Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

# A head-related transfer function model for fast synthesizing multiple virtual sound images

**Zhiqiang Liang, Bosun Xie**

Acoustic Lab, Physics Dept., School of Science, South China University of Technology, Guangzhou, China 510641

**PACS:** 43.66.Pn Binaural hearing 43.60.Hj Time-frequency signal processing, wavelets.

## ABSTRACT

In realtime rendering of a virtual auditory environment, multiple virtual sound images may be synthesized simultaneously, which cost a lot of computation resource. The present work proposes a head-related transfer function (HRTF) model for fast synthesizing multiple virtual sound images. The head-relate impulse response (HRIR) of the KEMAR artificial head in horizontal plane is decomposed by using two-level wavelet packet. To simplify the model, for each wavelet packet tree node (subband), the beginning and ending parts of the coefficients, which are close to zeros, are discarded, while the main part of the coefficients, which contribute most to the HRIR energy, are preserved. The results show that, when an appropriate wavelet function is selected, coeffients with only 25 samples are sufficient to reconstruct the original HRIR.. The average error across all azimuthes caused by simplification is about 2.5% with a maximal error below 4%. The present HRTF model is very easy to implement by using wavelet filters and sparse filters. Its computational load is M*S+W, where M is the number of the sound images, S and W are the computational load of the sparse filters and wavelet filters. The coefficients of sparse filters are the upsample (zeros insert) from the wavelet coefficients, hence the length of nonzero coefficients are much less than that of the original HRTF filter. This means that the present HRTF model can save much computational resource when M is large.

## INTRODUCTION

In a virtual auditory display (VAD), the input stimulus is convoluted with the head-relate impulse responses (HRIRs) to synthesize binaural signals. Usually, the length of a measured HRIR varies from 128 to 4096 points (at 44.1kHz sample frequency) [1]-[3]. If a complex auditory environment with multiple virtual sound images is rendered, the cost of computation is high. Therefore, it is required to simplify the HRIRs or head-realted transfer functions (HRTFs) model used in computation.

The common method is to design low order FIR or IIR HRTF filters[4]-[7]. It has been shown that a FIR filter with length of about 60~70 points, or a IIR filter with length of 40~50 points (at 44.1 kHz ~ 48 kHz sample frequency) can model the HRTF with slight error [4]. The required order of an IIR filter is often less than that of a FIR filter, but the IIR filter should be designed carefully to avoid the unstability.

Cesar et.al presented a HRTF model based on wavelet[8]. The model consisted of a set of sparse filters, followed by wavelet decomposed filters. The coefficients of the sparse filters were obtained by an adaptive filtering algorithm or an analytical formulation. According to the author, the sparse filters could be reduced to have only 30 coefficiets to model the original HRTF. The error (energy loss) is about 10%. Cesar's model is very efficient in synthesizing mulitple virtual sound images of the same input stimulus, but is not quite efficient in synthesizing multipile virtual sound images of different input stimuli, because the wavelet decomposion have to be carried out for each stimulus.

To address this problem, a new HRTF model based on wavelet is presented in this paper. The proposed model makes the signal processing for auditory environment auditory environment much more effective.

## THEORY

Figure 1(a) and 1(b) show the structure of multiple level descrete wavelet decomposition and reconstruction. Where $H_0(z)$ and $H_1(z)$ are the wavelet low-pass and high-pass decomposition filters, $G_0(z)$ and $G_1(z)$ are the wavelet low-pass and high-pass reconstruction filters, respectively. They all can be calculated from the corresponding wavelet function. $C_0, C_1, \cdots, C_m$ are the wavelet coefficients of the input signal $r(n)$.

In figure 1, the downsampling can be moved to the position after the decomposition filters and the upsampling can be moved to the position before the reconstruction filters[9], the resulting equivalent structure of figure 1 are shown in figure 2. In figure 2:

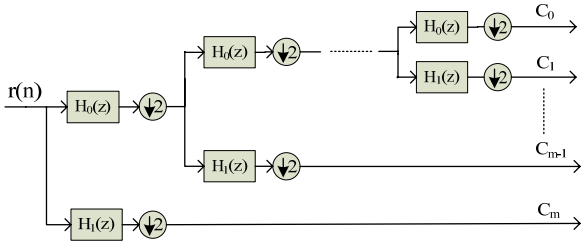
$$H^0(z) = \prod_{k=0}^{m-1} H_0\left(z^{2^k}\right) \tag{1}$$

$$H^i(z) = H_1\left(z^{2^{m-i}}\right) \prod_{k=0}^{m-1-i} H_0\left(z^{2^k}\right) \tag{2}$$

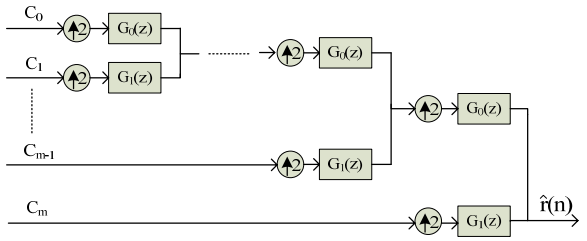
$$G^0(z)=\prod_{k=0}^{m-1}G_0\left(z^{2^k}\right) \quad (3)$$

$$G^i(z)=G_1\left(z^{2^{m-i}}\right)\prod_{k=0}^{m-1-i}G_0\left(z^{2^k}\right) \quad (4)$$

$$L_0=2^m,\ L_i=2^{m-i+1}\ \ (i=1,\cdots,m) \quad (5)$$




(a)




(b)

Figure 1. Discrete wavelet decompostion (a) and reconstruction (b)

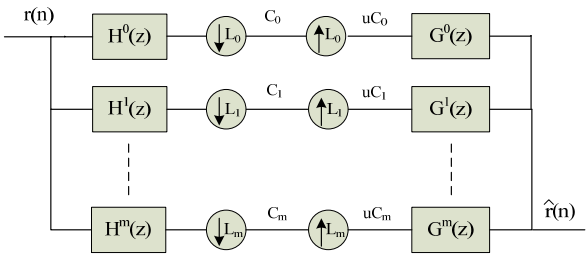



**Figure 2**. Equivalent structure of figure 1

In figure 2, $uC_0$,$uC_1\cdots uC_m$ are the upsampleing (zero interplotion) of the wavelet coefficients $C_0$,$C_1\cdots C_m$ , thus they have the same sampling frequency as the reconstructed signal $\hat{r}(n)$. Therefore,

$$\hat{r}(n)=uC_0(n)*g_0(n)+uC_1(n)*g_1(n)+\cdots+uC_m(n)*g_m(n) \quad (6)$$

$g_0(n)$,$g_1(n)$, $\cdots$ ,$g_m(n)$ are the time domain versions of the $G^0(z)$,$G^1(z)$,$\cdots$, $G^m(z)$ in figure 2. Most wavelet filter banks have the perfect reconstruction property. Hence the system output signal $\hat{r}(n)$ is the time delay of the input signal $r(n)$, that is:

$$\hat{r}(n)=r(n-N) \quad (7)$$

where $N$ is a constant. If the input signal $r(n)$ is a HRIR, it can be reconstructed according to equation (6). The reconstructed HRIR is the time delay version of the original HRIR. If the time delay is small, the reconstructed HRIR is indistinguishable from the original one by hearing.

As a result, according to equation (6), the proposed structure of the HRTF model is shown in figure 3. The model consists of two parts. The first part is the sparse filters $R_0(z^{L_0})$ , $R_1(z^{L_1})$ , $\cdots$ , $R_m(z^{L_m})$ , whose coefficients are $uC_0$,$uC_1\cdots uC_m$. The second part is the reconstructed filters $G_0(z)$,$G_1(z)$,$\cdots$,$G_m(z)$, which are identical to those in figure 2. And the corresponding coefficients can be calculated according to equations (3) and (4).

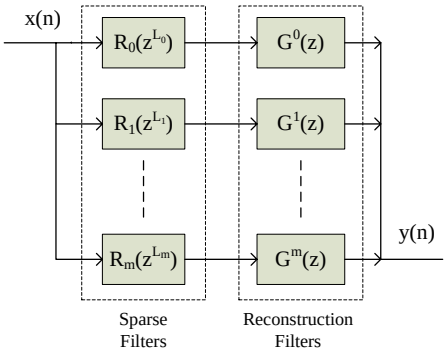



**Figure 3**. Structure of the proposed HRTF model

The coefficients of the spare filters $R_i(z^{L_i})$ ($I$ =0,$\cdots$,$m$) are the upsampling of the wavelet coefficients of the HRIR. Hence its nonzero coefficients are seperated by $L_i$-1 zeros. A spares filter can be implemented in an efficient way, as show in Figure 4. Where $b_0$,$b_1$, $\cdots$,$b_k$ are the nonzero coefficients of the sparse filter.

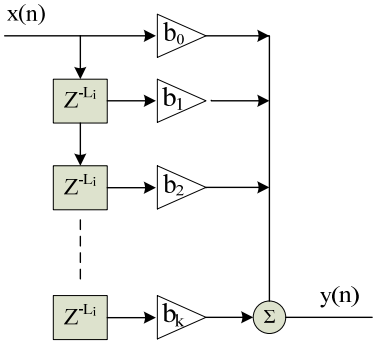



**Figure 4**. Implementation of the sparse filter $R_i(z^{L_i})$

The implementation of the sparse filter is very similar to the general FIR filter except for the delay unit of $L_i$ samples instead of just one sample. In actual programming, the delay operation is implemented by pointer shifting. Hence, it is very easy to modify the common FIR filter program to sparse filter program. Thus, the computation load of a sparse filter is not determined by its actual length, but is determined by the length of its nonzero coefficients. In the proposed HRTF model, the nonzero coefficients of the sparse filters are the wavelet coefficients of HRIR, thus the model can be simplified by compressing the wavelet coefficients of HRIR.

## WAVELET ANALYSIS

A HRIR can also be analysed by using wavelet packet. Modeling a HRTF by using wavelet packet analysis is almost identical to that by wavelet analysis.

There are many ways to analyze a HRIR by using wavelet or wavelet packet. To compress the wavelet coefficients effi-

ciently, it is needed to assign most energy in some subbands (nodes). The key is to select a proper wavelet tree and wavelet function. We tried some wavelet trees and wavelet functions. At the end, we choose the 2 level (4 subbands) wavelet packet, as shown in figure 5. Seveval orthogonal or biorthogonal wavelet functions, such as "Daubechies"," Coiflets"," Symlets" are good choice.

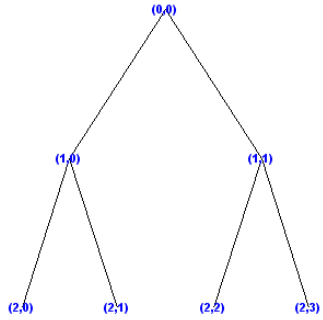



**Figure 5**. 2 level wavelet packet tree

In this study, the left-ear HRIRs of KEMAR at 72 horizontal directions from MIT media Lab were used [2]. The HRIRs were sampled at 44.1 kHz with length of 512 points. The initial delay in the measured HRIRs (detected by the leading edge of the 10% of the maximum) was removed at first10]. And then the preceding 128 samples of the data were remained and called "original HRIR" in this study. The original HRIR was then analyzed by using 2 level wavelet packet

Figure 6 shows the analysis result of the HRIR at azimuth $\theta$ = 90° . The wavelet function is 'Coif5'. It can be seen that the node(2,0) contains the most energy of the HRIR, and node (2,1) and (2,3) contain a little energy, while there is almost no energy in node(2,2). Moreover, in each node, most of the energy is represented by the largest coefficients. If these coefficiens are kept and the coefficients with small value in the begining and end part are discarded, the most energy of the original HRIR is preserved.

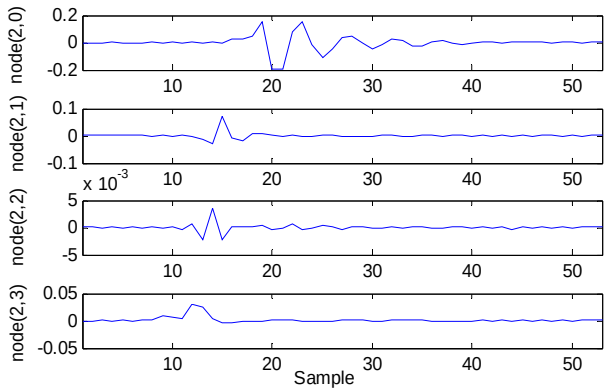



**Figure 6**. Wavelet coefficients of the HRIR at $\theta$ =90°

## SIMPLIFICATION OF THE MODEL

The nonzeros coefficients of the sparse filters in the model are the wavelet coefficients of HRIR. Thus, the model can be simplified by reducing the number of the wavelet coefficients.

A threshholding algorithm was used to simplify the model. At each node of the wavelet tree, the begining and end part of the wavelet coefficients, whose absolute value are smaller than the thresholded, are set to zero, and the middle part is unchanged, as show in figure 7.

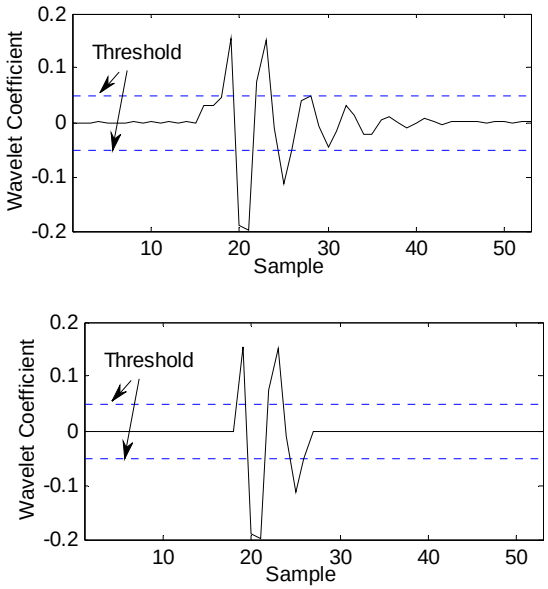



**Figure 7**. Wavelet coefficients that is before (up) and after (below) thresholding process

The threshold for each node is identical. However, it is not fixed and depends on the number of the wavelet coefficients that is preserved. For example, if more wavelet coefficients are preserved for a better reconstruction of the HRIR, a larger theshold is required. Hence, an algorithm for dynamic searching the threhold according to the length of wavelet coefficients that are preserved has been designed.

As an example, after thresholding process, the wavelet coefficients in Figure 6 become the results showed in figure 8. There are only 25 nonzero coefficients in the all nodes, while these 25 coefficients contain most information of the original HRIR. With these 25 coefficients, the error of the resulting model is only 0.7%.

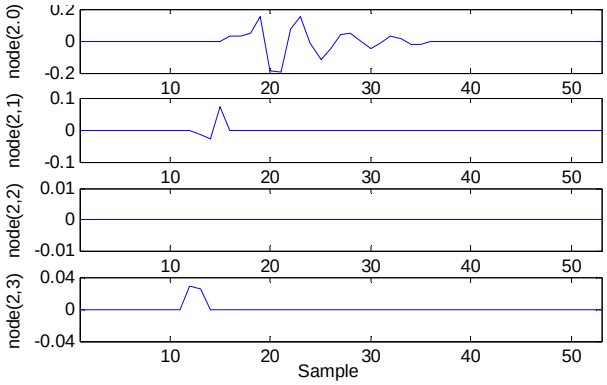



**Figure 8**. The result after thresholding process in figure 6

After thesholding process, the zeros at the beginning and ending can be cut. However the number of the zeros at the beginning must be saved because they are related to the time delay of each node (subband). They are modeled as the initial delay before each sparse filter.

## MODEL DELAY

As shown in equation (7), the model introduces a time delay compared with the original HRIR. The delay depands on the wavelet function. However, it is usually very small. Tabel I shows the delay and corresponding wavelet function. More-

over, the delay can be reduced greatly by reducing the initial delay before the sparse filters. But this operation should be done carefully . The reduction should be identical for all nodes (subbands) of all HRIRs.

**Table 1**. Model Delay (samples)

| *wavelet* | db4 | coif5 | sym9 |
|---|---|---|---|
| *Model dealy* | 18 | 84 | 48 |

## MODELING ERROR

The relative energy error is used here to evaluate the error between the original HRIR and the modeled HRIR[11], as defined by follows:

$$\varepsilon = \frac{\sum_n \left| h(n) - \hat{h}(n) \right|^2}{\sum_n \left| h(n) \right|^2} \times 100\% \qquad (7)$$

where $h(n)$ is the orginal HRIR and $\hat{h}(n)$ is the reconstruted HRIR from the model. The error depends on the number of wavelet coefficients preserved and the wavelet function. It also depends on the source directions. To give an overview, the average (across 72 horizontal source directions) and maximum error for all HRIRs are calculated and shown in table 2 and table 3

**Table 2**. Average relative energy error (%)

| *Coefficient number* | 20 | 25 | 30 | 35 |
|---|---|---|---|---|
| *db4* | 4.78 | 2.51 | 1.40 | 0.84 |
| *coif5* | 4.14 | 2.15 | 1.15 | 0.65 |
| *sym9* | 4.16 | 2.20 | 1.17 | 0.68 |

**Table 3**. Maximum relative energy error (%)

| *Coefficient number* | 20 | 25 | 30 | 35 |
|---|---|---|---|---|
| *db4* | 7.61 | 3.94 | 2.84 | 1.81 |
| *coif5* | 8.01 | 3.80 | 2.12 | 1.74 |
| *sym9* | 6.92 | 3.75 | 2.10 | 2.00 |

## COMPUTATIONAL LOAD

The computational load of the model comes from two parts: reconstruction filters $G^i(z)$ ($i$ from 1 to 4) and the sparse filters $R_i(z^{L_i})$ ($i$ from 1 to 4). The length of the reconstruction filter depends on the wavelet function, as show in table 4

**Table 4**. Length of the reconstruction filter $G^i(z)$ (i=1 to 4)

| *wavelet* | db4 | coif5 | sym9 |
|---|---|---|---|
| *Length* | 23 | 89 | 53 |

The computional load of the sparse filters depends on the preserved wavelet coefficients. For example, if 25 wavelet coefficients are preserved, the nonzero coefficients of all the 4 spares filters are 25 points totally. If db4 is chosen as the wavelet function, the total length of the reconstruction filters is 96 points. Thus, the computational load of the model is approximately equivalent to that of a 117 points FIR filter. However, this is the case of synthesizing a virtual sound image only. When multiple virtual sound images are synthesized, only the computational load of the sparse filters is multiplied, the computational load of the reconstruction filters is unchanged. This means that the model is very efficient in synthesizing multiple virtual sound images.

## CONCLUSION

A HRTF model is presented in this paper. The model consists of two parts: the reconstruction filters and the sparse filters. The coefficients of the reconstruction filters can be calculated from the wavelet reconstruction low-pass and high-pass filters. The coefficients of the spares filters are the upsampling of the wavelet coefficients of original HRIR. Therefore, the model is very easy to implement.

After analyzing the HRIR by using 2 level wavelet packet, some wavelet coefficients are preserved to bulid up the model. Error of the model depends on the wavelet function and the number of the wavelet coefficients preserved. Results indicate that when 25 coefficients are used to build the model, the relative energy error is about 2.5% for db4 wavelet.

When multiple virtual sound images are synthesized, only the computational load of the spares filters is multiple, the reconstruction filters can be operated once. Thus, the model is very efficient for synthesizing multiple virtual sound images.

ACKNOWLEDGMENTS: The author acknowledges the support of the National Nature Science Fund of China Grant No. 10774049

## REFERENCES

1 Blauert J, Brueggen M and Bronkhorst A W, et al “The AUDIS catalog of human HRTFs” *J. Acoust. Soc. Am*. **103(5)**, 3082 (1998)

2 Gardner W G, Martion K D “HRTF measurements of a KEMAR” *J. Acoust. Soc. Am*. **97(6)**, 3907-3908 (1995)

3 Genuit K, Xiang N “Measurements of artificial head transfer functions for auralization and virtual auditory environment” Proceeding of 15th International Congress on Acoustics (invited paper), Trondheim, Norway, II 469-472 (1995)

4 Sandvad J, Hammershøi D “Binaural Auralization: Comparison of FIR and IIR Filter Representation of HRIRs” AES 96th Convention, Amsterdam, The Netherlands, Preprint:3862 (1994)

5 Kulkami A, Colburn H.S “Efficient finite-impulse-resonse filter models of the head-related transfer function” *J. Acoust. Soc. Am*. **97(5)**, 3278 (1995)

6 Blommer M.A and Wakefield G.H "On the design of pole-zero approximations using a logarithmic error measure" IEEE Trans.Signal processing, **42(11)**, pp. 3245-3248 (1994).

7 Mackenzie J, Huopaniem J, et al “Low-order modelling of headrelated transfer functions using balanced model truncation” Signal Processing Letters, IEEE **4(2)**,39-41(1997)

8 Julio Cesar B. Torres, Mariane R. Petraglia, Roberto A. Tenenbaum “An efficient wavelet-based HRTF model for auralization” ACTA ACUSTICA UNITED WITH ACUSTICA **90**,108 – 120 (2004)

9 Strang G., Nguyen T. “Wavelet and Filter Banks” (Wellesley-Cambrigde Press, Cambrigde, 1997) pp.100-102

10 ZHONG Xiao-li “Criterion Selection in the Leading-edge Method for Evaluating Interaural Time Difference”(in Chinese) Audio Engineering **31(9)** 113-118(2007)

11 Xie, B. S., “Head Related Transfer Function and Vir-tual Auditory” (in Chinese), National Defense Industry Press, 2008 pp.106-107