Fast HRTF Measurement System with Unconstrained Head Movements for 3D Audio in Virtual and Augmented Reality Applications

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

[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[2]Maxim Integrated Products, Inc., Singapore

*Corresponding author

NDHAI@ntu.edu.sg, NKCHAUDHARY@ntu.edu.sg, SPEKSI@ntu.edu.sg,
RISHABH001@e.ntu.edu.sg, Jianjun.He@maximintegrated.com, EWSGAN@ntu.edu.sg

Abstract— Binaural audio plays an indispensable role in virtual reality (VR) and augmented reality (AR). Binaural audio recreates the sensation of the three dimensional auditory experience using Head-Related Transfer Functions (HRTFs). HRTFs are as unique as our fingerprint. To achieve an immersive audio experience, HRTFs measured from every particular user is required. Nowadays, the conventional methods for HRTF measurements requires a well-controlled environment, hardly any movement of the user, and projecting to the user a high level of unpleasant sound in a rather long duration. Such difficulties have greatly limited the use of individually measurement HRTFs and hinder the authenticity of immersive audio. To solve these problems, we proposed a fast and convenient HRTF measurement system that is an order of magnitude faster and more importantly, it does not place any constraints on the user's movement. With the help of a head-tracker and advanced adaptive signal processing algorithms, this system is able to achieve satisfactory HRTF measurement accuracy. In this demonstration, we will present a fast real-time HRTF acquisition system and show how the individualized HRTFs improve the audio experience in VR/AR applications.

Keywords—Fast and relaxed HRTF acquisition, Binaural rendering, Virtual reality (VR), Augmented reality (AR)

INTRODUCTION

With the advent of virtual and augmented reality headsets like Oculus Rift, Microsoft HoloLens, etc., natural 3D sound rendering has become an integral part of user experience along with the visual aspect. In this context, 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 the idiosyncratic anthropometric characteristics [3], HRTFs are unique to each person and thus, individualized HRTFs should be used for natural 3D rendering. As a result, individualized HRTFs are required to be measured for every individual from acoustical measurements [4]. Alternative ways to approximately individualize the HRTFs using non-individualized HRTF database includes subjective tuning [5], using frontal projection headphones [6], and anthropometric features [7, 8].

COVENTIONAL TECHNIQUES

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.

OUR PROPOSED TECHNIQUES

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, as shown in Fig. 1. 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

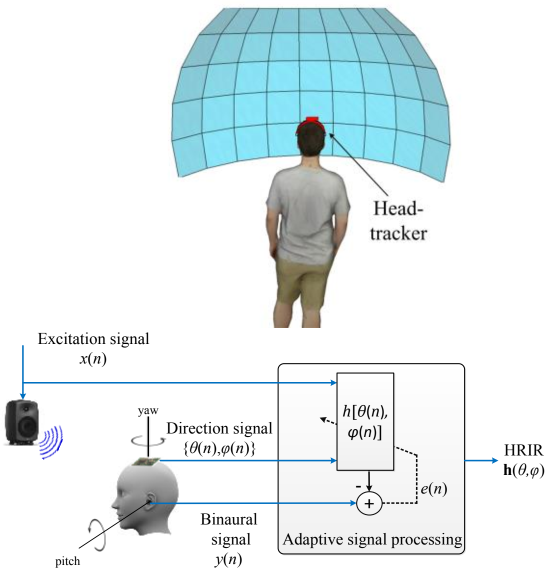

Fig. 1 System overview [14]

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.

PROPOSED DEMO SYSTEM

This demo consists of two steps. The first step is to measure the HRTFs for the user. As shown in Fig. 2, the user will be seated on a chair with binaural microphone worn on the ear and head tracker attached. The loudspeaker is placed about 1m right in front center of the user at ear level. As the measurement starts, the loudspeaker emanates the excitation signal, and the user moves his/her head horizontally and vertically without any constraints. An illustration will be shown in the laptop to feedback to the user the coverage of the directions and the user will be suggested to move to the less covered directions to obtain a balanced HRTF measurement accuracy. The software in the laptop will record the audio data and head-tracker data. This process will take around 1-2 minutes to cover entire frontal grid shown in Fig. 1. As soon as the measurement is finished, our software will compute the HRTFs for given resolution in both azimuth and elevation (say 5 degree). In the second step, we will validate the measurement using binaural auralization. We will synthesize audio tracks using the measured HRTFs as well as HRTFs from other users as a comparison. The user will listen to these tracks using headphones and evaluate how their sound localization can be improved using the personalized HRTFs from our measurements. The binaural auralization could be combined with visual images or animation using VR/AR headsets. For this demo, we would need some space to place our speaker

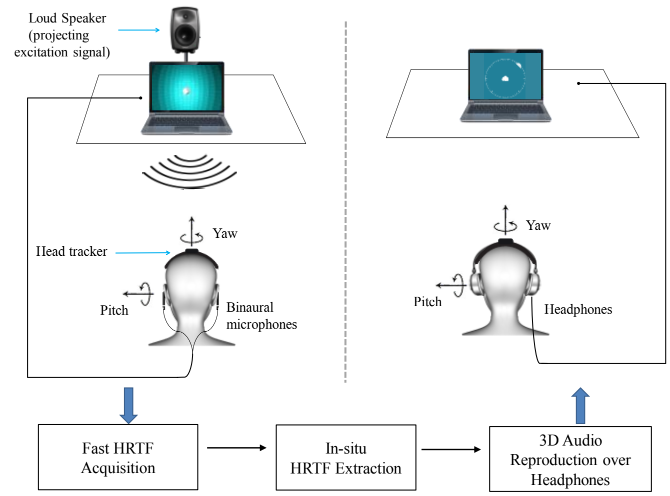

Fig. 2 Setup of the proposed demo system

system a chair for the user. Moreover, it would be ideal if the space is a separate room or a quiet corner of the room.

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.