

# LOW-ORDER MODELING OF HEAD-RELATED TRANSFER FUNCTIONS USING WAVELET TRANSFORMS

Julio C. B. Torres<sup>1</sup>

Mariane R. Petraglia<sup>2</sup>

Roberto A. Tenenbaum<sup>3</sup>

<sup>1,2</sup>Federal University of Rio de Janeiro

COPPE / Eletronic Eng. Department

CP 68564, ZIP 21945-970,

Rio de Janeiro, RJ - Brazil

*julio.torres@ufrj.br, mariane@pads.ufrj.br*

<sup>3</sup>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. 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, the low-order model has approximately 30% of the number of coefficients of the original HRTF, which represents an important reduction in the computational cost of an auralization system implementation.

## 1. INTRODUCTION

Nowadays, a large number of systems make use of 3D-audio technology to improve the realism of simulated virtual acoustic environments, such as home theaters, video games and room acoustic simulators. In this context, Head-Related Transfer Functions (HRTFs) describe the acoustic transmission from a sound source localized in a free field to the ear canal. This paper is concentrated on the modeling and reduction of the HRTFs used in room acoustic simulation for auralization [1]. The HRTF data consist of a set of directional transfer functions [2, 3], which code simultaneously the interaural differences of time and level as function of the frequency and their monaural cues are used to identify the source direction when the interaural differences are small [4, 5, 6].

The main problem for direct use of HRTF data resides in the large number of functions and in their large lengths. To keep the precision, a set of HRTFs must have approximately 1,400 functions, one for each direction and for each ear, as measured by the MIT Media Lab database [2], which is used in this paper.

In systems with 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 results in a simplified simulation of the acoustical environment, limited to simple effects of source positioning and reverberation.

A low-order HRTF model based on wavelet transforms and sparse filters is presented and analyzed in this paper. The HRTF model is reduced using an energy criterion to discard some coefficients of the sparse filters without affecting significantly the HRTF frequency response and allowing a faster auralization system.

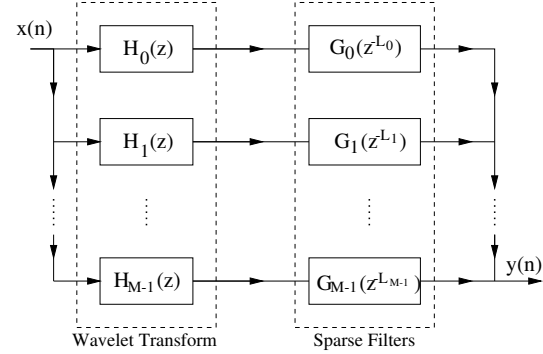


Fig. 1. Proposed structure for implementing the HRTFs.

## 2. HRTF MODELING USING WAVELETS

The proposed HRTF model consists of a wavelet transform, implemented as a non-uniform filter bank, and sparse filters, as illustrated in Fig. 1. 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 [7, 8]

$$H_0(z) = \prod_{k=0}^{J-1} H^0(z^{2^k}), \quad (1)$$

and

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

for  $m = 1, \dots, M-1$ , 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

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

The system transfer function  $P(z)$  implemented by the structure of Fig. 1 is related to the sparse subfilters  $G_m(z^{L_m})$  and anal-

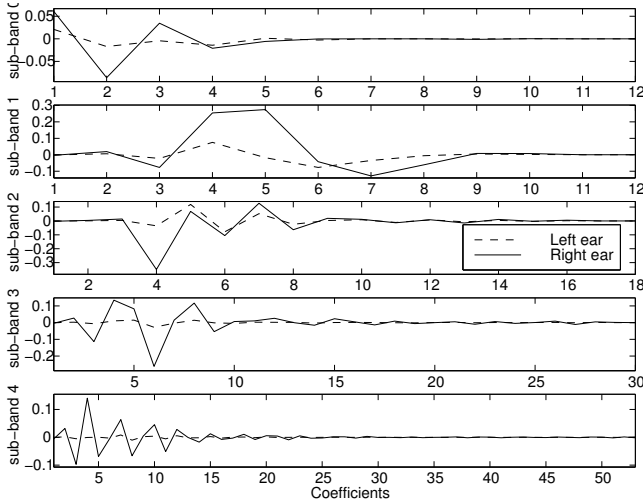
ysis filters  $H_m(z)$  by [9]

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

The Daubechies and Biorthogonal wavelets [10] were tested in the HRTF modeling, and the results which allowed a good compromise between transform computational load and selectivity was obtained with length-8 Daubechies wavelet and  $J = 4$  stages. The corresponding sparsity factors are  $L_0 = L_1 = 16$ ,  $L_2 = 8$ ,  $L_3 = 4$ ,  $L_4 = 2$ . All subsequent figures and results were obtained with such wavelet configuration.

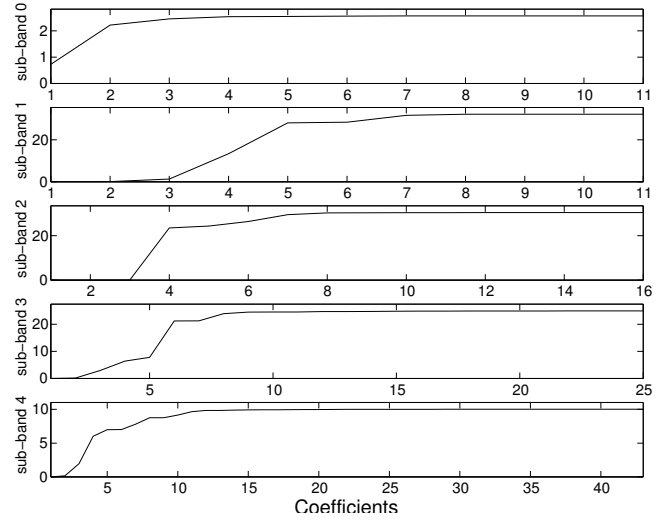
### 3. REDUCED-ORDER MODEL

The impulse responses provided by MIT originally had 512 coefficients. After an analysis of such data, it was possible to remove an initial delay present in all directions and shortened the HRIR length to 128 samples still preserving more than 95% of the total impulse response energy. The larger errors occur for azimuth angles  $\theta$  between  $250^\circ$  and  $300^\circ$ , when the decays of the impulse responses are the slowest ones, preserving from 90% to 95% of total energy. For the directions in the same side of the sound source, almost 99% of the energy is kept. Therefore, in the rest of the paper all Head Related Impulse Responses (HRIRs) will have 128 samples.



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

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 the  $0^\circ$  elevation and  $90^\circ$  azimuth direction. From this figure we can observe that, for 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, due to their low values when compared to the others of the same subband.



**Fig. 3.** Cumulative energy contribution of the subfilter coefficients of Fig. 2 for right ear.

An important issue when removing coefficients (or selecting the windowing of 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, by inspection, from the beginning of the original HRIR during the modeling stage and was used latter in the HRTF implementation.

#### 3.1. Energy Criterion for Order Reduction

The energy criterion applied to select the important coefficients should be defined not in terms of absolute values of energy, but in fractions of the energy for each direction, per subband, because the subband HRIR energy varies with the angle of incidence. The HRIR energy per subband  $E_m(\phi, \theta)$ , for a direction with elevation  $\phi$  and azimuth  $\theta$ , is given by

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

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  $\phi = 0^\circ$  and  $\theta = 90^\circ$ .

From Fig. 3, 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. This occurs due to the fast decay of the HRIR in each subband, which can be confirmed from Fig. 2. On other hand, in some subbands, the first subfilter coefficients also do not contribute significantly to the subband HRIR energy.

Based on the above 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  $\text{THR}_{\min}$  and  $\text{THR}_{\max}$ , defined as follows:

$$\text{THR}_{\min}(m) = \frac{\alpha_m}{2} E_m, \quad (6)$$

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

for  $m = 0, \dots, M - 1$ , where  $\alpha_m$  is the tolerated energy loss and  $E_m$  is the total energy for subband  $m$ .

Due to the wavelet decomposition, the low-frequency subbands have already a smaller number of coefficients than the high-frequency bands. Besides, a reduction in the number of coefficients in the low-frequency bands will affect mainly the low-frequency spectrum of the HRTF, interfering in the sound localization ability, if the interaural time cues were not present or were small [11]. In order to reduce these errors, the energy loss per subband  $\alpha_m$  should be smaller at low frequencies than at high frequencies.

Using the thresholds of Eqs. 6 and 7, considering a total energy loss of  $\alpha = 10\%$ , 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 subband energies vary with the HRTF direction, and so do the intervals of “relevant” coefficients. Since these variations are small, the mean values of the starting and ending points of the intervals obtained for all the directions were used. These intervals are applied to the HRTFs of

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

**Table 1.** Intervals of coefficients kept in the reduced-order model for each subband.

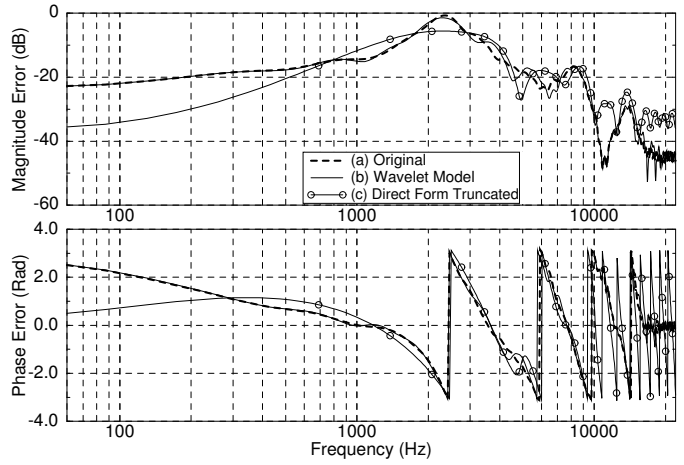
all directions. Since the individual intervals are different from the mean ones, we can not ensure that all HRIR models have energy loss of 10%. It will depend on the direction of the HRTF which is being modeled.

These coefficients intervals are applied to the HRIR, for the left ear, at direction  $\phi = 0^\circ$  and  $\theta = 90^\circ$ . The respective magnitude and phase frequency responses are shown in Fig. 4. In such figure the original HRTF frequency response (magnitude and phase) is compared to the reduced model and to the original HRIR truncated using the first 28 coefficients. For the same number of coefficients, the energy loss using the truncated HRIR is always higher than using the wavelet model, as well as the frequency error.

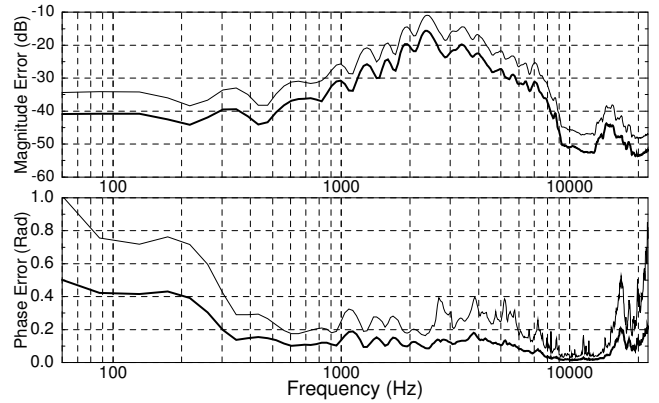
#### 4. EVALUATION OF THE MODELING ERROR

In this section, an evaluation of the interaural error in the frequency domain, introduced by the reduction in the number of coefficients of the wavelet model, is presented. The parameters chosen to quantify the influence of the model reduction on the HRTF characteristics are the mean and the standard deviation of the magnitude and phase frequency errors of the interaural differences between the original HRTF and the reduced model.

For all available directions, the interaural differences at a set of 512 frequencies were computed with the original HRTFs and with their reduced models, except for directions in the median plane ( $\phi = 0^\circ$ ), where the measured functions were identical and the interaural differences were zero. The mean magnitude and phase



**Fig. 4.** Magnitude and phase frequency responses of original HRIR, proposed reduced-order model and HRIR truncated with 28 coefficients ( $\phi = 0^\circ$ ,  $\theta = 90^\circ$ , left ear).



**Fig. 5.** Magnitude and phase errors of the interaural differences. Thick curve: average, thin curve: average plus standard deviation.

errors between the original and modeled interaural differences are shown in Fig. 5, as well as the corresponding standard deviation.

From Fig. 5, it can be observed that the average magnitude errors at low frequencies (up to 1 kHz) were small (less than -40 dB). At approximately 2.3 kHz, the curves presented a peak, exactly where the HRTFs have more energy. The phase errors at low frequencies (up to 1 kHz) were large, however, at such frequencies, the magnitude errors were small, minimizing the effects of the phase errors in the source direction localization. Around 2.3 kHz, where the magnitude errors presented the highest values, the phase errors were very small. The peaks of the phase errors curves at high frequencies (larger than 10 kHz) were mainly due to the discontinuities present in the phase responses of the HRTFs, which are slightly moved when the HRTF model is reduced. Again, such high-frequency phase errors are not noticed in the auralization, since the magnitude errors are small at this frequency band.

Figure 6 presents a graphical comparison between the frequency responses of the original and the proposed reduced-order model

for 19 incident angles over the horizontal plane ( $0^\circ$  elevation), from  $0^\circ$  to  $180^\circ$  azimuth at the left ear. From such figure, we observe that the proposed model preserves most of the frequency content of the original HRTF.

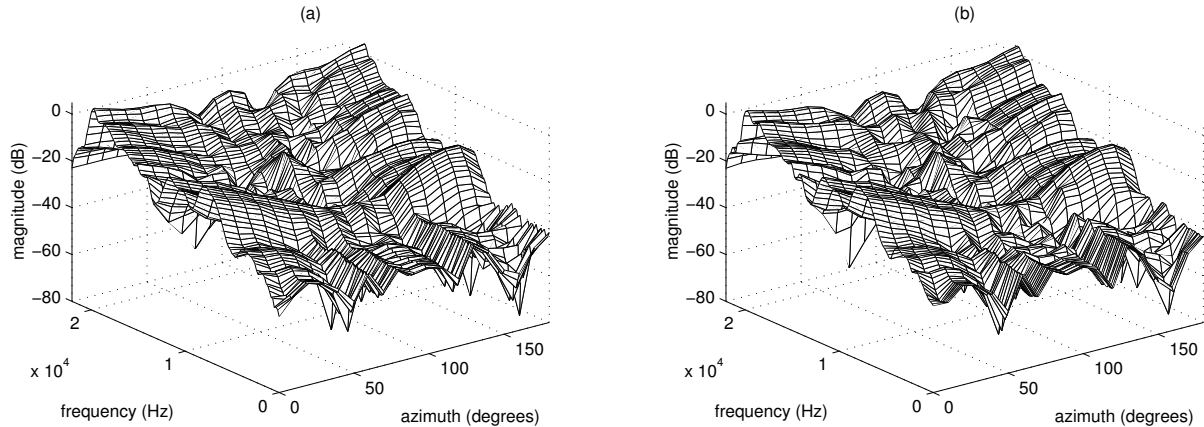
Finally, the proposed model is compared to IIR models. Tenth-order IIR models designed using Prony, Yule-Walker and balanced model truncation (BMT) [12] were considered. The average spectral signal-to-error power ratios (SERs) obtained for the 38 directions (from  $0^\circ$  to  $360^\circ$  and  $0^\circ$  elevation) were 26 dB with the wavelet reduced-order model, 19 dB with Prony and Yule-Walker methods, and 29 dB with the BMT process. Therefore, the proposed method had approximately the same average SER as the BMT model, with the advantage of not presenting the undesirable effects of the transient responses of IIR filters.

## 5. CONCLUSIONS

We have presented an effective method for modeling of head-related transfer functions using wavelet transforms. With the proposed reduced-order model, it is shown that with a total of 28 coefficients, the wavelet structure is able to model the 128-length HRTFs with very small errors. Such errors vary with direction and are larger in the ear opposite to the source side. In the frequency domain, the largest errors occur at frequencies from 1 to 4 kHz, where the HRTFs have more energy. Therefore, the corresponding percentage errors are small. The interaural time and level differences are preserved in the impulse responses of both ears. The modeling complexity can be further reduced by considering the redundancy of the reduced model coefficients for adjacent directions HRTFs.

## 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] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The cipic hrtf database," in *WASSAP '01 (2001 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics)*, Oct. 2001, CIPIC website: <http://interface.cipic.ucdavis.edu/>.
- [4] 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.
- [5] E. M. von Hornbostel and M. Wertheimer, *Über die Wahrnehmung der Schallrichtung [On the perception of the direction of sound]*, Sitzungsber. Akad. Wiss. Berlin, 1920.
- [6] F. L. Wightman and D. J. Kistler, "Monaural sound localization revisited," *J. Acoust. Soc. Am.*, vol. 101, no. 2, pp. 1050–1063, Feb. 1997.
- [7] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice-Hall, Englewood Cliffs, New Jersey, 1993.
- [8] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Prentice-Hall, Englewood Cliffs, New Jersey, 1995.
- [9] 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.
- [10] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge-Press, Cambridge, 1997.
- [11] F. L. Wightman and D. J. Kistler, "The dominant role of low-frequency interaural time differences in sound localization," *J. Acoust. Soc. Am.*, vol. 91, no. 3, pp. 1648–1661, Mar. 1992.
- [12] J. Mackenzie, J. Huopaniemi, V. Välimäki, and I. Kale, "Low-order modeling of head-related transfer functions using balanced model truncation," *IEEE Signal Processing Letters*, vol. 4, no. 2, pp. 39–41, Sept. 1997.



**Fig. 6.** Magnitude frequency responses of (a) 128-length original HRIR and (b) reduced-order wavelet model with 28 coefficients.