

A New Approach to Validate Computer Modeling Auralizations by Using Articulation Indexes

Melo, V.S.G.^{*1}; Tenenbaum, R.A.^{*2}; Lucio Naranjo, J.F.⁺

^{*}Laboratory of Instrumentation on Dynamics, Acoustics and Vibrations – LIDAV, IPRJ-UERJ, Nova Friburgo, RJ, ¹vivasgmelo@gmail.com, ²ratenenbaum@gmail.com

⁺Faculty of Computer Engineering, National Politechnic School, Quito, Ecuador, jose.lucio@epn.edu.ec

Abstract

In this work, a new approach to validate computer modeling auralizations by using articulation indexes is presented. The generation of acoustical virtual reality with the proprietary computer code RAIOS is briefly described. The code simulates the room acoustics, including specular and diffuse reflections, and a set of artificial neural networks provides the room's binaural impulse responses at selected points in the audience region. Then, these responses are convolved with anechoic signals to generate virtual acoustical realities. In the sequel, the impulse responses at some room's positions are measured and the convolution with the anechoic signals is provided to obtain a virtual articulation index. Finally, the computer modeling auralizations are also reproduced to the subjects and the obtained articulation index results are compared for validation purposes.

Keywords: Auralization. Articulation Indexes. Artificial Neural Networks.

1. Introduction

Virtual acoustic reality is a rather new branch of acoustics due to the recent development of digital computation technology. The most interesting and useful result of this technique is the so-called auralization. As stated by Vorländer (2008), 'auralization is the technique of creating audible sound files from numerical (simulated, measured or synthesized) data.' This means that the auralization of a sound can bring to the listener not only the sound contents but also the influence of the ambient where it is reproduced.

Acousticians have been developing many metrics to evaluate, for instance, room acoustic quality. All of them are based on the monoaural room's impulse responses (T_{30} , C_{80} , D_{50} , among others) or on the binaural room's impulse responses ($IACC$, for instance). However, the auralization in some room's position is, essentially, a signal in a .wav or similar format and there is no metrics, up to now, to *quantify* it.

To validate the acoustical virtual reality, however, it is not enough to reproduce the auralized sound to people and ask them to compare with the actual sound in the original room, since the answers cannot provide more than generic opinions about the verisimilitude of the two sounds. In other words, this subjective evaluation is not sufficient to validate the adopted technique to generate the virtual acoustical reality. This raises, then, an aspect to investigate: Which metrics to use in order to *validate* an auralization?

A new methodology to validate auralizations is proposed in this work: To use the *articulation index* (AI) as a parameter to evaluate the auralizations quality. The AI is a metrics adopted to assess the word intelligibility in rooms. Very briefly, the AI is measured in a room by applying articulation tests to a group of subjects in the room. The average percentage of hits indicates the room's AI (MELO et al., 2013).

The proposed technique presented here is to use lists of monosyllables —recorded at an anechoic chamber in order to suppress the

influence of the recording room—to measure the articulation index at different positions in the room. Then, the room binaural impulse responses (BIR's) at some positions are measured with a dummy head (DH) through the room excitation with compensated sweeps (MÜLLER & MASSARANI, 2001). These BIR's are then convolved with the anechoic lists of monosyllables to produce what will be called as *auralizations of first kind*. The auralizations are then applied to subjects through equalized headphones to evaluate the (virtual of first kind) articulation index.

In the sequel, the room is simulated with the computer code RAIOS and the impulse responses are obtained in the corresponding points (receivers). Then, a set of artificial neural networks (ANN's) that models the head-related impulse responses (HRIR's) (BLAUERT, 1997) reads the simulator output and generates the BIR's at the selected points in the room. The modeled BIR's are then convolved with the anechoic lists of monosyllables to produce what will be called *auralizations of second kind*. It is worth noting that these last auralizations are genuinely numerical. Finally, the articulation tests are applied to the same subjects through equalized headphones, measuring the (third) AI. Finally, the three AI's are compared to *validate* the numerical auralization.

2. Room Acoustic Simulation

The room acoustic simulation is performed with the aid of the computational code RAIOS (Room Acoustics Integrated and Optimized Software), developed at the Laboratory of Instrumentation in Dynamics, Acoustics and Vibration – LIDAV, State University of Rio de Janeiro. The code RAIOS has proved to furnish reliable acoustic data, when it participated in the international intercomparison of room acoustics computer

simulation, the Round Robin 3 (BORK, 2005; TENENBAUM et al., 2007a and 2007b).

The code RAIOS provides the impulse response at selected points inside the simulated room. It also computes the main acoustical parameters according to ISO 3382 (1997). One of the main computer code's features is that the room simulation is performed by a hybrid procedure. The specular reflections are computed by a modified ray-tracing algorithm (CAMILO & TENENBAUM, 2002) and the diffuse reflections are simulated with a sound energy transition approach (KRUZINS & FRICKE, 1982). As stressed by Dalenbäck et al. (DALEMBÄCK et al., 1994), the diffuse reflections—that makes thicker the reverberant tail of a room impulse response—are particularly important in the auralization phase of a room simulation.

Figure 1 presents the main screen of the computational code RAIOS. It is seen, in the bigger area, the perspective of a simulated room, presenting its boundaries, the sound source and the receivers. At the right, it is shown the area of entries, as sound sources data, receiver's data, surface finishing data, and atmospheric conditions. At the bottom, it is shown the area for results output, presenting, in this case, the decays curves per octave band.

The output files of computer code RAIOS of interest in this research are the ones with extension .ray. Their contents include all wavefronts that reach each receiver, with the following information: Direction of arriving; global acoustic power; spectrum per octave band; and arriving time. This output will be the input of the artificial neural network, briefly described in the next section. It is worth noting that, in a simulation with good accuracy, the number of wavefronts that reach each receiver is of the order of 10^4 – 10^5 .

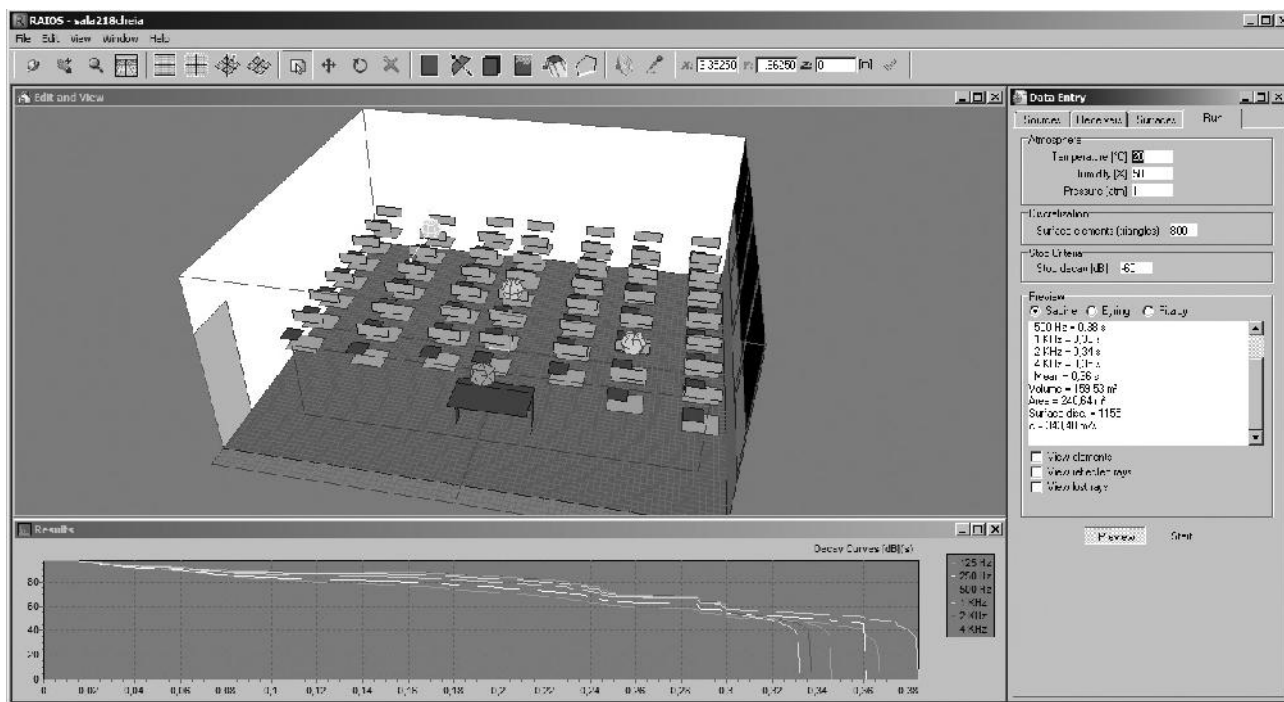


Figure 1: Main screen of computational code RAIOS showing the tested room. In the lower window it is shown one of the outputs, namely the decay curves per octave bands.

3. Generating the Virtual Bir's

Once the output of the computational code RAIOS is obtained, the binaural impulse responses (BIR's) is the next step to be achieved. To provide these responses, the head-related impulse responses (HRIR's) must be modeled precisely since they represent the kernel of the auralization procedure. There are several well known data banks with measurements of these functions, as in (GARDNER & MARTIN, 1995), for instance. However, all of them are discrete and very rarely one measured direction will fit to the actual arrival direction of a wavefront. This means that the HRIR's must be interpolated (AJDLER et al., 2005). Furthermore, since the wavefronts present a filtering of the acoustical ray's spectrum, due to the multiple reflections with absorption on the room boundaries and the natural frequency dependent attenuation due to the sound wave propagation, a spectral modification must be provided for each wavefront.

These procedures can be provided by a set of artificial neural networks (ANN's) (HAYKIN,

2009). The proposed method is capable to reconstruct the desired HRIR functions by means of spatial interpolation and spectral modification. 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, is used. 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 adopted (LUCIO NARANJO et al., 2010, 2013). In the final configuration, the global area around the head is divided into 1898 sections, the time delays are preserved, and the architecture is the one shown in Fig. 2.

The ANN's entries are the wavefront direction, given by the azimuth and elevation angles, and the nine octave-bands spectrum between 32,5 Hz and 16 kHz. The outputs are the 128 coefficients considered as sufficient to model the HRIR's (TORRES et al., 2004). The best architecture — a balance between accuracy and computational cost — is the one

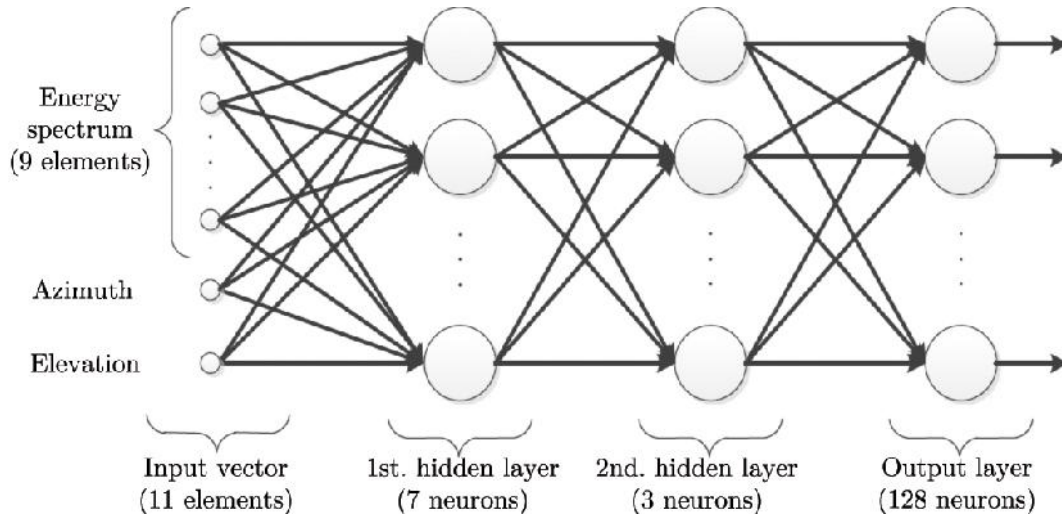


Figure 2: Artificial neural network basic architecture.

shown in Fig. 2, with seven neurons in the first hidden layer and three neurons in the second one

Figure 3 illustrates how the ANN model reproduces superbly the actual HRIR's. Among several tested directions, two directions were selected to be shown here. The graphics show that for the two measured

directions, the modeled HRIR's are very accurately recovered, being almost impossible, at naked eye, to find out the differences. The head related impulse response functions have a counterpart in the frequency domain, the so-called *head related transfer functions* (HRTF's). These are complex functions and the ANN model showed to fit both amplitude and phase.

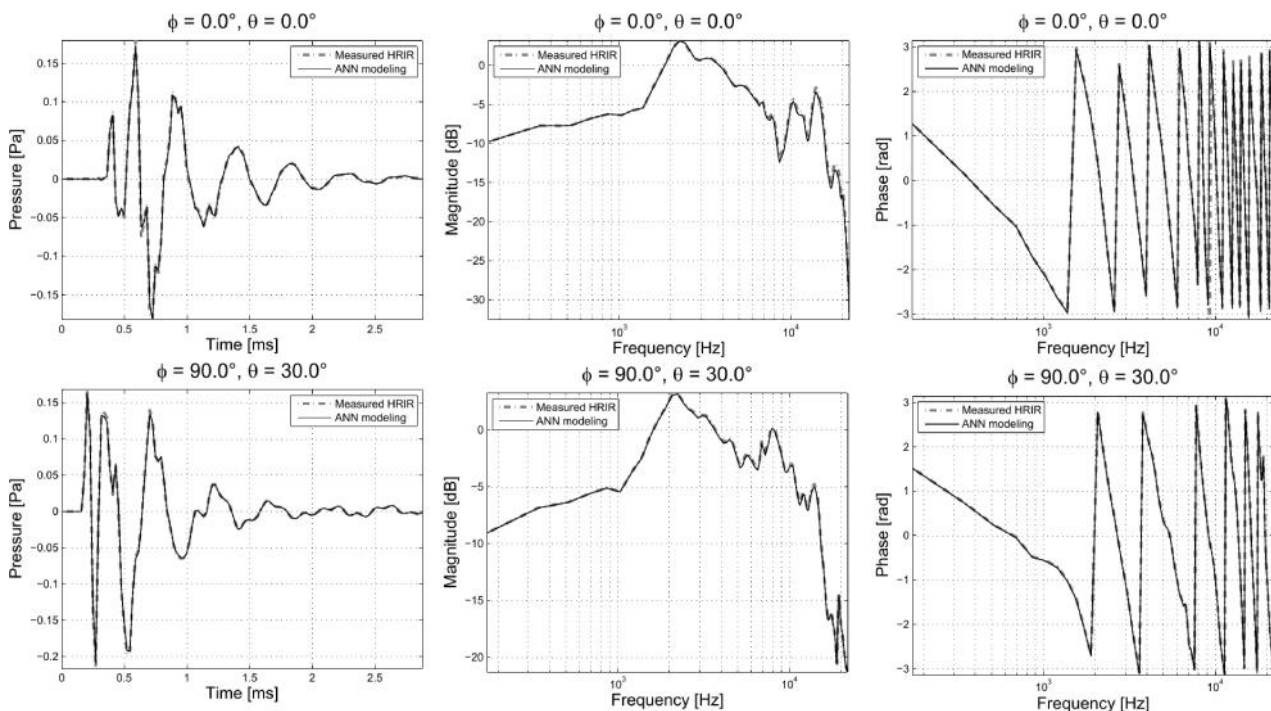


Figure 3: HRIR's measured and modeled by ANN's at two distinct directions, showing the very good fitting. The dashed lines (measured) are almost undistinguishable from the modeled ones (solid line). At left, the HRIR's; at the middle, the modulus of the HRTF's; at right, the corresponding phase.

4. The Methodology

With the aim at validating the numerically generated auralizations, the following steps were provided for the measurements and numerical modeling, constituting the adopted methodology, briefly described in the sequel. This methodology is similar to the one proposed by Melo et al. (2013), with the main difference that now auralizations generated by computer model are included.

1. Preparing a set of three lists with 100 monosyllables each, from 226 recordings registered in an anechoic chamber with a female professional speaker. It is worth noting that the lists of monosyllables are in the Portuguese language.

2. Measuring the transfer function of the dodecahedron used as sound source.

3. Measuring the headphones transfer functions used in the auralizations phase.

4. Conducting articulation tests (AT's) in the room with voluntary university student public. The results of these articulation tests are treated statistically and the actual average articulation index is computed.

5. Emitting compensated frequency sweeps through the dodecahedron, in the presence of public, and recording the binaural impulse responses (BIR's) in the room for three dummy head (DH) positions. Ten sweeps are emitted in order to extract an average binaural impulse response for each DH position in the tested room.

6. The background noise level present in each one of the tested rooms is recorded, with the DH, as well as other data such as the ambient temperature and the air relative humidity, for simulation purposes.

7. At the laboratory, the signals recorded by each DH position are deconvolved with the emitted sweep in order to obtain the binaural impulse responses (BIR's).

8. The BIR's are then convolved with the anechoic signal that contains the lists of monosyllables presented to the students in the classroom, in order to obtain the sound that would be heard in the rooms during the dictations, in the absence of the background noise. This signal contains the measuring room characteristics (its reverberation, for instance) but not the background noise.

9. The background noise registered during the articulation tests is extracted from the DH recordings, in a procedure which will be referred to as cutting (MELO et al., 2013).

10. The background noise extracted in the previous step is added to the signals obtained in Step 8 and the result is convolved with the headphones impulse response. This signal will be called *virtual auralization of the first kind*.

11. The virtual auralizations of first kind are reproduced through headphones to the same public. The results of these articulation tests are treated statistically and the average virtual articulation index (of the first kind) is computed.

12. The same anechoic lists of monosyllables are convolved with the output of the ANN's, that means, the computationally modeled BIR's, for each DH position, and the background noise obtained in Step 9 is added. Then, the signal is convolved with the headphones impulse responses. These results will be called *virtual auralizations of the second kind*.

13. The virtual auralizations of second kind are reproduced through the headphones to the same public. The results of these articulation tests are treated statistically and the average virtual articulation index (of the second kind) is computed.

14. The original articulation index results are compared with those of the two virtual ones.

5. Main Findings

In this preliminary study, all tests were performed in a university classroom (seen in Fig. 1) with a public of 40 volunteers. Three lists of 100 monosyllables each were presented to the students during the actual articulation tests. The same 300 monosyllables were presented to the students during the virtual articulation tests of the first and second kinds (a total of 900 monosyllables). However, the time interval between these three articulation tests (more than one week) turns almost impossible to the subjects to memorize the words.

The average room simulation quality produced by the computer code RAIOS, at the considered case, can be estimated by comparing some of its outputs, i.e., some of the computer simulated acoustic quality parameters with the measured ones, in the actual room. Table 1 presents the comparative result, for two microphone positions, of the following parameters: Definition, D_{50} ; Clarity index (for words), C_{50} ; Clarity index (for music), C_{80} ; and Center time T_s (BERANEK, 1996).

Table 1: Comparative results among some of the measured and simulated acoustical parameters.

	Parameter	D_{50} (%)	C_{50} (dB)	C_{80} (dB)	T_s (ms)
Position 1	Measured	80,2	6,1	9,9	32,5
Position 1	Simulated	81,3	6,4	9,8	32,6
	Abs. Diff.	1,1	0,3	0,1	0,1
Position 2	Measured	85,2	7,6	11,9	26,6
Position 2	Simulated	84,9	7,5	10,9	26,3
	Abs. Diff.	0,3	0,1	1,0	0,3

As can be seen in Table 1, the computer code RAIOS provided a very good simulation result, when compared with the measured ones. For the four parameters shown, the obtained absolute difference is very slight. However, this is not an unexpected result since the

classroom under consideration (see, Fig. 1) is a very simple ‘shoebox’ one.

The comparative results among the actual average articulation index, the average articulation index obtained with the virtual auralization of first kind, and the one obtained with the virtual auralization of the second kind are displayed on Table 2. The values are presented for each list of monosyllables and the ensemble average, in boldface. The error columns indicate the difference between the results of the articulation index obtained with the virtual auralizations and the articulation index effectively measured in the room.

Table 2: Consolidated summary of the obtained articulation indexes (%).

List	Actual AI	Virtual AI (1 st kind)	Error 1	Virtual AI (2 nd kind)	Error 2
1	86,62	86,62	0,00	87,14	0,52
2	94,86	91,54	-3,32	92,21	-2,65
3	95,82	91,31	-4,51	91,29	-4,53
Average	92,52	89,82	-2,70	90,21	-2,32

As can be seen in Table 2, there is a distribution in the articulation indexes that depends on the list of monosyllables used. For the actual AI the values go from a minimum of 86.62% to a maximum of 95.82%, a difference of 9.20%. In practice, this means that List 1 seems to the listeners to be more "difficult" to recognize than List 2 and List 3.

Comparing the results by list, it is seen that, for List 1, the Error 2 becomes 0.52%, for List 2, the Error 2 becomes -2.65% and for List 3, the Error 2 becomes -4.53%, the biggest one. In any case, all errors are below $\pm 5\%$, a value considered acceptable for uncertainty in articulation indexes (MELO et al., 2013), and approximately one half of the difference found among the actual AI's for the three lists. Furthermore, it is worth noting that the absolute value of the ensemble average error,

i.e., the error for the 300 monosyllables, was lower than 2.5%.

6. Conclusions and Remarks

A new approach to validate acoustical virtual reality obtained by numerical simulation was presented. The adopted metrics is the articulation index obtained in the room with actual articulation tests. Two virtual articulation indexes were determined. The first one comes from binaural impulse responses measurements in the room under consideration, resulting in the auralizations of the first kind. The second one was performed strictly through computational simulation, resulting in the auralizations of the second kind. The main subject of this research is then to find out if the numerical auralization can be validated by using articulation indexes.

The results showed that the proposed metrics to validate computer modeling auralizations are reliable, with an average absolute error lower than 2.5%.

Of course, not all dimensions of an auralized sound are covered by the proposed technique. Indeed, there are aspects in the auralization, such as sound coloration and sound localization, that are not evaluated by the proposed metrics. Nevertheless, it was presented a simple way to quantify computer auralizations quality.

Similarly to the room acoustics sound quality parameters, other metrics are to be found. For example, the directional sound characteristics of a virtual auralization could be evaluated by using a specially prepared room with a matrix array of points to be selected by the headphones user. However, many details of such experiments are still under investigation (WERSNYI, 2009).

The proposed methodology, including the headphones equalization and the correction of the non-linearity of the sound source, among others, must be strictly followed to ensure that

the bulk of the spurious influences in the measurements are suppressed.

Usually, the output of a numerical simulator of room acoustics — and, as a consequence, the output of the artificial neural networks — brings the influence of the internal room acoustics, but the background noise is void. The inclusion of the background noise in the auralizations of the second kind (Step 12 of the methodology) proved to be necessary to introduce verisimilitude in the heard sounds and the actual disturbance that the background noise introduces in the articulation tests.

There are some studies that might accompany this one. The first one is to extend the tests to a great number of rooms, preferentially with different acoustic conditions. The second one is to introduce controlled noise in the room to check its effect on the articulation indexes. The third one is to verify if there is an important influence on the language in the results. Since this last can only be exercised by native language listeners, the authors leave here the suggestion to another research group to reproduce our tests in a different language.

7. Acknowledgments

The authors acknowledge the FAPERJ, for its financial support. We would like also to thank to the cooperation of the undergraduate student Lucas R. Santos.

References

- AJDLER, T., FALLER, C., SBAIZ, L. AND VETTERLI, M., 2005. Sound Field Analysis along a Circle and Its Applications to HRTF Interpolation. *J. Audio Eng. Soc.*, Vol. 56(3), 156–175.
- BERANEK, L. 1996. **Concert and Opera Halls: How They Sound**. Acoustical Society of America, Woodbury.
- BLAUERT, J., 1997. **Spatial Hearing**. The MIT Press, Cambridge.
- BORK, I. 2005. Report on the 3rd Round Robin on room acoustics computer simulation – Part 2: Calculations. *Acustica united with Acta Acustica*, Vol. 91(4), 753–763.

- CAMILO, T.S. AND TENENBAUM, R.A. 2002. Método híbrido para simulação numérica de acústica de salas - Parte 1: combinação dos métodos de traçado de raios e de transição de energia. **XX Encontro da Sociedade Brasileira de Acústica**, Rio de Janeiro. (In Portuguese).
- DALEMBÄCK, B.I., KLEINER, M. AND SVENSSON, P. 1994. A macroscopic view of the diffuse reflection. **J. Audio Engin. Soc.**, Vol. 42(10), 973–987.
- GARDNER, B., MARTIN, K., 1995. HRTF Measurements of a KEMAR Dummy-Head Microphone. **J. Acoust. Soc. Am.**, Vol. 97(6), 3907–3908.
- HAYKIN, S., 2009. **Neural Networks and Learning Machines**. Prentice Hall, New Jersey.
- ISO 3382**, 1997. Acoustics – Measurement of the reverberation time of rooms with reference to other acoustical parameters.
- KRUZINS, E., FRICKE, F.R., 1982. The prediction of sound fields in non-diffuse spaces by a ‘random walk’ approach. **J. Sound and Vib.**, Vol. 81(4), 549–564.
- LUCIO NARANJO, J.F.L., TENENBAUM, R.A. AND TORRES, J.C.B., 2010. Using Artificial Neural Networks to generate virtual acoustic reality applied on escape training in blind conditions **Int. Rev. Chem. Eng.**, Vol. 2(6), 754–759.
- LUCIO NARANJO, J.F.L., TENENBAUM, R.A., TORRES, J.C.B., 2013. Accuracy validation of the ANN’s interpolation technique through the use of acoustic quality parameters. **XVI Computer Modeling Meeting**, Ilhéus.
- MELO, V.S.G., TENENBAUM, R.A. AND MUSAFIR, R.E., 2013. Intelligibility assessment in elementary school classrooms from binaural room impulse responses measured with a childlike dummy head. **Applied Acoustics**, Vol. 74, 1436–1447.
- MÜLLER, S. & MASSARANI, P.M., 2001. Transfer-functions measurements with sweeps. **J. Audio Eng. Soc.**, Vol. 2001, 443–471.
- TENENBAUM, R.A., CAMILO, T.S., TORRES J.C.B. AND GERGES, S.N.Y., 2007a. Hybrid method for numerical simulation of room acoustics with auralization: Part 1 - Theoretical and numerical aspects. **J. Brazilian Soc. Mech. Sci. Engin.**, Vol. 29(2), 211–221.
- TENENBAUM, R.A., CAMILO, T.S., TORRES J.C.B. AND STUTZ, L.T., 2007b. Hybrid method for numerical simulation of room acoustics: Part 2 - Validation of the computational code RAIOS 3. **J. Brazilian Soc. Mech. Sci. Engin.**, Vol. 29(2), 222–231.
- TORRES, J.C.B., PETRAGLIA, M.R., AND TENENBAUM, R.A., 2004. An Efficient wavelet-based HRTF for auralization **Acta Acustica united with Acustica**, Vol. 90(1), 108–120.
- VORLÄNDER, M., 2008. **Auralization: Fundamentals of Acoustics, Modeling, Simulation, Algorithms and Acoustic Virtual Reality** Springer, Berlin
- WERSNYI, G., 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.