A SPATIAL HEARING MODEL BASED ON RECONSTRUCTION FROM WAVELET TRANSFORM MODULUS MAXIMA

Jie Zhang *Zhenyang Wu*

Department of Radio Engineering, Southeast University, Nanjing, 210096, P. R. China

Email: zjhf1978@sohu.com zhenyang@seu.edu.cn

ABSTRACT

In the research of spatial hearing and realization of virtual auditory space, it is important to accurately model the acoustical characteristics of HRTFs (Head-related Transfer Functions) or HRIRs (Head-related Impulse Responses). In our study, for the sake of modeling HRIRs more truthfully, aiming at HRTFs' characteristics in time domain, we managed to carry through adaptive non-linear approximation in the field of wavelet transformation and introduced a new spatial hearing approximation model based on translation-invariant a' trous algorithm and reconstruction from modulus maxima. And the simulation results show that, this is a more effective HRIRs' model, averagely 7.7dB better than the traditional PCA (Principal Component Analysis) approximation based on relative MSE criterion.

1. INTRODUCTION

In the research field of spatial hearing, people have long ago taken cognizance of the cues implied in a little grotesque spectrum. Thus the systemic description of the entire relevant information, HRTF, has been introduced to describe them, where interaural time difference (ITD) can be seen as the delay between lateral HRTF and low-amplitude contralateral HRTF in transmission time lags; and interaural intensity difference (IID) can be their difference in magnitudes. And people have also constructed different mathematical models to simulate this delicate function, and try to give a reasonable interpretation for this fancy phenomenon [1-4].

As a popular model structure, PCA model has been greatly studied in this field [2-4]. Basically, [2] applied PCA to logarithmic magnitudes of HRTFs measured for 10 subjects and at 256 positions. In this model, HRTFs for each position were represented as a weighted combination of a set of basis functions in low-dimensional subspaces, which were obtained from the measured logarithmic magnitudes of HRTFs; whereas the phase was approximated from the magnitude related to certain position based on minimum phase characteristics. The spatial feature extraction and regularization model put forward in [3] contains the phase information during the HRTFs' approximation representation. Furthermore, a thin-plate spline using regulation procedures was also introduced to obtain a mathematical representation for sampled spatial positions. As a result, unmeasured positions' HRTFs can be approximately reestablished from the mathematical equation, which effectively conquers the spatially discrete limitation of HRTFs' measurement. However, these two PCA models were all used to process complex-valued (including amplitude and phase) HRTFs, and needed a lot of complex and logarithmic operations. And in [4], Wu applied Karhunen-Loève transformation to HRIRs in time domain to obtain another PCA model, which only required real-valued operations and behaved effectively for real-time implementation of VAS (Virtual Auditory Space).

However, traditional orthogonal bases generally cannot give attention to two aspects of signals' temporal and spectral characteristics; while due to the compact support wavelet functions' capability in time-frequency localization and multi-scale analysis, they have extensive applications in signal's nonlinear optimal and all-sided approximation [5]. Our work just finds a basis from this, and considers carrying through adaptive non-linear approximation in the field of HRIRs' wavelet transformation. Recently, there are also some reports of HRTFs' modeling by wavelet transformation, for example, [6] constructed a group of sparse filters to model HRTF based on wavelet's multi-scale characteristics, whose results showed the excellence of filter-bank's model than HRTF's conventional filter design algorithms, Prony, Yule-Walker and BMT. Whereas based on the distributing characteristic identification for all the HRIRs' data in KEMAR package, our work in this paper points out that, because HRIRs cannot accord with Gaussian distribution commendably, the PCA approximation model based on Karhunen-Loève transform isn't optimal and non-linear method can work well. As for this, we carry out detailed data processing experiments and validate our opinion.

To sum up, the paper is organized as follows. In Section 2, some characteristics of HRTFs are given as the bases for our work. And in following Section 3, the method of HRIRs' adaptive non-linear approximation based on wavelet transformation modulus maxima is introduced in detail. And then the contrastive results with HRIRs' PCA model are presented in Section 4. After that, several next-step directions are brought forward for future study.

2. ANALYSIS OF HRTFs' CHARACTERISTICS

For this kind of signals just as HRTFs, Batteau made some investigations on the equivalent HRIRs in time domain [7]. And he found that, the impulse response could be seen as superposition of signals from echoes reflected on outer ears,

$$h(t) = \delta(t) + a_1\delta(t-\tau_1) + a_2\delta(t-\tau_2), \qquad (1)$$

where a_1 and a_2 are the reflection coefficients; τ_1 and τ_2 are the time delay of reflected signals. Furthermore, when the elevation angle moves towards lower latitudes, the time delays usually increase. Hiranaka and Yamasaki further validated the above phenomena; and found that, all the reflection delays are less than 350μs for human through experimentation. Moreover, they also found that, when sound source locates at frontage of human body, there are at least two reflection waves; at the backside, there is only one; while on top of head, there are scarcely any reflection components [8]. Hebrank and Wright also confirmed the relationship between HRTFs' characteristics and delays of sound's reflections [9].

All these reveal the facts that there probably exist some singularities of HRTFs in time domain, and suggest us to utilize different singularities of different signal components in modeling

This work is supported by 973-Project of China under Grant No. 2002CB312102.

HRTFs. Our work is just based on the above contents, and introduces a new spatial hearing approximation model utilizing the translation-invariant *a' trous* algorithm and reconstruction of modulus maxima [5, 10].

Furthermore, it is well known that, PCA is the optimal approximation representation of a group data under linear condition [5]. If $H(n)$ is a stochastic vector composed of all the HRIRs of KEMAR package, the HRIR $h_l(n)$ on a certain position can be expressed as [4],

$$h_l(n) = \mathbf{Q}\mathbf{w} + \mathbf{h}_{av} = \sum_{i=1}^{N} w_i(n,\theta_l,\varphi_l)\mathbf{q}_i + \mathbf{h}_{av} + \varepsilon_l \qquad (2)$$

where $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_N]$, satisfying the normalized condition: $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = I$, is composed of the eigenvectors $\mathbf{q}_i$ of $H(n)$'s autocovariance matrix $\mathbf{R}_h$; $\mathbf{R}_h = \frac{1}{P}\sum_{l=1}^{P}(\mathbf{h}_l - \mathbf{h}_{av})(\mathbf{h}_l - \mathbf{h}_{av})^T$, $\mathbf{h}_{av} = \frac{1}{P}\sum_{l=1}^{P}\mathbf{h}_l$; $\mathbf{w}_i = \mathbf{Q}^T(\mathbf{h}_i - \mathbf{h}_{av})$; P is the measured HRIRs' number in KEMAR package; and then ε_l is the approximation error of $h_l(n)$'s PCA approximation model with N ranks.

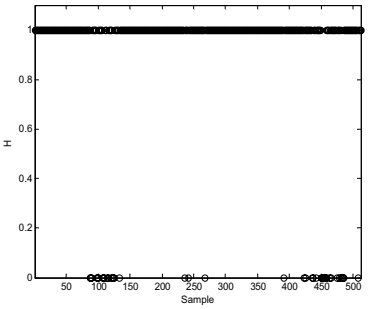


Figure 1. Lilliefors hypothesis testing for good fit to a normal distribution of HRIRs

Such results can be made clear in the below geometrical explanation [5]: the HRIR connected to each direction is a realization of stochastic vector H of HRIRs; bases vectors q_i of Karhunen-Loève transformation present the principal axes of HRIRs data distribution; and the biggest eigenvalues correspond to the directions with dense distributing of the data. As a consequence, HRIRs' reconstruction from projection onto these directions can results in minimum average errors. However, the presupposition is that, if the stochastic vector H of HRIRs has Gaussian distribution, its probability density is uniform along the ellipsoids whose axes are proportional to the eigenvalues σ_i in directions of vector q_i. However, if H is not a Gaussian process, a non-linear approximation may be much more accurate than the linear approximation; and the Karhunen-Loève transformation basis is no longer optimal. Here, by performing a Lilliefors test for goodness of fit to a normal distribution of HRIRs, shown in Figure 1, it is observable that, most of the HRIRs' samples don't satisfy the hypothesis of normal distribution very well (The result of the hypothesis test is a Boolean value, which is 0 when you do not reject the null hypothesis, and 1 when you do reject that hypothesis.). Therefore, the PCA model based on Karhunen-Loève transformation cannot optimally approximate non-Gaussian HRIRs' data.

3. ADAPTIVE NON-LINEAR APPROXIMATION ALGORITHM BASED ON WAVELET TRANSFORMATION MODULUS MAXIMA

First of all, we introduce the basic theory of wavelet transformation and analysis [5, 10]. We all know that, traditional Fourier transformation can depict stationary signal very well all through the time or space domain; but some essential information in a signal, such as singularities and irregular structures, is usually blurred in its transformation domain. However, these local characteristics can be detected and represented by using wavelet transformation at different fine scales. Suppose $\forall h(t) \in L^2(R)$, where $L^2(R)$ denotes the vector space of measurable, square-integrable one-dimensional function or signal $h(t)$, the continuous wavelet transformation of a signal $h(t)$ is defined below,

$$WT_h(a,b) = \langle h, \psi_{a,b} \rangle = |a|^{-1/2} \int_{-\infty}^{\infty} h(t)\overline{\psi\left(\frac{t-b}{a}\right)}dt, a \neq 0 \qquad (3)$$

where $\psi_{a,b}(t) = |a|^{-1/2}\psi\left(\frac{t-b}{a}\right)$, and $\psi(t)$ is the basic or mother wavelet. Then the inverse transformation is

$$h(t) = C_\psi^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi_{a,b}(t)\,WT_h(a,b)\,db\frac{da}{|a|^2} \qquad (4)$$

where $C_\psi = \int_{-\infty}^{\infty}\frac{|\hat{\psi}(\omega)|^2}{|\omega|}d\omega < \infty$, and $\hat{\psi}(\omega)$ is the Fourier transformation of $\psi(t)$.

Practically, the parameters a and b are usually digitized as $b = \frac{k}{2^j}, a = \frac{1}{2^j}; j,k \in Z$, then $\psi_{a,b}(t) = \psi_{\frac{1}{2^j},\frac{k}{2^j}}(t) = 2^{j/2}\psi(2^j t - k)$, which can be abbreviated as $\psi_{j,k}(t)$. And then the wavelet transformation can be written in the following inner product between $h(t)$ and $\psi_{j,k}(t)$, i.e. $WT_h\left(\frac{1}{2^j},\frac{k}{2^j}\right) = \langle h, \psi_{j,k} \rangle$. Because wavelet transform has different time-frequency resolution, it can meet and measure a signal's different local variations in time or frequency domain. This characteristic has broad applications in detection and representation of signal's time-frequency structures.

Then for purpose of interpreting this problem clearly, we introduce a measure index, Lipschitz exponent (L.E.), which characterizes the local regularity of a signal at any time point. Suppose $\forall h(t) \in L^2(R)$, the Lipschitz exponent of a signal $h(t)$ at point x_0 is defined as,

$$|h(t_0 + \varepsilon) - P_n(t_0 + \varepsilon)| \leq A|\varepsilon|^\alpha, n < \alpha < n+1 \qquad (5)$$

where ε is a sufficiently small quantity; $P_n(t)$ is a polynomial of degree n across point $h(t_0)$; and then the Lipschitz exponent of the signal $h(t)$ at point x_0 is α.

It can also be proved that, at the interval $[t_1, t_2]$, if there is

$$|WT_h(a,b)| \le Ka^{\alpha} \text{ i.e. } \log|WT_h(a,b)| \le \log K + \alpha \log a, \quad (6)$$

the signal $h(t)$ is well-proportioned Lipschitz α at interval $[t_1, t_2]$, where K is a constant related to the given wavelet function. Moreover, if there is $a = 2^j$, the formula (6) has the following expression,

$$|WT_h(2^j,b)| \le K2^{j\alpha} \text{ or } \log_2|WT_h(2^j,b)| \le \log_2 K + \alpha j. \quad (7)$$

Thus the term αj makes connection between scale j and Lipschitz exponent α, and reveals some variational regulations between a (or j) and Lipschitz α: when $\alpha > 0$, the wavelet transformation modulus maxima, i.e. the expression before the sign of inequality, increase with a (or j); when $\alpha = 0$, it remains unchangeable; while $\alpha < 0$, the modulus maxima decrease with a (or j); and the variational degree is close pertinent to the absolute value of α. Our work also bases on the above premise, and carries through adaptive non-linear approximation in the field of HRIRs' wavelet transformation modulus maxima at different scales.

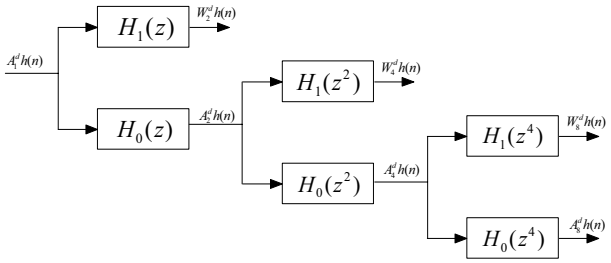


Figure 2. Decomposition flow chart of a'trous algorithm

Moreover, the usually applied fast numerical computations of wavelet transform—Mallat algorithm is subsampling after wavelet analysis filtering, which is not translation-invariant and unfit for the singularities' detection of a signal [5]. Therefore, a fast dyadic wavelet transform using a filter bank algorithm without subsampling, introduced by Holschneider and Morlet and called the *algorithme a'trous* (in French), is calculated to construct signal's translation-invariant representations. Because this algorithm is translation-invariant, it is especially suitable for the computation of wavelet transformation modulus maxima to improve singulaties' detection precision, which is specially propitious for applications of signal's pattern recognition, and so on. Its decomposition flow chart is illustrated in Figure 2, where we choose the Biorthonormal Quadratic Spline Filter (bior 3.1) as our wavelet function [10]. In our following experiments of HRIRs' approximation modeling, the comparative results respectively using a'trous and Mallat algorithm are also given, from which we can clearly see the superiority of a'trous algorithm.

Suppose that, $h(n)$ is the HRIR signal to be analyzed; h_0 and h_1 are the low-pass and high-pass filter related to our given wavelet function, bior3.1; $S = A_1^d h(n), (n \in Z)$ is the sampling sequence of HRIR signal $h(n)$; $W_{2^j}^d h(n), (n \in Z)$ is the wavelet transformation; and $A_{2^j}^d h(n), (n \in Z)$ is the relevant approximation of S at scale j, the a'trous algorithm of wavelet transform is shown below:

$$W_{2^{j+1}}^d h(n) = (A_{2^j}^d h * h_1^j)(n) \quad (8)$$

$$A_{2^{j+1}}^d h(n) = (A_{2^j}^d h * h_0^j)(n) \quad (9)$$

where $\left\{\left\{A_{2^j}^d h(n)\right\}; \left\{W_{2^j}^d h(n)\right\}, j = 1,2,\cdots,J\right\}$ are the wavelet transformation of signal S; j is the analytical scale; and J is the maximum decomposition scale. Then the modulus maxima are just the below $W_{2^j}^d h(n)$ satisfying the inequations (10), indicated as $W_{2^j}^d h(n_{k_j}), j = 1,2,\cdots,J; k_j = 1,2,\cdots,K_j$, where K_j is the quantity of modulus maxima; and k_j is the sequence number of modulus maximum.

$$\begin{aligned} &|W_{2^j}^d h(n)| \ge |W_{2^j}^d h(n-1)| > |W_{2^j}^d h(n-2)| \\ &|W_{2^j}^d h(n)| \ge |W_{2^j}^d h(n+1)| \end{aligned} \quad \text{or} \quad \begin{aligned} &|W_{2^j}^d h(n)| \ge |W_{2^j}^d h(n-1)| \\ &|W_{2^j}^d h(n)| \ge |W_{2^j}^d h(n+1)| > |W_{2^j}^d h(n+2)| \end{aligned} \quad (10)$$

The reconstruction of a signal $h(n)$ from its wavelet transformation modulus maxima is just to utilize the selective modulus maxima $W_{2^j}^d h(n_{k_j})$ to recover the corresponding wavelet transformation and resulting original signal's approximation [5].

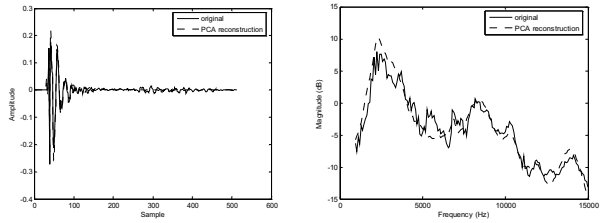
On all accounts, the realization method of HRIRs' spatial hearing approximation model from wavelet transformation modulus maxima is concluded as follows. Firstly, we apply translation-invariant a'trous decomposition algorithm to represent HRIR signal $h(n)$ at different scales. Where as for the 512 samples long HRIR data of KEMAR, when the decomposition layer is 2, the approximation precision usually satisfy the requirement in our simulation experiments, and the improving of effect is not obvious with the increase or decrease of layers. Secondly, based on the above processing results, we get the modulus maxima of HRIR's wavelet transformation [10]. After that, under the circumstance with approximately equal threshold with PCA model and wavelet model based on Mallat algorithm, the threshold T of modulus selection is also set as 2% of HRIR's Euclid norm at certain position (T is related to individual HRIR signal on this location. We compare every modulus maxima $|W_{2^j}^d h(n_{k_j})|$ with the threshold, if the modulus is a smaller amount than T, which is discarded; otherwise, it is reserved.). At last, during the reconstruction of HRIR signal, we make use of relevant synthesis algorithm to recover the HRIR's approximation version from its singularities once more [5, 10].

4. SIMULATION RESULTS

Here, the data of KEMAR's HRTFs, which offered by MIT Media Lab [11], are used in our simulation work. The measurements of the data were made with speaker on the discrete positions every 10^o in elevation, and $5^o \sim 30^o$ unequally in azimuth. Moreover, the measured data may be contaminated by deficient factors, and thus it is necessary to get rid of the contamination before further processing. Some other meticulous and important processing refers to [11]. In view of all the 710 measurement positions of KEMAR, we firstly process the HRIRs data according to PCA method in time domain, where 18 principal components, accounting for 98% of the variance in the original HRIRs, are selected as the basic basis vectors to approximate the measured HRIRs. And the resulting PCA approximation model totally has 21996 values. Figure 3 is a demonstration (Azimuth=0^o, Elevation=90^o) of the HRIRs' PCA model.

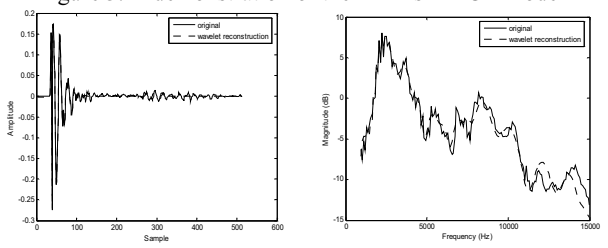
Figure 4 give a demonstration of the HRIRs' wavelet adaptive approximation model based on Mallat algorithm [12].

The demonstration of HRIRs' spatial hearing model based on wavelet transformation modulus maxima is given in Figure 5. From the quantitative results, we can see that, the error of HRIR's PCA model at this position is 0.2; the error of wavelet adaptive approximation model based on Mallat algorithm is 0.02; while the HRIR's model based on wavelet transformation modulus maxima has error 0.006, which excels the above two results.



(a) HRIR (b) HRTF

Figure 3. A demonstration of the HRIRs' PCA model



(a) HRIR (b) HRTF

Figure 4. A demonstration of the HRIRs' wavelet adaptive approximation model based on Mallat algorithm

As for all the measurement positions, shown in Figure 6, under the circumstance with approximately equal threshold, the HRIRs' wavelet adaptive model based on Mallat algorithm preserves 24572 parameters values (denoted as "threshold selection" in the legend box of Figure 6), where we select Daubechies10 wavelet; and its model's error -18.1dB is better than the PCA's -13.1dB. But both of them are inferior to the 17644 values and -20.8dB error of HRIRs' wavelet modulus maxima approximation model. Of course, the improved approximate error is achieved by a little redundancy of processing time. On the notebook PC with Intel Celeron M 1.5GHz and 224M memory, PCA needs about 10s, while the wavelet approximation model based on modulus maxima needs 32s. However, fortunately, these processing time are all within the bounds of VAS's implementation [1].

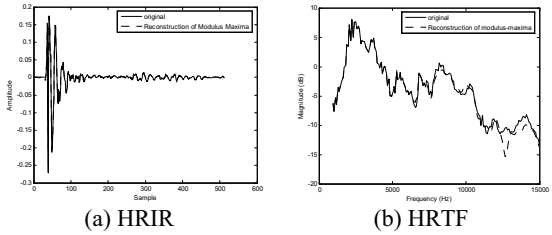
Here, we use the following error formulation [3]:

$$e = \left\| h - \hat{h} \right\|^2 \Big/ \left\| h \right\|^2 \qquad (11)$$

where h is the measured HRIR; and $\hat{h}$ is the approximation model's HRIR.

5. CONCLUSIONS AND FUTURE WORK

As seen from the results of different approximation models, this adaptive one based on wavelet transformation modulus maxima and singularities' detection of HRIRs achieves better effects than that using Mallat algorithm and the PCA model, which show the non-linear methods' validity in HRIRs' approximation modeling. As for future work, some other practical utilities should be considered in detail, and listening tests will also be performed to evaluate subjective performance of these models.



(a) HRIR (b) HRTF

Figure 5. A demonstration of the HRIRs' wavelet transformation modulus maxima model

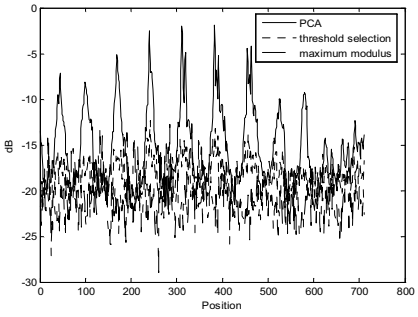


Figure 6. Comparison of three approximation models on all the positions

6. REFERENCES

[1] J. P. Blauert, *Spatial Hearing*, revised edition, Cambridge, MA: MIT, 1997.

[2] D. J. Kistler, F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *J. Acoust. Soc. Am.*, vol.91, no. 3, pp. 1637-1647, 1992.

[3] J. Chen, B. D. Van Veen, K. E. Hecox, "A spatial feature extraction and regularization model for the head-related transfer function," *J. Acoust. Soc. Am.*, vol. 97, no. 1, pp. 439-452, 1995.

[4] Z. Y. Wu et al, "A time domain binaural model based on spatial feature extraction for the head-related transfer function," *J. Acoust. Soc. Am.*, vol. 102, no. 4, pp. 2211-2218, 1997.

[5] S. Mallat, A wavelet tour of signal processing, Academic Press, 1997.

[6] Julio Torres, Mariane Petraglia, Roberto Tenenbaum, "Low-order modeling of head-related transfer functions using wavelet transforms," *ISCAS*, vol. 3, pp. 513-516, 2004.

[7] D. W. Batteau, "The role of the pinna in human localization," *Proc. Royal Society London*, 168(series B), pp. 158–180, 1967.

[8] Y. Hiranaka, H. Yamasaki, "Envelope Representations of Pinna Impulse Responses Relating to Three-Dimensional Locatization of Sound Sources," *J. Acoust. Soc. Am.*, vol. 73, no. 1, pp. 291-296, 1993.

[9] J. Hebrank, D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *J. Acoust. Soc. Am.*, vol. 56, no. 6, pp. 1829-1834, 1974.

[10] David Donoho, Mark Duncan, Xiaoming Huo etc, *The WaveLab package*, Stanford University, 1999.

[11] Bill Gardner, Keith Martin, "HRTF Measurements of a KEMAR Dummy-Head Microphone," *Technical Report #280*, MIT Media Lab Perceptual Computing, Cambridge, MA, May 1994.

[12] Zhang Jie, Ma Hao, "A spatial hearing model based on wavelet adaptive non-linear approximation theory," Submitted to *Journal of Southeast University*.