Accepted by ICASSP 2016

FAST CONTINUOUS HRTF ACQUISITION WITH UNCONSTRAINED MOVEMENTS OF HUMAN SUBJECTS

Jianjun He, Rishabh Ranjan, and Woon-Seng Gan

Digital Signal Processing Lab, School of Electrical and Electronic Engineering,
Nanyang Technological University, SINGAPORE
{jhe007@e.ntu.edu.sg, rishabh001@e.ntu.edu.sg, ewsgan@ntu.edu.sg}

ABSTRACT

Head related transfer function (HRTF) is widely used in 3D audio reproduction, especially over headphones. Conventionally, HRTF database is acquired at discrete directions and the acquisition process is time-consuming. Recent works have been proposed to improve HRTF acquisition efficiency via continuous acquisition. However, these HRTF acquisition techniques still require subject to sit still (with limited head movement) in a rotating chair. In this paper, we further relax the head movement constraint during acquisition by using a head tracker. The proposed continuous HRTF acquisition technique relies on the activation based normalized least-mean-square (ANLMS) algorithm to extract HRTF on the fly. Experimental results validated the accuracy of the proposed technique, when compared with the standard static acquisition technique.

***Index Terms*— Head related transfer function, least-mean-square, fast acquisition , unconstrained movements**

1. INTRODUCTION

Binaural rendering is an important and popular technique for virtual auditory display and 3D audio reproduction [1], [2]. Head related transfer functions (HRTF, or its time-domain representation, HRIR), are fundamental in binaural rendering [3]. However, the measured HRTFs vary from person to person [4], [5], [6]. Therefore, to recreate an immersive listening experience using binaural rendering, individualized HRTF is required [7], [8], [23]. In general, individualized HRTFs can be obtained from direct acoustic measurements [21] with interpolation [24], [25], perceptual feedbacks [26]-[28], using special frontal projection emitters [29], and anthropometry measurements [30]-[36].

Obtaining individualized HRTF requires individual acoustic measurements with subjects sitting in an anechoic chamber and a loudspeaker, located at a particular direction and distance, emits an excitation signal. This signal is captured by a binaural microphone placed at the entrance of the block ear canal [1], [4] of a human. Conventional techniques to measure HRTFs are conducted at discrete positions in a “start-stop-move” manner [9]. Consider 360° of each plane and HRTF measurements require an azimuthal resolution of 5° [10], this process could take hours to complete. Therefore, there are extensive efforts to reduce the measurement duration. Interpolation has been widely used to reduce the angular resolution of the HRTF measurements [11]. In [12], Pulkki *et al.* proposed to use a rotating loudspeaker rather than a static loudspeaker at discrete positions. Since the directions in HRTF acquisition are relative between the loudspeaker position and orientation of human subjects, rotating the human subject is equivalent to rotating loudspeaker. Considering the compactness of the facilities of the rotating setup, rotating the human subject is more advantageous than rotating the loudspeaker [13]. In [14], Enzer employed a least-mean-square (LMS) adaptive technique to estimate HRTF from 1-D (horizontal plane) continuous measurements, which is also extended to 3-D [15] and continuous HRTF fields [16]. Frequency domain adaptive filtering techniques could also be employed, as reported [17]. Studies in [18] validated the efficacy of continuous HRTF measurements.

However, the existing continuous HRTF measurements still require a rotating machine and subjects have to sit still on the chair, which sometimes could be painstaking or impractical [19]. Also, it can only measure the HRTFs of sitting posture rather than a conventional standing posture. To further relax the human subjects in HRTF measurement, we propose a new, fast and unconstrained continuous HRTF measurement system in this paper. In this system, the subject is free to rotate his/her head (or body), and its movement is automatically captured by a head-tracker equipped on the head of the subject. Binaural recordings are also simultaneously measured and synchronized with the excitation signal, as well as the subject movement data. We propose activation based normalized LMS (ANLMS) technique to estimate HRTF from this system. In the next section, we will formulate the problem and present the algorithm, which is followed by the simulation results.

2. PROBLEM STATEMENT

As shown in Fig. 1, in the proposed system, we emit an input excitation signal $x(n)$ from a loudspeaker, and record

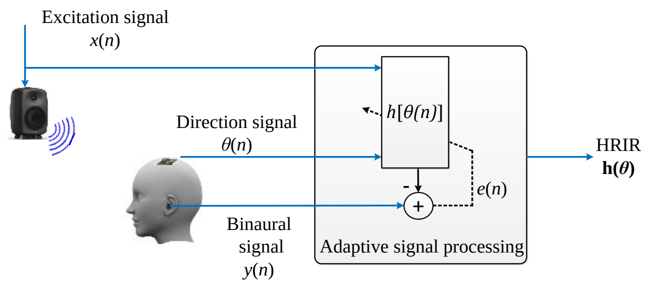

Fig. 1 An overview of the proposed fast and unconstrained continuous HRTF measurement system

the binaural signal $y(n)$ at the blocked ear (left and right) position of the listener and obtain the direction information of the head (or even torso) movement $\theta(n)$ through a head-tracking device. It is important that the three signals are captured synchronously. Given these three signals as input, our aim is to obtain the HRTF of the subject via an adaptive filter. Note that the HRTF can only be derived at the measured directions that the listener has rotated to. In order to achieve denser HRTF measurements, subject has to move his/her head to cover wider locations. In the following, we will only show the derivation of one side of the HRTF, as the other side can be obtained similarly.

Next, we introduce our signal model. Based on the linear time-invariant (LTI) system, we can express the binaurally recorded signal as

$$\begin{aligned} y(n) &= \sum_{l=0}^{L} x(n-l)h\left[l,\theta(n)\right]+v(n) \\ &= \mathbf{h}^T\left[\theta(n)\right]\mathbf{x}(n)+v(n), \end{aligned} \quad (1)$$

where $h\left[l,\theta(n)\right]$ represents one sample of the HRIR (left or right) for the dynamically varying direction $\theta(n)$. Here, $\theta(n)$ has been linearly interpolated to match the sampling rate of the audio signal since the head-tracker usually employs a much lower sampling rate, compared to the audio signals. The total length of HRIR is L. $v(n)$ refers to the measurement and model noise. And the vectors are expressed as follows:

$$\begin{aligned} \mathbf{x}(n) &= \left[x(n), x(n-1), \ldots, x(n-L+1)\right]^T \\ \mathbf{h}\left[\theta(n)\right] &= \left\{h\left[0,\theta(n)\right], h\left[1,\theta(n)\right], \ldots, h\left[L-1,\theta(n)\right]\right\}^T. \end{aligned} \quad (2)$$

To account for the free movement of the head and the resulting direction $\theta(n)$, we can rewrite (1) as:

$$y(n) = \mathbf{d}^T(n)\mathbf{H}^T\mathbf{x}(n)+v(n), \quad (3)$$

where $\mathbf{H} = \left[\mathbf{h}(\theta_1), \mathbf{h}(\theta_2), \ldots, \mathbf{h}(\theta_K)\right]$ is a matrix representation of K discrete HRIRs to be estimated, and $\mathbf{d}(n) = \left[0,\ldots,0,1,0,\ldots,0\right]^T_{1\times K}$ is the direction activation vector, which has a value of one only at position that corresponds to the direction $\theta(n)$, and zero for the other directions. The advantage of (3) is that simplifies the change of HRTF filters into the change of the direction activation vector. This could better account for the free movements and the associated randomly updated directions.

Table I Settings of the five experiments

#	Rotation style	Step size μ	Rotation duration(s)	SNR_x (dB)	Azimuthal resolution
1	Random and fixed	0.1	10	30	5°
2	Random	[0.01, 0.05, 0.1, 0.5, 1]	10	30	5°
3	Random	0.1	[1, 5, 10, 15]	30	5°
4	Random	0.1	10	[20, 30, 40]	5°
5	Random and fixed	0.1	10	30	5°, 1°

3. ACTIVATION BASED NLMS ALGORITHM

To solve the problem in (3), we extended the NLMS algorithm used in [14] that estimates HRIR from the measurement based on a fixed speed rotating setup. The adaptive filter updating equation is expressed as

$$\hat{\mathbf{H}}_{n+1} = \hat{\mathbf{H}}_n + \mu \frac{e(n)}{\left\|\mathbf{x}(n)\right\|_2^2}\mathbf{x}(n)\mathbf{d}^T(n), \quad (4)$$

where $e(n) = y(n) - \mathbf{d}^T(n)\hat{\mathbf{H}}_n^T\mathbf{x}(n)$. We refer this algorithm as activation based NLMS (ANLMS). In fact, at each time n, only one HRIR filter is updated since $\mathbf{d}(n)$ has only one non-zero value each time. This implies that our ANLMS exhibits the same computational complexity as NLMS [14], but requires more memory (K times) to store all HRTFs at each iteration.

To evaluate the performance of the proposed system, we compute the normalized mean-square error (NMSE) of HRIR estimation [14], [15], which is written as:

$$\text{NMSE}(\theta) = 10\log_{10}\left(\frac{\left\|\hat{\mathbf{h}}(\theta)-\mathbf{h}(\theta)\right\|_2^2}{\left\|\mathbf{h}(\theta)\right\|_2^2}\right). \quad (5)$$

Lower NMSE values indicate a better HRIR accuracy. Note that other factors that affect the performance include: step size, signal–to-noise ratio, and rotation speed (or the rotation duration for a certain range of directions). Next, we conduct simulations to evaluate the performance of the proposed system.

4. SIMULATION RESULTS AND DISCUSSIONS

To evaluate the proposed HRTF measurement system, we conducted a simulation experiment. The experiment simulates binaural recording using signal model (3), i.e., applying dynamic filtering to the input signal. The HRTF in this experiment is selected from the CIPIC HRTF database [21] (subject 008). The input signal is a white noise, though according to studies in [37], there are other choices, such as

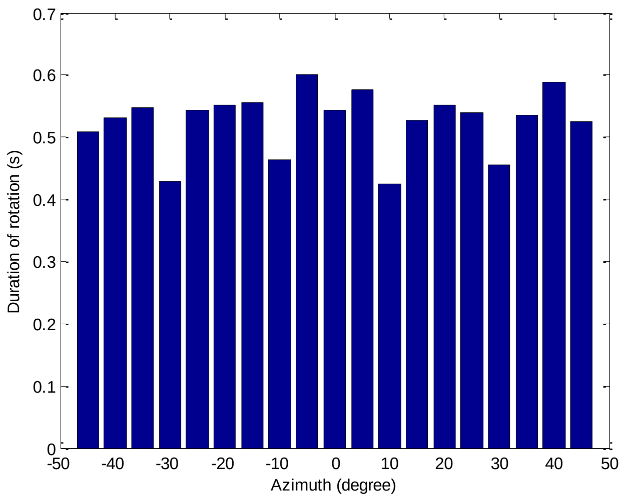

Fig. 2 Histogram of the random rotation at different directions

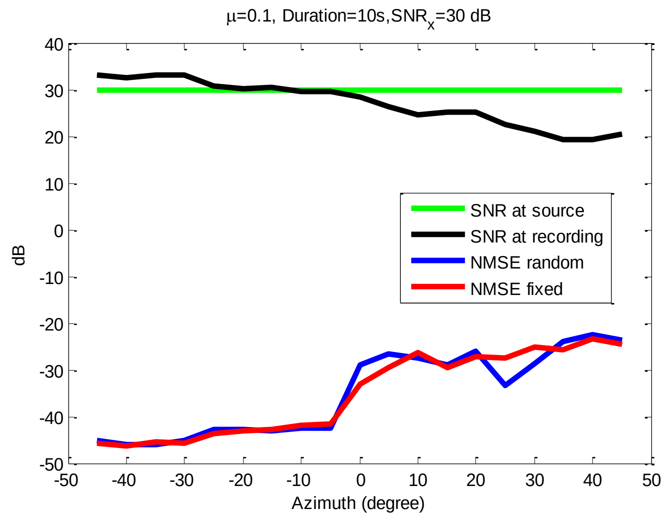

Fig. 3 NMSE and SNR results of random and fixed rotation

multiple exponential sweep, perfect sequences, etc. The directions considered is from -45° to 45° azimuth with elevation set as 0°. The length of HRIR is L = 200 samples, at a sampling rate of 44,100 Hz. The sampling rate of the head tracked is simulated at 250 Hz. To simulate the binaural recording, we convolve the excitation signal with the HRTF from -45° to 45° (in step of 5°), which results in a total of K=19 HRTFs. The measurement and model noise $v(n)$ is also generated using white noise, at different level of SNR. We consider the SNR at the source as $\text{SNR}_x = 10\log_{10}\left[E\|x(n)\|_2^2 / E\|v(n)\|_2^2\right]$. By default, we set SNR_x at 30 dB. Other default settings include step size of μ = 0.1, rotation duration = 10s. In the following, we consider five experiments, where the settings are specified in Table I. The NMSE results only consider the HRTFs of the left ear.

In the first experiment, we compare the random movement (currently, only random rotation style is considered) with the fixed constant speed rotation (as used in previous continuous HRTF measurements [14]). Fig. 2 shows the duration of each direction from a randomized rotation. In Fig. 3, we plot the NMSE results for different directions. It can be seen that the NMSE is quite low, from -45 dB at -45° to -25 dB at 45°. The inferior performance at the positive direction is because those HRTFs have lower amplitude compared to the HRTF of the negative directions. To validate this, we compute the SNR at the binaural recording as $\text{SNR}_y(\theta) = 10\log_{10}\left[E\left\|\left[x * h(\theta)\right](n)\right\|_2^2 / E\|v(n)\|_2^2\right]$, which is also plotted in Fig. 3. Clearly, the lower the SNR_y, the worse the NMSE. Furthermore, we also illustrated the results of the fixed speed rotation case (as used in [14]) in Fig. 3, and observed that the NMSE is quite close. This implies that either using a random rotation or fixed speed rotation, the accuracy of HRTF measurement is comparable, and thus, it validates the efficacy of the proposed system. A further analysis shows that the HRIR estimation errors for the 19 directions are insignificant.

In the second experiment, we examine the effect of the step size, as shown in Fig. 4. Clearly, the performance is quite poor when the step size is too small (μ = 0.01) or too large (μ = 0.5, 1). Better performance are achieved with step size of μ = 0.05, 0.1.

In the third experiment, we examine the effect of the rotation speed or rotation time. In any HRTF measurement, it is more desirable to reduce the measurement time. However, reducing the measurement time could have an adverse effect on the HRTF measurement accuracy. Based on the results in Fig. 5, at least 5s is needed to cover a 90° range. That is to say, for 360°, this is equivalent to the revolution time of 20s.

In the fourth experiment, we examine the effect of the SNR. As shown in Fig. 6, higher SNR leads to lower NMSE, especially at the ipsilateral directions.

In actual binaural recording, the azimuth resolution is very small. In order to simulate the actual binaural recording better, a smaller azimuthal resolution is required for binaural synthesis. In the last experiment, we evaluate the effect of the azimuthal direction resolution in the binaural synthesis. It can be observed from Fig. 7 that using a more coarse resolution in the ANLMS model (5°) than in the simulated binaural recording (1°) would result in 20-30 dB worse of NMSE, compared to the matched resolution of 5°. Increasing the azimuth resolution in ANLMS model (1°) improves the NMSE by 5-10 dB and obtains HRTFs at more directions (every 1°). Notably, this observation is common in both random and fixed rotation cases. Thus, we conclude that including more HRTFs (i.e., increase K) with finer resolution in the signal model stated in (3) could better match the actual binaural recording and lead to improvements on the HRTF measurement accuracy.

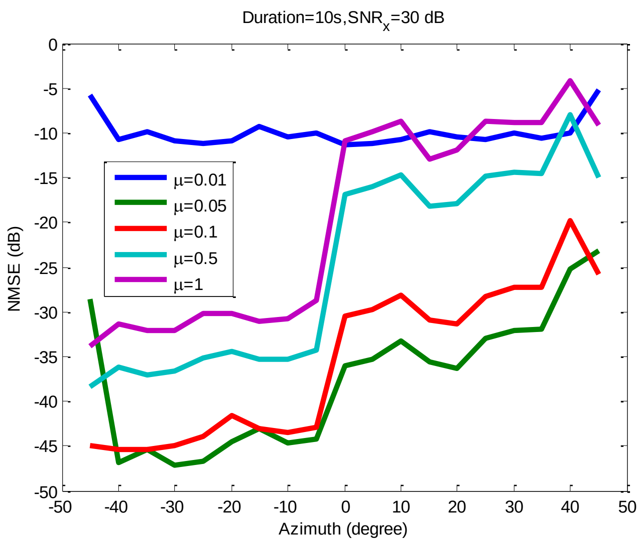

Fig. 4 Effect of the step size

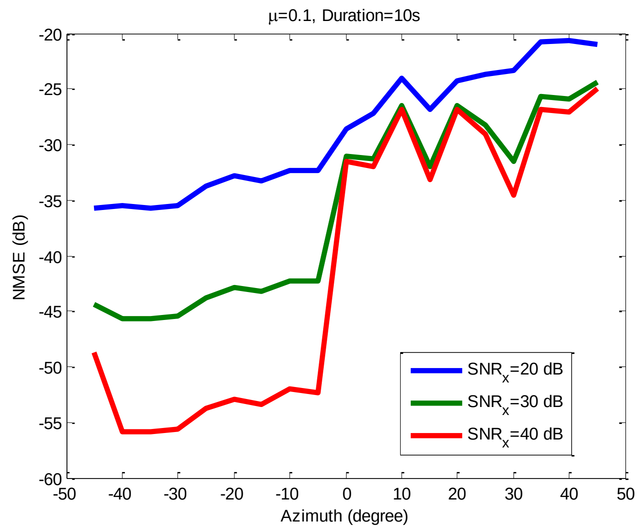

Fig. 6 Effect of SNR

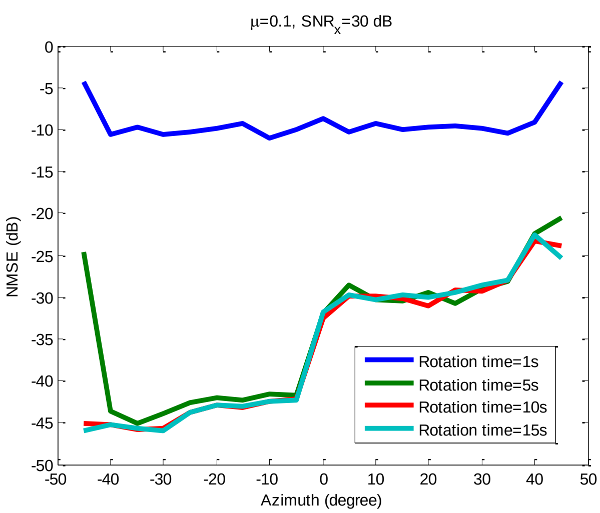

Fig. 5 Effect of the rotation time

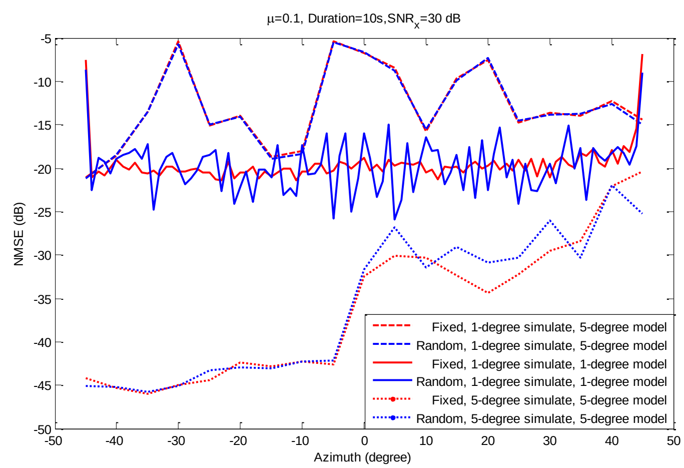

Fig. 7 Effect of azimuthal direction resolution in binaural synthesis

5. CONCLUSIONS

In this paper, we introduced a novel, fast and unconstrained continuous HRTF measurement system, where the subjects are free to move (rotate) their head. With a head-tracker, the head movement information is automatically recorded, together with the binaural signals and the excitation signal. Employing the activation based NLMS algorithm, we can obtain a set of HRTFs with accuracy comparable to the conventional continuous HRTF measurement with fixed speed rotation (negligible difference). Simulations using synthesized binaural signals validate the effectiveness of the proposed system. Future work includes the subjective evaluation (using ABX test) of the proposed method.

REFERENCES

[1] J. Blauert, *Spatial Hearing*. Cambridge, MA: The MIT Press, 1997.

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

[3] H. Møller, “Fundamentals of binaural technology,” *Applied Acoustics*, vol. 36, no.3/4, pp. 171–218, Mar./Apr. 1992.

[4] 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.

[5] E. Wenzel, M. Arruda, D. Kistler, and F. Wightman, “Localization using non-individualized head-related transfer functions,” *J. Acoust. Soc. Amer.*, vol. 94, no. 7, pp. 111–123, Jul. 1993.

[6] W. Gardner and K. Martin, “HRTF measurements of a KEMAR,” *J. Acoust. Soc. Amer.*, vol. 97, no. 6, pp.3907–3908, Jun. 1995.

[7] 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.

[8] H. Møller, M. Sorensen, C. Jensen, and D. Hammershøi, “Binaural technique: do we need individual recordings?” *J. Audio Eng. Soc.*, vol. 44, no. 6, pp. 451–469, Jun. 1996.

[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] T. Ajdler, L. Sbaiz, and M. Vetterli, “Plenacoustic function on the circle with application to HRTF interpolation,” in *Proc. ICASSP*, Philadelphia, PA, Mar. 2005.

[11] P. Minnaar, J. Plogsties, and F. Christensen, “Directional resolution of head-related transfer functions required in binaural synthesis,” *J. Audio Eng. Soc.*, vol. 53, no. 10, pp. 919–929, Oct. 2005.

[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] K. Fukudome, T. Suetsugu, T. Ueshin, R. Idegami, and K. Takeya, “The fast measurement of head related impulse responses for all azimuthal directions using the continuous measurement method with a servo-swiveled chair,” *Applied Acoustics*, vol. 68, no. 8, pp. 864–884, Aug. 2007.

[14] G. Enzner, “Analysis and optimal control of LMS-type adaptive filtering for continuousazimuth acquisition of head related impulse responses,” In *Proc. ICASSP*, Las Vegas, NV, Apr. 2008.

[15] G. Enzner. “3D-continuous-azimuth acquisition of head-related impulse responses usingmultichannel adaptive filtering,” in *Proc. WASPAA*, New Paltz, NY, Oct. 2009.

[16] G. Enzner, M. Krawczyk, F. M. Hoffmann, and M. Weinert. 3D reconstruction of HRTF-fields from 1D continuous measurements,” in *Proc. WASPAA*, New Paltz, NY, Oct. 2011.

[17] P. Y. Ren, and Z. H. Fu, “Fast and accurate high-density hrtf measurements using generalized frequency-domain adaptive filter,” in *Proc. ICSV*, Beijing, China, Jul. 2014.

[18] S. Malik, J. Fligge, and G. Enzner. “Continuous HRTF acquisition vs. HRTF interpolation for binaural rendering of dynamical auditory virtual environments,” In *Proc. ITG Conf. Speech Comm.*, Bochum, Germany, Oct. 2010.

[19] G. Wersényi, and J. Wilson, “Evaluation of Head Movements in Short-term Measurements and Recordings with Human Subjects using Head-Tracking Sensors,” *Acta Technica Jaurinensis*, vol. 8, no. 3, pp. 218-229, Mar. 2015.

[20] S. Haykin, *Adaptive Filter Theory, 4th ed.* Upper Saddle River, NJ: Prentice-Hall, 2002.

[21] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, “The CIPIC HRTF database,” in *Proc. WASPAA*, New Paltz, NY, USA, Oct. 2001.

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

[23] 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.

[24] H. Gamper, “Head-related transfer function interpolation in azimuth, elevation, and distance,” *J. Acoust. Soc. Amer.*, vol. 134, pp. EL547–554, Dec. 2013.

[25] 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.

[26] 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.

[27] B. U. Seeber and H. Fastl, “Subjective selection of nonindividual head-related transfer functions," in *Proc. ICAD*, Boston, MA. Jul. 2003, pp. 259-262.

[28] K. J. Fink, and L. Ray, “Individualization of head related transfer functions using principal component analysis,” *Applied Acoustics*, vol. 87, no. 1, pp. 162-173, Jan. 2015.

[29] 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.

[30] 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.

[31] H. Hu, L. Zhou, H. Ma, and Z. Wu, “HRTF personalization based on artificial neural network in individual virtual auditory space,” *Applied Acoustics*, vol. 69, no. 2, pp. 163– 172, Feb. 2008.

[32] L. Li and Q. Huang, “HRTF personalization modeling based on RBF neural network,” in *Proc. ICASSP*, Vancouver, British Columbia, Canada, May 2013.

[33] G. Grindlay and M. A. O. Vasilescu, “A multilinear approach to HRTF personalization,” in *Proc. ICASSP*, Honolulu, Hawaii, USA, Apr. 2007.

[34] S. Spagnol, M. Geronazzo, and F. Avanzini, “On the relation between pinna reflection patterns and head-related transfer function features,” *IEEE Trans. Audio, Speech, Lang. Process*., vol. 21, no. 3, pp. 508–519, Mar. 2013.

[35] P. Bilinski, J. Ahrens, M. R. P. Thomas, I. Tashev, and J. C. Plata, “HRTF magnitude synthesis via sparse representation of anthropometric features,” in *Proc. ICASSP*, Florence, Italy, pp. 4501-4505, May 2014.

[36] 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.

[37] C. Antweiler and M. Dörbecker, “Perfect sequence excitation of the NLMS algorithm and its application to acoustic echo control,” *Annales des Telecommunications*, vol. 49. Springer, 1994, pp. 386–397.