Proceedings of the XII International Symposium on Dynamic Problems of Mechanics (DINAME 2007),
P.S.Varoto and M.A.Trindade (editors), ABCM, Ilhabela, SP, Brazil, February 26 - March 2, 2007

Efficient Auralization by Grouping Directions and Modeling HRTFs Using Wavelets

Julio Cesar B. Torres [1], Roberto A. Tenenbaum[2], and Mariane R. Petraglia [1]

[1] Universidade Federal do Rio de Janeiro, Escola Politécnica/COPPE. juliotorres@ufrj.br, mariane@pads.ufrj.br

[2] Universidade do Estado do Rio de Janeiro, Instituto Politécnico. tenenbaum@iprj.uerj.br

Abstract: In the last ten years, it can be observed an increasing number of immersive audio systems. This growing is due mainly to new technologies trying to simulate the human sensation to be immersed in a real ambient. An example is the inclusion of audio tracks in DVDs recorded with dummy heads, trying to recreate the recording space in the reproduction. However, systems of acoustical virtual reality, also called auralization, require a very high computational complexity to reproduce the 3-D characteristics of the actual sound. One of the best ways to reduce this computational complexity is to model, in an efficient and realistic way, the transfer functions related to the human head, the HRTFs, using wavelets and sparse filters, which is reported in this paper. Furthermore, since the HRTFs are not much sensitive to subtle changes in directions (depending of course on the frequency range), a new scheme to group directions in a judicious way is described, with simulation results, to reduce even more the computational complexity. The main idea is that, since the selectivity of the wavelet filters is high, it is possible to use a common set of coefficients for the same band of all HRTFs of a region. An efficient auralization scheme, exploring the similarity among the model coefficients is presented, and some different grouping settling are tested. The main conclusion is that the adopted techniques improve noticeably the auralization with an error lower than 2 dB.

Keywords: Auralization, Room Acoustic Simulation, Wavelet Transforms

NOMENCLATURE

Greek Symbols
ϕ = elevation angle
θ = azimuth angle

INTRODUCTION

In the past ten years, a considerable growth of immersive audio systems has been observed, using loudspeakes or headphones. Such growth is mainly due to the development of new technologies and to the necessity of the human being to feel itself immerse in the audio-visual program. An example is the inclusion of audio tracks, recorded with dummy heads, in DVDs, which allows the listener to perceive the tridimensional characteristics of the sound at the recording event. However, this type of recording does not allow the listener to modify its position inside the sound field.

In order to allow the listener to interact with the audio system, modifying its position, orientation and even characteristics of the sound field, the acoustical virtual reality systems (AVR) had been created. These systems demand high degree of complexity to produce a sound equivalent to the one recorded with artificial heads. Even with the current technological development, is not possible the use of such systems in real time. Its use in realtime only becomes possible if simplifications were accepted in the system. However, such simplifications imply in the reduction of the quality and the faithfulness of the audio, when compared to non-simplified systems.

The complexity reduction of the acoustic virtual reality systems can be obtained by modeling more efficiently the sound field behavior. The receiver modeling is made through the Head-Related Transfer Functions (HRTFs) [1, 2], which correspond to pairs of impulse responses (HRIRs) measured for many directions around the receiver. In order to simulate that a sound source states in a given space position around the listener, an anechoic signal must be convolved with the HRIRs relative to this direction. Removing the influence of the reproduction system, such as performing a headphone equalization, the perceived sound should be identical to that one recorded in a free field or anechoic chamber.

An acoustical virtual reality system can simulate several sound sources. Even with a single source, the emitted sound waves may suffer multiple reflections in the room surfaces. Thus, for each possible wave-front direction arriving to the receiver, the sound source signal would have to be convolved with the respective HRIR

direction. Therefore, the more reverberant is a room, the greater is the number of directions necessary to generate the three-dimensional audio signal.

The human being has a limited capacity in recognizing accurately the direction of a sound source [16]. The average capacity of the human being to identify the direction of a sound source varies between 5° and 20° [2] and, therefore, a discrete set of directions can be used to measure the HRTFs without loss of the capacity of direction recognition. Generally, approximately 700 directions are used around the head, with the sound source placed between 1.0 and 1.2 meters, resulting in a set of 1400 HRTFs [4, 1]. The computational cost of a system with simultaneous processing of diverse directions can be reduced by diminishing the number of directions and/or reducing the length of the HRIRs. The reduction in the number of the processed directions can lead to degradation of the 3D sound field perception, since some directions in which the sound could reach the receiver would not be used in the simulation.

The reduction of the HRIR length would also intervene with the direction perception. However, if the spectral characteristics of each direction were kept, it would be possible to reduce its length without loss of the auralization quality. This reduction was carried out successfully through the modeling of the HRTFs with wavelet transforms and sparse filters [7, 6, 9], where a reduction of approximately 70% in the HRIR implementation was obtained. Thus, a HRIR that had originally 100 coefficients in the time domain, could be implemented by a set of 30 coefficients, plus the wavelet transform computational cost. Although this considerable computational profit, gotten with the wavelet modeling, the high redundancy of information of the set of HRTFs can also be used to reduce even more the computational load. In this direction, it was verified that, at the low-frequency sub-bands, HRTFs of near directions presents very similar behaviors. This similarity is due to the large wavelengths of the low-frequency sounds, which are not subjected to the diffraction produced by the head/torso. This difficulty in recognizing the direction of low-frequency sounds is reflected in the module of the HRTFs up to, approximately, 1 kHz, were the frequency response is almost flat.

Based on this HRTF modeling with wavelets, this article presents an analysis of how the signal processing can be reduced when sound is arriving from closed directions. This performance gain is obtained by considering the similarity of the sparse coefficients responsible for the low frequencies of the HRTFs. Through the analysis of the error generated with the proposed simplification and considering its application in a acoustical virtual reality system, the aperture angles of azimuth and elevation and the number of directions that can be grouped, without the affecting the 3D audio quality, are discussed in this paper.

HRTF CHARACTERISTICS

The HRTFs are functions whose frequency responses depends on the direction of the sound source. Figure 1 presents the magnitude of the frequency response of a set of HRTFs pertaining to the horizontal plane at the level of the ears as a function of the azimuth angle. This plane is equivalent to an elevation of 0° in a spherical coordinates system.

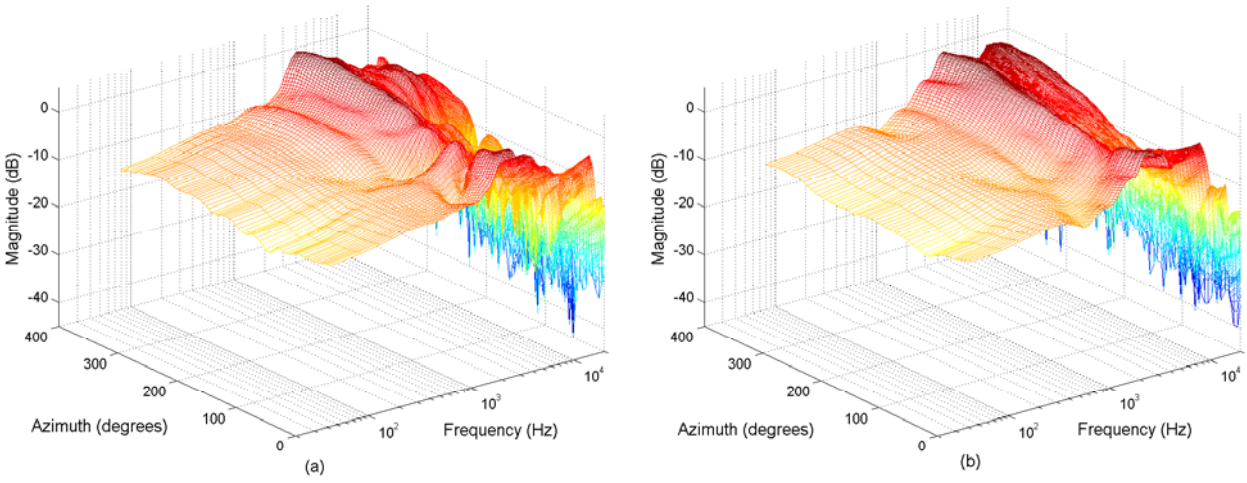

Figure 1 – Frequency response (magnitude) of the HRTFs for left ear and elevations (a) 0° and (b) 40°.

From Fig. 1 it can be observed that the low-frequency area (from 20 Hz to 1 kHz) does not have practically variation in the HRTFs magnitude as a function of the azimuth angle. The analysis for different elevations

presented a similar behavior. The variations in module and phase of the HRTFs and the differences between the HRTFs of distinct directions allow the identification of the localization of the sound source. Since the low-frequencies do not present significative differences, this part of the HRTFs does not supply information for recognition of the direction. For low-frequency sounds, prevails the interaural time and sound pressure level differences for a weak direction discrimination [14, 15].

HRTF MODELING WITH WAVELET TRANSFORMS

In this approach, the HRIR is seen as a finite impulse response (FIR) system and its modeling is based on the polyphase decomposition of the transfer function [12, 5, 13], as shown in Fig. 2.

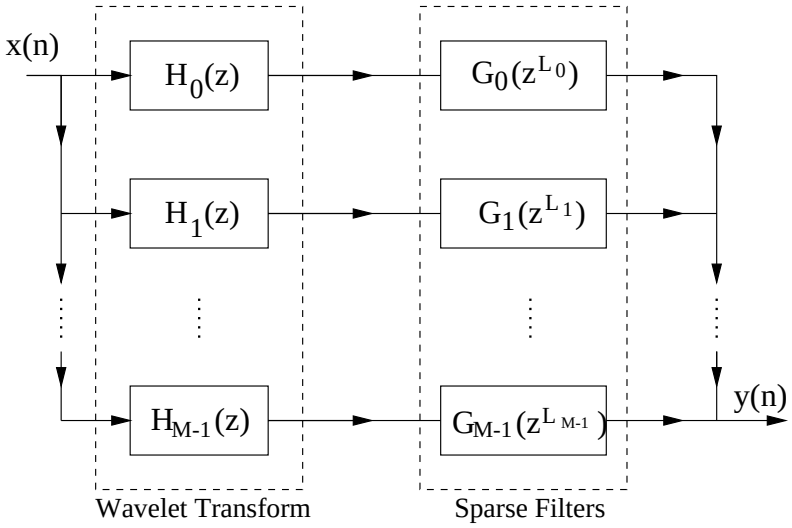

Figure 2 – FIR system for modeling a single HRTF with wavelet transforms.

In Fig. 2, the analysis filter bank $H_m(z)$, which implements a discrete wavelet transform, and the sparse filters $G_m(z^{L_m})$ provide an impulse response equal to the HRIR direction which is being modeled [8]. The analysis filters used for implementation of wavelet transform had been selected by presenting the best relation cost/benefit between the selectivity and transform length [9].

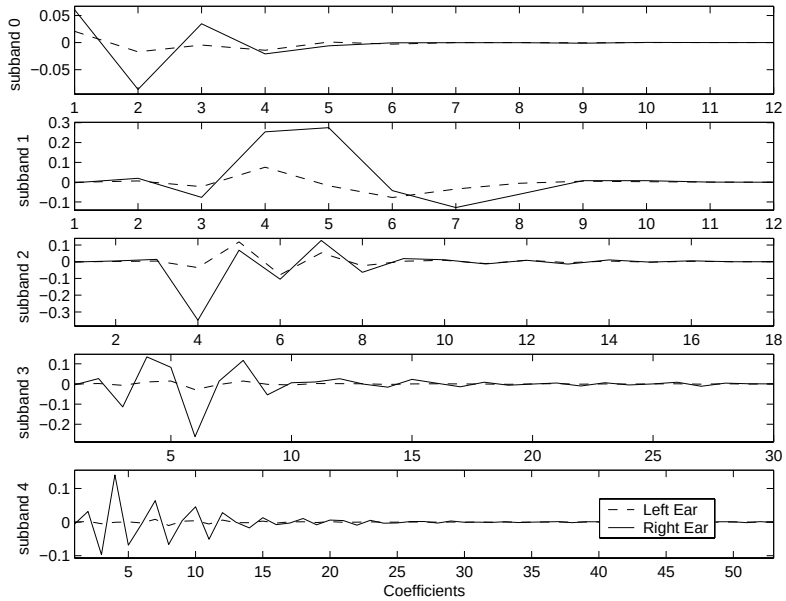

Figure 3 – Coefficients of the sparse filters for elevation 0° and azimuth 90°.

After tests with different filters, including biorthogonal ones, the prototype filters from Daubechies family

with length 8 (daub8) [3] had been used in four stages in a octave decomposition structure. Figure 3 presents an example of the modeled coefficients $G_m(z^{L_m})$ for each ear and for direction defined by elevation angle of 0° and azimuth angle of 90° (sound source positioned at 90° to the right of the listener).

COMPUTATIONAL LOAD REDUCTION

In this section, two techniques based on the spectral characteristics of the HRTFs and on the energy of the sparse coefficients are presented, in order to reduce the computational cost and to make the acoustical virtual reality systems more efficient. First it will be used a procedure to reduce the total number of sparse coefficients, considering an energy loss criterion. After that, the implementation cost of the HRTFs for near directions will be reduced, considering the similarity of the coefficients.

Reduction of the number of coefficients

The reduction of the number of coefficients is obtained through an analysis of the coefficients accumulated energy in each sub-band. However, the energy of each HRTF varies with the direction. The maximum and minimum values of energy occur for azimuth angles of 90° and 270°, respectively. In such way, an energy criterion may not have to be defined in absolute terms, but in percentages of energy in each sub-band, for each direction.

The energy of the HRIR $E(\phi,\theta)$ is given by

$$E(\phi,\theta) = \sum_{n=0}^{N-1} p^2_{\phi,\theta}(n), \tag{1}$$

where N is the length of the HRIR $p_{\phi,\theta}(n)$. The energy sub-band $E_m(\phi,\theta)$ is given by

$$E_m(\phi,\theta) = \sum_{k=0}^{K_m-1} g^2_{m,k}(\phi,\theta), \tag{2}$$

where K_m is the number of sparse coefficients of sub-band m.

The cumulative contribution of each sparse coefficient, in each sub-band, can be observed in Fig. 4, for the right ear and direction $\phi = 0°$ and $\theta = 90°$. The sum of the energies accumulated in each sub-band supplies the total energy of the HRIR.

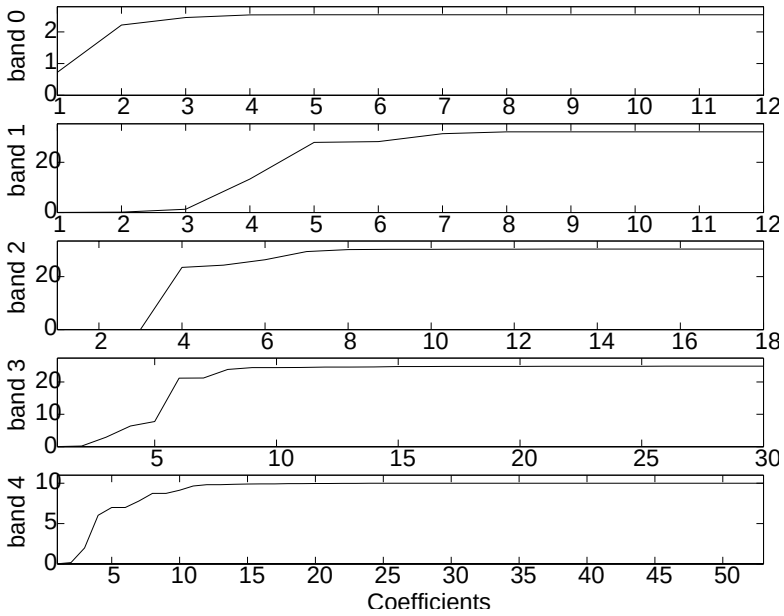

Figure 4 – Cumulative energy of the sparse coefficients for direction $\phi = 0°$ and $\theta = 90°$, right ear.

As observed in Fig. 4, the cumulative energy in the third band, for instance, only reaches considerable value after the third coefficient and has practically all energy accumulated up to the seventh coefficient. Thus, if

the coefficients up to the third position and after the seventh position were discarded, in this band, only five coefficients in this sub-band will remain. This same analysis can be applied to all sub-bands, however, defining limits in such that the total energy loss with the non-significant coefficients are at most 10% of the original HRIR energy. Applying the criterion described in [9], for all directions, the intervals (windows) described on Table 1 are obtained. These intervals guarantee that the maximum energy loss produced by reduction of the number of coefficients will be of 10%. However, for several directions the loss is not maximum. As shown in [9], the loss of 10% of the total energy by removing sparse coefficients lead to smaller errors in the frequency response than the loss produced by directly removing coefficients of the HRIRs (time domain). An analysis of the error due to reduction of the coefficients is presented in [11].

Prototype Filter	sub-band					total
Daub8	0	1	2	3	4	$\tilde{K}$
Intervals	1-6	3-7	4-7	3-9	3-8	
Number of coefficients	6	5	4	7	6	28

Table 1 – Intervals and number of kept coefficients for each sub-band.

Therefore, the number of coefficients can be reduced to approximately 30% of the total, if in each sub-band only the coefficients with more significance were considered. The energy loss with the discarding of coefficients is at most 10% of the total energy of the HRTF and does not modify significantly its spectral content. In the example presented in Fig. 4, the energy loss is only 4%, since the intervals from table 1 were obtained as an average of all available HRTF directions.

Reduction of the number of directions

The coefficients of each sub-band are responsible for a region of spectrum of the HRTF and the influence of these coefficients in adjacent sub-bands depends on the selectivity of the prototypes filters used in the octave structure (wavelet decomposition). Considering that the prototype used (daub8) presents a satisfactory relationship between selectivity and implementation cost (lengths of the filters $H_m(z)$ and delays produced), small variations in the values of the coefficients of bands 0 and 1 (lower frequencies) do not produce significant alterations in other sub-bands. On the other hand, variations in the coefficients of the last band produce alterations in the all sub-bands, due to low selectivity of the analysis filter of this band (all high-frequency content is concentred in this sub-band).

If one considers a region of the space around the receiver (defined by intervals of elevation and azimuth angles) [10], inside these regions, all HRTFs will be replaced by its reduced version. Analyzing the coefficients obtained in a given sub-band, for all directions pertaining to this region of space, it is observed that the coefficients relative to the low and middle frequencies present small variations. For higher sub-bands, the variation of the coefficients is larger. This is expected by two reasons: the low selectivity of the prototype filters of high-frequency sub-bands and the large variations in the spectrum of the HRTFs at high-frequencies.

Considering the direction $\phi = 0^\circ$ and $\theta = 90^\circ$ as the main direction and using an aperture angle of 40° for both elevation and azimuth, a region whose limits are $-20^\circ < \phi < 20^\circ$ and $70^\circ < \theta < 110^\circ$ is defined. Fig. 5 presents in the first column the coefficients of all the HRTFs belonging to this region, by sub-band. In this figure, the coefficients variation due to the direction change can be observed. In the second column, the curves corresponding to the average of the coefficients and to the average plus standard deviation are presented, by sub-band.

Analyzing the coefficients variation, it is verified that the largest deviations occur in the last two sub-bands. Considering that the variation of the coefficients at low-frequencies is very small, and that small variations are not capable to introduce considerable distortions in the frequency response, due to the wavelet filters selectivity, then it is possible to use a common set of coefficients for the same band of all HRTFs pertaining to a region, for low-frequencies.

Substituting the original coefficients of the first sub-band of a given HRTF inside a region by the average of the coefficients of all first sub-bands of the same region, is verified that this modification introduces very small variations in the magnitude and phase, which probably does not affects significantly the perception of the direction of the processed sound. This can be observed in Fig. 7, where the magnitude and phase of the frequency response of the original HRTF (original coefficients) are compared to the frequency response obtained by replacing the coefficients of the first sub-band by the average of the coefficients of all the first bands. Figure 6 presents the results obtained for the direction $(0^\circ, 90^\circ)$, for both ears. This behavior is similar to the other directions of this region.

Applying the average of the coefficients in the two first sub-bands, the same comparison is presented in Fig. 7.

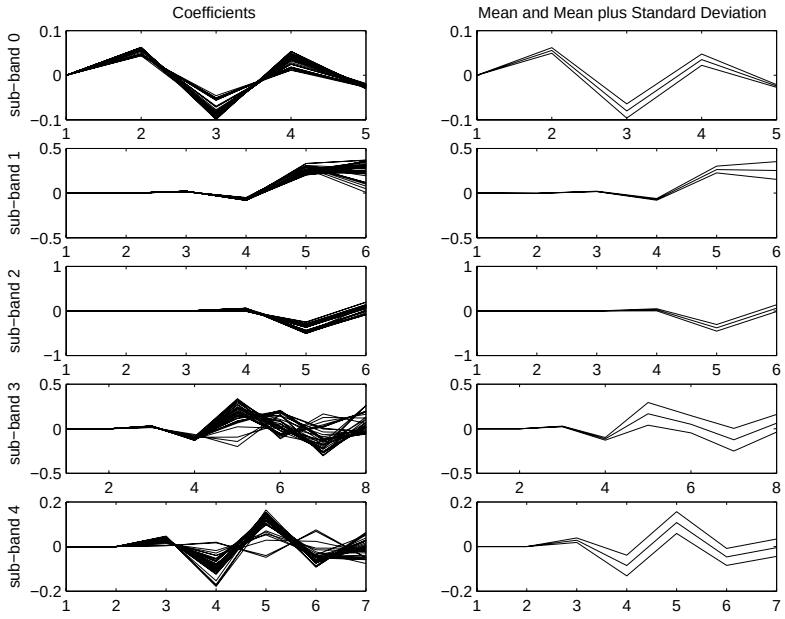

Figure 5 – (a) Variation of the coefficients for each sub-band inside a region and (b) mean and mean plus standard deviation of the coefficients.

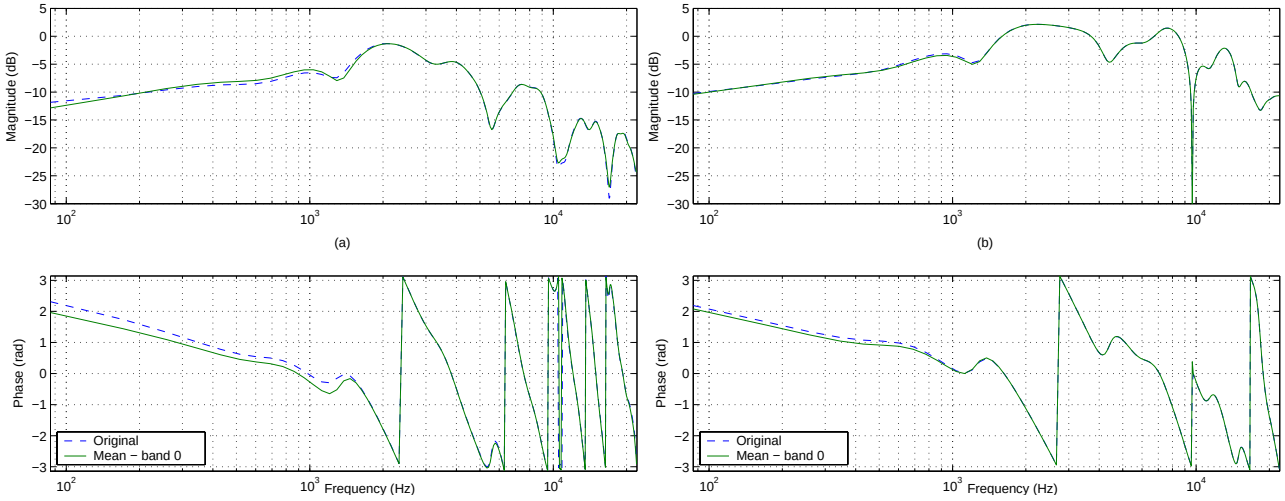

Figure 6 – Magnitude and phase comparison between frequency responses for direction $(0^\circ, 90^\circ)$ of region $-20^\circ < \phi < 20^\circ$ and $70^\circ < \theta < 110^\circ$, substituting the coefficients of the first sub-band by the mean coefficients: (a) Left ear and (b) Right ear.

Figure 8 presents the results obtained using the average coefficients on the three first sub-bands.

From the graphs presented in Figs. 6 to 8 it can be verified that the removal of some coefficients from the sparse filters and substituting these coefficients by the mean coefficients over the first sub-bands do not affect significantly the frequency response pertaining to the HRTFs to a given region of the space.

Therefore, a considerable computational gain can be obtained if only the last sub-bands were processed individually. Since the first sub-bands are equal for all directions inside the region, they can be processed once. Let us take as an example a region with 25 directions and each direction with 28 sparse coefficients, as shown in Table 1. Without using the proposed method, $25 \times 28 = 700$ operations of addition and multiplication would be necessary. Using average in bands 0 and 1 in substitution of original coefficients, it would be necessary only $11 + 25 \times 17 = 436$ operations, providing a reduction of 37,7 % in the computational load.

It is evident that how large will be the region (wide solid angles) larger will be the computational gain. The analysis presented in this article refers to regions with an a aperture angle of 40° around of a main direction. It is important to notice the relation of commitment between computational gain and the auralization quality, which

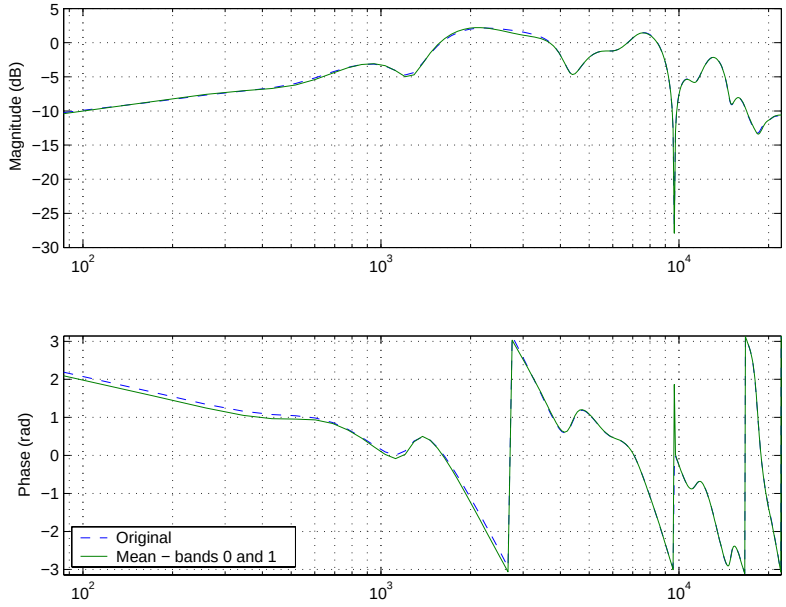

Figure 7 – Comparison between the frequency responses for direction $(0^\circ, 90^\circ)$, substituting the coefficients of the two first sub-bands by the mean values (region defined by $-20^\circ < \phi < 20^\circ$ and $70^\circ < \theta < 110^\circ$).

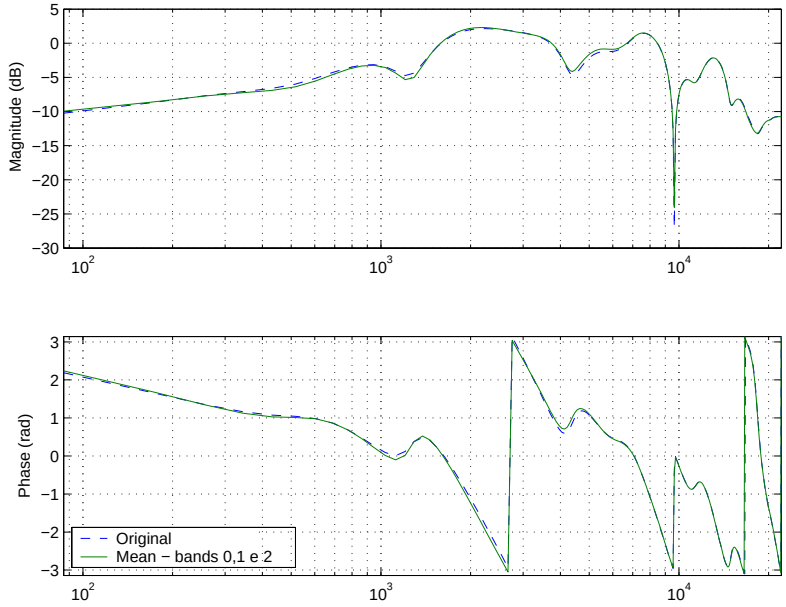

Figure 8 – Comparison between the frequency responses for direction $(0^\circ, 90^\circ)$, substituting the coefficients of the three first sub-bands by the mean values (region defined by $-20^\circ < \phi < 20^\circ$ and $70^\circ < \theta < 110^\circ$).

will be influenced by deviations in the frequency response of the HRTFs as function of the number of directions inside a region of the space. Thus, subjective tests still will be necessary in order to evaluate, psicoacoustically, what are the main angles of opening and directions that provide the best relationship between quality and computational gain.

CONCLUSIONS

In this paper a system for auralization with reduced computational complexity was presented, based on efficient model for the HRTFs and on the grouping of these functions for near directions. This grouping is possible due to similarity of the corresponding coefficients of the model at low-frequencies. Through the analysis of error generated by the proposed simplification, the solid angles (azimuth and elevation limits) can be derived, without introducing considerable loss in the quality of the 3D audio system, considering its application

in a reality system virtual acoustics (acoustics of rooms).

REFERENCES

[1] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano. The cipic hrtf database. In WASPAA '01 (2001 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics), October 2001. CIPIC website: http://interface.cipic.ucdavis.edu/.

[2] J. Blauert. Spatial Hearing. The MIT Press, Cambridge, 1997.

[3] I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. IEEE Trans. Inform. Theory, 36:961–1005, September 1990.

[4] W. G. Gardner and K. D. Martin. HRTF measurements of a kemar. J. Acoust. Soc. Am., 97(6):3907–3908, 1995. MIT website: http://sound.media.mit.edu/KEMAR.html.

[5] G. Strang and T. Nguyen. Wavelets and Filter Banks. Wellesley-Cambrigde-Press, Cambrigde, 1997.

[6] J. C. B. Torres and M. R. Petraglia. Performance analysis of an adaptive filter employing wavelets and sparse subfilters. In EUSIPCO 2000, volume II, pages 997–1001, Sep 2000.

[7] J. C. B. Torres, M. R. Petraglia, and R. A. Tenenbaum. Auralização de salas utilizando wavelets para modelagem das HRTFs. Seminário de Engenharia de Áudio, 2002.

[8] J. C. B. Torres, M. R. Petraglia, and R. A. Tenenbaum. HRTF modeling using wavelet decomposition. XIV Congresso Brasileiro de Automática, pages 2208–2213, Sep 2002.

[9] J. C. B. Torres, M. R. Petraglia, and R. A. Tenenbaum. An efficient wavelet-based HRTF model for auralization. Acta Acustica united with Acustica, 90(1), Jan 2004.

[10] J. C. B. Torres, M. R. Petraglia, and R. A. Tenenbaum. Low-order modeling and grouping of hrtfs for auralization using wavelet transforms. ICASSP 2004, 2004.

[11] J. C. B. Torres, M. R. Petraglia, and R. A. Tenenbaum. Low-order modelling of head-related transfer functions using wavelet transform. ISCAS 2004, 2004.

[12] P. P. Vaidyanathan. Multirate Systems and Filter Banks. Prentice-Hall, Englewood Cliffs, New Jersey, 1993.

[13] M. Vetterli and J. Kovacevic. Wavelets and Subband Coding. Prentice-Hall, Englewood Cliffs, New Jersey, 1995.

[14] F. L. Wightman and D. J. Kistler. The dominant role of low-frequency interaural time differences in sound localization. J. Acoust. Soc. Am., 91(3):1648–1661, March 1992.

[15] F. L. Wightman and D. J. Kistler. Monaural sound localization revisited. J. Acoust. Soc. Am., 101(2):1050–1063, February 1997.

[16] F. L. Wightman and D. J. Kistler. Resolution of front-back ambiguity in spatial hearing by listener and source movement. J. Acoust. Soc. Am., 105(5):2841–2853, May 1999.

RESPONSIBILITY NOTICE

The author(s) is (are) the only responsible for the printed material included in this paper.