Proceedings of 20th International Congress on Acoustics, ICA 2010

23–27 August 2010, Sydney, Australia

Interpolation method of head-related transfer functions in the z-plane domain using a common-pole and zero model

Yukio Iwaya (1)(2), **Wataru Sato**(2), **Takuma Okamoto**(1)(3), **Makoto Otani** (1)(4), **and Yôiti Suzuki** (1)(2)

(1) Research Institute of Electrical Communication, Tohoku University, 2–1–1 Katahira, Aoba-ku, Sendai, Miyagi, Japan
(2) Graduate School of Information Sciences, Tohoku University, 6–3–09 Aramaki Aza Aoba, Aoba-ku, Sendai, Miyagi, Japan
(3) Graduate School of Engineering, Tohoku University, 6–3–09 Aramaki Aza Aoba, Aoba-ku, Sendai, Miyagi, Japan
(4) Faculty of Engineering, Shinshu University, 4–17–1 Wakasato, Nagano, Nagano, Japan

PACS: 43.66Pn,43.66Qp

ABSTRACT

Sound image localization can be controlled by convolving a listener's head-related transfer functions (HRTFs) corresponding to each sound source position to be rendered. Using this technique, a virtual auditory display (VAD) can be constructed, which can display sound images at arbitrary positions. The VAD applying this architecture should have a set of HRTFs of a listener to be given in advance. However, it is difficult to measure HRTFs in all directions around the listener. Therefore, an interpolation method of HRTFs measured at discrete positions is needed for such a VAD system. Previous studies have investigated methods in which linear interpolation is used in the time or frequency domain. These methods provide good accuracy when directions corresponding to the HRTFs used in interpolation are sufficiently close to each other. In contrast, when the directions are not close, the accuracy of the interpolation decreases markedly. In particular, the frequencies of spectral peaks and notches in HRTFs are inaccurate in the interpolated HRTFs because the frequencies of such peaks and notches vary according to the sound source position. On the other hand, the frequencies of notches are important cues of sound localization in elevation localization. Therefore, an interpolation method that can represent the frequencies of peaks and notches of HRTFs accurately is necessary to realize high-definition VAD based on the HRTF synthesis technique. This paper presents a proposal of a novel method for HRTF interpolation. In the method, HRTFs are first modeled using the common-pole and zero model in the z-plane. The interpolated HRTF is obtained by transforming the z-plane to the frequency domain. The accuracy of the proposed method was evaluated, demonstrating that the accuracy of reproduced spectral notches can be improved.

INTRODUCTION

We can perceive a three-dimensional position of a sound source in daily life: humans have sound localization ability [1]. It is known that the basic cues of sound localization are interaural level difference, interaural time difference, and spectral cues. The former two cues are caused by physical differences of the propagation paths of sound waves between two ears. They are used mainly in lateral localization [1]. On the other hand, spectral cues are mainly used when interaural differences are small, for example in a median plane localization [2]. Head-related transfer functions (HRTFs) of both ears include all cues. We can control perceived sound image positions by convolving HRTFs to a sound source signal. Using this theorem, a virtual auditory display (VAD) can be constructed [3, 4], which can display sound images at arbitrary positions. The VAD applying this theorem should have a set of HRTFs of a listener to be given in advance. However, measurement of a set of HRTFs requires complicated apparatus. Moreover, a listener must remain motionless during measurements [5]. Therefore, the measurement process often becomes annoying to the listener. Recently, numerical simulation of HRTFs has been performed [6, 7, 8, 9], by which a set of HRTFs is obtainable from a three-dimensional model of a listener. However, it is difficult to construct a three-dimensional model in usual cases. Several researchers investigated a selection method of a set of HRTFs from a HRTF-database with a hearing test [5, 10]. This technique is useful to select the best set of HRTFs from a database. Nevertheless, a set of HRTFs that is selected should be adjusted to each listener. In any case, HRTF measurements of continuous whole positions have not been realized. We can obtain HRTFs from limited discrete source positions. Therefore, an interpolation method of HRTFs using measured functions at discrete positions is needed for such a VAD system.

Several methods for interpolation of HRTFs in the time or frequency domain have been investigated in previous studies. These methods use a simple linear interpolation [11, 12, 13, 14]. Therefore, these methods have good accuracy if directions corresponding to the HRTFs used in the interpolation are mutually close to a sufficient degree. However, the accuracy of the interpolation decreases markedly when the directions are not close. Distinct spectral peaks and notches are present in HRTFs. These peaks and notches shift in frequency domain according to position of the sound source. Furthermore, Iida *et al.* [15] pointed out that these peaks and notches at around 5–10 kHz are important cues for elevation localization. However, the linear interpolation method cannot express the frequency shift of peaks and notches. Therefore, an interpolation method that can represent the frequencies of peaks and notches in HRTFs accurately is necessary to realize a high-definition VAD system based on the HRTF synthesis technique.

This paper presents a proposal of a novel method for HRTF interpolation. Using the proposed method, HRTFs are first modeled using the common-pole and zero model [17] in the z-plane.

Then the loci of zeros as a function of elevation are traced using the dynamic programming (DP) [16] method based on distances in the z-plane. Then the zeros are linearly interpolated in the z-plane for the desired directions. Consequently, the interpolated HRTF is obtained by transforming the z-plane to the frequency domain. The accuracy of the proposed method was evaluated to demonstrate that the spectral distortion can be improved. Furthermore, both the frequency and depth of notches were evaluated.

MODELING METHOD OF HEAD-RELATED TRANSFER FUNCTIONS IN THE Z-PLANE

Several interpolation methods have been proposed based on the modeled transfer function. The moving average model (MA model) [2] and the autoregressive moving average model (ARMA model) [11] are often used. However, these methods require numerous parameters because these parameters are independently calculated among different positions. In contrast, Haneda *et al.* [17] proposed the common-pole and zero model. This model is based on a simple concept: The poles in the transfer functions of a specific room are common irrespective of positions in the room and the zeros are associated to a sound source position. The common-pole and zero model is derived from these physical concept and poles are common among positions. Therefore, the transfer functions in the room can be represented by fewer parameters than the simple ARMA model has. Watanabe *et al.* [18] proposed an interpolation method of HRTFs in a horizontal plane based on the common-pole and zero model. However, the three-dimensional interpolation method has not been investigated. The authors apply the common-pole and zero model to three-dimensional interpolation of HRTFs.

Modeling of HRTFs based on the common-pole and zero model

An HRTF from a specific direction θ_m is modeled based on the common-pole and zero model[17, 18]. The modeled HRTF $H_{CAPZ}(\theta_m, z)$ can be described as

$$\begin{aligned} H_{CAPZ}(\theta_m, z) &= \frac{C_C z^{-Q_1} \prod_{i=1}^{Q_2} [1 - q_i(\theta_m) z^{-1}]}{\prod_{i=1}^{P} (1 - p_{C_i} z^{-1})} \\ &= \frac{\sum_{i=0}^{Q} b_i(\theta_m) z^{-i}}{1 - \sum_{i=1}^{P} a_{C_i} z^{-i}}, \end{aligned} \quad (1)$$

where C_C, p_{C_i} and $q_i(\theta_m)$ respectively represent a constant value, i-th common pole, and i-th zero. In addition, P and Q respectively denote the order of common-pole and zeros and $Q = Q_1 + Q_2$. The impulse response $h_{CAPZ}(\theta_m, k)$ is obtainable with the inverse z-transform as

$$h_{CAPZ}(\theta_m, k) = \sum_{i=1}^{P} a_{C_i} h_{CAPZ}(\theta_m, k - i) + \sum_{i=0}^{Q} b_i(\theta_m) \delta(k - i), \quad (2)$$

where k denotes the discrete time. Both a_{C_i} and $b_i(\theta_m)$ can be estimated so that the cost function $e_{eq}(\theta_m, k)$ becomes minimum in all measured directions, as

$$\begin{aligned} e_{eq}(\theta_m, k) &= h(\theta_m, k) - \sum_{i=1}^{P} a_{C_i} h(\theta_m, k - i) \\ &\quad + \sum_{i=0}^{Q} b_i(\theta_m) \delta(k - i). \end{aligned} \quad (3)$$

We can obtain poles and zeros from a_{C_i} and $b_i(\theta_m)$ with the relation expressed by Eq. 1. An HRTF to be synthesized can be estimated only by interpolating zeros using neighbor zeros at different positions if common poles can be estimated adequately.

The orders of poles and zeros are decided so that spectral distortion (SD) between measured and modeled HRTFs is smaller than a threshold value (0.4 dB), as

$$\mathrm{SD} = \sqrt{\frac{1}{N} \sum_{k=1}^{N} \left(20 \log \frac{|H(f_k)|}{|\hat{H}(f_k)|} \right)^2}, \quad [\mathrm{dB}] \quad (4)$$

where $H(f_k)$ and $\hat{H}(f_k)$ respectively denote measured and modeled HRTFs. In this case, the order of zeros is made as small as possible so that the parameters in interpolation become small. Consequently, the order of poles is made as large as possible.

PROPOSED METHOD

Using the modeled HRTFs based on the common-pole and zero model described in the previous section, a new interpolation is proposed. Our new method consists of four stages as follows.

1. Modeling HRTFs measured at discrete positions based on the common-pole and zero model
2. Associating zeros of different positions with the dynamic programming technique or with Euclidean distance in z-plane
3. Estimating zeros' positions with linear interpolation in the z-plane
4. Synthesizing interpolated HRTF with common poles and interpolated zeros

The first stage was described in the previous section. Details of the second stage are discussed in following subsections.

Association of zeros among different positions

The modeling of HRTFs produces a set of Q_2 zeros for each discrete source position. They are expected to be associated among different positions. In the proposed method, the dynamic programming method (DP) is applied to associate zeros in different positions. A cost function is defined as

$$\begin{aligned} c(i, j, j', j'') &= w_1 |z_{i,j} - z_{i-1,j'}| \\ &\quad + w_2 |z_{i,j} - 2z_{i-1,j'} + z_{i-2,j''}|, \end{aligned} \quad (5)$$

where $z_{i,j}$, $z_{i-1,j'}$, $z_{i-2,j''}$ are, respectively, the j-th zero at position i, j'-th zero at position $i-1$, and j''-th zero at position $i-2$. When zeros at two previous positions such as $z_{i-1,j'}$, $z_{i-2,j''}$ have been associated in advance, the associated zero at i-th position $z_{i,j}$ is decided recursively so that the cost function results in the minimum value. We used $w_1 = w_2 = 1$ for the performance evaluation that is explained later in this report.

An example of zero-association is presented in Fig. 1. A set of HRTFs measured in an anechoic room at Tohoku University with a spherical speaker array was used for these analyses. The HRTFs of a dummy head (Samrai; Koken Co. Ltd.) in a median plane were measured from frontal –80 deg. to frontal +90 deg. with an interval of 10 deg. An optimized Aoshima's Time-Stretched Pulse (OATSP) [19] (sampling frequency: 48 kHz, 8192 samples) was used for measurements. Measurement results were modeled with common-pole and zero model with orders such as $P = 23$ for poles and $Q = 18$ for zeros. Results of the zero-association are shown in Fig. 1. Figure 1 portrays a z-plane expression of zeros for all elevation directions. A unit circle is also shown in Fig. 1 with a frequency scale. Small white and black circles respectively denote zeros and associated zeros. The results demonstrate that zeros can be associated

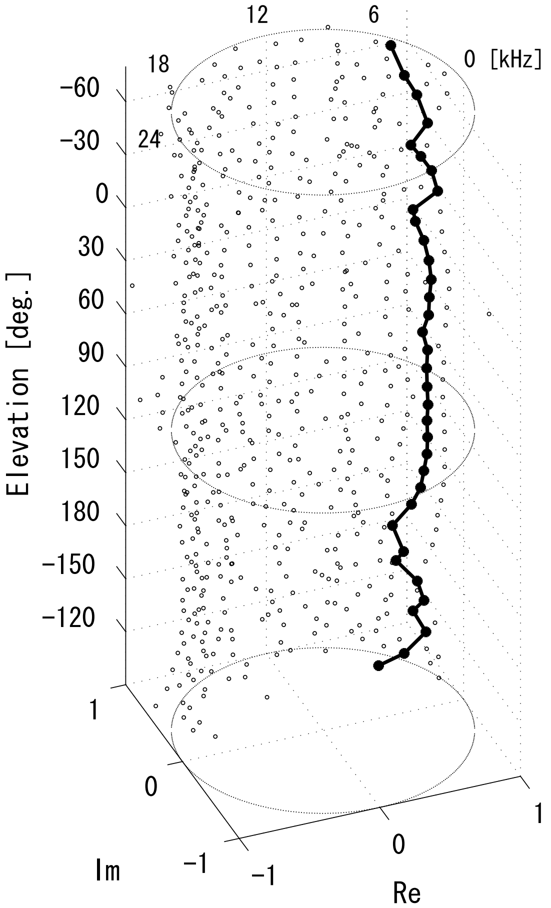

Figure 1: Example of association of zeros

adequately in a low-frequency region. However, the distances among zeros decreased at high frequencies and the association with the proposed method does not work well.

Additionally, we found that the orders of zeros increase when the positions of HRTF to be modeled are in a shadowed area, where a sound source is located at the contra-side of the concerned ear. Therefore, association is more difficult in the side. It might be better that the orders of zeros and poles in modeling are not common among all positions and that they should be decided in its position adequately. Then the association and interpolation in z-plane should be performed with the neighbor positions, where the orders might be the same. Therefore, we simply use four positions neighboring the target position to be interpolated in zero-association (Fig. 2).

Interpolation of zero's position in z-plane

The neighboring four positions shown in Fig. 2 are separated by $\|\phi_1 - \phi_2\|$ in elevation and by $\|\theta_1 - \theta_2\|$ in the horizontal direction. In fact, $h_k(n)$ $(n = 1,2,3,4)$ is the time-domain representation of HRTFs, i.e., head-related impulse responses (HRIR), for each position. Furthermore, $h(\theta,\phi,n)$ is the HRIR to be interpolated. When the associated zeros in two positions are given as Z_{p1} and Z_{p2}, an interpolated zero Z_{pE} is estimated as

$$Z_{pE} \quad = \quad (1-r)\cdot Z_{p1} + r\cdot Z_{p2}, \tag{6}$$

where r denotes weighting determined by the following equations:

$$r_\theta = \frac{\theta - \theta_1}{\theta_2 - \theta_1}, \tag{7}$$

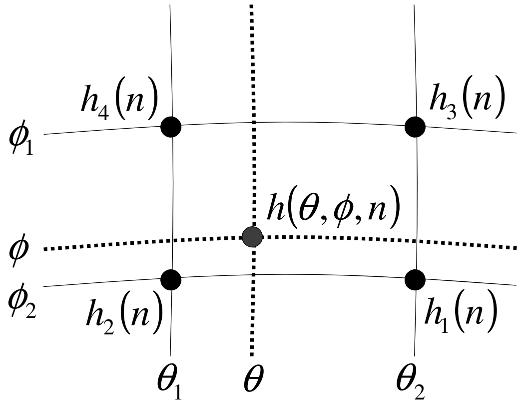

Figure 2: Four positions selected for the association and interpolation of zeros

for the horizontal direction, and

$$r_\phi = \frac{\phi - \phi_1}{\phi_2 - \phi_1}, \tag{8}$$

for the elevation direction.

Synthesizing interpolated HRTF

After interpolation of zeros Z_{pE} in the target position, we can obtain the HRTF at the position using Z_{pE} and common pole p_{C_i} in Eq. 1.

EVALUATION OF THE PROPOSED METHOD

A computer simulation was performed to compare the interpolation accuracy between the linear interpolation and the proposed method. A set of HRTFs of the dummy head (Samrai; Koken Co. Ltd.) was calculated using the boundary element method (BEM) simulation. The HRTFs were calculated with an interval of one degree in both horizontal and elevation angles. In all, 64,442 directions of HRTFs were obtained. The calculated frequencies were up to 20 kHz with a frequency step in the BEM simulation of 93.75 Hz.

Examples of the simulation are shown in Fig. 3. Three panels are shown in the figure. The target position of each is the same (horizontal angle 152 deg, elevation angle -43 deg). However, intervals among the four neighboring positions described above for interpolation differ, such as 5, 10, and 20 deg. Solid dashed, and dashed lines respectively represent the target HRTFs, those of the proposed method, and the linear interpolation in a frequency domain.

When the source positions of the four neighbor HRTFs are located close (5 deg) to each other, the results of both methods agree well with the target HRTF. However, the accuracy of both methods decreases as the interval among the neighboring sources increases. The reason is that the correlation among the neighboring HRTFs decreases when the interval increases and that the association becomes difficult in the proposed method.

We can see several spectral notches in the target HRTF. These notches can be well synthesized in our method because the zeros are adequately associated. In contrast, when the interval increases, the accuracy of linear interpolation decreases because frequency positions of notches among the neighbor HRTFs are not identical. In the linear interpolation, it is difficult to synthesize a notch whose frequency is between those of notches of the neighbor HRTFs. These notches are known as important cues in elevation localization [15]; reportedly, the depths of notches are also important[20]. In the linear interpolation, these notches might disappear with a large interval. Therefore, our proposed method is expected to improve the sound localization accuracy greatly in a VAD system.

CONCLUSION

We proposed an interpolation method of HRTF based on the common-pole and zero model in this paper. Using the proposed method, the interpolation accuracy can be maintained even when directions among HRTFs used in the interpolation are not close. On the other hand, the accuracy of the simple linear interpolation in the frequency domain decreased as the directional discrepancy among HRTFs increased, indicating that the number of directions of HRTFs in VAD system that must be prepared in advance can be reduced significantly using the proposed method.

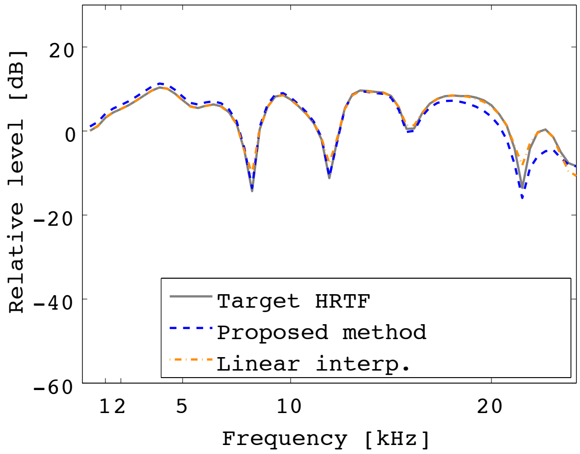

Fig. 3-(a) intervals = 5 deg

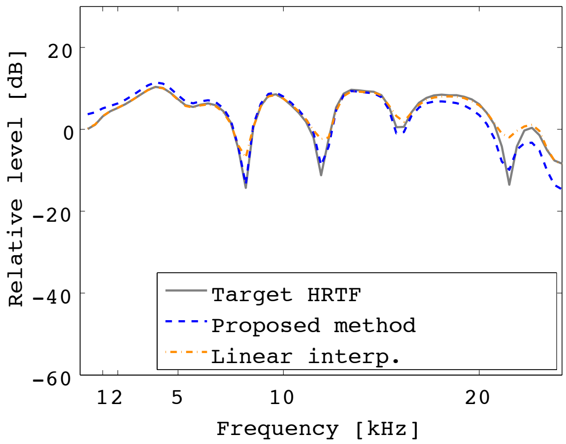

Fig. 3-(b) intervals = 10 deg

ACKNOWLEDGMENT

This study was partially supported by a Kakenhi Grant-in-Aid for Specially Promoted Research (to Y. Suzuki, 19001004) and a Kakenhi Grant-in-Aid for Scientific Research (C)(General) (to Y. Iwaya, 20500110).

REFERENCES

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

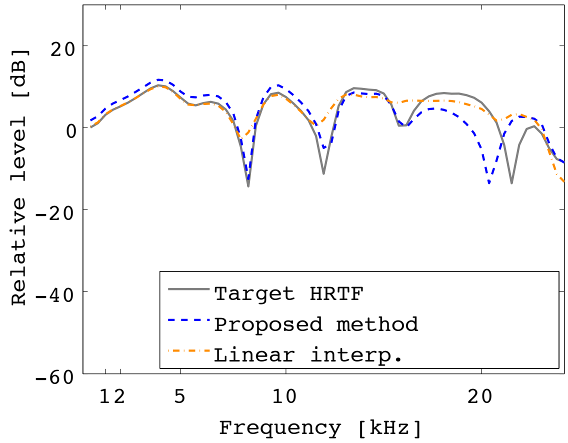

Fig. 3-(c) intervals = 20 deg

Figure 3: Comparison of interpolation methods. The proposed method and linear interpolation in frequency domain are compared. Solid, dashed, and dotted lines respectively indicate the target HRTF (horizontal angle 135 deg, elevation angle -43 deg) of the proposed method, and linear interpolation. Three panels show differences among the intervals of four neighbor HRTFs used in the interpolation, i.e. 5, 10, and 20 deg.

[2] F. Asano, Y. Suzuki and T. Sone, "Role of spectral cues in median plane localization," *J. Acoust. Soc. Am.*, **88**, pp.159-168, 1990.

[3] A. Honda, H. Shibata, S. Hidaka, Y. Iwaya, J. Gyoba, and Y. Suzuki, "Transfer effects on communication and collision Avoidance behavior from playing a three-dimensional auditory game based on a virtual auditory display," *Applied Acoustics*, **70**, 868-874, 2009.

[4] M. Otani and T. Hirahara, "Auditory artifacts due to switching head-related transfer functions of a dynamic virtual auditory display," *IEICE Trans. on Fundamentals*, **E91-A**(6), pp. 1320-1328, 2008.

[5] Y. Iwaya, K. Saito, and Y. Suzuki, "Individualization method of HRTFs using a tournament-style listening test of virtual sound images," *Proc. of WESPAC IX* (in 8 pages in Conference CD-ROM), 2006

[6] M. Otani and S. Ise, "Fast calculation system specialized for head-related transfer function based on boundary element method," *J. Acoust. Soc. Am.*, **119**(5), pp. 2589–2598, 2006.

[7] B.F.G. Katz, "Boundary element method calculation of individual head-related transfer function – I. Rigid model calculation," *J. Acoust. Soc. Am.*, **110**(5), pp. 2440–2448, 2001.

[8] Y. Iwaya and Y. Suzuki, "Numerical analysis of the effects of pinna shape and position on the characteristics of head-related transfer functions," *Proc. of Forum Acusticum 2008*, 2008.

[9] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Comparison of simulated and measured HRTFs: FDTD simulation using MRI head data," *Audio Engineering Society 123rd Convention* Paper 7240, New York, NY.

[10] Y. Park, S. Hwang and B. Kwon, "Overview on 3D Sound Research at VASL, NoViC, KAIST," *Proc. of International Workshop on the Principles and Applications of Spatial Hearing*, 4 pages pdf files, 2009.

[11] M. A. Blommer and G. H. Wakefield, “Pole-Zero Approximation for Head-Related Transfer Functions Using a Logarithmic Error Criterion," *IEEE Trans. on Speech and Audio Processing*, **5**, pp.278-287, 1997.

[12] F. P. Freeland, L. W. P. Biscainho and P. S. R. Diniz, “Efficient HRTF Interpolation in 3D Moving Sound," *Proc. AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio*, Espoo, Finland, June 2002.

[13] M. Matsumoto, M. Tohyama and H. Yanagawa, “A method of interpolating binaural impulse responses for moving sound images," *Acoust. Sci. & Tech.*, **24**, pp.284-292, 2003.

[14] L. Wang, F. Yin and Z. Chen, “Head-related transfer function interpolation through multivariate polynomial fitting of principal component weights," *Acoust. Soc. & Tech.*, **30**, pp.395-403, 2009.

[15] K. Iida, M. Itoh, A. Itagaki and M. Morimoto, “Median plane localization using a parametric model of the head-related transfer function based on spectral cues," *Applied Acoustics*, **68**, pp.835-850, 2007.

[16] L. Rabiner and B. H. Juang, “Fundamentals of Speech Recognition,” Prentice Hall, U. S. (New Jersey), 1993.

[17] Y. Haneda, S. Makino, Y. Kaneda and N. Kitawaki, “Common–Acoustical–Pole and Zero Modeling of Head-Related Transfer Functions," *IEEE Transactions on Speech and Audio Processing*, **7**, pp.188-196, 1999.

[18] K. Watanabe, S. Takane and Y. Suzuki, “A Novel Interpolation Method of HRTFs Based on the Common–Acoustical–Pole and Zero Model,” *Acta Acustica United with Acustica*, *91*, pp. 958–956, 2005.

[19] Y. Suzuki, H. Y. Kim, F. Asano and T. Sone, “An optimum computer generated pulse signal suitable for the measurement of very long impulse responses," *J. Acoust. Soc. Am.*, **97**, pp. 1119-1123, 1995.

[20] Y. Iwaya, M. Otani, T. Magariyachi and Y. Suzuki, ”Effect of low-frequency spectral notch on sound image localization,” *Proc. of the 10th Western Pacific Acoustics Conference* (WESPAC X), 6 pages in CDROM, 2009.