LOW-ORDER MODELING AND GROUPING OF HRTFS FOR AURALIZATION USING WAVELET TRANSFORMS

Julio C. B. Torres[1]   *Mariane R. Petraglia*[2]   *Roberto A. Tenenbaum*[3]

[1,2]Federal University of Rio de Janeiro
COPPE / Electronic Eng. Department
CP 68504, ZIP 21945-970,
Rio de Janeiro, RJ - Brazil
julio.torres@ufrj.br, mariane@pads.ufrj.br

[3]Federal University of Rio de Janeiro
COPPE / Mechanical Eng. Department
CP 68503, ZIP 21945-970,
Rio de Janeiro, RJ - Brazil
roberto@serv.com.ufrj.br

ABSTRACT

In this paper, an efficient method for modeling Head-Related Transfer Functions (HRTFs) of an auralization system is presented. The proposed model is based on the decomposition of the impulse response of the HRTFs by wavelet transforms and on the grouping of such functions for close directions. Through an analysis of the HRTF energy content per subband it is shown how the model can be reduced without introducing considerable error in the magnitude and phase frequency responses. As a result of the proposed technique, a significant reduction in the processing time of the auralization process is obtained by a low-order model which has approximately 30% of the number of coefficients of the original HRTF and by a reduction in the number of HRTF directions to less than 10%.

1. INTRODUCTION

Nowadays, with the growth of the number of systems that use 3D-audio technology, the auralization process has become a very important tool for improving the realism of simulated virtual acoustic environments and designing real rooms. With auralization, it is possible to generate by numerical simulation the sound that would be heard anywhere in a room. Through such method, spatial sound perception is obtained without physically building the room nor measuring its acoustical characteristics. It only requires information about the walls' dimensions and absorption coefficients, and about the source and receiver characteristics (position, orientation and directionality).

To achieve high fidelity and realism of the auralized sound, the human audible characteristics must be preserved in the room simulation and in the sound reproduction steps. In this context, Head-Related Transfer Functions (HRTFs) [1] describe the acoustic transmission from a sound source located in a free field to the ear canal. The HRTF data consist of a set of directional transfer functions [2], which simultaneously convey the interaural time and level differences as functions of the frequency. Their monoaural cues are used to identify the source direction when the interaural differences are small [3, 4].

The source sound that propagates in the room is absorbed, diffracted and reflected on each surface, generating hundreds of new waves. The sound waves can be represented by rays, which indicate the propagation directions and contain the information about propagation time and absorbed energy as a function of the frequency. The propagation modeling of this paper is obtained by a hybrid method combining ray tracing and energy transition methods [5]-[7]. In simulations of typical rooms, such as a classroom or a music hall, the number of rays reaching a receptor cell (usually a sphere) can be very large, varying from 50,000 up to 500,000 rays. With such number of rays, the auralization process becomes extremely time costing, because of the number of convolutions needed to produce the binaural sound. The HRTF corresponding to the direction of each incoming ray has to be modified by the energy spectrum of the ray and delayed according to its arrival time.

The main problem for direct use of HRTF data resides in the large number of functions and in their large lengths. The set of HRTFs used in this work has approximately 1,400 functions, one for each direction and for each ear, as measured by the MIT Media Lab database [2].

In systems with a real-time auralization output or real-time updating, such as video-games, cinema sound effects and even in some room acoustic simulators, the complexity of such functions needs to be drastically reduced. This reduction usually leads to simplified simulations of the acoustical environment, and hence is limited to trivial effects of source positioning and reverberation.

The goal of this work is to investigate the influence of the substitution of the original HRTF for a given direction by a representative HRTF obtained by averaging a group of such functions. The representative HRTF is modeled by a low-order wavelet-based structure, which allows a computational gain without affecting considerably the original perception of the sound source. The wavelet modeling method, the grouping scheme and the auralization results are described in the next sections.

2. HRTF MODELING USING WAVELETS

Illustrated in Fig. 1, the proposed HRTF model consists of a wavelet transform, implemented as a non-uniform filter bank, and sparse filters. For tree structured filter banks, the analysis filters $H_m(z)$ of Fig. 1 are related to the low-pass and high-pass prototype filters, $H^0(z)$ and $H^1(z)$, by [8, 9]
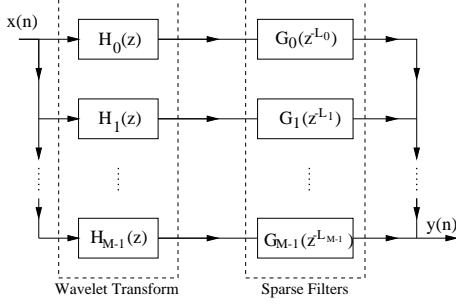
$$H_0(z) = \prod_{k=0}^{J-1} H^0(z^{2^k}), \qquad (1)$$

Fig. 1. Proposed structure for implementing the HRTFs.

and

$$H_m(z) = H^1(z^{2^{J-m}}) \prod_{k=0}^{J-1-m} H^0(z^{2^k}), \qquad (2)$$

for $m = 1, \ldots, M - 1$, where $M = J + 1$ is the number of subbands and J is the number of stages of the multi-resolution decomposition. The sparsity factors of the subfilters $G_m(z^{L_m})$ are given by

$$L_m = \begin{cases} 2^J, & m = 0 \\ 2^{J-m+1}, & m = 1, \ldots, J. \end{cases} \qquad (3)$$

The system transfer function $P(z)$ implemented by the structure of Fig. 1 can be written in terms of the analysis filters $H_m(z)$ and sparse subfilters $G_m(z^{L_m})$ as follows [10]:

$$P(z) = \begin{bmatrix} G_0(z^{L_0}) & \cdots & G_{M-1}(z^{L_{M-1}}) \end{bmatrix} \begin{bmatrix} H_0(z) \\ \vdots \\ H_{M-1}(z) \end{bmatrix}. \qquad (4)$$

The Daubechies and Biorthogonal wavelets [11] were here applied in the HRTF modeling, and the best results were obtained with length-8 Daubechies wavelet and $J = 4$ stages. The corresponding sparsity factors are $L_0 = L_1 = 16$, $L_2 = 8$, $L_3 = 4$, $L_4 = 2$. In the following section, all figures and results were obtained with such wavelet configuration.

3. REDUCED-ORDER MODEL

The impulse responses provided by the MIT database had originally 512 coefficients. After an analysis of such data, it was possible to remove the initial delay present in all directions and shortened the HRIR length to 128 samples, and still preserve more than 95% of the total impulse response energy. The largest errors occur for azimuth angles θ between $250°$ and $300°$, where the slowest decays of the Head-Related Impulse Responses (HRIRs) take place.

Figure 2 shows the coefficients of the subfilters $G_m(z)$ of the proposed model for the 128-length HRIRs of the left and right ears, considering a $0°$ elevation and $90°$ azimuth direction. In this figure we observe that, in all subbands, the coefficients with highest amplitudes appear at the beginning of the impulse responses, after a sequence of very small values. These large coefficients are the most important in each band for HRTF reconstruction, since they hold most of the impulse response energy in the corresponding subband. The remaining coefficients can be discarded.
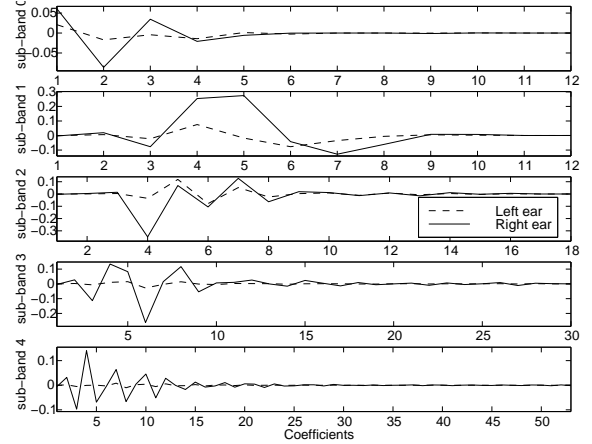

Fig. 2. Subfilters coefficients for HRIRs of the left and right ears for elevation $\phi = 0°$ and azimuth $\theta = 90°$.

An important issue when removing coefficients (or selecting the window in each subband HRIR) is the initial time delay related to the distance from the sound source to the entrance of the ears' canals, where the probe microphones are placed. In this work, such initial time delay was removed from the beginning of the original HRIR during the modeling stage and was used latter in the HRTF implementation.

The criterion applied to select the important coefficients of each subband m is based on their contribution to the subband HRIR energy $E_m(\phi, \theta)$, defined for a direction with elevation ϕ and azimuth θ by

$$E_m(\phi, \theta) = \sum_{k=0}^{K_m - 1} [g_{m,k}(\phi, \theta)]^2, \qquad (5)$$

where $g_{m,k}(\phi, \theta)$ is the k-th coefficient of the subfilter $G_m(z)$ and K_m is the total number of coefficients of such filter.

The cumulative energy contributions of the sparse subfilters coefficients to the total subband HRIR energies (obtained varying the number of coefficients K_m in Eq. 5) are shown in Fig. 3, using the right ear measurements for direction $\phi = 0°$ and $\theta = 90°$. From this figure, we verify that after a certain number of coefficients, the introduction of another coefficient in the energy calculus does not contribute effectively to the HRIR subband energy. On other hand, in some subbands, the first subfilter coefficients also do not contribute significantly to the subband HRIR energy. Based on these observations, the energy contributions of the sparse subfilters coefficients in each subband were computed for all HRTFs of the databank and the "relevant" coefficients were selected. The index interval corresponding to the coefficients which will be kept in the reduced model was obtained by comparing the energy contributions to two thresholds, THR$_{\min}$ and THR$_{\max}$, defined as follows:
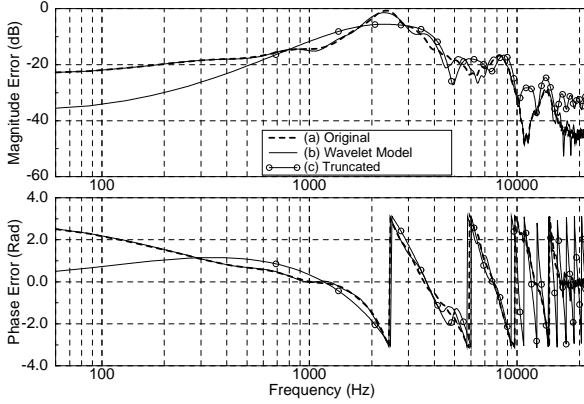
$$\text{THR}_{\min}(m) = \frac{\alpha_m}{2} E_m, \qquad (6)$$

$$\text{THR}_{\max}(m) = \frac{(1 - \alpha_m)}{2} E_m, \qquad (7)$$

for $m = 0, \ldots, M - 1$, where α_m is the tolerated energy loss and E_m is the total energy for subband m. Using the above thresh-
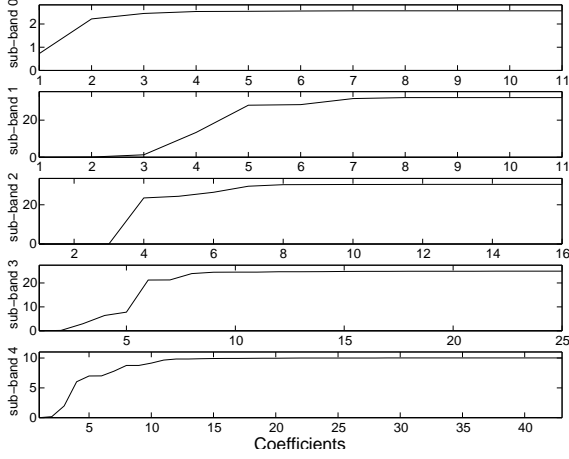
Fig. 3. Cumulative energy contribution of the subfilter coefficients of Fig. 2 for right ear.

Fig. 4. Frequency responses of original HRTF, proposed reduced-order model and HRTF truncated to 28 coefficients ($\phi = 0°$, $\theta = 90°$, left ear).

olds with a total energy loss $\alpha = 10\%$, the intervals corresponding to the "relevant" coefficients are shown in Tab. 1. The total number of coefficients of the resulting model is $\tilde{K} = 28$. The

| subbands	0	1	2	3	4
intervals	1-6	3-7	4-7	3-9	3-8

Table 1. Coefficients kept in the reduced-order model for each subband.

frequency responses (magnitude and phase) for the left ear at direction $\phi = 0°$ and $\theta = 90°$ of the original HRTF, of the proposed reduced-order model, and of the HRTF truncated to 28 coefficients are shown in Fig. 4. From this figure, we observe that for the same number of coefficients, the deviation in the frequency-domain using the truncated HRTF is always higher than using the proposed reduced-order wavelet model.

4. HRTF GROUPING

In this section we compare several HRTF grouping configurations and evaluate their effects on the auralized sound when using complete and reduced models for the HRTFs.

The space around the listener is divided in regions, with each region defined by an azimuth and an elevation angle range. Four grouping schemes were select to this investigation, shown in Fig. 5 (A) - (D). The representative HRTF of each group is obtained by taking
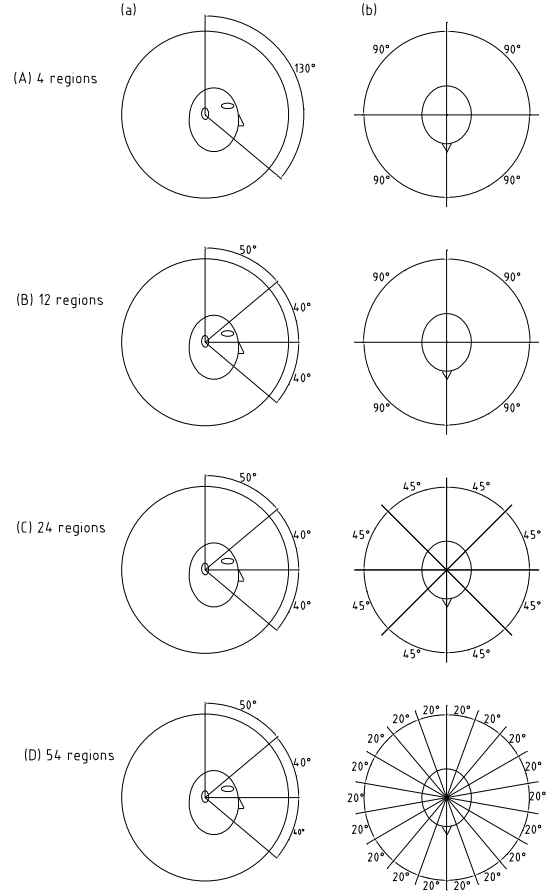
Fig. 5. Four grouping schemes for HRTFs: (a) side view and (b) top view.

the mean of all HRTFs of directions in that region. The representative HRTF model is composed by a set of sparse filters with the same number of coefficients of each HRTF model of that region.

To evaluate the influence of each grouping scheme and of the HRTF reduction in the auralization process, all rays which reached the receptor in the room acoustic simulator were processed in four different ways: (*i*) each ray was processed by its own HRTF direction without any model reduction nor grouping scheme; (*ii*) each ray was processed by its own HRTF direction applying model reduction; (*iii*) all rays in a region were processed by the representative HRTF of that region without model reduction using the grouping schemes of Fig. 5; and (*iv*) the same as in (*iii*) but also applying the model reduction to the representative HRTF.

For each case, the binaural impulse responses (BIRs) and the

normalised inter-aural cross correlation functions (IACF) were obtained. The IACF is defined as

$$\text{IACF}(\tau)_{t_1, t_2} = \frac{\int_{t_1}^{t_2} p_l(t) p_r(t + \tau) d\tau}{\sqrt{\int_{t_1}^{t_2} p_l(t) dt \int_{t_1}^{t_2} p_r(t) dt}} \qquad (8)$$

where $p_l(t)$ and $p_r(t)$ are the BIRs of the left and right ears, respectively. Usually the evaluation interval is given by $t_1 = 0$ ms and $t_2 = 80$ ms, for -1 ms $\leq \tau \leq 1$ ms. The IACF allows a preliminary evaluation of how the BIRs were affected in terms of correlation. Since the best case is the first one, the deviation of the IACF obtained with the model reduction or grouping from the IACF of the best case indicates the resulting degradation in the auralization process.

Fig. 6 shows the cross correlation functions for BIRs obtained with the simulation of a room with approximately 1.2 s of reverberation time. The first column of Fig. 6 shows the IACFs using HRTFs without reduction. The second column shows the IACFs applying the reduced HRTFs. The first row shows the results when no grouping scheme is used. From the second row to the last one the results for the grouping schemes described in Fig. 5 are presented.
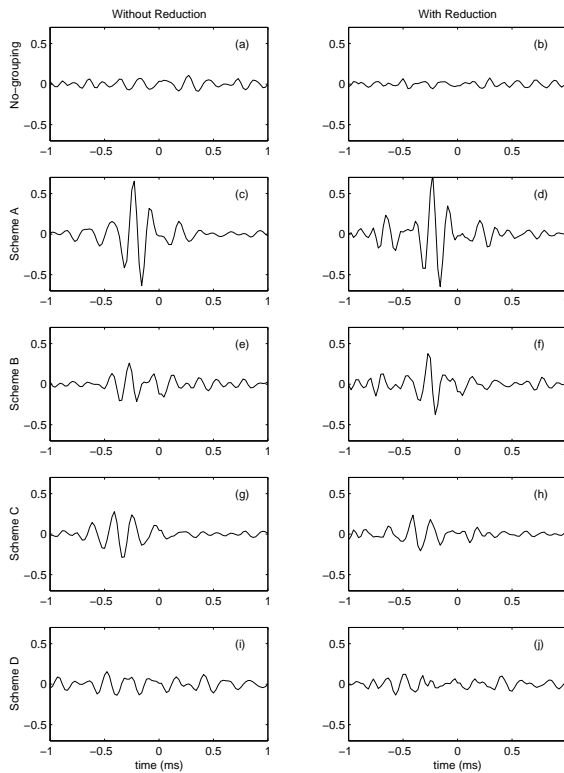

Fig. 6. IACFs for no-grouping and grouping schemes of Fig. 5.

Comparing the first and second columns of Fig. 6, it can be observed that the IACF does not change significantly when the reduction in the number of HRTF coefficients is applied, indicating that the HRTF characteristics are preserved. The IACF deviation introduced by the HRTF grouping is predominant, because a large number of rays are being processed by a small number of representative HRTFs. As the number of regions grows, the IACFs

approximate closely to the IACF obtained when no grouping or model order reduction is used. The IACFs obtained with grouping scheme D (Figs. 6(i)-(j)) are very similar to the those obtained with the non-simplified scheme (Figs. 6(a)-(b)), indicating that the corresponding BIRs will produce auralized sound very similar to the no-grouping scheme.

5. CONCLUSION

In this paper, a wavelet transform based model and grouping schemes have been applied to the HRTFs of an auralization system. The proposed wavelet model is able to implement an 128-length HRTF with only 28-coefficients. The interaural time and level differences are preserved with the reduced-order model applied to the right and left ears impulse responses, despite the small deviations in their frequency responses. It has been also shown, through evaluation of the inter-aural cross correlation function, that HRTF grouping together with the reduced-order wavelet model can be used to simplify the auralization process, without affecting significantly the source localization perception and the resulting sound quality.

6. REFERENCES

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

[2] W. G. Gardner and K. D. Martin, "HRTF measurements of a kemar," *J. Acoust. Soc. Am.*, vol. 97, no. 6, pp. 3907–3908, 1995, MIT website: http://sound.media.mit.edu/KEMAR.html.

[3] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 111–123, July 1993.

[4] F. L. Wightman and D. J. Kistler, "Monaural sound localization revisited," *J. Acoust. Soc. Am.*, vol. 101, no. 2, pp. 1050–1063, Feb. 1997.

[5] A. M. Ondet and J. L. Barby, "Modeling of sound propagation in fitted workshops using ray tracing," *J. Acoust. Soc. Am.*, vol. 85, no. 2, pp. 787–796, 1989.

[6] U. Stephenson, "Comparison of the mirror image source method and the sound particle simulation method," *Applied Acoustics*, vol. 29, pp. 35–72, 1990.

[7] D. Alarcão, J. L. B. Coelho, and R. Tenenbaum, "On modelling of room acoustics by a sound energy transition approach," in *Proc. of EEA symposium on architectural acoustics. Madrid*, 2000.

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

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

[10] J. C. B. Torres and M. R. Petraglia, "Performance analysis of adaptive filter structure employing wavelet and sparse subfilters," *IEE Proceedings - Vision, Image and Signal Processing*, vol. 149, no. Issue 02, pp. 115–119, April 2002.

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