Such differences require higher network complexity to deal with such variety of data. Therefore, the investigation of the aperture areas is a key element for the success of the proposed model. However, such similarity is not uniform around the listener space. It depends on the sound source location and may lead to a non-uniform space discretization.

c. **Time shifts**: The HRIRs present initial delays that

depend on the position of sound source relative to the head (ears). These time-shifts are caused by the distance between the sound source and the ears entrance. The presence of such delays may increase the network performance if the functions were highly correlated in terms of spectral characteristics (magnitude and/or phase). This occurs generally for smaller areas. However, for wider areas, these delays may decrease the network ability to learn patterns, since there are high spectral fluctuation between input samples and the time-shift are also larger due to the possibility of sources located far from each other. Therefore, for such situation, by removing these initial time delay, a more homogeneous set of data may aid the training process.

Therefore, it was observed that the influence of combined effects of the above parameters requires a more detailed analysis. As stated before, the MSE (training error) is not suitable for correctly evaluating the network performance under the human hearing point of view. In this section, one error criteria is presented in order to evaluate the accuracy of the proposed interpolation system.

A. Mean Magnitude Absolute Error (MAE)

The Mean Magnitude Absolute Error is a difference measure between the magnitudes from target and output functions. It is a scalar computed for a given direction. Equation (2) presents the average of the absolute differences between magnitudes.

$$\mathrm{MAE}(\theta,\phi) = 10\log\left(\frac{1}{\Omega}\sum_{\omega=1}^{\Omega}\left|\mathrm{M}_{\theta,\phi}(\omega) - \widehat{\mathrm{M}}_{\theta,\phi}(\omega)\right|\right), \quad (2)$$

where $\mathrm{M}_{\theta,\phi}(\omega)$ is the magnitude of the frequency response $\mathrm{H}_{\theta,\phi}(\omega)$ of the target function, given by

$$\mathrm{M}_{\theta,\phi}(\omega) = \left|\mathrm{H}_{\theta,\phi}(\omega)\right|, \quad (3)$$

while $\widehat{\mathrm{M}}_{\theta,\phi}(\omega)$ is the same for the output function for a given direction (θ,ϕ) and Ω is the number of discrete frequency bins used in the FFT.

IV. MODEL PERFORMANCE ANALYSIS

In order to determine the system behavior as a function of the mentioned parameters, several networks were trained for different configurations. Networks were trained from aperture angles varying from $5° \times 5°$ ($\Delta\phi \times \Delta\theta$, azimuth and elevation) to $40° \times 40°$, covering the whole auditory listener space. For each reception area's size (a network's operation area), the train parameters distribution density was kept constant.

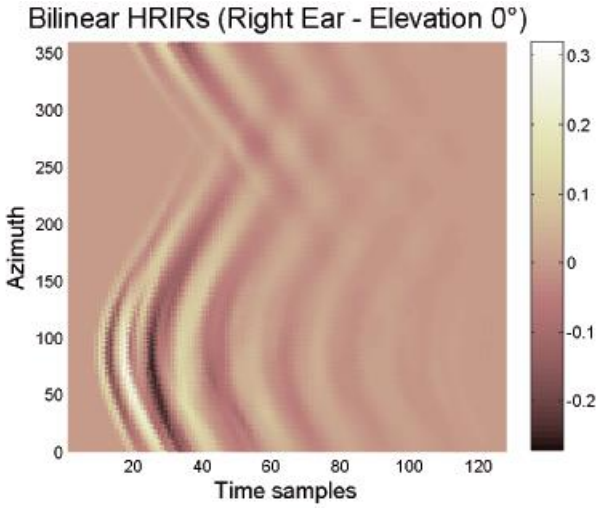
Also, for every trained area, two situations were considered: HRIR functions with time-shifts (original measured delays) and functions whose delays were removed. In such case, the delay were removed according to the assumption of a spherical shape head, where the distance $\mathrm{D}_{\theta,\phi}$ traveled by the sound wave were calculated considering the duplex theory [19], stated below.

$$\mathrm{D}_{\theta,\phi} = 2r\cos^{-1}(\cos(2\pi\theta)\cos(2\pi|\phi - a|)), \quad (4)$$

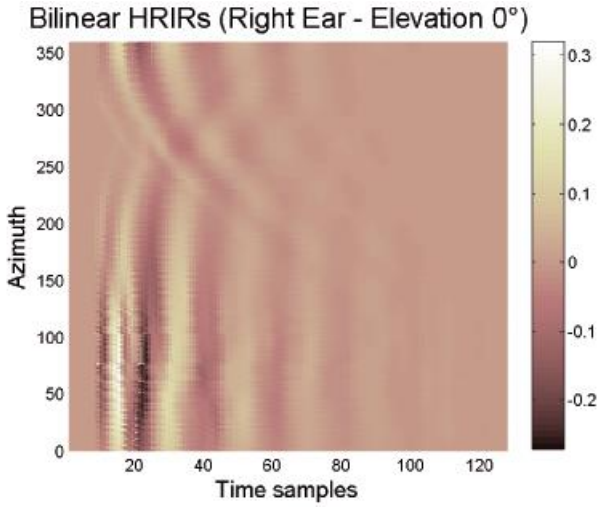
where θ and ϕ are the elevation and azimuth of the considered HRIR, r is the radius of the sphere and a is the azimuth position of the considered ear (90° for the right ear and 270° for the left one).

A comparison between the original HRIR and its time-shift removed version can be seen in Fig. 5 for the horizontal plane, covering all azimuth angles.

From Fig. 5 one may observe that the initial energy is concentrated at approximately the same time, except for contralateral sound source locations (about 270°), where there is not a well-defined starting energy point, due to the hear barrier and torso diffraction effects.



(a)



(b)

Figure 5: Comparison of delay removal for a fixed elevation in $\theta = 0°$ and azimuth variations of 1° from 0° to 360°: (a) Original HRIRs (b) Delayless HRIR versions.

A. Overall averaged error

As mentioned, the MSE is not suitable for evaluating the interpolation quality. The HRIR energy varies according to the sound source location, which may present lower errors for functions with lower energy levels. In this case, the mismatch between target and output can be large, but the MSE may be smaller than when those functions have higher energy levels and smaller differences.

The performance of each network (N_p) can be measured by averaging the MAE for representative directions inside the operating area.

$$N_p = \frac{1}{S}\sum_{s=1}^{S} \mathrm{MAE}_s, \qquad (4)$$

where S is the total number of considered directions, s is a sequential index which represents an individual direction. The considered functions for this calculation had fixed an elevation (0°) and azimuth steps of one degree ($\Delta\phi = 1°$), i.e. the HRIRs computed in the horizontal plane in which the human resolution is more accurate. With this, the optimum number of neurons in the hidden layer can be determined by increasing that number from 2 to 10, for both cases: with and without delays. The results of simulations are shown in Fig. 6, where the overall average errors MAE is presented and compared for several aperture angles and targets.

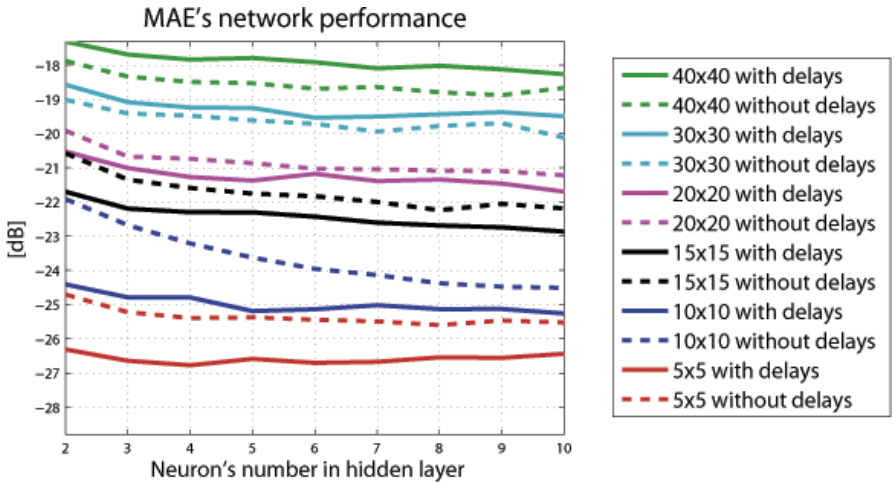


Figure 6: MAE overall average results for different reception areas

Figure 6 presents the MAE for two situations: with original initial delays and when delays were removed. Solid lines represent simulation with original delays. From this figure, one observes that for both metrics the lower errors were achieved by networks with smaller areas. Increasing the areas the error also increases. The smallest error is obtained for targets with original delays. When delays were removed from targets of the networks with the smallest areas (dashed lines) the error becomes 2-3 dB higher. On the other hand, as the aperture size increase, the difference between the errors with and without delays decreases. The approximate size when the errors are similar occurs for networks between $20° \times 20°$ and $30° \times 30°$.

From this error behavior, one may conclude that if the training is performed with smaller areas, the original delays should be preserved, since they contribute with slight time differences in highly correlated targets. For larger areas these delays does not contribute to the learning process. The functions already present variations and lower correlation than when small areas are used. Therefore, by removing such delays, the error for larger areas decreases.

Another conclusion that can be extracted from Fig. 6 is related to the number of neurons in the hidden layer. It is observed that the best performance is found when L=4 is used,

for a 5° × 5° working area and preserving the original delays of the HRIRs. The overall behavior of networks with 5° × 5° with original delays is almost flat and presented variation smaller than 1 dB. Therefore, for such small areas the number of neurons in the hidden layer has a very slight influence in the global behavior.

Errors at higher frequencies may increase objective error metrics, but will not present significant disturbance in the perception of the sound source location [2]. The same occurs for very low frequencies, as stated by Fletcher and Munson in the Equal-loudness contours curves [20] and the ISO 226 standard [21].

V. NUMERICAL RESULTS

In order to present few examples of the proposed interpolation scheme, focusing on specific directions instead of averages of several functions and errors, four measured HRIRs were chosen. From Fig. 7 to Fig. 10, it is presented in the five plots:

a) time domain comparison between output and target;
b) magnitude and (d) phase frequency response comparison;
c) absolute errors in time (upper plot) and magnitude (lower plot).

Such directions – elevation $\theta = 0°$ and azimuths 0°, 90°, 180° and 270° – were chosen due to the their strong variation in spectra, energy and initial delays.

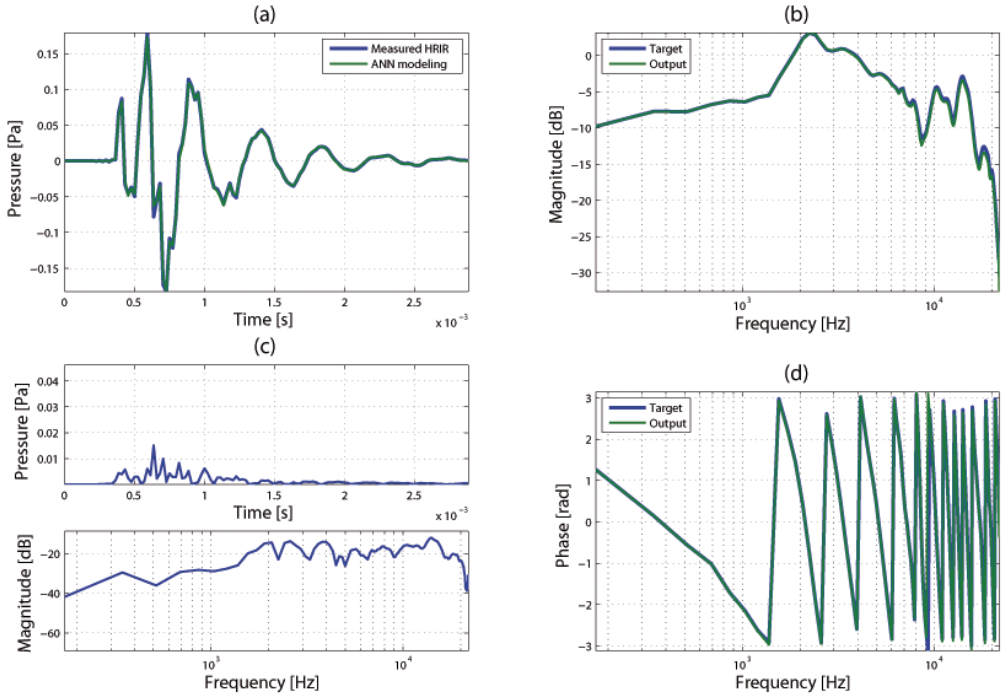


Figure 7: Comparative results between ANN modeling and HRIR measurement: (a) Time domain. (b) Magnitude in the frequency domain. (c) Absolute error in time and frequency domains. (d) Phase in the frequency domain phase

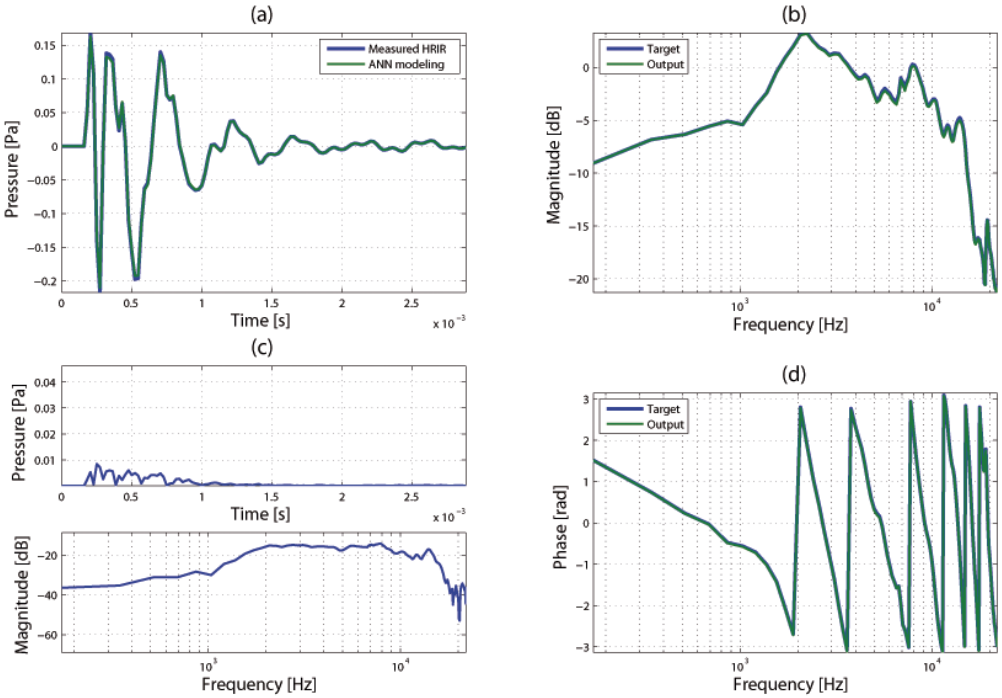


Figure 8: Comparative results between ANN modeling and HRIR measurement: (a) Time domain. (b) Magnitude in the frequency domain. (c) Absolute error in time and frequency domains. (d) Phase in the frequency domain phase

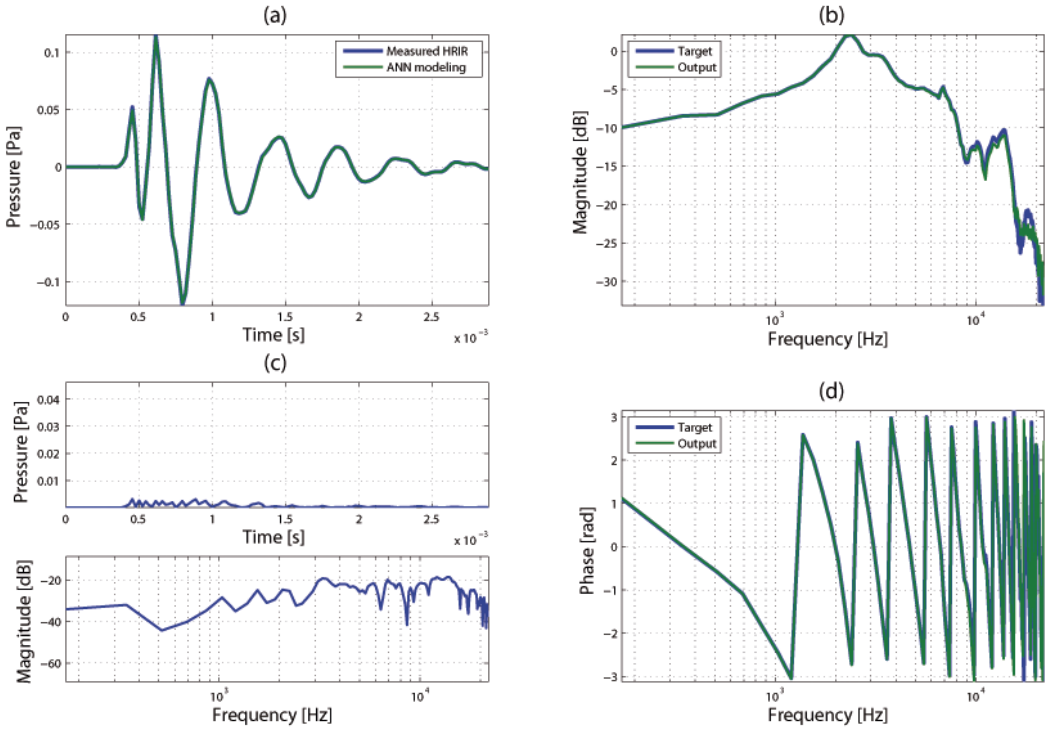


Figure 9: Comparative results between ANN modeling and HRIR measurement: (a) Time domain. (b) Magnitude in the frequency domain. (c) Absolute error in time and frequency domains. (d) Phase in the frequency domain phase

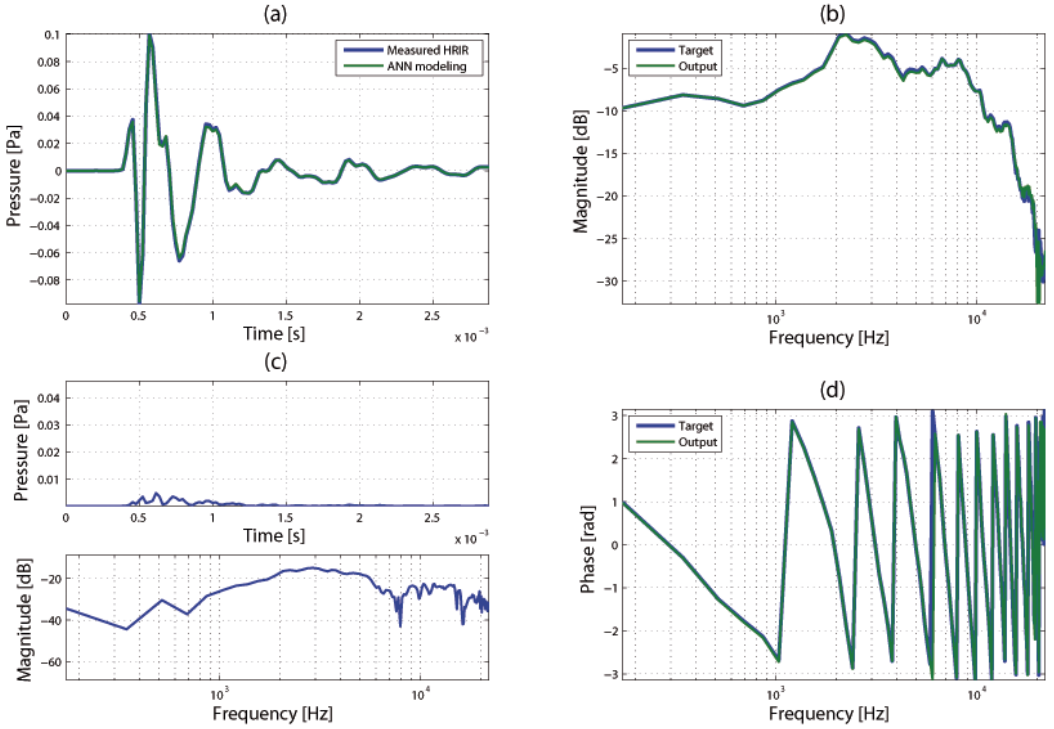


Figure 10: Comparative results between ANN modeling and HRIR measurement: (a) Time domain. (b) Magnitude in the frequency domain. (c) Absolute error in time and frequency domains. (d) Phase in the frequency domain phase

A general comparison between the bilinear technique and the ANN model, for fixed elevation in $\theta = 0°$ and azimuth variations of 1°, are presented in the time domain in Fig. 11 and in the frequency domain (magnitude) in Fig. 12. No noticeable differences are observed.

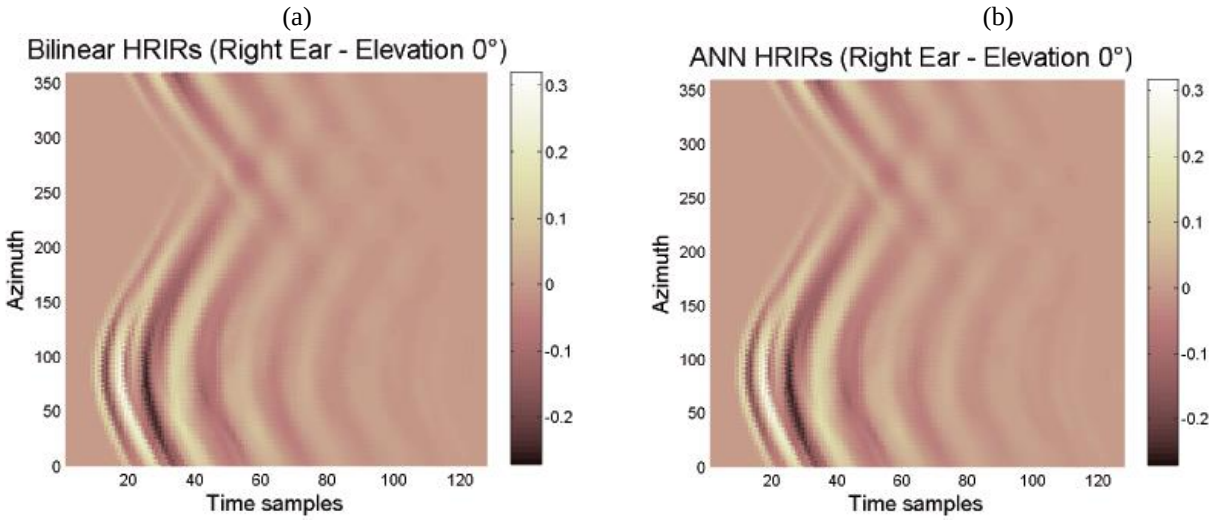


Figure 11: Time domain interpolation results with time shifts for fixed elevation in $\theta = 0°$ and azimuth variations of 1°: (a) Bilinear (b) ANN.

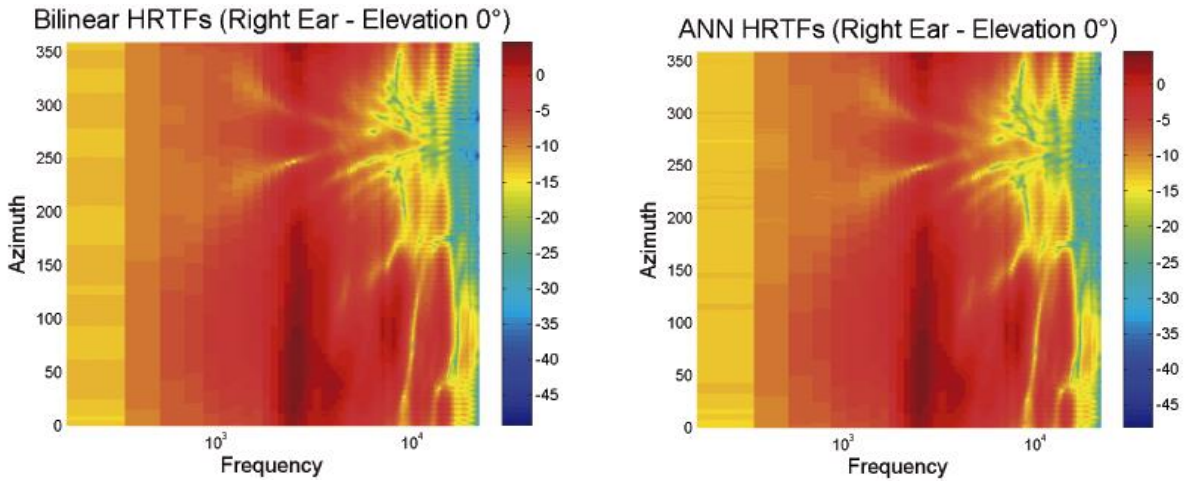


Figure 12: Frequency domain interpolation results with time shifts for fixed elevation in θ = 0° and azimuth variations of 1°: (a) Bilinear (b) ANN.

VI. COMPUTATIONAL COST

This section presents a comparison between the implementation costs of the typical bilinear interpolation method (BIM) and the ANN interpolation scheme presented in this work. The established parameter for such comparison is the required number of elementary arithmetic operations for each method. The ANN execution does not consider the training computational load, since there is no synaptical weights actualization during this phase. Therefore, the number of arithmetic operations A_0 is given by

$$A_0 = 2\left[e \cdot n_1 + \left(\sum_{j=1}^{m-1} n_j \cdot n_{j+1}\right) + n_m \cdot s\right], \qquad (5)$$

where e is the input vector size, n_j is number of neurons in the jth intermediate layer, m is the number of intermediate layers and s is the number of neurons in the output layer.

The results presented in the previous section came from an ANN set with two elements input vector, one intermediate layer with 2 neurons and a 128 neurons output layer. The computational load for the BIM will cost $4 + (6 \times L)$ multiplications and $6 + (3 \times L)$ additions. For L=128, the values indicated in the left column of Table 1 were obtained while the figures corresponding to the ANN, computed in Eq. 5 with $s = 128$, m = 1, $e = 2$ and three differents n_m (2, 3, 4) are presented in the third, fourth and fifth columns of Table 1. By comparing these results it can be stated that the artificial neural network model presents a computational complexity reduction of almost 49.83%.

VII. APPROACHES FOR OBSTACLE DETECTION

Once the technique for interpolate HRIRs had proven its accuracy and computational convenience, the next step is to generate a 3D sound emitter. This can be achieved by implementing a convolution product between the ANN's result and an arbitrary anechoic sound. This procedure can be found in the broadly majority of signal processing signals books.

Nevertheless, in order to have means to apply this solution, the ANNs inside the 3D sound emitter must be feed with spherical coordinates that pin point the obstacle position. Therefore, a device with obstacle detection capabilities must be provided. Two approaches will be developed for this end.

A. Computational Vision for obstacle detection and proximity

Artificial vision is one of artificial intelligence techniques applied for many years for the detection of different types of objects. Therefore, it is proposed to develop an artificial vision system for the detection and classification of obstacles. In this sense, several techniques could be applied. For instance, Stereo Vision [22] allows the identification of objects and proximity in three-dimensional space. This information can be transformed in distance and position of the detected obstacles. In a second phase, an ANN could be used for classification and interpretation of relevant obstacles characteristics. Figure 13 shows how obstacles can be represented in real time through computer vision.

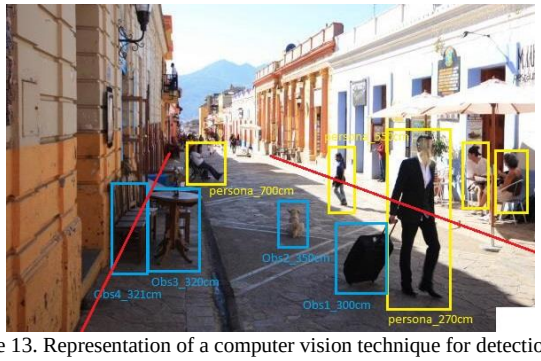


Figure 13. Representation of a computer vision technique for detection and classification of objects.

TABLE 1
COMPARISON OF THE COMPUTATIONAL COST BETWEEN BIM AND ANN INTERPOLATION OF FUNCTION WITH L = 128. THREE ARCHITECTURES ARE CONSIDERED.

Detail	BIM	n_1=2	n_1=3	n_1=4
Number of additions (+)	390	260	390	520
Number of multiplication (×)	772	260	390	520
Computational gain (+)	-	33.33%	0.00%	-33.33%
Computational gain (×)	-	66.32%	49.48%	32.64%
Mean computational gain	-	49.83%	24.74%	-0.35%

The phases for obstacle detection and classification are described in Fig. 14.

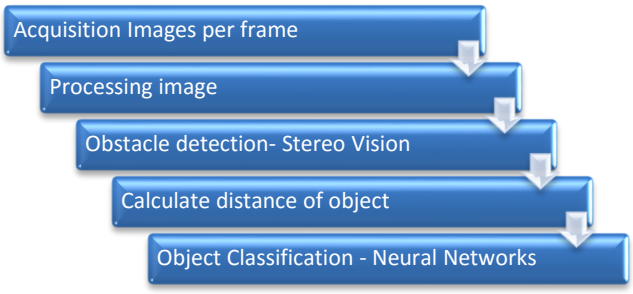


Figure 14. Phases for detection and classification of objects.

B. *Laser scanning for obstacle detection and proximity measurement*

Modern electronic sensor systems for obstacle detection are widely used in robotics applications and assistance to people.

One of the most effective approaches to apply this relies on laser proximity sensors. Such approach exhibit excellent benefits for obstacle detection and proximity measurement without contact or friction with the objects in question. Distance measurement by these devices is done in the following ways:

- Flight time [23]
- Phase shift

Based on the aforementioned methods it is essential that the sensor has the ability to scan the scene taking *n* samples as shown in Fig. 15.

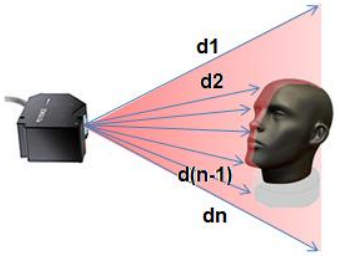


Figure 15. Laser Scanning [24]

Depending on the opening angle of the laser sensor it is possible to detect a point cloud to identify the presence of objects. Adding a sensor that has the ability to perform a vertical scan can have more information and a more detailed point cloud.

The point cloud provides a wealth of information to the user, allowing to pin point the object's position and distance relative to user that acts as a reference frame.

The steps to identify the obstacles through the use of this type of sensor are shown in Fig. 16.

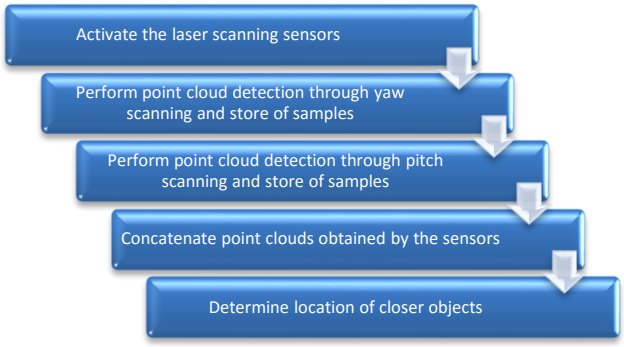


Figure 16. Phases for laser scanning for obstacle detection and proximity measurement

VIII. CONCLUSIONS

The main goal of this study was to present an interpolation procedure for HRIRs by using a committee of artificial neural networks. The Mean Absolute Error (MAE) results show that the obtained accuracy refrains the use of reception areas with smaller angle apertures. The time shifts, although more subtle in smaller reception areas than those found in bigger ones, still constitutes a learning pattern for networks with the adequate capacity to learn from it. Such capacity is defined by the number of neurons in the hidden layer.

The model performance presents larger deviations in high frequency components which are faded away because of the generalization property in the ANN modeling. Nevertheless, these deviations can be neglected due to the low energy present in such frequencies and also to the limited characteristics of the human hearing in high frequencies.

In summary, the size of the reception area is a critical element to be considered for ensuring a high precision interpolator. A classical interpolation technique, such as a bilinear one, could be substituted by an ANN whose output presented very small errors if compared with the corresponding target functions.

Preliminary comparisons performed in time and frequency domains with actual measured functions, show that an ANN committee of reduced architecture (with 1 hidden layer with at most 2-4 neurons), working inside $5° \times 5°$ reception areas and trained with functions with original delays, is able to substitute an interpolation method with a computational reduction of almost 50% while keeping a similar precision.

Of course, by using such small reception areas, the system must deal with a considerable increment of the number of networks to be used and involves a larger spent in memory resources. Nevertheless, current computers have enough memory to work with such memory requirements without compromising the result's generation speed.

The next step of this research is to implement a convolution product process in order to insert the 3D sound effect of the interpolated HRIR into an arbitrary signal (3D sound emitter). At the same time, the obstacle detection approaches will be implemented in proper hardware architecture in order to feed spatial coordinates to the 3D sound emitter.

REFERENCES

[1] M. Vorländer, 2008. Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality. Springer, Berlin, pp. 86–87

[2] J. Blauert, 1997. Spatial Hearing. The MIT Press, Cambridge, pp. 372–392.

[3] [3] G. Wersnyi, 2009. Effect of emulated head-tracking for reducing localization errors in virtual audio simulation. IEEE Transactions On Audio, Speech, And Language Processing, vol. 17, n. 2, pp. 247–252.

[4] T. Ajdler, C. Faller , L. Sbaiz and M. Vetterli, 2005. Sound Field Analysis along a Circle and Its Applications to HRTF Interpolation. The Journal of the Audio Engineering Society, vol. 56, n. 3, pp. 156–175.

[5] H. Hacihabiboglu, B. Gunel and A. M. Kondoz, 2005. Head-related transfer function filter interpolation by root displacement. In Proc. IEEEWorkshop on Applications of Signal Processing to Audio and Acoustics (WASPAA-05), New Paltz, NY, USA, pp. 134–137.

[6] T. Nishino, N. Inoue, K. Takeda and F. Itakura, 2007. Estimation of HRTFs on the horizontal plane using physical features. Applied Acoustics, vol. 68, pp. 897–908.

[7] N. H. Adams and G. H. Wakefield, 2008. State-space synthesis of virtual auditory space. IEEE Transactions On Audio, Speech, And Language Processing, vol. 16, n. 5, pp. 881–890.

[8] G. Enzner, 2009. 3D-continuous-azimuth acquisition of head-related impulse responses using multi-channel adaptive filtering. In Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA-2009), New Paltz, NY, USA, pp. 325–328.

[9] M. Pollow and M. Vorl¨ander, 2011. Deriving continuous HRTFs from discrete data points. Fortschritte der Akustik: DAGA 2011; 37. Jahrestagung f¨ur Akustik; 21. Dsseldorf / DEGA.Wiss. Ed. J. Becker-Schweitzer. pp. 641–642.

[10] J. Ahrens, M. R. P. Thomas and I. Tashev, 2012. HRTF magnitude modeling using a non-regularized least-squares fit of spherical harmonics coefficients on incomplete data. In Proc. Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), Asia-Pacific, Dec. 2012, pp. 1–5.

[11] J. F. Lucio Naranjo, R. A. Tenenbaum and J. C. B. Torres, 2010. Using Artificial Neural Networks to generate virtual acoustic reality applied on escape training in blind conditions. International Review of Chemical Engineering (I.RE.CH.E.), Vol. 2, No. 6, pp. 754–759.

[12] Y. Liu and B. Xie, 2012. Analysis on the stability of spatial interpolation schemes for head-related transfer function. The Journal of the Acoustical Society of America, vol. 131, n. 4, pp. 3305–3305.

[13] J. Tebelskis, 1995. Speech Recognition using Neural Networks. Ph.D. Thesis - Carnegie Mellon University, Pittsburgh, pp. 4–7.

[14] D. Rumelhart, J. McClelland and the PDP Research Group, 1986. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. MIT Press, pp. 328–330.

[15] S. Haykin, 2009. Neural Networks and Learning Machines. Prentice Hall, New Jersey, pp. 21–24.

[16] B. Gardner, K. Martin, 1995. HRTF Measurements of a KEMAR Dummy-Head Microphone. The Journal of the Acoustical Society of America, vol. 97, n. 6, pp. 3907–3908.

[17] J. C. B. Torres, M. R. Petraglia and R. A. Tenenbaum, 2004. An Efficient wavelet-based HRTF for auralization. Acta Acustica united with Acustica, vol. 90, n. 1, pp. 108–120.

[18] Z. Haraszy, D. Ianchis and V. Tiponut, 2009 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, pp. 114–118.

[19] E. A. Macpherson and J. C. Middlebrooks, 2002 Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. The Journal of the Acoustical Society of America, vol. 111, n. 5, pp. 2219–2236.

[20] H. Fletcher and W. A. Munson, 1933. Loudness, its definition, measurement and calculation. The Journal of the Acoustical Society of America, vol. 5, pp. 82–108.

[21] ISO226, 2003. Acoustics – Normal equal-loudness-level contours.

[22] D. Marr and T. Poggio, 1979. A computational theory of human stereo vision. Proceedings of the Royal Society of London B: Biological Sciences, 204(1156), 301-328.

[23] J. Corso, 2012. Escáner de Tiempo de vuelo y de triangulación. [Online]. Available: http://www.surveyterra.com/2012/06/escaner-de-tiempo-de-vuelo-y-de.html

[24] L. Ilbay. “Reconstrucción activa de objetos 3d mediante escaneo láser y generación de la vista por medio del software MatLab”. Engineering Monography. Dept. Control and Automatization. Eng., National Polytechnic School, Quito, 2014.

José Francisco Lucio Naranjo was born in Quito – Ecuador, in June 1st of 1979. He has an engineer degree in Computer Systems by the Pontificia Universidad Católica del Ecuador (2005). In 2010 and in 2014, he obtained, respectively, a M.Sc. and a Ph.D degrees, both in Computer Modeling by IPRJ, Rio de Janeiro State University, Brazil. His major field of study is computational modeling focus on acoustic numeric simulation applied in virtual reality using artificial neural networks.

Dr. Lucio is currently acting as computer science titular professor at the National Polytechnic School (EPN) of Ecuador. He has also acted as substitute teacher at the Rio de Janeiro State University (UERJ), distance education tutor in the Federal Fluminense University (UFF), curricular professor at the Universidad de las Américas (UDLA) and auxiliary professor at the Central University of Ecuador (UCE). He was also an associate member of the Acoustical Society of Brazil (SOBRAC). (UCE). También ha sido miembro asociado de la Sociedad Brasileña de Acústica (SOBRAC).

Roberto A. Tenenbaum is a Mechanical Engineer by the Federal University of Rio de Janeiro, in 1972; obtained his M.Sc. in the field of Thermoelasticity by COPPE, Federal University of Rio de Janeiro, in 1975 and his D.Sc. in the field of Acoustics by the same university, in 1987. His major field of interest is dynamics, acoustics and vibration.

He has worked as Engineer at the Nuclear Technology Company, in Rio de Janeiro, from 1973 to 1974. Then, as a Professor and Researcher at the Acoustics and Vibration Lab. (LAVI) in the Graduate Program for Mechanical Engineering, in the Federal University of Rio de Janeiro, in Rio de Janeiro, Brazil, from 1974 to 2004. Now, he works as a Professor and Researcher at the Dynamics, Acoustics and Vibration Lab. (LIDAV) in the Graduate Program in Computer Modeling, in the State of Rio de Janeiro University, in Nova Friburgo, Brazil. He is the author of three books on dynamics: Dinâmica, 1997, in Portuguese; Fundamentals of Applied Dynamics, 2004, in English, published by Springer-NY; and Dinâmica Aplicada, 2006, also in Portuguese. He is also the author or co-author of more than 150 articles published in journals and conferences.
Professor Tenenbaum is an honorary and founder member of the Brazilian Society of Mechanical Sciences and Engineering, ABCM; a founder member of the Acoustical Society of Brazil, SOBRAC; is member of The Acoustical Society of America, ASA; and the International Society for Inverse Problems in Science and Engineering, ISIPSE, among others.

Luis Alberto Morales Escobar was born in Quito – Ecuador, in January 14th of 1985. He has an engineer degree in Electronics by the Escuela Politécnica Nacional (2010). In 2012, he obtained a M.Sc. in Automatics and Robotics, in Universitat Politecnica de Catalunya, Barcelona- Spain. His major field of study is Computer Vision and Robotic Systems. MSc. Luis Morales is currently acting as titular professor at the Escuela Politécnica Nacional of Ecuador, and he is manager of Electronic Instrumentation Laboratory of the Faculty of Electrical and Electronic Engineering from the EPN.

Henry Patricio Paz Arias was born in Zamora – Ecuador, in Abril 14st of 1986. He has an engineer degree in Systems by the Universidad Nacional de Loja Ecuador (2010). In 2012 he obtained the Master in Computer Science in the Area of Artificial Intelligence. Currently he is pursuing a PhD in computer science in the area of intelligent systems of the National Polytechnic School (EPN) of Ecuador. Ing. Paz is currently acting as computer science titular professor at the National Polytechnic School (EPN) of Ecuador. He has also acted as teacher in the Universidad Interglogal - Pachuca - México, National University of Loja - Ecuador teaching materials artificial intelligence.

Carlos Efraín Iñiguez Jarrín was born in Quito – Ecuador, in November 15th of 1979. He got an engineer degree in Systems Computer at Escuela Politécnica Nacional (2005). In 2013, he obtained a M.Sc. in Web Engineering, at Universidad Politécnica de Madrid, Spain. His major field of study is Software Engineering. MSc. Carlos Iñiguez is acting as titular professor at Escuela Politécnica Nacional of Ecuador (EPN), and he is currently leading the software development for the research project ¨Simulation of Acoustic Wave Propagation in closed areas¨, at EPN.