DEPENDENCY OF THE FINITE-IMPULSE-RESPONSE-BASED HEAD-RELATED IMPULSE RESPONSE MODEL ON FILTER ORDER

Jian Zhang, Chundong Xu, Risheng Xia, Junfeng Li, Yonghong Yan

Institute of Acoustics, Chinese Academy of Sciences
No. 21, Beisihuan Xilu, Haidian, Beijing, 100190, China

ABSTRACT

Various approaches have been reported on HRIR modeling to lighten the high computation cost of the 3-D audio systems without sacrificing the quality of the rendered sounds. The performance of these HRIR models have been widely evaluated usually in terms of the objective estimation errors between the original measured HRIRs and the modeled HRIRs. However, it is still unclear how much these objective evaluation results match the psychoacoustic evaluations. In this research, an efficient finite-impulse-response (FIR) model is studied as a case study which is essentially based on the concept of the minimum-phase modeling technique. The accuracy dependency of this modeling approach on the order of FIR filter is examined with the objective estimation errors and the psychoacoustic tests. In the psychoacoustic tests, the MIT HRIR database are exploited and evaluated in terms of sound source localization difference and sound quality difference by comparing the synthesized stimuli with the measured HRIRs and those with the FIR models of different orders. Results indicated that the measured hundred-sample-length HRIRs can be sufficiently modeled by the low-order FIR model from the perceptual point of view, and provided the relationship between perceptual sound localization/ quality difference and the objective estimation results that should be useful for evaluating the other HRIR modeling approaches.

1. INTRODUCTION

Head-related impulse responses (HRIRs) play an important role in binaural 3-D audio rendering, which is generally realized by convolving the input stimulus with HRIRs. The direct way to obtain the HRIR from a given source location is to measure the HRIR at the ear drum for the impulse placed at the source [1]. However, the measured HRIRs are always a couple of hundred-sample lengths, which results in the high computational cost for real-time applications especially when simultaneously rendering multiple sound sources. To overcome these problems, various approaches have been reported to model HRIRs in the temporal and/or spectral domain [2][3]. The HRIR modeling approaches that have been reported in the literatures can be roughly classified into three categories: the physical-based computational approach [4][5], the parametric modeling approach [6][7], and the filter-based modeling approach [8][9].

As the simplest physical computational approach, the structural model is composed of different basic filters each of which is used to model the acoustic effects of each component of human body on wave propagation in an anechoic environment. Though this model is conceptually simple to implement, it is very difficult to estimate the model parameters from the geometrical measurements, especially for pinna [10]. Moreover, it is still unknown how to deal with the sources coming from the back half space [4]. Another well-known computational model is the boundary-element method (BEM), which provides an elegant way of the partial differential equations that describe acoustic wave propagation around a physical object [5]. The disadvantage of this approach include the difficulty in getting accurate surface meshes (especially for pinna) and the high computational cost. The negative aspect of these physical computational approaches is that they made a number of assumptions, which may remove some essential information which are necessary for producing a realistic acoustic environment simulation.

The parametric modeling approaches, which attempt to functionally represent the HRIRs, generally first model the measured HRIRs using a set of parameters that are further used for HRIR synthesis. For instance, Evans et al. suggested a form of continuous orthogonal representation in which the HRTFs were expressed as a weighted sum of surface spherical harmonics [6], and Kistler and Wightman reported to approximate the minimum-phase HRTFs with principal components analysis (PCA) [7]. But the use of such models in systems still has many draw backs. One of them is the HRIR implementation which, even if greatly compressed, requires a large computational cost to uncompress or recover the data.

The FIR and IIR filters are also widely used for HRIR modeling in the temporal and/or spectral domain. Bolmmer and Wakefield presented to design the IIR filter based on the error criteria of log-magnitude spectrum differences [9]. Asano et al. investigated the abilities of IIR filters with different orders for modeling individual HRIRs, and showed that a 40th-order IIR filter yielded good approximation of individual HRIRs in terms of sound localization difference, with the exception of increased front-back confusions in frontal incident angles [11]. Though IIR is able to model HRIRs with a quite low order, it is very difficult for IIR filter to be interpolated between discrete positions.

The performance of most HRIR models have been widely evaluated usually in terms of the objective estimation errors between the original measured HRIRs and the modeled HRIRs. However, it is still unclear how much these objective evaluation results match the psychoacoustic evaluations. In this paper, we focus on a FIR approach for HRIR modeling on the concept of the minimum-phase approximation technique. Main attention was paid to investigate the accuracy dependence of this FIR-based modeling approach on the order of FIR filters through psychoacoustic tests in terms of sound source localization difference and sound quality difference. And one objective evaluation errors was used to investigate the effectiveness of this FIR models. Psychoacoustic test results demonstrated that the measured HRIRs can be sufficiently modeled by the low-order FIR model. And the relationship between perceptual sound localization/ quality difference and the objective estimation results is useful for evaluating the other HRIR modeling approaches.

2. METHOD

The implementation of the filter-based modeling approaches can be in the time or frequency domain. The modeling approach studied here is in the frequency domain. More specifically, the frequency responses of the measured HRIRs are approximated on the minimum phase theory, which consists of the following steps (We use the front HRIR (0°, 0°) of MIT database as an example):

1. *ITD estimation of HRIRs.*
   The commonly used approach for ITD estimation is based on the cross-correlation between left and right channels of HRIRs [7]. The ITD is calculated in this work through estimate the time delay of each HRIR, which is determined as the time at which the HRIR becomes non-zero using the threshold-based detection technique, and yielded the almost identical results with cross-correlation method [2]. As shown in the figure 1, the time delay is 33 samples (the red part).

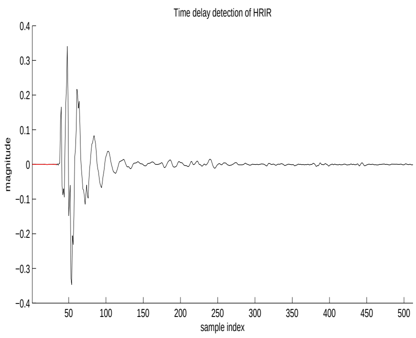

Figure 1: Example of time delay detection of HRIR.

2. *Amplitude approximation of the minimum-phase HRIRs with the different FIRs of different orders.*
   The minimum-phase HRIRs, h'(t), are first derived by removing the zeros at the beginning of HRIRs based on the estimated ITDs. The corresponding transfer functions are denoted as H'(k) in the frequency domain. Then the amplitude responses of H'(k) are approximated by the FIR filters. The coefficients of the FIR filters to be estimated are eventually determined by using a linear least square error function [12]. Note that given different FIR orders, different coefficients of FIR filters can be derived. In other words, the FIR models with different orders should yield different abilities in approximating the amplitude response H'(k) of the minimum-phase HRIRs. Figure 2 depicts the frequency response of minimum-phase HRIR and the frequency response of FIRs with the order of 47 and 68.
3. *Modeled HRIRs synthesis by adding ITD cues.*
   The time domain modeled HRIRs h(t) can be obtained using the FIR filters, followed by supplementing the ITD cues. Actually, it is needn't add the zeros to the FIR directly. This procedure is essentially implemented by adding the zeros to the output signal when the convolving of input signal and

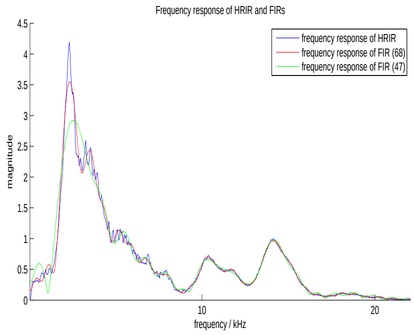

Figure 2: Frequency response of HRIR and FIR.

FIR is done. So the number of zeros will not affect the efficiency of this model. Figure 3 shows the original HRIR and the FIR with the order of 47. (For easily compare, the start point of FIR in the figure was moved 33 samples.)

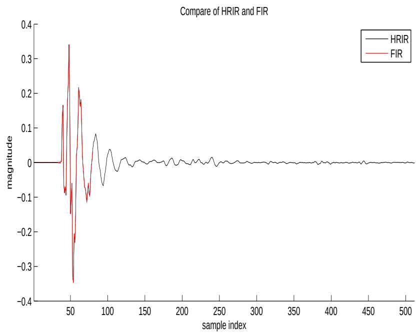

Figure 3: Compare of original HRIR and modeled FIR.

3. PERCEPTUAL EVALUATIONS

3.1. Data

The HRIRs were provided by MIT measured using a KEMAR dummy head [1]. Gardner and Martin made an assumption that the dummy head is perfectly symmetrical, so the HRIRs need be collected for only one ear. This assumption allowed them to mount two different pinnae on the KEMAR, and the HRIRs associated with both pinna types could be collected simultaneously. The HRIRs used in this paper are measured with the right ear.

To investigate the dependence of the ability of the considered HRIR modeling approach on the order of FIR filters and find out the minimum order of the FIR for modeling the measured HRIRs, various FIR orders were designed according to the E series standard [13]. Since the length of the measured HRIRs were 512 samples, the examined FIR orders were eventually determined as 10, 15, 22, 33, 47, 68, 91, 121, 178, 200, 261, 383, 464. Given a FIR order, the modeled HRIRs can be computed using the approach described in the previous section. In our tests, the directions of horizontal plane and vertical plane are evaluated. The horizontal angles of the HRIRs are from -90° to 90° at the interval of 30°, and the vertical angles were from -30° to 90°, the interval is 30°.

To generate the stimuli for psychoacoustic tests, three signals were exploited, including one broad-band white noise, one male speech signal and one telephone ring signal, with the length of 2s, 3s and 2s respectively. The stimuli were eventually generated by convolving three types of signals with the HRIRs modeled with the FIRs of different orders and adding time delay. The generated stimuli were subsequently presented to subjects for psychoacoustic tests.

3.2. Subjective evaluation

A total of 10 subjects (5 male and 5 female) with normal hearing were recruited and paid for their participation in psychoacoustic tests. The subjects were aged from 23 to 28. In tests, stimuli were presented to the subjects at a comfortable listening level with Sennheiser HD 280 Pro headphones. Two psychoacoustic tests were carried out for comparing the stimuli generated by the modeled HRIRs and those by the measured HRIRs, in terms of sound quality difference and sound localization difference, respectively. In each evaluation (for quality or localization difference), each subject listened to a total of 1512 stimuli (3 signals $\times$ 12 DOAs $\times$ (13 orders + 1 original) $\times$ 3 times), where each stimuli were presented three times. The stimuli were grouped into 108 sets in each of which 14 stimuli (processed by the FIR filters with 13 different ordered and the measured original HRIRs) were randomly presented to subjects. In listening tests, the 108 sets were further divided into 6 sessions with 18 sets in each session. The subjects were asked to have a break after one session.

For the perceptual tests, the paired comparison evaluation was used, in which one stimulus was generated by convolving the dry signal with the measured HRIR and the other with the FIRs of different orders. The presentation order of the stimuli sets and the stimuli in each set were randomized for each subject. The stimuli could be listened to several times until the subject made a decision. For each paired comparison, subjects provided a score on the five-grade scale based on his/her preference in terms of the degree of difference in sound quality or sound localization. The detail specification of the five-grade scale is shown in Table. 1. During the test, no feedback information was given to the subjects.

Table 1: Five-grade score scale using in the psychoacoustic listening tests and its description.

Score	Description
1	Exactly different
2	Different
3	Uncertain
4	Almost same
5	Exactly same

3.3. Objective evaluation

In order to objectively evaluate the model against the original measured HRTF, and further find the relationship between the objective and subjective evaluations, the spectral distortion (SD) was considered as error measure [10].

$$SD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(20\log_{10}\frac{|H(f_i)|}{|\tilde{H}(f_i)|}\right)^2}\,[dB] \quad (1)$$

where H is the frequency response of original HRIR, $\tilde{H}$ is the frequency response of modeled HRIR (FIR), and N is the number of available frequencies in the considered range, that limited between 500 Hz and 16 kHz. For calculate $\tilde{H}$, the zeros are added to the FIRs.

3.4. Results

The overall psychoacoustic results in terms of sound quality difference and localization difference are plotted in figure 4. As shown in figure 4, a FIR filter with 68 coefficients (Note that the zeros that represent the time delays are not considered) is sufficient in both sound quality and localization. That is, the FIRs with 68 coefficients (not 512 samples as in the measured HRIRs) are able to provide the quite similar perceptual sensation. The FIR-based HRIR modeling approach greatly reduced the length of HRIRs, which further reduced the computational cost and speed up the synthesis procedure of binaural signals. Furthermore, if only sound source localization performance is considered as evaluation criterion, the minimum order of FIRs can be further decreased to 47, at which the sound quality will be slightly different from that by the measured HRIRs.

As the three signals used in our psychoacoustic tests exhibited different energy distribution in the time-frequency domain, the dependency of the modeling ability of FIR-based approach on the FIR order was further investigated by looking at its relationship with the stimuli type. The perceptual results in terms of sound quality difference and sound source localization difference are plotted in figure 5 and figure 6, respectively. From these results, it is noted that the minimum order of FIR that provided the acceptable sound quality and sound localization is dependent on the stimuli type. For example, the FIR order that yielded sufficiently acceptable sound quality and localization for the telephone ring signal is lower than that for the male speech and white noise signals. This difference might come from the difference in energy distribution of each stimulus in the time-frequency domain.

Figure 7 depicts the SD of FIRs with different orders. As shown in figure 7, the SD is nearly 2 dB where the order of FIR is 47, which obtain acceptable source localization performance. If the sound quality is concerned, which means the order of FIR would be 68, the SD is nearly 1 dB. When the SD is larger than 4, correspond to the FIR order is small than 22, both the sound quality and source localization performances are poor.

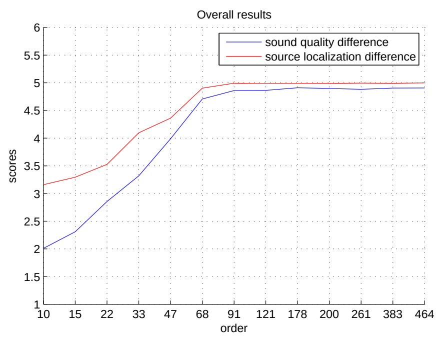

Figure 4: Overall results.

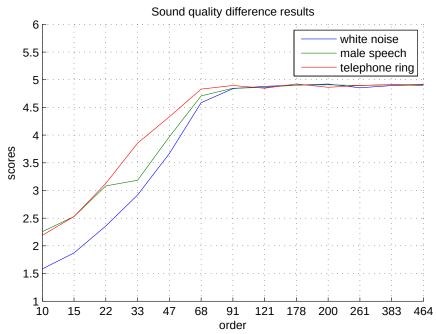

Figure 5: Sound quality difference results of different stimuli.

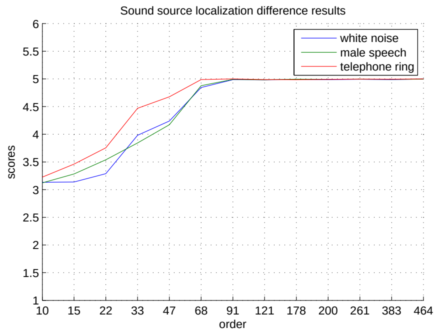

Figure 6: Sound source localization difference results of different stimuli.

4. CONCLUSION

In this paper, a FIR-based HRIR modeling approach based on minimum phase theory was studied. In this paper, the HRIR is approximated by a FIR filter and ITD cues. The main attention in this paper was paid to the performance dependence of this modeling approach on the order of FIRs through psychoacoustic tests in terms of sound quality difference and sound source localization difference. Psychoacoustic test results indicated that the measured HRIRs with the length of hundred samples can be perceptually modeled by the low-order FIRs with a dozen of coefficients. And the performance is further evaluated by objective quantity, the relationship of the objective and subjective evaluation would be helpful for other HRIR modeling methods.

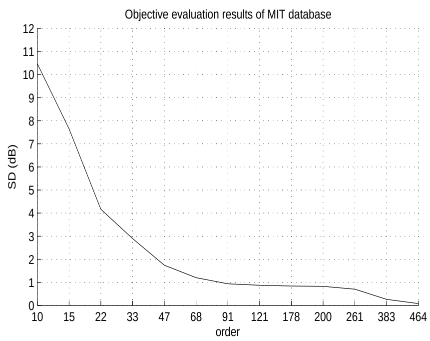

Figure 7: objective evaluation results of MIT database.

Acknowledgment

This work is partially supported by the National Natural Science Foundation of China (Nos. 10925419, 90920302, 61072124, 11074275, 11161140319, 91120001), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant Nos. XDA06030100, XDA06030500) and the National 863 Program (No. 2012AA012503).

5. REFERENCES

[1] W. G. Gardner and K. D. Martin, "Hrtf measurements of a kemar," *J. Acoust. Soc. Am.*, vol. 97, no. 6, pp. 3907–3908, 1997.

[2] J. Sandvad and D. Hammershoi, "Binaural auralization, comparison of fir and iir filter representation of hirs," in *Audio Engineering Society Convention 96*, Feb 1994.

[3] J. Huopaniemi and M. Karjalainen, "Review of digital filter design and implementation methods for 3-d sound," in *Audio Engineering Society Convention 102*, Mar 1997.

[4] C. P. Brown and R. O. Duda, "A structural model for binaural sound synthesis," *IEEE Trans. on Speech and Audio Processing*, vol. 6, no. 5, pp. 476–488, 1998.

[5] B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function, impedance effects and comparisons to real measurements," *J. Acoust. Soc. Am.*, vol. 110, no. 5, pp. 2449–2455, 2001.

[6] M. J. Evans, J. A. S. Angus, and A. I. Tew, "Analyzing head-related transfer function measurements using surface spherical harmonics," *J. Acoust. Soc. Am.*, vol. 104, no. 4, pp. 2400–2411, 1998.

[7] D. J. Kistler and 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, pp. 1637–1647, 1992.

[8] A. Kulkarni and H. S. Colburn, "Efficient finite-impulse-resonse filter models of the head-related transfer function," *J. Acoust. Soc. Am.*, vol. 97, no. 5, pp. 3278, 1995.

[9] M. A. Blommer and G. H. Wakefield, "On the design of pole-zero approximations using a logarithmic error measure," *IEEE Trans. on Speech and Audio Processing*, vol. 42, no. 11, pp. 3245–3248, 1994.

[10] S. Spagnol, M. Geronazzo, and F. Avanzini, "On the relation between pinna reflection patterns and head-related transfer function features," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 21, no. 3, pp. 508–519, 2013.

[11] F. Asano, Y. Suzuki, and T. Sone, "Role of spectral cues in median plane localization," *J. Acoust. Soc. Am.*, vol. 88, no. 1, pp. 159–168, 1990.

[12] E. C. Levi, "Complex-curve fitting," *IRE Trans. on Automatic Control*, vol. AC-4, pp. 37–44, 1959.

[13] IEC, "Preferred number series for resistors and capacitors," in *Standard IEC 60063, International Electro technical Commission*, 1963.

[14] J. Chen, B. D. V. Veen, and 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.