PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Room acoustics modeling using a hybrid method with fast auralization with artificial neural network techniques

Roberto A. TENENBAUM[1], Filipe O. TAMINATO[1], Viviane S.G. MELO[2]

[1] State University of Rio de Janeiro, Brazil

[2] Federal University of Santa Maria, Brazil

ABSTRACT

This work presents a new technique to produce fast and reliable auralizations with the computer code RAIOS, a room acoustics simulator, product of research development at the Instrumentation in Dynamics, Acoustics and Vibrations Lab, State University of Rio de Janeiro. It discusses briefly the hybrid model used in the room simulation and the binaural room impulse responses generation technique using artificial neural networks, with a significant reduction in computational cost of around 85%. It is shown that the binaural impulse responses generated by the classical convolution method and the artificial neural network technique are almost indistinguishable. In the sequel, the measured and simulated binaural impulse responses for one of the rooms used in the Round Robin 4 simulation code´s inter-comparison is presented, showing a quite good agreement.

Keywords: Fast auralization, Room acoustics simulation, Artificial neural networks

1. INTRODUCTION

This work deals with computer simulation of room acoustics and techniques to generate auralization at selected points in the room. In general, the computational simulation of room acoustics presupposes the requirements of geometric acoustics (1). This means that the sound wavefronts can be treated as acoustic rays, which leave the sound source and propagate in the room, reflecting on their internal surfaces (2). There are two main ways of modeling acoustic rays: the ray tracing method (3,4) and the virtual sources method (5). There are also hybrid algorithms that use the virtual sources method for the calculation of the direct sound and first reflections and the ray tracing method for the computation of the remaining reflections (6). Some commercial software is essentially based on the ray tracing method (7), presenting satisfactory results for the monoaural acoustic field. However, as already pointed out by many authors (8), the diffuse reflection plays an important role in room acoustics by providing a greater uniformity in the sound field (9), but also essential when it is desired the auralization for a source-receiver pair. In this case, it seems essential to have a good model to deal with diffuse reflections, since the ray tracing technique cannot do so. One of the ways to properly model diffuse reflections is the radiosity technique (10). The most effective implementation of this model is to store the portion of sound energy diffusely reflected after each specular reflection in a matrix whose lines correspond to discrete times and whose columns represent the spatial discretization of the room's contour surfaces in triangular elements. These so-called *transition matrices*, then, evolve by transferring diffuse energy between surface elements of the room, according to the solid angle that each one is seen and, of course, the respective scattering coefficients (11).

Among the most impressive applications of room acoustics computer simulation it is precisely the possibility of generating auralization (12) the most powerful branch of acoustical virtual reality (AVR). Reliable auralization – usually to be heard with properly equalized headphones – can provide the listener with an authentic sense of sound immersion in the simulated environment. This provides the acoustical designer, for instance, with information about the sound quality of the room under analysis that would not be possible by only observing the acoustic quality parameters (AQPs) provided by the simulator. One could say that the difference between looking to AQPs of a room or alternatively hearing its auralization from a given anechoic signal is similar to the difference between reading the menu of a good restaurant or tasting its dishes. It is worth noting, however, that the evaluation of an auralization, being subjective, i.e., depending on the human interpretation, is still a subject of research in progress. Perhaps the most complete work in this area is the spatial audio quality inventory – SAQI (13).

[1]ratenenbaum@gmail.com

Although 42 different terms in English are considered to qualify the different human sensations related to an audio content, there was no pretense of finding objective parameters, i.e., a measurable metric estimated by a number. Some alternatives have been published in this direction, using articulation scores to evaluate speech intelligibility, comparing actual rooms with their corresponding computational auralizations (14-16).

The Instrumentation in Dynamics, Acoustics and Vibrations Lab – LIDAV, State University of Rio de Janeiro, Brazil, has been developing a research computational code for room acoustics simulation aimed at AVR. This code, known as Room Acoustics Integrated and Optimized Software (RAIOS), uses a ray-tracing method to model the specular reflections and a transition matrix method to compute the diffuse reflections (11). It also produces auralizations for selected source-receiver pairs. In Section 2, a new technique for generating auralizations is presented, replacing the convolution method with a methodology that models the head-related impulse responses (HRIRs) through artificial neural networks (ANNs) (17) of the radial basis functions kind (18) with a very significant computational gain. In Section 3, the computational cost of the ANN technique is discussed.
In Section 4, a brief discussion on the results obtained for the computation of filtered HRIRs through the convolution method and through artificial neural networks, is compared, showing that the two methods present results that are practically indistinguishable from each other. In Section 5, the simulation results obtained in one of the rooms used in Round-Robin 4 – the 4th international inter-comparison on room acoustics simulation codes, with auralization – with code RAIOS, are compared with the measured results, furnished by the competition organizers. Finally, in Section 6, the main conclusions are presented.

2. FAST AURALIZATION WITH ANN TECHNIQUE

Once the acoustic field in the room simulation step is completed, the goal in sequel consists in the determination of the room impulse mono (RIRs) and binaural (BRIRs) responses at selected points. As regards the calculation of RIRs, it is a question of convert the energy arrival, via Hilbert's transform (19) and the adequate filtering in octave bands, in filtered impulse responses, whose computational cost is relatively small. In order to compute the BRIRs, however, it is necessary to take into account the head-related impulse responses (HRIRs) – or their corresponding in frequency domain, the so-called head-related transfer functions (HRTFs). In the computational codes that generate auralization, this is done via convolution procedure. Each wavefront – or acoustic ray – that reaches a given receiver carries three main information: (a) its spectrum, in octave bands; (b) its direction of arrival, in azimuth and elevation; and (c) its arrival time. Each wavefront, when leaving the sound source, has a flat spectrum. Nevertheless, due to the several reflections with absorption and scattering at the room's boundary surfaces, as well as the attenuation due to its propagation in the air during its trajectory, reaches the receiver with a filtered spectrum. This must then be multiplied by the HRTF of the direction closest to its arrival direction and to the resulting spectrum an inverse fast Fourier transform (IFFT) must be applied in order to obtain the filtered HRIR for that wavefront. In the sequel, every filtered HRIR must be delayed by its arrival time and the sum of all the filtered and delayed HRIRs that reach the receiver will constitute the BRIR for that source-receiver pair. The procedure is equivalent to the convolution of the acoustic ray with the HRIR of the corresponding direction and henceforth it is called convolution method (CM), or classical method. Figure 1 illustrates the procedure.

The described procedure has the drawback of its computational cost. In fact, the complex product between the HRTF of the considered direction with the spectrum of the wavefront that reaches the receiver, followed by the fast inverse Fourier transform in two channels to obtain the filtered HRIRs presents a high computational cost, mainly taking into account that the number of wavefronts that reaches each receiver can be of the order of 10^5, in a good simulation. However, the delay and addition procedures of the filtered HRIRs for the BRIR calculation demand low computational costs.

The RAIOS computational code uses another technique to generate the BRIRs. The idea is to replace the HRIRs/HRTFs database with a set of trained and tested artificial neural networks that "learned" previously the HRIRs patterns. In this way, the procedure is all carried out in time domain. The radial basis function networks have as input the nine wavefront spectral components in octave bands (between 63 Hz and 16 kHz) that reach the receiver. Ten neurons are used in the hidden network layer and, at the output, 128 neurons are used, which correspond to the amplitude values of the

resulting filtered HRIR. Then, the delay procedure of each filtered HRIR and addition of all these functions to obtain the BRIR is identical to that of the convolution method. However, the computational cost of the artificial neural network technique is about 15% of the cost of the convolution method, as it is showed in next section. Figure 2 illustrates the procedure.

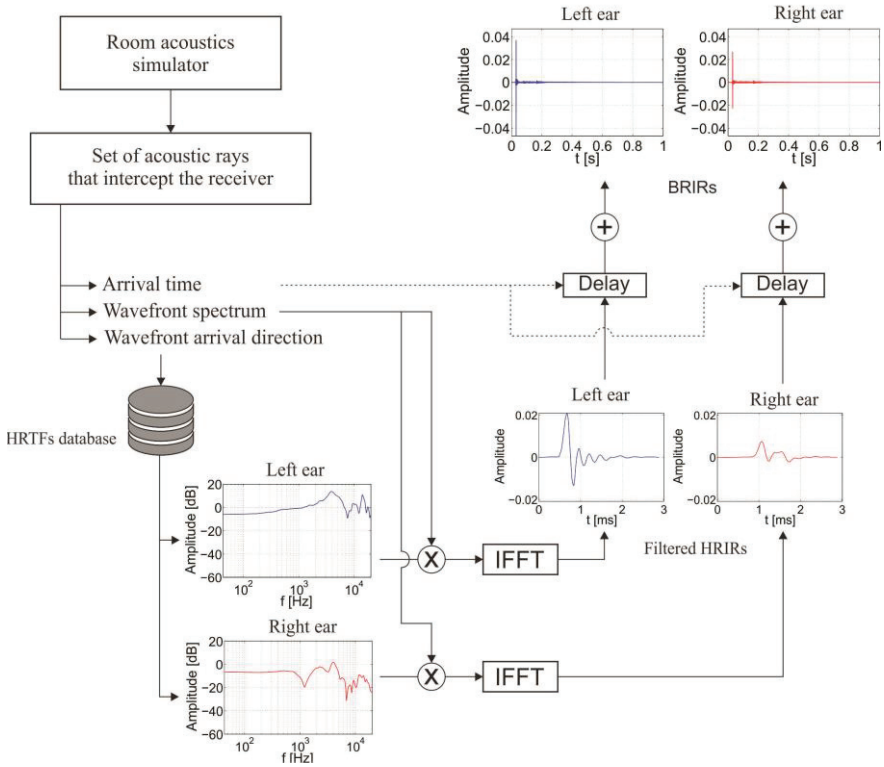

Figure 1 – General procedure to produce a BRIR by the convolution method.

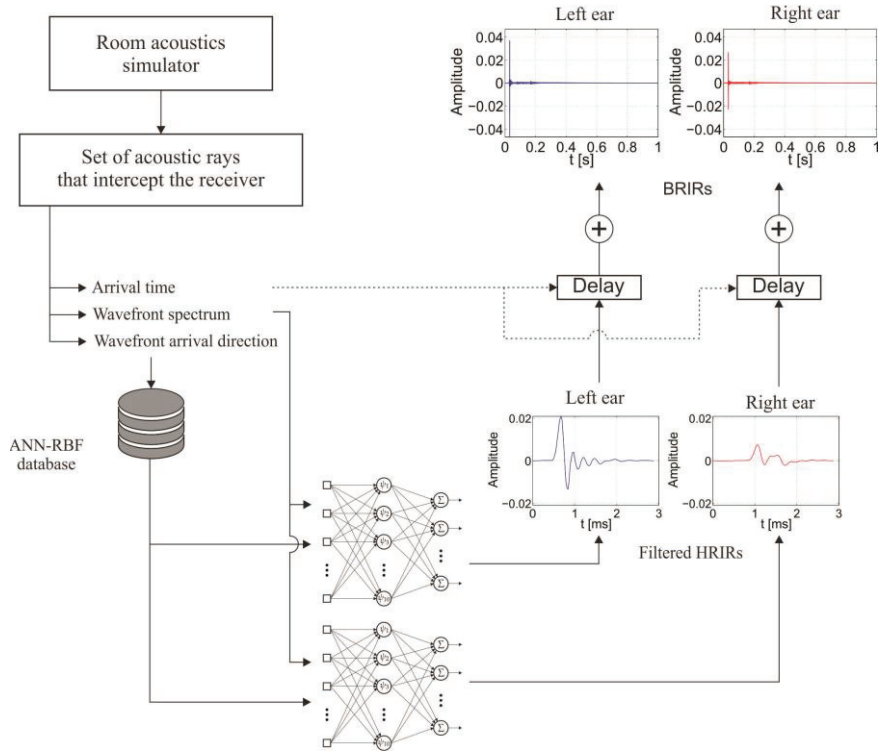

Figure 3 – General procedure to produce a BRIR by the ANN-RBF technique.

At the bottom of Fig. 3 two previously trained RBF-kind ANNs are displayed, one for each ear, for the given wavefront direction. In each network besides the nine entries, it is considered an intermediate layer with N_1 and an output layer with 128 neurons, which are the considered time domain samples of the filtered HRIR.

3. COMPUTATIONAL COST

In order to examine the numerical efficiency of the convolution method (CM) with that of the RBF-kind artificial neural network method (ANNM), a comparison is made as to the number of arithmetic operations that each technique requires. The number of operations in the convolution method to compute the filtered HRIRs equals the sum of two parts. The first one corresponds to the number of multiplications between the ray spectrum (in octave bands) and the HRTF. Note, however, that due to the HRTF symmetry, only $l/2$ products, with l being the number of samples, are necessary. The second one corresponds to the number of operations for calculating the inverse Fourier transform. The total number of mathematical operations in the convolution method is then given by (21)

$$O_{CM} = \frac{l}{2} + 5l \log_2 l. \tag{1}$$

It is worth noting that, since the convolution method deals with Fourier transforms, $l = 512$ samples are necessary to preserve information in all octave bands. So, for each filtered HRIR, the computational cost in number of arithmetic operations is

$$O_{CM} = 23{,}296. \tag{2}$$

The training and testing procedures of artificial neural networks are known to be computationally costly. However, the execution phase – particularly the one of the radial basis function networks – is quite efficient. The number of operations of the RBF network during the implementation phase is also the sum of two parts. The first one is associated to the sum of the number of operations performed to compute the output of the activation function of each neuron of the intermediate layer. It can be shown (22) that N_0 being the number of input entries of the RBF network, N_1 the number of neurons of the intermediate layer and N the number of retained terms of the Taylor series in which the activation function is truncated (23), the number of operations of this part is

$$O_1 = N_1(3N_0 + 2N). \tag{3}$$

The second one corresponds to the sum of the number of operations to calculate each neuron of the output layer. Being N_2 the number of neurons of this layer, the number of sums and products of this second part is

$$O_2 = 2N_1N_2. \tag{4}$$

In addition, it is necessary to take into account the cost of normalizing the inputs and the subsequent multiplication of the network output by the inverse factor that was used to normalize the input. These procedures require only $N_0 + N_2$ multiplications. One can conclude than that the total number of operations to generate the output of the RBF network in the ANN method is

$$O_{ANNM} = N_1(3N_0 + 2N_2 + 2N) + N_0 + N_2. \tag{5}$$

Taking now $N_0 = 9$ (nine octave bands), $N_2 = 128$ (minimum acceptable number of time samples) and $N = 35$ (20) one would have

$$O_{ANNM} = 353N_1 + 137. \tag{6}$$

The computational cost of the artificial neural network method, therefore, is a linear function of the number of neurons in the intermediate layer of the RBF network, N_1. There is no way other than trial and error to check the minimum number of neurons in this layer that still provides good binaural room impulse responses. Several tests were performed with N_1 ranging from 2 to 30. The best cost-benefit ratio was found for $N_1 = 10$. That means that the difference between the filtered HRIRs calculated by the two methods (CM and ANNM) becomes, for that value of the number of neurons in the intermediate layer of the network, imperceptible. The comparative results presented in Section 4 are computed with $N_1 = 10$. The computational cost in terms of the number of arithmetic operations for the method of the RBF-kind artificial neural networks is, then,

$$O_{ANNM} = 3667, \tag{7}$$

which corresponds to 15.74% of the number of arithmetic operations calculated in Eq. (2). The main conclusion is that by working directly in the time domain – and avoiding the convolution procedure –, the computational gain is approximately 84,3%.

4. COMPARATIVE RESULTS FOR FILTERED HRIRs

As mentioned, the convolution technique is the classic BRIR generation method and it is present in all acoustic field simulation software with auralization, to our knowledge. Therefore, in order to verify the reliability of the method of generating the BRIRs with the artificial neural networks of the radial basis function type, a comparison between the two methods is presented in the sequel. Since once the filtered HRIRs are generated the procedure is identical, involving the delays and their addition to generate the BRIR, the comparison between the two methods will be done among the filtered HRIRs. Figures 3 to 10 indicate the filtered HRIRs computed via the convolution method (CM) compared with the same functions calculated by the artificial neural networks method (ANNM), for eight randomly chosen directions, each one representative of the eight octants that the sphere can be divided around the head (φ = azimuth; θ = elevation). In addition to the confrontation of the filtered HRIRs, in time domain, their counterpart in frequency domain, both in amplitude and phase, are shown. To evaluate the similarity between the functions, it is adopted the normalized correlation coefficient, r (21). The similarity between the two functions is better as r is closest to 1.

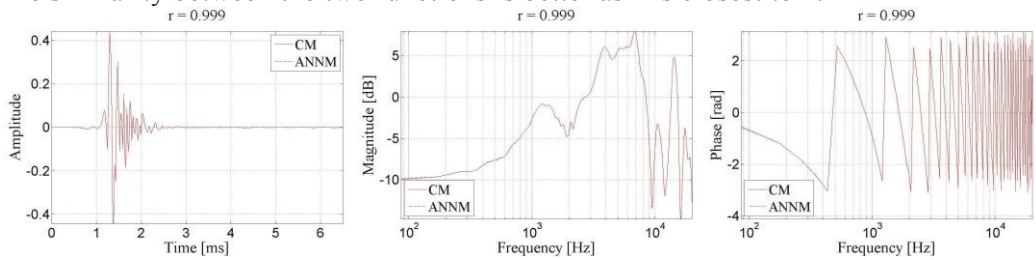

Figure 3 - Filtered HRIRs obtained with CM and ANNM, for $\varphi = 37^0$ and $\theta = 31^0$.

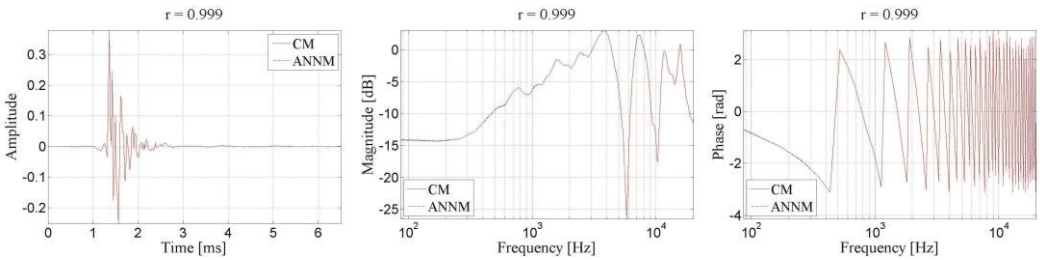

Figure 4 - Filtered HRIRs obtained with CM and ANNM, for $\varphi = 12^0$ and $\theta = -45^0$.

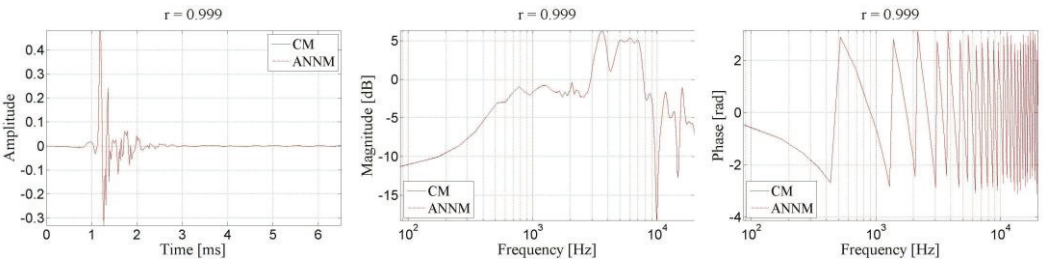

Figure 5 - Filtered HRIRs obtained with CM and ANNM, for $\varphi = 104^0$ and $\theta = 19^0$.

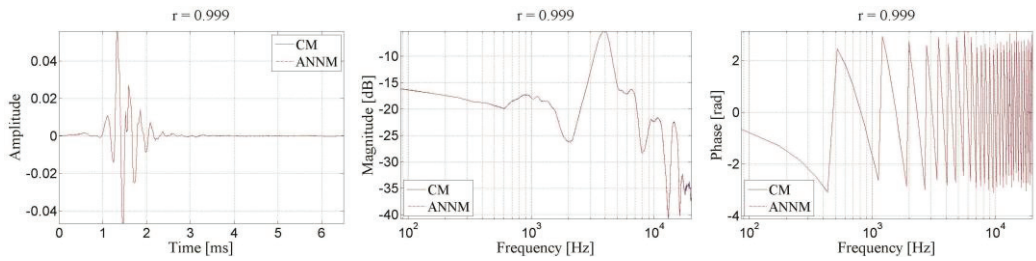

Figure 6 - Filtered HRIRs obtained with CM and ANNM, for $\varphi = 156^0$ and $\theta = -22^0$.

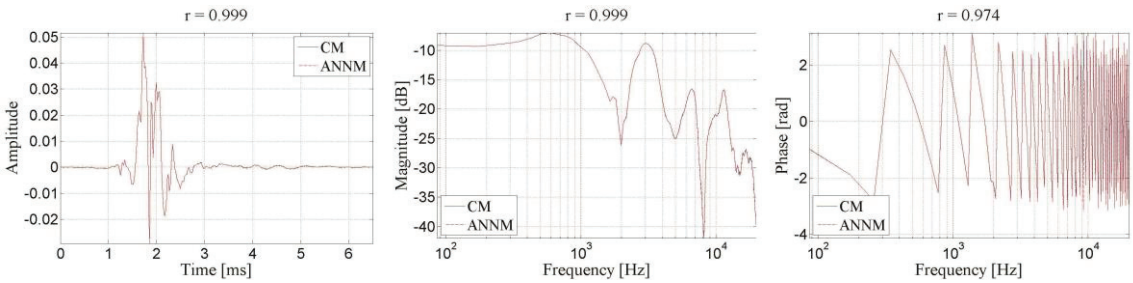

Figure 7 - Filtered HRIRs obtained with CM and ANNM, for $\varphi = 243^0$ and $\theta = 9^0$.

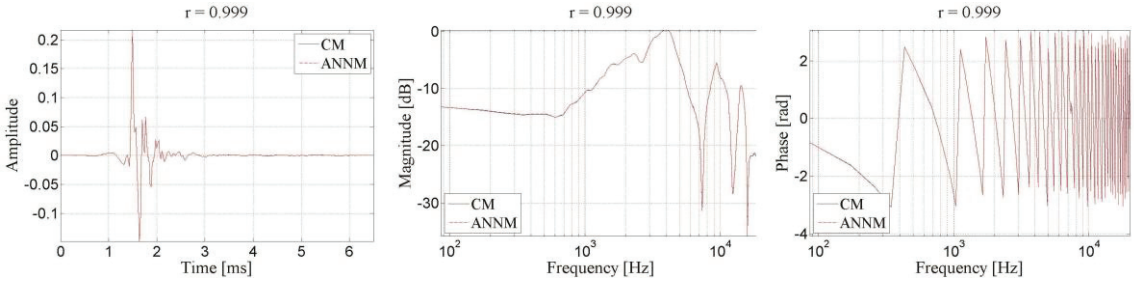

Figure 8 - Filtered HRIRs obtained with CM and ANNM, for $\varphi = 195^0$ and $\theta = -28^0$.

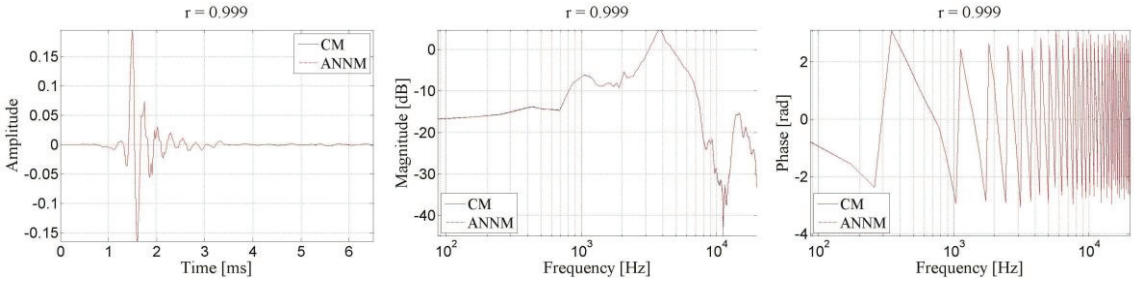

Figure 9 - Filtered HRIRs obtained with CM and ANNM, for $\varphi = 331^0$ and $\theta = 55^0$.

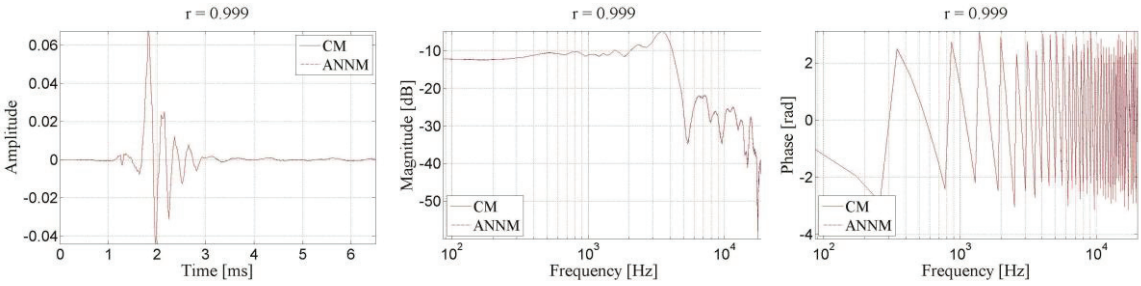

Figure 10 - Filtered HRIRs obtained with CM and ANNM, for $\varphi = 272^0$ and $\theta = -18^0$.

It can be seen in Figs. 3 to 10 a significant similarity between the graphs, both in time and frequency domain, with values of the correlation coefficient very close to one. This indicates that the technique of producing filtered HRIRs via artificial neural networks is, aside from faster, very reliable.

5. MEASURED AND SIMULATED BRIRs IN ROUND ROBIN 4

Round Robin 4 (RR4) was the last international inter-comparison, so far, involving room acoustics computer simulation codes, and the first one with auralization. It was jointly organized by the Technical University of Berlin and the University of Aachen and launched in 2016. In this project, information was provided on nine rooms, in a total of 25 different configurations, with geometry,

materials, absorption and scattering coefficients of internal surfaces. In addition, databases with the HRIRs of the dummy head used in the binaural measurements, for 45 angular positions of rotation about the vertical axis of the head above the torso (azimuth) were given. A variable number of sound sources and mono and binaural receivers were provided in each room. A database was also provided with the directivities (in azimuth and elevation) of two sound sources used. The data are available in (24). In all, 114 RIRs and 949 BRIRs were requested as output. Possibly due to the heavy data requirement to be calculated in RR4, the number of participant teams (eight) was much lower than in RR3 (21). In this section, the comparative results between the measured BRIRs – provided by the organizing staff of RR4, after sending simulation results by all participants – and those simulated by the LIDAV team using the RAIOS computational code for the room called "chamber music hall" are presented. Figure 11 shows part of the room with the positions of sources (LS03–06) and binaural receiver (MP6).

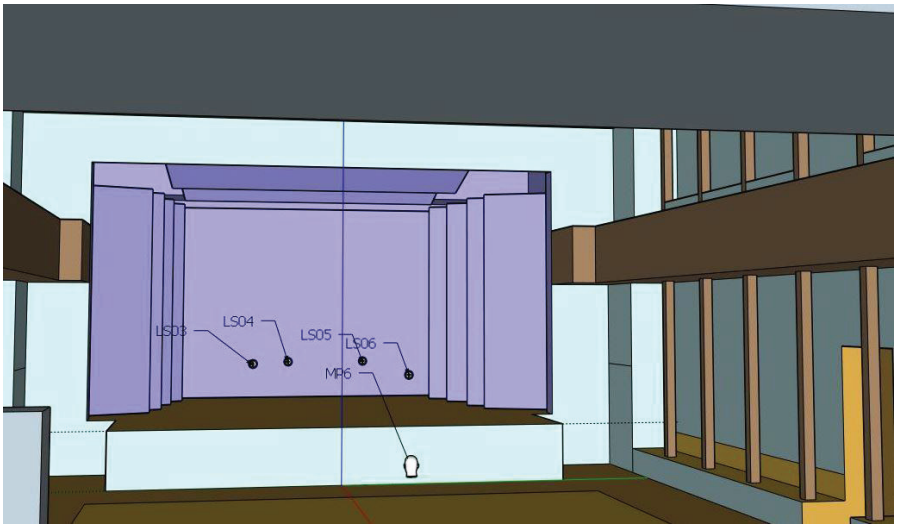

Figure 11 – Room called "chamber music hall" of RR4 with its source and receiver positions. Partial view.

In the sequel, the interaural cross-correlation coefficient, IACC, is computed for both measured (RR4 organizers) and simulated (RAIOS) BRIRs for the pairs LS03-MP6 (Pair 1), LS04-MP6 (Pair 2), LS05-MP6 (Pair 3) and LS06-MP6 (Pair 4). The results are shown in Table 1.

Table 1 – Interaural cross-correlation coefficient computed for the source-receiver pairs

S-R Pairs	Pair 1	Pair 2	Pair 3	Pair 4
Measured	0.41	0.43	0.36	0.34
Simulated	0.49	0.38	0.29	0.29
Difference	–0.08	0.05	0.07	–0.07

As can be withdraw from Table 1, the obtained values show to be close (measured versus simulated) with two of them greater for the measured BRIRs and the other two for the simulated ones. In all cases, the percentual error is below 20%.

6. CONCLUSION REMARKS

A new technique was proposed to model the filtered head-related impulse responses, which are the core of binaural room impulse responses generation with the purpose of auralization production. The model is based on artificial neural networks of the radial basis functions kind and was implemented in RAIOS code, resulting in a computational time saving of around 85% in the BRIRs computation.

As in any procedure of room acoustics simulation, there are two blocks of higher computational cost, namely: The calculation of all acoustic wavefront arriving at each receiver; and the computation

of the binaural room impulse responses for each receiver. This second computationally costly block benefited from the modeling through artificial neural networks, with a significant computational time saving, especially if many receivers are under consideration.

The comparative results between the filtered HRIRs/HRTFs obtained by the classical convolution method and by the new model using artificial neural networks showed that the functions are, for all practical purposes, indistinguishable, presenting a normalized cross-correlation coefficient practically equal to unity, both in time and frequency domains. Although only eight of the 64,442 directions have been shown here, for all of them the comparative results are equally accurate.

The measured and simulated with computational code RAIOS results for one binaural receiver and four directional sound source positions were compared for one room of RR4. The simulation was performed with the ANN technique and the comparative data of the inter-aural cross correlation coefficient show that there is good agreement. Nevertheless, the discrepancies cannot be attributed specifically to any one of the many uncertainties in the simulation, such as relative positions, scattering and absorption coefficients used, directivity of sources data etc.

REFERENCES

1. Kuttruff H. Room Acoustics. 5. ed. London: Spon Press, 2009.
2. Savioja L, Svensson UP. Overview of geometrical room acoustic modeling techniques. J. Acoust. Soc. Am., v. 138(2), 708–730, 2015.
3. Kulowski A. Algorithmic representation of the ray tracing technique. Applied Acoustics, 18, 449–469, 1984.
4. Ondet M, Barbry JL. Modeling of sound propagation in fitted workshops using ray tracing. J. Acoust. Soc. Am. 85(2), 787–796, 1989.
5. Borish J. Extension of the image model to arbitrary polyhedra. J. Acoust. Soc. Am. 75, 1827, 1984.
6. Vorländer M. Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm. J. Acoust. Soc. Am. 86(1), 172–178, 1989.
7. Rindel JH, Christensen CL. Room acoustic simulation and auralization – how close can we get to the real room. WESPAC 8, 2003.
8. D'Antonio P, Cox TJ. Diffusor application in rooms. Applied Acoustics, 60, 113–142, 2000.
9. Cox TJ, Dalenbäck BI, D'Antonio P, Embrechts JJ, Jeon JY, Mommertz E, Vorländer M. A tutorial on scattering and diffusion coefficients for room acoustic surfaces. Acta Acust. Acust. 92(1), 1–15, 2006.
10. Alarcão D, Bento Coelho JL, Tenenbaum RA. On modeling of room acoustics by a sound energy transition approach. Proceedings of EEA Symposium on Architectural Acoustics, 2000.
11. Tenenbaum RA, Camilo TS, Torres JCB, Gerges SNY. Hybrid method for numerical simulation of room acoustics: Part 1 - theoretical and numerical aspects. J. Braz. Soc. Mech. Sci. Eng., 29(2), 211–221, 2007.
12 Vorländer M. Auralization: Fundamentals of Acoustics, Modeling, Simulation, Algorithms and Acoustic Virtual Reality. Berlin: Springer-Verlag, 2008.
13. Lindau A, Erbes V, Lepa S, Maempel HJ, Brinkmann F, Weinzierl S. A spatial audio quality inventory for virtual acoustic environments (SAQI). Acta Acustica united with Acustica, 100(5), 984–994, 2014.
14. Peng J. Feasibility of subjective speech intelligibility assessment based on auralization. Applied Acoustics, 66: 591–601, 2005.
15. Hodgson M, York N, Yang W, Bliss M. Comparison of predicted, measured and auralized sound fields with respect to speech intelligibility in classrooms using CATT-Acoustic and ODEON. Acta Acustica united with Acustica, 94(6): 883–890, 2008.
16. Tenenbaum RA, Taminato FO, Melo VSG, Torres JCB. Auralization generated by modeling HRIRs with artificial neural networks and its validation using articulation tests. Applied Acoustics, 130, 260–269, 2018.
17. Haykin S. Neural Networks and Learning Machines. 3. ed. New Jersey: Prentice Hall, 2009.
18. Mulgrew B. Applying radial basis functions. IEEE Signal Processing Magazine, 50–65, 1996.
19. Kak S. Number theoretic Hilbert transform. Circuits Systems Signal Processing, 33, 2539–2548, 2014.
20. Taminato FO. Artificial neural networks applied to model head-related impulse responses to generate auralization. Ph.D. Thesis, State University of Rio de Janeiro, 2018.
21. Willcox RR. Introduction to robust estimation and hypothesis testing. Academic Press, New York, 2005.
22. Proakis JG, Manolakis DG. Digital Signal Processing: Principles, Algorithms and Applications. 3rd ed., Prentice-Hall, New Jersey, 1996.
23. Conte SD, Boor C. Elementary numerical analysis: An algorithmic approach. 3rd ed., McGraw-Hill, New York, 1980.
24. Brinkmann F, Aspöck L, Ackerman D, Opdam R, Vorländer M, Weinzierl S. A ground truth for room acoustics simulation: Concept and database. http://dx.doi.org/10.14279/depositonce-6726, 2018.