A local representation of the head-related transfer function

Shichao Hu,[a] Jorge Trevino, Cesar Salvador, and Shuichi Sakamoto
Research Institute of Electrical Communication and Graduate School of Information Sciences, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai, 980-0813, Japan
shichao@ais.riec.tohoku.ac.jp, jorge@ais.riec.tohoku.ac.jp, salvador@ais.riec.tohoku.ac.jp, saka@ais.riec.tohoku.ac.jp

Junfeng Li
Institute of Acoustics, Chinese Academy of Sciences, Beijing, 100190, China
junfeng.li.1979@gmail.com

Yôiti Suzuki
Research Institute of Electrical Communication and Graduate School of Information Sciences, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai, 980-0813, Japan
yoh@riec.tohoku.ac.jp

Abstract: Spatial descriptions of the head-related transfer function (HRTF) using spherical harmonics, which is commonly used for the purpose, consider all directions simultaneously. However, in perceptual studies, it is necessary to model HRTFs with different angular resolutions at different directions. To this end, an alternative spatial representation of the HRTF, based on local analysis functions, is introduced. The proposal is shown to have the potential to describe the local features of the HRTF. This is verified by comparing the reconstruction error achieved by the proposal to that of the spherical harmonic decomposition when reconstructing the HRTF inside a spherical cap.

© 2016 Acoustical Society of America

[CFG]
Date Received: March 15, 2016 **Date Accepted:** July 11, 2016

1. Introduction

The head-related transfer function (HRTF) characterizes sound transmission from any source position to the listener's ear.[1–3] HRTFs are, therefore, functions of the frequency and direction of the source. While these two parameters are independent, this letter focuses on the behavior of the HRTFs along different directions at a certain frequency.

A common method to analyze HRTFs covering all directions is spherical harmonic decomposition.[4–6] Rather than examining the data direction-by-direction, this approach characterizes the HRTFs using a set of expansion coefficients. The spherical harmonic functions, which are called global functions, take significant values from all spherical directions. Therefore, each coefficient corresponding to a particular spatial frequency includes information from all directions and for a given temporal frequency. This is suitable for global representation of the target data; however, it requires knowledge of the HRTFs for a sampling that covers all directions. In addition, perceptual studies suggest that the minimum audible angle that can be used to characterize human sound localization depends on the source's direction.[7] The directional resolution of HRTF required in binaural synthesis varies for all directions on the sphere around the head.[8] For these reasons, methods are needed for analyzing the HRTFs at different resolutions for different directions.

This letter considers a method to represent HRTFs using local analysis functions that take significant values over a local region; these functions are intended to replace the global functions used in conventional methods. Inspired by the wavelet transform, HRTFs are locally analyzed by a set of local functions, which were derived from a generating function by changing its central directions and spatial resolutions. In this manner, local HRTF features for different spatial frequencies can be modeled. This proposal offers the possibility of a local representation of HRTF by choosing the corresponding local analysis functions.

[a]Author to whom correspondence should be addressed.

2. Modeling of the HRTFs

A sound source direction, at a certain distance in spherical coordinates, is specified by its azimuthal angle $\theta \in [-180°, 180°]$ and elevation angle $\phi \in [-90°, 90°]$. The spatial part of an HRTF data set $H(\theta, \phi)$ is generally represented by using the following weighted sum:

$$H(\theta, \phi) = \sum_{i=1}^{\infty} c_i \cdot W_i(\theta, \phi), \quad (1)$$

where $W_i(\theta, \phi)$ denotes the ith analysis function for $i = 1, 2, 3,\ldots$; c_i is the corresponding coefficient in the decomposition. In practice, i must be truncated to a maximum integer. The resulting decomposition can be written in matrix form as follows:

$$\mathbf{H} = \mathbf{WC} + \epsilon, \quad (2)$$

where $\mathbf{H}$ is a vector containing the samples for the HRTF at a given frequency and all available directions. $\mathbf{W}$ is a matrix formed by sampling the analysis functions at these directions. Vector $\mathbf{C}$ contains the corresponding expansion coefficients, and ϵ is the truncation error. The expansion coefficients can be obtained using the least squares method, as follows:

$$\hat{\mathbf{C}} = \mathbf{W}^{+}\mathbf{H}, \quad (3)$$

where $\mathbf{W}^{+}$ is the generalized inverse of matrix $\mathbf{W}$.

2.1 Proposed local analysis functions on the sphere

Traditional methods to define the analysis functions are based on spatial principal component analysis[9] or spherical harmonic decomposition, in which matrix $\mathbf{W}$ contains a set of spatial principal components or spherical harmonic functions correspondingly. As an alternative, this study proposes a new set of local analysis functions that take significant values over a local region on the sphere. These local functions consist of an isotropic spatial window to select a local region, and a cosine on the sphere as an oscillating function. The proposed function is defined as follows:

$$W_0(\theta, \phi) = \cos\left[\alpha \cdot D_0(\theta, \phi)\right] \cdot e^{-D_0^2(\theta,\phi)/2\sigma^2}, \quad (4)$$

where $W_0(\theta, \phi)$ is one of the proposed local analysis functions, with a center at (θ_0, ϕ_0). $D_0(\theta, \phi)$ denotes the great circle distance between two points on the unit sphere, at angles (θ, ϕ) and (θ_0, ϕ_0). The parameters α and σ control the oscillation rate and the width of the Gaussian window, respectively. By introducing a scaling factor S, a scaled analysis function is defined as follows:

$$W_{0,S}(\theta, \phi) = \sqrt{S}\cos\left[\alpha \cdot S \cdot D_0(\theta, \phi)\right] \cdot e^{-S^2 \cdot D_0^2(\theta,\phi)/2\sigma^2}, \quad (5)$$

where a small value of S gives a coarse approximation capturing low spatial frequencies, while large scales show the details capturing the high spatial frequencies.

In this study, HRTF data are represented as a weighted sum of the local functions as defined in Eq. (5), for a set of scaling factors S_ℓ, with scale index $\ell = 1, 2,\ldots$, in such a manner that

$$H(\theta, \phi) = \sum_{\ell=1}^{\infty} \sum_{m \in \mathbb{D}(\ell)} c_{\ell,m} \cdot W_{m,\ell}(\theta, \phi). \quad (6)$$

Here, $c_{\ell,m}$ denote the expansion coefficient of scale index ℓ and center direction (θ_m, ϕ_m). Sets $\mathbb{D}(\ell)$ are samplings of all directions with an average angular separation defined according to the corresponding scaling factor. The expansion coefficients $c_{\ell,m}$ can be calculated using a least squares method, as described in Eq. (3).

3. Evaluation

3.1 Conditions of evaluation

The proposal is tested by applying it to a target HRTF data set calculated using the Boundary Element Method for the SAMRAI (Koken) dummy head.[10] The HRTF data for sound sources at 1.5 m were calculated at frequencies between 93.75 and 20 000 Hz, with intervals of 93.75 Hz and samples at every 2° in azimuth and elevation angles, for a total of 16 022 directions. This number of samples is high enough to recover HRTFs at all directions within the audible frequency range.[5]

The parameters in Eq. (5) are empirically set to $\alpha = 0.667$ and $\sigma = 1$. There are numerous methods to define the sets $\mathbb{D}(\ell)$;[11] this letter uses a sampling of all directions starting from the vertexes of an icosahedron that are regularly distributed on the sphere. Next scale samplings are derived by adding a new vertex at the center of every edge of the previous scale. In addition, a dyadic step is used to define the scaling factor, using the following equation:

$$S_\ell = 2^{\ell-1}. \tag{7}$$

For an objective evaluation, a conventional measure of approximation error in the frequency domain is defined using the Root-Mean-Squared (RMS) value.[12] Here a similar RMS value for measuring the approximation error in the spatial domain is defined as

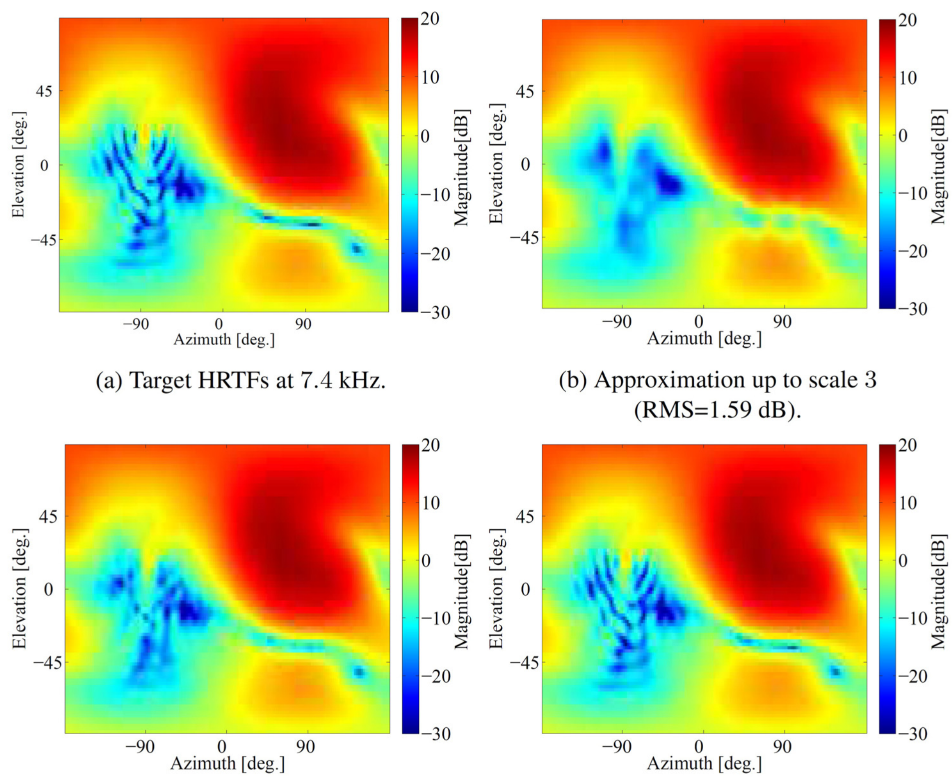
$$E_{\mathrm{RMS}} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left[20\log_{10}\frac{|H_{\mathrm{synth}}(\theta_j,\phi_j)|}{|H_{\mathrm{target}}(\theta_j,\phi_j)|}\right]^2}, \tag{8}$$

where $H_{\mathrm{synth}}(\theta_j,\phi_j)$ and $H_{\mathrm{target}}(\theta_j,\phi_j)$ are the reconstructed HRTF and target HRTF at direction (θ_j,ϕ_j), respectively; N is the total number of HRTF samples under study.

3.2 Local representation of HRTF

In this section, magnitudes of target HRTFs are approximated up to a certain scale using the method proposed in Sec. 2. In Fig. 1, panel (a) illustrates the spatial patterns of HRTF magnitude for all directions at 7.4 kHz; panels (b), (c), and (d), respectively, show the results of the proposed analysis up to scales of 3, 4, and 5. The E_{RMS} is 1.59, 1.07, and 0.43 dB, respectively. This indicates that high-scale approximations capture finer details and generate smaller errors. More details regarding this representation for all directions are provided in a previous work of this study.[13]

The expected advantage of the proposed function is its ability to accurately represent a locality, which may allow for the local modeling of HRTFs over a region. Here, a local representation of the HRTF is realized by selecting the local analysis



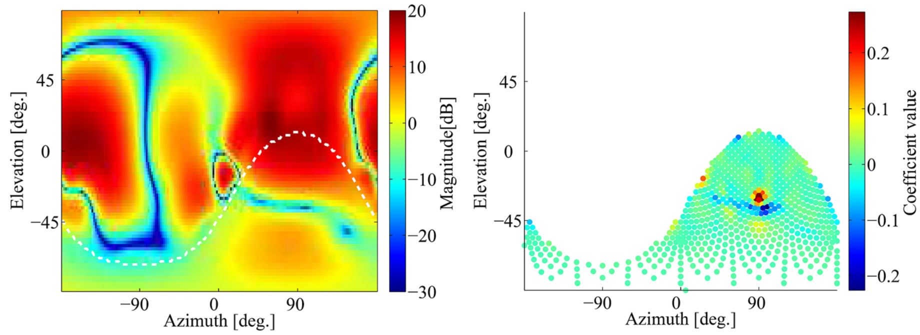
(a) Target HRTFs at 7.4 kHz.

(b) Approximation up to scale 3 (RMS=1.59 dB).

(c) Approximation up to scale 4 (RMS=1.07 dB).

(d) Approximation up to scale 5 (RMS=0.43 dB).

Fig. 1. (Color online) Target HRTF of the left ear (a) and its approximations [(b), (c), (d)] using the proposed method for different maximum scales with every 2° in azimuth (from −180° to 180°) and elevation (from −90° to 90°) angles.



(a) Local representation of the HRTF. (b) Values for the coefficients of scale 5.

Fig. 2. (Color online) Panel (a) shows the reconstructed left ear HRTF for a local region center at (90°, −48°) with a size of 2.888 steradians on the sphere (below the white dotted line); panel (b) shows the corresponding coefficients of scale 5.

functions corresponding to the region. Because functions of lower scales have an influence over a larger area on the sphere, the low-scale functions are selected across a wider area while only the high-scale functions close to the area under study are used. More specifically, all the analysis functions whose radius of influence intersects the area under study are selected for this local representation. The radius of influence of a local function is defined as the distance between its center position (where it gets its maximal value) and the position where its amplitude decreases to a low value (in this study, this was set to 10% of the peak value).

An example of this regional representation is illustrated in Fig. 2. Panel (a) of this figure shows an approximation up to scale 5 of a local region center at (90°, −48°) on the sphere. The radius of this local region on the sphere is chosen to be a spherical distance of 1.5 m, whose size is 2.888 steradians. Panel (b) shows the corresponding coefficients of scale 5. There is a correspondence between the narrow deep notch of the HRTF and the values of the scale 5 coefficients. This indicates that the expansion coefficients have the potential to describe the local features of HRTFs. The approximation error of this regional representation results in an E_{RMS} of 0.32 dB.

HRTFs of two local regions (both are spherical caps with a size of 1.345 steradians) for the ipsilateral [center at (90°, 0°)] and contralateral [center at (−90°, 0°)] sides are also represented with the proposed method. As shown in Table 1, when modeling up to a same scale, approximation at the contralateral side yields a bigger RMS error than that at the ipsilateral side. The results clearly show the different difficulties in modeling for these two sides.

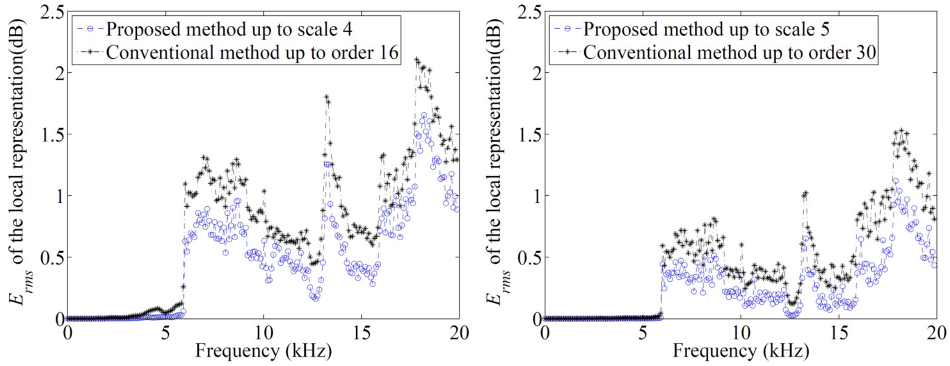
A comparison is made between the E_{RMS} generated by the proposed method based on the local functions, and that generated by the conventional method[4] based on spherical harmonics for the local representation of the HRTFs at 7.4 kHz over the region used in Fig. 2. Given a comparable number of analysis functions, modeling with the proposed local functions up to scale 4 (number of local functions = 275) and scale 5 (number of local functions = 937) yielded E_{RMS} values of 0.68 and 0.32 dB, respectively. These E_{RMS} values are smaller than the corresponding values yielded when using spherical harmonics up to order 16 (number of harmonics = 289) and order 30 (number of harmonics = 961) with E_{RMS} values of 1.14 and 0.60 dB, respectively.

3.3 Approximation error for multiple frequencies

The proposed method is now applied to HRTFs of multiple frequencies, linearly distributed between 93.75 and 20 000 Hz. To compare the local representability of the proposed method with the conventional method[4] based on spherical harmonics, these two methods are now applied to HRTFs at multiple frequencies. Although the method that uses spherical harmonics is suitable for global representation, in nature, only the

Table 1. Approximation errors in terms of RMS values of local representations for the ipsilateral and contralateral side.

Approximation	Ipsilateral side error (dB)	Contralateral side error (dB)
Up to scale 3	1.41	3.73
Up to scale 4	0.26	3.07
Up to scale 5	0.06	1.26



(a) Comparison with spherical harmonics up to order 16.

(b) Comparison with spherical harmonics up to order 30.

Fig. 3. (Color online) Comparisons for the local representation performance in terms of E_{RMS} between the proposed method and the conventional method based on spherical harmonics, using a comparable number of analysis functions. The local region for consideration is that shown in Fig. 2. Panel (a) shows the E_{RMS} with the proposed method up to scale 4 (number of local functions = 275) and that with the conventional method up to order 16 (number of spherical harmonics = 289), while panel (b) shows the E_{RMS} values with the proposed method up to scale 5 (number of local functions = 937) and the conventional method up to order 30 (number of spherical harmonics = 961).

local region center at (90°, −48°) is evaluated by the two methods. Moreover, for fairness in the comparison, the cases were configured so that the number of parameters was comparable. As a result, the proposed method up to scale 4 (number of local functions = 275) and scale 5 (number of local functions = 937) is compared with the conventional method up to, respectively, order 16 (number of spherical harmonic functions = 289) and order 30 (number of spherical harmonic functions = 961). Figure 3 shows the result. It can be seen from the figure that the proposed method using the local functions consistently yields smaller E_{RMS} values than those yielded by the spherical harmonics method when the number of analysis functions is comparable.

4. Discussion and conclusion

In this letter, a new method to represent the spatial patterns of HRTFs is proposed; this new method is based on a set of local analysis functions. These analysis functions are centered in different directions and scaled so as to capture the local variations at different resolutions. Numerical experiments confirm that high-scale approximations can capture finer details. The proposal can reconstruct HRTFs over a compact region by appropriately choosing the analysis functions. Therefore, it can achieve the direction-dependent spatial resolution needed for accurate sound localization. A local representation of HRTF at 7.4 kHz shows a correspondence between the values for high-scale coefficients and the local features of the HRTFs. This suggests that the expansion coefficients have the potential to describe the local features of HRTFs. The correspondence may be improved by defining analysis functions with lower redundancy; this will be a future step in this study. Finally, given a comparable number of analysis functions, the proposed method yields smaller RMS error values when representing an HRTF spatial pattern of a local region than those yielded by a conventional modeling technique based on spherical harmonics.

Acknowledgments

This study was partly supported by Grant-in-Aid of JSPS for Scientific Research (Grant No. 24240016), the Foresight Program for "Ultra-realistic acoustic interactive communication on next-generation Internet." The BEM solver used was developed by Dr. Makoto Otani.

References and links

[1]V. Mellert, K. Siebrasse, and S. Mehrgardt, "Determination of the transfer function of the external ear by an impulse response measurement," J. Acoust. Soc. Am. **56**, 1913–1915 (1974).

[2]E. Shaw, "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," J. Acoust. Soc. Am. **56**, 1848–1861 (1974).

[3]J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA, 1997).

[4]M. J. Evans, J. A. Angus, and A. I. Tew, "Analyzing head-related transfer function measurements using surface spherical harmonics," J. Acoust. Soc. Am. **104**, 2400–2411 (1998).

[5]W. Zhang, T. D. Abhayapala, R. A. Kennedy, and R. Duraiswami, "Insights into head-related transfer function: Spatial dimensionality and continuous representation," J. Acoust. Soc. Am. **127**, 2347–2357 (2010).
[6]G. D. Romigh, D. S. Brungart, R. M. Stern, and B. D. Simpson, "Efficient real spherical harmonic representation of head-related transfer functions," IEEE J. Sel. Topics Signal Process. **9**, 921–930 (2015).
[7]A. W. Mills, "On the minimum audible angle," J. Acoust. Soc. Am. **30**, 237–246 (1958).
[8]P. Minnaar, J. Plogsties, and F. Christensen, "Directional resolution of head-related transfer functions required in binaural synthesis," J. Audio Eng. Soc. **53**, 919–929 (2005).
[9]B.-S. Xie, "Recovery of individual head-related transfer functions from a small set of measurements," J. Acoust. Soc. Am. **132**, 282–294 (2012).
[10]M. Otani and S. Ise, "Fast calculation system specialized for head-related transfer function based on boundary element method," J. Acoust. Soc. Am. **119**, 2589–2598 (2006).
[11]E. B. Saff and A. B. Kuijlaars, "Distributing many points on a sphere," Math. Intell. **19**, 5–11 (1997).
[12]K.-S. Lee and S.-P. Lee, "A relevant distance criterion for interpolation of head-related transfer functions," IEEE Trans. Audio, Speech, Lang. Process. **19**, 1780–1790 (2011).
[13]J. Trevino, S. Hu, C. Salvador, S. Sakamoto, J. Li, and Y. Suzuki, "A compact representation of the head-related transfer function inspired by the wavelet transform on the sphere," in *2015 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)* (2015), pp. 372–375.