Audio Engineering Society

Convention e-Brief 331

Presented at the 142nd Convention
2017 May 20–23 Berlin, Germany

Personalized HRTF measurement and 3D Audio Rendering for AR/VR Headsets

Woon-Seng Gan[1], Santi Peksi[1], JianJun He[2], Rishabh Ranjan[3], Nguyen Duy Hai[1], Nitesh Kumar Chaudhary[1]

[1] *DSP Lab, School of EEE, Nanyang Technological University, Singapore,*

[2] *Maxim Integrated Products Inc, Singapore*

[3]*Immerzen Labs Pte Ltd, Singapore*

Correspondence should be addressed to Woon-Seng Gan (ewsgan@ntu.edu.sg)

ABSTRACT

This e-Brief describes our recent work in acquiring a fast, personalized head related transfer function (HRTF) and a personalized 3D audio rendering headsets for augmented and virtual reality (AVR) headsets. Binaural signal acquisition and rendering are important tasks in capturing the idiosyncratic acoustics of the pinnae, head and torso, and playback via headphones to the left and right ears. We will highlight a personalized HRTF binaural acquisition cum 3D audio headphone playback system that can take advantage of our individual ear-head anthropometry information in 3D sound acquisition and rendering.

1 Introduction

For augmented and virtual reality (AVR) to be truly immersive, personalized 3D audio must be accompanied with 3D video. In this paper, we show how we can make use of embedded head tracking unit in AVR headsets, together with binaural microphones to capture in-situ head related impulse response (HRIR) or binaural room impulse response (BRIR) without the constraint to head and torso movement. Adaptive signal processing algorithm has been developed specifically for this personalized HRIR/BRIR acquisition system. A patented 3D audio headsets is also being developed to extract important cues from channel based audio signal to reproduce in a multi-driver headsets that can create the idiosyncratic frontal spectral cues. This paper outlines the key processing steps to realize this personalized 3D audio acquisition-rendering system and highlight new applications to take advantage of this headsets. This paper is organized as follows. In Section 2, we outline a novel approach in acquiring fast HRIR/BRIR using a pair of binaural microphone and a virtual reality head gear that displays the 3D spatial locations. This is followed by a set of experimental results that display the differences between HRIR/BRIR captured under different head movements and orientations in Section 3. Section 4 describes our patented 3D audio headphones and its usage in AVR application. The final section concludes with several application examples that will leverage these technologies to improve the immersive 3D audio experience.

2 Fast HRIR/BRIR Acquisition

Head related transfer function (HRTF) plays a vital role in binaural rendering for virtual auditory displays and 3D audio reproduction over headphones, creating an immersive listening experience for interactive VR/AR applications [1, 2]. Due to our idiosyncratic anthropometric characteristics [3], HRTFs are unique to each person and thus, individualized HRTFs should be used for natural 3D audio rendering. As a result, individualized HRTFs are required to be measured for every individual from acoustical measurements [4]. Alternate ways to approximately individualize the HRTFs using non-individualized HRTF database includes subjective tuning [5], using frontal projection headphones [6], and study anthropometric features [7, 8].

The measurements of HRTFs are generally conducted in an anechoic chamber with loudspeakers playing an excitation signal and recording the response at subject's blocked ear canal using binaural microphones [1,3]. Discrete stop-and-go method [9] is commonly used, where loudspeakers at different positions are played and recorded one by one. This method could take hours to complete and hence is very cumbersome for human subjects. Several techniques have been proposed to reduce the total measurement time, including interpolation that reduces angular resolution of HRTF measurements [10], multiple exponential sweep method [11] that uses multiple loudspeakers play exponential sine sweep tones in either interleaving or overlapped manner. Furthermore, continuous acquisition methods substantially reduce the measurement time as compared to the static methods. Continuous acquisition methods usually require an additional rotation facility, e.g., a track for moving loudspeaker [12], a rotating chair for the users [9]. Additionally, they still require constrained head movements.

To relax the constraint for users, we proposed a novel fast and continuous HRTF acquisition with unconstrained head movements [13, 14]. With this system, subjects are free to move his/her head in both horizontal and median plane to cover the entire 2D grid as much as possible. Figure 1 shows the system diagram to acquire the personalized HRIR/BRIR on the fly, while Fig. 2 shows the details of adaptive estimation of the HRIR/BRIR in the VR gear with head tracker. Head-tracker continuously tracks the user head orientation and is synchronized with the binaural recording of the excitation signal from the loudspeaker. Therefore, this system allows fast, continuous and personalized HRTF measurement of human subjects. To accurately extract the HRTFs on-the-fly from such continuous and random head movement measurement data, we have developed and evaluated several adaptive signal processing algorithms. Our studies reveal that progressive based NMLS algorithm [9] is only good for "non-return" type of movements, whereas activation based NLMS algorithm [13] work best with random movements with multiple returns. In practice, a user tends to move in regular movement pattern but with some uncertainties resulting in multiple revisits at certain positions. Additionally, every individual can move in unique manner requiring HRTF estimation method to account for any random unconstrained head movements. Furthermore, the initial conditions of the adaptive filter weights are optimized for the proposed hybrid progressive-activation based NLMS approach. Preliminary testing results have shown that this system is capable to achieve high accuracy and the measured HRTFs are almost perceptually indistinguishable from those measured using standard methods. A real-time demonstration of this system has been presented in [15].

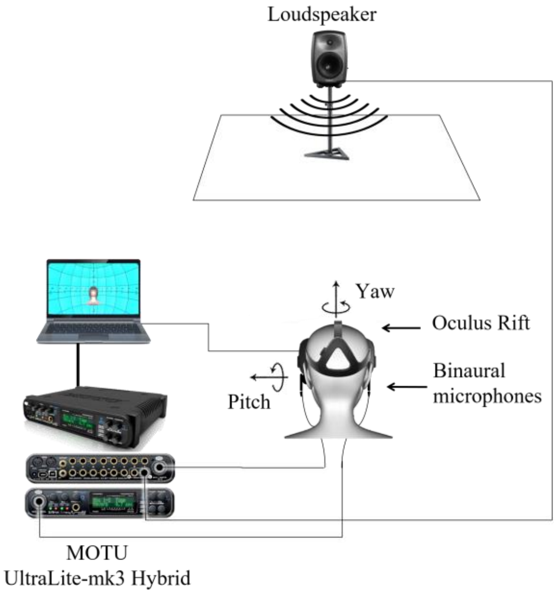

Figure 1. System diagram for personalized HRIR measurements.

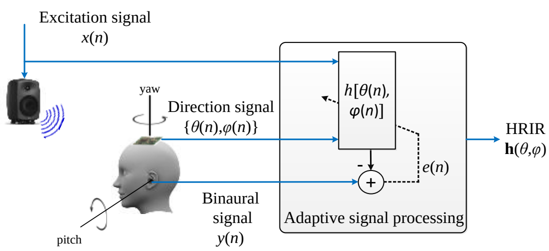

Figure 2. Adaptive system to estimate the HRIR/BRIR using binaural microphones and head tracker.

3 Measurement Results

We have compiled some preliminary measurement results of HRIR/BRIR captured under different conditions:

(i) Static versus dynamic movement

The measurement of HRIR/BRIR can be obtained using conventional start-stop (static) measurement compared to those measured under continuous movement (dynamic). Figures 3(a) and (b) show the left and right HRIR measured using static (above) and dynamic (below) at azimuth 0 degree and -90 degree, respectively. It is observed that close similarities of their impulse responses, and spectrum variation can be better observed in HRTF plots.

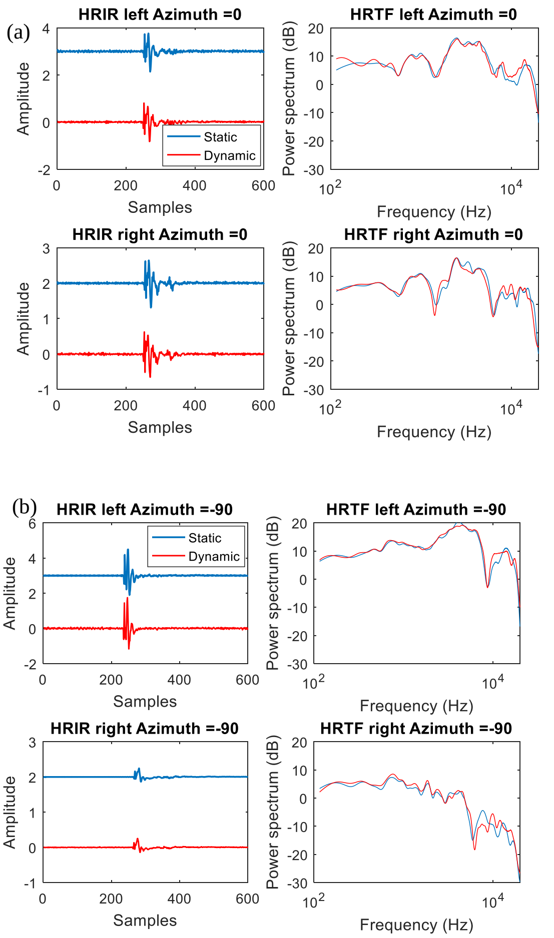

Figure 3(a) and (b). Show the left/right HRIR/HRTF measured using static and dynamic movement at 0 degree and -90 degree azimuth.

(ii) With and without wearing VR gear

Similar measurement can be carried out by acquiring the HRIR/BRIR wearing VR gear (with built-in head tracker) and without wearing VR gear (using a standalone head tracker with external screen to indicate visited location) in dynamic movement. Figure 4 shows the HRIR/BRIR measured with and without VR Rift headsets at an azimuth of -60 degree. Slight variation in the high frequency can be observed, but in general both spectrum are similar.

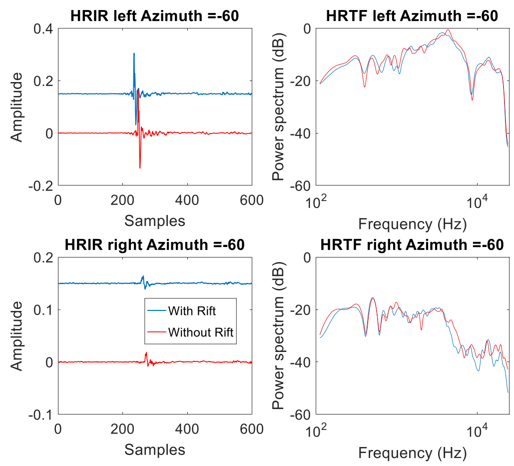

Figure 4. Show the left/right HRIR/HRTF measured with and without VR headgear at -60 degree azimuth.

(iii) Tandem versus non-tandem head and torso movement

The flexibility of our fast HRIR measurement system is to allow head movement (just head rotation), while keeping our torso facing in front. This measurement can be compared to the conventional HRTF measurement in anechoic chamber, where both head and torso movements are usually rotated in sync. Figure 5 shows some variations of the HRIR/BRIR between tandem and non-tandem head-and-torso movement. The usual peak and null patterns of the two HRTFs have some variations across mid-high frequency range.

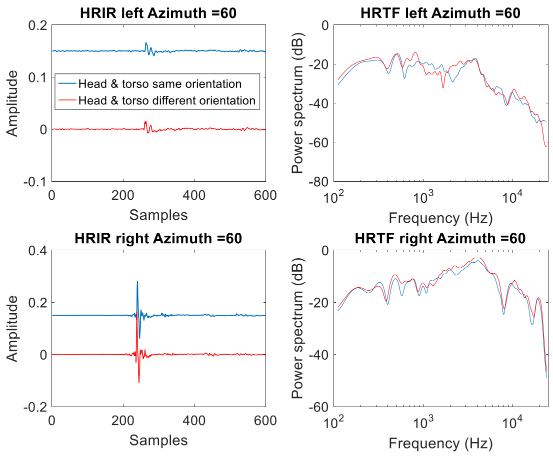

Figure 5. Show the left/right HRIR/HRTF measured with and without head-torso movement at 60 degree azimuth.

A comprehensive study of the measured HRIR/BRIR is currently carried out to compare its measurement accuracy with the reference HRIR/BRIR measurement (physical placement of speaker with reference to the human subject) at different azimuths and elevations in 3D space.

4 3D Audio Rendering and Headphones

The measured HRIR/BRIR data can be used in synthesizing personal 3D audio effect and playback through a pair of headphones. The 3D audio rendering module is linked seamlessly with the previous HRIR/BRIR acquisition module (in Figure 1) through a sharable database between the two modules. However, depending on the user acquisition rate for HRIR/BRIR acquisition, some locations may not be captured to a certain degree of error performance, and will require further processing, such as spatial interpolation to obtain a complete HRIR/BRIR database that covers all required locations. Other processing algorithms include refining HRTF resolution, headphone equalization, cross fading, and mixing of multiple virtual sound sources can all be carried out in the 3D audio rendering module.

In this e-Brief, we will also introduce our patented 3D audio headphones [16,17] that can be used to enhance the frontal perception of sound images. This new type of 3D audio headphones was derived from an intensive study by Sunder *et al.* [18].

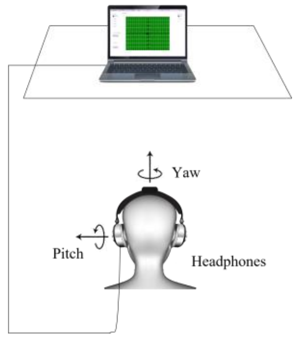

Figure 6. Show the additional processing required in reproducing the 3D audio effect based on the personalized HRTF measured in the same environment

By specifically projecting sound to the ear, we can inherently capture the idiosyncratic frontal pinna spectral cues of the listener during the 3D sound playback. Subjective studies [18] had shown that the front-back reversal can be reduced by almost 50% using this type of 3D audio headphones, thus improving the veracity of the 3D audio.

The advantage of this technique is that it does not require any measurements, training nor collecting anthropometric data of the listener. However, the frontal projection individualization technique has been limited to horizontal plane visualization, this new type of 3D audio headphones requires special kind of headphone equalization [18].

5 Applications and Conclusions

The ability to obtain quick personalized HRIR/BRIR opens up opportunity for AVR user to listen to accurate-rendered virtual sound objects on the move and better able to fuss the virtual sound images to the real audio environment. This continuous 3D audio acquire and render system will greatly help in interactive gaming by able to synthesize virtual sound objects in 3D space and render over the headsets on the fly. The 3D audio headphones provide additional personal frontal spectral cues that will project sound images that improves externalization in the frontal direction. Applications such as navigation, simulator, virtual museum and virtual tourists can provide demands for these personalized audio capture/render tool to create a more immersive AVR experience. However, there is a need to further customize various tools for different applications.

Acknowledgement

This work was partially supported by the NTU-NHG Ageing Research Grant, under the project code ARG/14003. We would like to thank Oculus for donating two sets of Oculus Rift for us to carry out this reserach.

References

[1] D. Begault, *3D Sound for Virtual Reality and Multimedia*. Cambridge, MA: Academic Press, 1994.

[2] R. Ranjan, and W. S. Gan, "Natural Listening over Headphones in Augmented Reality Using Adaptive Filtering Techniques," *IEEE/ACM Trans. Audio, Speech, Lang. Process.,* vol. 23, no. 11, pp. 1988-2002, Nov. 2015.

[3] H. Møller, M. Sorensen, D. Hammershøi, and C. Jensen, "Head-related transfer functions of human subjects," *J. Audio Eng. Soc.*, vol. 43, no. 5, pp.300–321, May 1995.

[4] K. Sunder, J. He, E. L. Tan, W. S. Gan, "Natural sound rendering for headphones: integration of signal processing techniques," *IEEE Signal Process. Mag.*, vol. 23, no. 2, pp. 100-114, Mar. 2015.

[5] C. J. Tan and W. S. Gan, "User-defined spectral manipulation of HRTF for improved localisation in 3 D sound systems," *Electronics letters,* vol. 34, no. 25, pp. 2387-2389, Dec. 1998.

[6] K. Sunder, E. L. Tan, and W. S. Gan, "Individualization of binaural synthesis using frontal projection headphones," *J. Audio Eng. Soc.*, vol. 61, no. 12, pp. 989-1000, Dec. 2013.

[7] D. N. Zotkin, J. Hwang, R. Duraiswaini, and L. S. Davis, "HRTF personalization using anthropometric measurements," in *Proc. WASPAA*, New York, pp. 157-160, Oct. 2003.

[8] J. He, W. S. Gan and E. L. Tan, "On the preprocessing and postprocessing of HRTF individualization based on sparse representation of anthropometric features," in *Proc. ICASSP*, Brisbane, Australia, pp. 639-643, Apr. 2015.

[9] G. Enzer, Chr. Antweiler, and S. Spors, "Trends in acquisition of individualized head-related transfer functions," *in J Blauert (ed) The Technology of Binaural Listening, Chapter 3*, Springer, 2013.

[10] G. D. Romigh, "Individualized head-related transfer functions: efficient modeling and estimation from small sets of spatial samples," PhD dissertation, School of Electrical and Computer Engineering, Carnegie Mellon University, 2012.

[11] P. Majdak, P.Balazs, and B. Laback. "Multiple exponential sweep method for fast measurement of head-related transfer functions," *J. Audio Eng. Soc.*, vol. 55, no. 7/8, pp. 623–637, Jul./Aug. 2007.

[12] V. Pulkki, M.. V. Laitinen, and V. P. Sivonen, "HRTF measurements with a continuously moving loudspeaker and swept sines," in *Proc. 128th AES conv*., London, UK, May 2010.

[13] J. He, R. Ranjan, and W. S. Gan, "Fast continuous HRTF acquisition with unconstrained movements of human subjects," in Proc. ICASSP, Shanghai, China, Mar. 2016, pp. 321-325.

[14] R. Ranjan, J. He, and W. S. Gan, "Fast continuous acquisition of HRTF in 2D for human subjects with unconstrained random head movements," in Proc. AES Headphone conference, Aalborg, Denmark, Aug. 2016.

[15] N Hai et al., "Fast HRTF personalization system with unconstrained head movements for 3D Audio in Virtual and Augmented Reality Applications," in demo presentation of 2017 International conference in Acoustic, Speech and Signal Processing, New Orleans, USA, Mar. 2017.

[16] W.S. Gan et al., Listening Device and Accompanying Signal Processing Method *US patent (US 9,357,282 B2) granted on May 31, 2016.*

[17] K. Sunder*, J.J. He*, E.L. Tan*, *W.S. Gan,* "Natural Sound Rendering for Headphones: Integration of Signal Processing Techniques," *IEEE Signal Processing Magazine, Vol 32, No. 2, pp 100-113, March 2015.*

[18] K. Sunder, E. L. Tan, and W. S. Gan, "Individualization of binaural synthesis using frontal projection headphones," *J. Audio Eng. Soc.,* vol. 61, no. 12, pp. 989-1000, Dec. 2013.