Buenos Aires – 5 to 9 September, 2016
Acoustics for the 21st Century…

PROCEEDINGS of the 22nd International Congress on Acoustics

Virtual Acoustics: Paper ICA2016-342

An asynchronous HRTF measurement method based on phase alignment

Mengfan Zhang, Fengyun Zhu, Tianshu Qu, Xihong Wu

Key Laboratory on Machine Perception (Ministry of Education), Speech and Hearing Research Center, Peking, University, Beijing, China, qutianshu@pku.edu.cn

Abstract

The virtual sound technology based on the Head-Related Transfer Function (HRTF) is important in many applications such as gaming, education and military. Currently, the most important and accurate method to obtain HRTF is experimental measurement. In experimental measurement, the Maxim Length Sequence (MLS), the sweep signal, and impulse signal are usually used as the exciting signal and played in loop to generate several HRTFs and then the HRTFs are averaged to improve the Signal to Noise Ratio (SNR) of the results. However, the inconsistency of the timing modules (Oscillator) used in the recording system and in the playback system results the time non-alignment between the measured HRTFs in different played loop. The time non-alignment destroys the function of the average process and makes the average results distortion. For solving this problem, this paper proposed an asynchronous HRTF measurement method based on the phase alignment. Firstly, HRTFs are measured using the MLS signal as the exciting signal which are played in loops; then the phase alignment algorithm is applied to HRTFs in each loop to compensate the difference of the timing modules (Oscillator) of the recording system and the playback system; lastly, the aligned HRTFs are averaged to generate the result HRTF. The evaluation experiment results show that the Peak-SNR of the proposed HRTF measured method are improved about 4.5 dB compare to that of the traditional method.

Keywords: HRTF; MLS; Sweep signal; Oscillator; Peak SNR

An asynchronous HRTF measurement method based on phase alignment

1 Introduction

Head-related transfer function (HRTF) describes the acoustic transmission characteristics that sound waves from a sound source to a listener's binaural eardrums in frequency domain, and in time domain, it is named as Head-related impulse response (HRIR). It is mainly a linear time-invariant filter system that sound waves filtering by human physical structure (such as the head, ear and torso, etc.). Experimental measurements are the most important and most accurate means to obtain HRTF. In recent 20 years, there are many research groups carried out the experimental measurement of HRTF in the anechoic room [1-4] and in the reverberation room [5-6]. Either in the anechoic room or in the listening room, there are some background noise when the HRTF are measured, which destroy the result of HRTF. For solving this problem, the sound pressure of exciting signal is increased to raise the SNR of measured HRTF. But this approach is limited. For example, it can easily make listeners discomfort, and at the same time the excessive sound pressure can lead to nonlinear distortion of the electro-acoustic system. On the other hand, due to the randomness characteristics of noise, its mathematical expectation is zero, and the multiple responses are averaged to enhance the SNR [7]. Thus, the exciting signal is played in loops to generate several HRIRs. By averaging the HRIRs, the SNR is increased.

The average of the experimental measurement results can reduce the background noise. But the inconsistency of the timing modules (Oscillator) used in the recording system and in the playback system causes the time non-alignment between playback signal and recording signal. It will lead to the measured HRIRs in different play loops unaligning, as shown in Figure 1. Therefore, the traditional method of directly averaging the measured HRIRs not only reduce the measurement noise, but also bring the average results distortion.

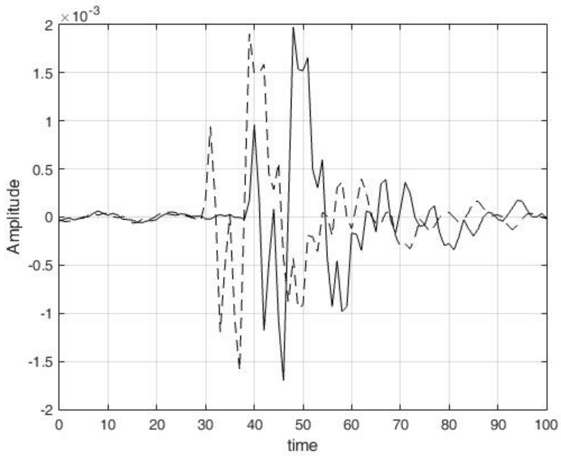

Figure 1: The time unaligning HRIRs measured in different play loops.

This paper is organized as follows. In section II the whole asynchronous HRTF measurement method scheme is presented. It contains an overview of HRTF measurement using MLS, the specific progress of phase alignment for each loop, and averaging all measured HRIRs. In section III, the experiments are designed and the measurement results are shown and evaluated. Finally, conclusions are drawn in section IV.

2 Method

This paper proposed an asynchronous HRTF measurement method based on phase alignment. The specific technological scheme is shown in Figure 2. Firstly, the HRTFs/HRIRs are measured from different loops of the input signal. Secondly, the phase alignment module is done on the measured HRTFs/HRIRs to reconcile the inconsistency of the timing modules (Oscillator) used in the recording system and in the playback system. Finally, the alignment HRTFs/HRIRs are averaged to improve SNR.

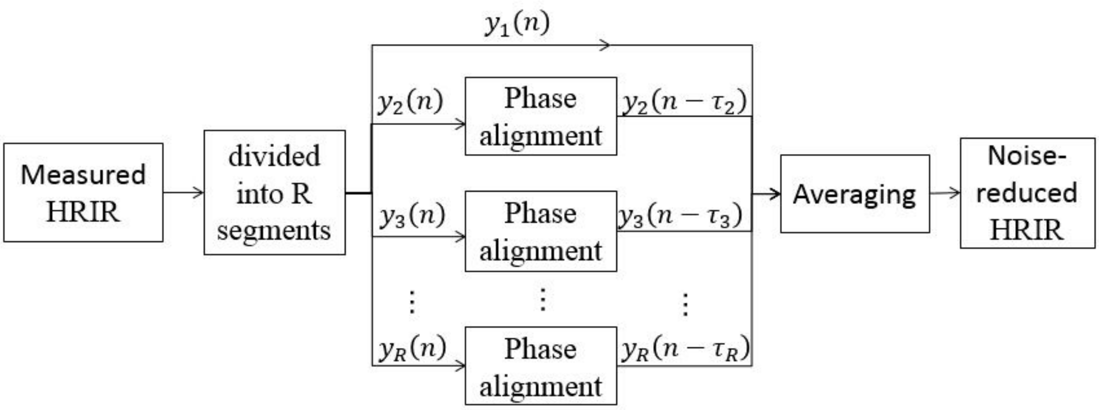

Figure 2: Specific technological scheme of the asynchronous HRTF measurement based on phase alignment.

2.1 HRTF measurement

The physical structure of human's head, torso and pinna comprehensively filters the sound waves getting into the ears. The transfer process from sound source to ears meets the condition of linear time-invariant (LTI). Thus, people usually use a LTI system to describe this process. The transfer function of the LTI system is HRTF.

In this paper, we use MLS as the exciting signal. The relation between input signal and output signal of the LTI system is shown below:

$$y(n) = h(n) * x(n) \tag{1}$$

Where, MLS is the input signal represented by $x(n)$, the transfer function of test system is expressed as $h(n)$, and the measured output signal is $y(n)$. If the input signal and the output signal are known, the transmission characteristics of the system can be calculated using the input signal and the output signal.

The MLS is most commonly used as a pseudo-random noise signal, which is a kind of deterministic discrete time series signal, and can be used in the method of averaging the multiple measurements to enhance the SNR. Moreover, pseudo-random noise has certain characteristics similar to white noise. The autocorrelation function of L-length MLS is calculated as:

$$\varphi_{xx}(n) = xcorr\big(x(n)\big) = \sum_{k=0}^{k=L-1} x(k)x(n+k) = x(n)\otimes x(-n) = \begin{cases} L, n = 0 \\ -1, n \neq 0 \end{cases} \tag{2}$$

Which means that the autocorrelation function of $x(n)$ divided by the sequence length is approximately a pulse signal. Then the HRTF can be obtained as,

$$\hat{h}(n) = \frac{1}{L}\varphi_{xy}(n) = \frac{1}{L}xcorr\big(x(n), y(n)\big) = \frac{1}{L}[y(n)\otimes x(-n)] = \frac{1}{L}[h(n)\otimes x(n)]\otimes x(-n) = \frac{1}{L}h(n)\otimes[x(n)\otimes x(-n)] = h(n)\otimes\frac{1}{L}\varphi_{xx} = h(n) - \frac{1}{L}\sum_{k=0}^{k=L-1} h(k) \tag{3}$$

Where, $\hat{h}(n)$ is the desired HRTF obtained by experimental measurements, and $h(n)$ is the actual HRTF of the test system.

2.2 Phase alignment

Suppose the MLS plays in R loops, and all HRIRs measured from the loops are segmented. The phase alignment algorithm is applied to R-1 HRIR segments to reconcile the time non-alignment between the first segment and the fourth segments. Figure 3 describes the phase-frequency response of the first segment and the posterior segment in the direction of elevation of 0 degrees and horizontal angle of 225 degrees, which measured in the anechoic chamber of Speech and Hearing Research Center at Peking University, China. The sampling rate is 48 kHz. The dashed line is the value of the difference between two segments, and it's shaped like an line reflected phase difference.

The dashed line shown in Figure 3 can be described as,

$$\varphi_1(\omega) - \varphi_2(\omega) = \omega \cdot \tau \tag{4}$$

Where, $\varphi_1(\omega)$ and $\varphi_2(\omega)$ are phase response of HRIRs in the first and the fourth segment, respectively. τ is the time difference between the first and the fourth segment of HRIR.

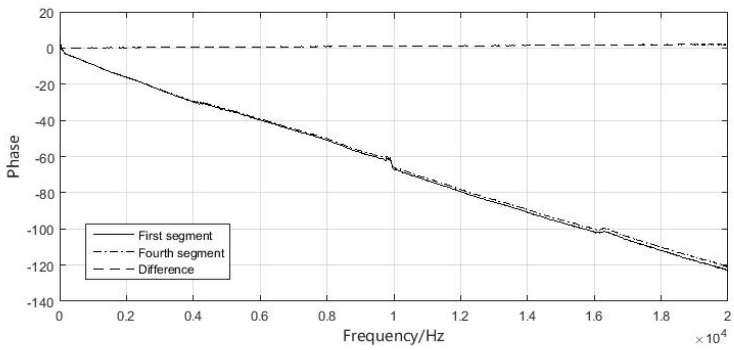

Figure 3: Phase-frequency response of HRIRs in the first segment and the fourth segment.

Randomized Hough Transform (RHT) [8] is used to get the slope of the dashed line. For the line is a kind of two-parameter curve, two points of a line are mapped into one point (k, b), as shown in Figure 4. Under this method, different points on the same line $y = kx + b$ will hit the same point. And if there is a line consisting of N points, by calculating $N(N-1)/2$ times, which means each two-point is assembled into a line and mapped into a point in the (k, b) space, we can find the right k which has been mapped the most times, and this k is the slope. Therefore, the slope of the dashed line is calculated.

After the slope of the dashed line is got, the time difference is also obtained. Then, the posterior segment, represented by the chain line in Figure 3, minus the time difference τ to be aligned to the first segment, represented by the solid line. The aligned phase and its original amplitude response are reassembled into new frequency response, and then it is transferred by the inverse fast Fourier transform to get the aligned HRIR.

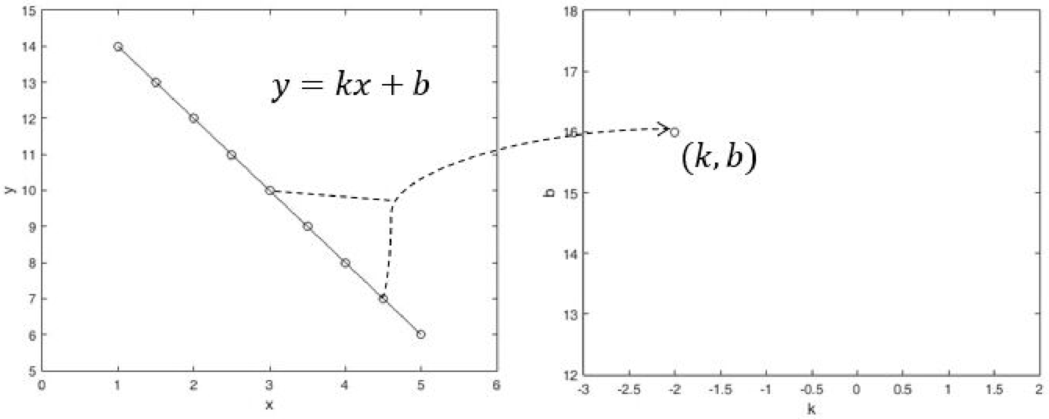

Figure 4: The principle of Randomized Hough Transform [9].

2.3 Averaging in time domain

By using phase alignment algorithm, the time non-alignments between the first segment and posterior segments are reconciled. Then, all R aligned HRIRs are averaged to improve the SNR.

The SNR of HRIR obtained by monocyclic measurement (there is only one loop of input sequence) :

$$SNR = \frac{[h_0(n)]^2}{[e_0(n)]^2} = \frac{e_s^2}{e_n^2} \tag{5}$$

The SNR of HRIR obtained by R loops measurement:

$$SNR' = \frac{[R \cdot h_0(n)]^2}{[\sum_{i=1}^{R} e_i(n)]^2} = \frac{[R \cdot h_0(n)]^2}{\sum_{i=1}^{R} [e_i(n)]^2} = \frac{[R \cdot h_0(n)]^2}{R \cdot [e_0(n)]^2} = R \cdot \frac{e_s^2}{e_n^2} \tag{6}$$

Where, $e_0(n)$ is the random noise in monocyclic measurement, $e_i(n)$ is the random noise in i th segment, and $h_0(n)$ is the ideal HRIR without system noise. e_s^2 is the power of HRIR in the noise-free condition, which is the signal power, and e_n^2 is the noise power.

From the above deduction, after averaging R-segment HRTFs which are time-aligned, the SNR is R times of the SNR in monocyclic measurement.

3 Experiments and results

3.1 Measurement environment and apparatus

The HRTF measurements were performed in a normal office room at PKU. The HRTFs were measured on the KEMAR artificial head. The KEMAR was a Knowles Electronics model DB-4004, and was configured with two neck rings and a torso. The KEMAR was equipped with two matched G.R.A.S. 40AG microphones at the positions of the left and the right ear canal, with matched 26 AC preamplifiers and connected to the DB-100 occluded ear simulators with DB-050 ear canal extensions. The BEQ II (HEAD Acoustics, Germany) is used as recording device. Fireface UC sound card and Dynaudio acoustics DA-BM6A speaker are used as playback devices.

3.2 The measurement results

The HRTFs were measured in eight azimuth positions, varying from 0^o to 315^o in steps of 45^o at the same elevation 0^o. The MLS was played in 30 loops, and recorded at the eardrum of the KEMAR. Then, the HRIRs were calculated using the method above.

Figure 5 shows the HRIRs measured from the left ear of the KEMAR in 270^o. The upper image describes the HRIRs obtained by monocyclic measurement, the middle one is the HRIRs measured using the proposed phase alignment method, and the bottom picture depicts the HRIRs by directly averaging the HRIRs without alignment. The noise is decreased in the middle and bottom pictures, which means the method of averaging can reduce noise. But the directly averaging method can cause distortion which is showed in the bottom figure.

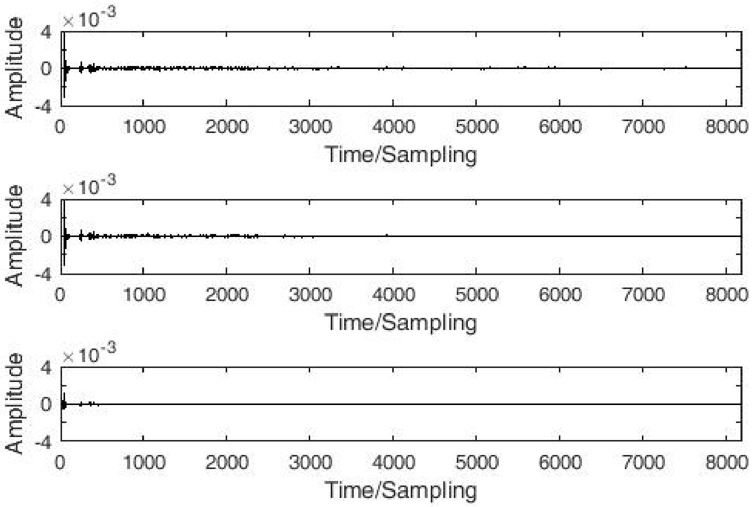

Figure 5: The HRIRs obtained from three measurement methods.

3.3 The evaluation of experiment results

Peak signal-to-noise ratio (PSNR) is the ratio between the maximum possible power of a signal and the power of the noise [10]. Because many signals have a very wide dynamic range, PSNR is usually expressed in terms of the logarithmic decibel scale.

$$PSNR = 10 \cdot log_{10}(\frac{S_{max}{}^2}{E_{noise}{}^2}) \quad (7)$$

Where, E_{noise} is the average of the noise, and S_{max} is the maximum value of the signal, i.e., HRIR.

The PSNR of HRIRs using three methods in all eight azimuth positions are shown in Table 1. From the results, the proposed method has the highest PSNR, which is around 8dB higher than the monocyclic measurement and around 4.5dB higher than the traditional method.

Table 1 The PSNR of HRIR measured by three methods.

Azimuth positions	Monocyclic measurement (dB)	Traditional method (dB)	Proposed method (dB)	Improvement (dB)
0	37.78	42.02	46.05	4.03
45	32.26	37.56	39.99	2.43
90	36.81	36.96	42.80	5.84
135	33.60	36.13	42.67	6.54
180	36.23	40.44	42.82	2.38
225	38.41	42.02	46.35	4.33
270	40.45	44.72	49.52	4.80
315	40.77	44.36	50.58	6.22

4 Conclusions

The asynchronous HRTF measurement method based on the phase alignment is proposed in this paper. Firstly, the HRTFs/HRIRs are measured using the MLS signal as the exciting signal which are played in loops; then the phase alignment algorithm, RHT(Randomized Hough Transform), is applied to HRTFs/HRIRs in each loop to compensate the difference of the timing modules (Oscillator) of the recording system and the playback system; lastly, the aligned HRTFs/HRIRs are averaged to improve the SNR of the result HRTF. The evaluation experiment results show that the Peak-SNR of the proposed HRTF measured method are improved around 8dB higher than the monocyclic measurement and around 4.5dB higher than the traditional method.

Acknowledgments

This work was supported by the National High Technology Research and Development Program of China (863 Program: 2015AA016306), and the National Natural Science Foundation of China No.61175043, No.61421062, No. 31170985)

References

[1] Gardner W. G.; Martin K. D.. HRTF measurement of a KEMAR, *J. Acoust. Soc. Am.*, Vol 97(6), 1995, pp 3907-3908.

[2] Algazi V.R.; Duda R.O.; Thompson D.M. and Avendano C.. The CIPIC HRTF database. *Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, October 21-24, 2001, pp 99-102.

[3] Xie B.; Zhong X.; Rao D.; Liang Z.. Head-related transfer function database and its analyses, *Science in China Series G: Physics, Mechanics and Astronomy*, Vol 50 (3), 2007, pp 267-280.

[4] Qu T.; Xiao Z.; Gong M.; Huang Y.; Li X.; Wu X.. Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap. *Audio, IEEE Transactions on Speech, and Language Processing*, Vol 17(6), 2009, pp 1124-1132.

[5] Kayser H.; Ewert S.D.; Anemuller J.; Rohdenburg T.; Hohmann V.; Kollmeier B.. Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses, *EURASIP Journal on Advances in Signal Proc.*, Vol 2009(1), 2009, pp 1-10.

[6] Jeub M.; Schafer M.; Vary P.. A binaural room impulse response database for the evaluation of dereverberation algorithms, *16th International Conference on Digital Signal Processing*, Santorini, Greece, July 5-7, 2009, pp 1–5.

[7] Riederer K. A. J.. Repeatability analysis of head-related transfer function measurements, *105th Audio Eng. Soc. Conv.*, San Francisco, California, September 26-29, 1998. In CD-ROM.

[8] Kultanen, P.; Xu L.; and Oja E.. Randomized Hough transform (RHT). *10th International Conference on Pattern Recognition*, Atlantic City, New Jersey, June 16-21, 1990, pp 1354-1361.

[9] Xu L.; Oja E.. Randomized Hough transform, *Encyclopedia of Artificial Intelligence*, Vol 3, 2009, pp 1354-1361.

[10] Huynh-Thu Q.; Ghanbari M. Scope of validity of PSNR in image/video quality assessment, *Electronics Letters*, Vol 44 (13), 2008, pp 800-801.