Modeling head-related transfer functions with spherical wavelets

Shichao Hu, Jorge Trevino, César Salvador, Shuichi Sakamoto, and Yôiti Suzuki

Research Institute of Electrical Communication and Graduate School of Information Sciences, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai, 980-8577, Japan

Abstract

The head-related transfer function (HRTF) describes the sound transmission characteristics from a sound source to a listener's ears. Recently, spherical harmonic decomposition has been extensively used for modeling the HRTF spatial patterns. Despite its advantage of approximating the coarse structure of HRTF spatial variations with modeling up to a low order, there are still some limitations since spherical harmonics take significant values in all directions. First, rapidly changing HRTF spatial variations in some local regions may require modeling up to a rather high order; this is not wise in terms of the modeling efficiency. Second, the expansion coefficients of the spherical harmonics describe the spatial frequency of the target dataset in all directions, and thus have difficulties in revealing the direction dependent HRTF characteristics. In this study, a method for locally modeling HRTF spatial patterns is proposed based on spherical wavelets, which take significant values only over a local region on the sphere. Results of numerical experiments show that our proposed method yields smaller approximation errors than the conventional method when representing HRTFs inside the local regions under evaluation. Furthermore, the expansion coefficients in the proposed method could well correspond to the HRTF local features on the sphere, which makes it a useful tool for the analysis and visualization of HRTF spatial patterns.

Keywords: Spherical acoustics, head-related transfer function, spherical wavelets, spherical harmonics, binaural audio, auditory display

1. Introduction

Spatial hearing ability is mainly attributed to the discrimination of acoustic characteristics and their interaural differences arising from the reflections and diffraction of the sound waves by the human body (i.e., pinnae, head, and torso) during the transmission from the source positions to the listener's eardrums. These transmission characteristics are described by the head-related transfer function (HRTF) [1, 2, 3]. With a knowledge of the HRTFs of a certain position, a corresponding virtual sound source in space can be synthesized [4, 5]. HRTFs have been utilized as an important way to generate virtual auditory stimuli that are used in spatial hearing experiments, as well as the development of next generation audio systems [6, 7].

HRTFs are generally obtained through measurement of the acoustic impulse responses [3, 8, 9], or are calculated based on three-dimensional model acquisition and numerical simulation methods [10, 11, 12] at discrete directions around the listener's head. There are several HRTF datasets that include HRTFs of a dummy head or real subjects [13, 14, 15, 16, 17, 18, 19]. The originally obtained HRTF dataset, in which the HRTF samples are represented point-by-point for each temporal or frequency bin and direction, is inconvenient for the analysis of HRTFs. An HRTF analysis method consists of representing the original HRTFs using mathematical models with a small number of coefficients. Generally, the HRTF is represented as a function with the indices of frequency and direction, which are independent variables. Therefore, HRTF modeling can normally be classified as the frequency domain approach [20, 21] and the spatial domain approach [22, 23], which model the HRTF along the frequency in each direction, and along the direction at each frequency bin, respectively. These modeling methods can effectively reduce the HRTF data size, as well as manipulate the HRTF data.

This study focuses on representing HRTFs in the spatial domain, which has many advantages. The directional representation allows for HRTF interpolation in space, which is beneficial for efficient representation of the HRTF. Recently, spatial descriptions of HRTFs have been studied based on spherical harmonics [22, 24, 25]. In this method, the HRTF is represented by a weighted sum of spherical harmonics up to a truncated order. Thus, an HRTF dataset can be represented by a set of expansion coefficients that accounts for the spatial spectrum at a certain temporal frequency. This popular modeling method provides a good approximation of the HRTF coarse structure, even using a small number of coefficients [26], and thus is suitable

for the efficient representation of the HRTF. In addition, the decomposition based on spherical harmonics, which are continuous functions on a sphere, allows for convenient interpolation of HRTFs [25, 27].

Despite the success of modeling HRTFs using spherical harmonics, there are still some limitations since spherical harmonics are global functions, that is, they take significant values in all directions on a sphere. Each expansion coefficient driven by the global function, the spherical harmonic, is an overall description of the HRTF spatial spectrum in all directions. However, the spatial peaks and notches of HRTFs that reveal the acoustic filtering effects are direction dependent, that is, they are distributed in different regions on the sphere. Some suddenly changing variations of HRTF spatial patterns in local regions require modeling with spherical harmonics up to a high order, which is not efficient for the purpose of data modeling and compression. In addition, previous perceptual studies suggest that the minimum audible angle that can be used to characterize human sound localization depends on the source direction [28]. Furthermore, the directional resolution of HRTFs required in binaural synthesis varies for all directions on the sphere around the head [29]. For these reasons, methods are needed for analyzing the HRTFs at different resolutions for different directions.

To analyze the direction–dependent features of HRTFs, we previously proposed a method that uses local functions on the sphere [30, 31] for representing HRTF spatial patterns. The results showed that the proposed method yields a smaller approximation error than the spherical harmonics method for a local region when providing a comparable number of analysis functions. In addition, there is a correspondence between the local features and some expansion coefficients of the local analysis functions, which may provide another way for the visualization and analysis of HRTF spatial patterns. However there is some redundancy among the analysis functions due to the lack of orthogonality on the sphere. If the orthogonality is further improved, the target HRTF can be more sparsely represented, and the level of detail in a local region can be better controlled with the expansion coefficients. The correspondence between the HRTF local features and the expansion coefficients may be improved as well.

Recently, P. Bates et al. proposed to use Slepian functions which are orthogonal inside a certain local region for the purpose of HRTF representation; this method is reasonable if HRTFs only inside that region are of interest [32]. However the generation of the Slepian functions depends on the predefined local region and do not consider other regions. This makes it difficult to lo-

cally control the spatial resolution of the HRTFs over the entire sphere, and wavelet analysis on the sphere may be a solution to this limitation. In the classic wavelet theory, a series of wavelets for different scales and positions is constructed by dilation and translation of a mother wavelet [33]. However, the extension of this construction method to the sphere is not straightforward. Although many previous attempts have been made to construct continuous wavelets or localized functions on a sphere [34, 35], modeling discrete data efficiently on a sphere still remains difficult, and it is also difficult for the purpose of HRTF modeling. To construct discrete wavelets for representing functions on a sphere, Schröder and Sweldens proposed the lifting scheme and validated its effectiveness [36]. Although spherical wavelets based on the lifting scheme have been widely applied to computer graphics, to the best of the authors' knowledge, these spherical wavelets have not yet been applied to HRTF modeling. The present study proposes to represent HRTFs using biothorgonal spherical wavelets based on the lifting scheme. Unlike spherical harmonics that consider all directions simultaneously, spherical wavelets take significant values locally and have different scales on the sphere. The proposed method is expected to provide a good description of HRTF local features with the expansion coefficients, which reveals the direction-dependent filtering effects at different spatial frequencies. This paper is organized as follows. Section 2 introduces the spherical wavelets based on lifting scheme. Our proposed method is formulated in Section 3. In Section 4, the proposed method is evaluated by numerical experiments. Here, the approximation errors for all directions and a local region of HRTF are compared between the spherical harmonic-based modeling and the proposed method. Our conclusions are presented in Section 5.

2. Introduction of spherical wavelet analysis based on the lifting scheme

This section gives a brief introduction into multiresolution analysis using spherical wavelets based on lifting schemes [36], which splits a signal spectrum into two parts: the approximation part and the detail part, as if the signal is processed by a low-pass and a high-pass filter, respectively. This multiresolution analysis based on lifting scheme is very similar to that of classic wavelet theory. As shown in Fig. 1, a high-pass filter (High) and a low-pass filter (Low) calculate the wavelet coefficients W_ℓ and scaling coefficients S_ℓ at a certain scale level $\ell = 1, ..., L-1$, which corresponds to the

detail and approximation parts, respectively.

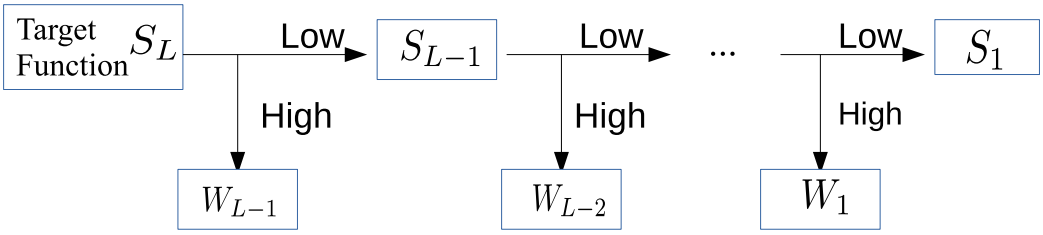

Figure 1: Multiresolution analysis using spherical wavelets based on lifting schemes.

Before introducing the lifting scheme, let us first design some variables. Considering the case of an icosahedron and its subdivisions, a vertex set at level ℓ is defined as $V_\ell = \{\vec{v}_{\ell,i}\}_{i=1,...,I_\ell}$, where $\vec{v}_{\ell,i}$ denotes a vertex at level ℓ, and I_ℓ is the total number of vertices at this scale. The next level vertex set $V_{\ell+1}$ can be obtained by adding midpoints at every edge of the previous level and then projecting them on the sphere; this midpoint set is denoted as $M_\ell = \{\vec{m}_{\ell,j}\}_{j=I_\ell+1,...,I_{\ell+1}}$, where $\vec{m}_{\ell,j}$ denotes the midpoint at level ℓ.

A vertex set can be divided into a lower-scale vertex set and the corresponding midpoint set, namely $V_{\ell+1} = V_\ell \cup M_\ell$. For example, as shown in Fig. 2, an icosahedron sampling set can be defined as a root level set V_1, which has a total of 12 points. By adding a midpoint at every edge, the vertex set of the following levels ($\ell = 2, 3, 4, ...$) can be generated.

Spherical wavelet transforms are calculated by using a local naming scheme, where each $\vec{m}_{\ell,j} \in M_\ell$ is only filtered over a small neighborhood. For example, in Fig. 3, the neighborhood $N(\vec{m})$ around a midpoint $\vec{m} \in M_\ell$ consists of 8 surrounding points ($N(\vec{m}) = \vec{v}^{(k)}, k = 1, ..., 8$)[36]. To perform the forward transform of a target function, the vertex level goes from the leaf level to the root level as ℓ decreases. The inverse transform can be performed by the above two steps, but backwards. The general form of this process can be described as follows.

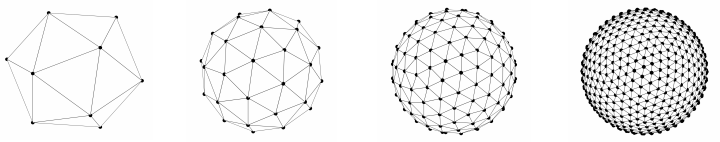

Figure 2: Sampling of an icosahedron V_1 and its following subdivisions V_2, V_3, and V_4 from left to right, with the vertex number of 12, 42, 162, and 642, respectively.

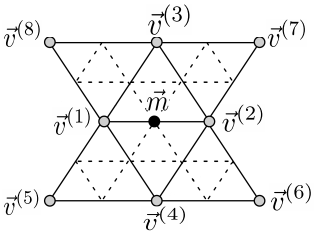

Figure 3: Local neighborhoods ($N(\vec{m}) = \vec{v}^{(k)}, k = 1, ..., 8\}$) of a point $\vec{m} \in M_\ell \subset V_{\ell+1}$ using the local naming scheme. The dashed-lines are the edges of next subdivision.

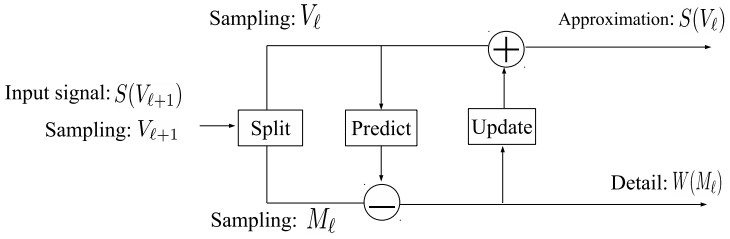

Figure 4: Lifting scheme forward wavelet transform.

- Forward step 1:
  $\forall \vec{v}_{\ell,i} \epsilon V_\ell : \quad S(\vec{v}_{\ell,i}) \leftarrow S(\vec{v}_{\ell+1,i})$
  $\forall \vec{m}_{\ell,j} \epsilon M_\ell : \quad W(\vec{m}_{\ell,j}) = S(\vec{v}_{\ell+1,j}) - \sum_{\vec{v}_{\ell,i} \epsilon N(\vec{m}_{\ell,j})} z^w_{\ell,i,j} S(\vec{v}_{\ell,i})$

  Here, $S(\vec{v}_{\ell,i})$ is the scaling coefficient corresponding to a point $\vec{v}_{\ell,i}$ on V_ℓ, and $W(\vec{m}_{\ell,j})$ is the wavelet coefficient corresponding to the point $\vec{m}_{\ell,j}$ on M_ℓ; $\vec{v}_{\ell,i}$ and $\vec{m}_{\ell,j}$ are included in $\vec{V}_{\ell+1}$. The forward transform starts from the target function H which is set to be the scaling coefficients at the finest level L, namely $H(\vec{v}_{L,i}) = S(\vec{v}_{L,i})$; $z^w_{\ell,i,j}$ are the filtering weights for the neighborhood set $N(\vec{m}_{\ell,j})$ around a midpoint $\vec{m}_{\ell,j} \epsilon M_\ell$. In this study, spherical wavelets with a butterfly subdivision scheme [36] are used; thus, $z^w_{\ell,1,j} = z^w_{\ell,2,j} = 1/2$, $z^w_{\ell,3,j} = z^w_{\ell,4,j} = 1/8$, and $z^w_{\ell,5,j} = z^w_{\ell,6,j} = z^w_{\ell,7,j} = z^w_{\ell,8,j} = -1/16$.

- Forward step 2:
  After calculating $W(\vec{m}_{\ell,j})$ in step 1, the scaling coefficients $S(\vec{v}_{\ell,i})$ are updated as:
  $\forall \vec{m}_{\ell,j} \epsilon M_\ell : \quad S(\vec{v}_{\ell,v_1}) = S(\vec{v}_{\ell,v_1}) + z^s_{\ell,1,j} W(\vec{m}_{\ell,j}), \quad S(\vec{v}_{\ell,v_2}) = S(\vec{v}_{\ell,v_2}) + z^s_{\ell,2,j} W(\vec{m}_{\ell,j})$.
  Here, $\vec{v}_{\ell,v_1}$ and $\vec{v}_{\ell,v_2}$ are the two endpoints on the parent edge of $\vec{m}_{\ell,j}$;

$z^s_{\ell,i,j}$ are weights chosen so that the resulting wavelet has a vanishing integral: $z^s_{\ell,i,j} = I_{\ell+1,j}/2I_{\ell,i}$, where $I_{\ell,i}$ denotes the integral of the scaling function corresponding to $\vec{v}_{\ell,i}$.

Fig. 4 shows the float chart of this forward transform, where the sampling points $V_{\ell+1}$ corresponding to the scaling coefficients $S(V_{\ell+1})$ are split into two parts, the lower-scale vertex set V_ℓ and the midpoint set M_ℓ. The coefficients of V_ℓ are used to "Predict" the coefficients of midpoint set M_ℓ and the prediction difference is used as the wavelet coefficients $W(M_\ell)$, which corresponds to the "Forward step 1". In "Forward step 2", the calculated wavelets coefficients $W(M_\ell)$ update the coefficients of V_ℓ and result in $S(V_\ell)$ which represent the approximation of $S(V_{\ell+1})$. The forward step 1 and 2 actually correspond to the high-pass filter and low-pass filter in Fig. 1 which calculate the wavelet coefficients and scaling coefficients, respectively.

The inverse transform can be performed by the above two steps, but backwards.

- Backward Step 1:
  Calculate $S(\vec{v}_{\ell,i})$:
  $\forall \vec{m}_{\ell,j} \epsilon M_\ell : \quad S(\vec{v}_{\ell,v_1}) = S(\vec{v}_{\ell,v_1}) - z^s_{\ell,1,j} W(\vec{m}_{\ell,j}), \;\; S(\vec{v}_{\ell,v_2}) = S(\vec{v}_{\ell,v_2}) - z^s_{\ell,2,j} W(\vec{m}_{\ell,j})$.

- Backward Step 2: $\forall \vec{v}_{\ell,i} \epsilon V_\ell : \quad S(\vec{v}_{\ell+1,i}) \leftarrow S(\vec{v}_{\ell,i})$
  $\forall \vec{m}_{\ell,j} \epsilon M_\ell : \quad S(\vec{v}_{\ell+1,j}) = W(\vec{m}_{\ell,j}) + \sum_{\vec{v}_{\ell,i} \epsilon N(\vec{m}_{\ell,j})} z^w_{\ell,i,j} S(\vec{v}_{\ell,i})$

In spherical wavelet analysis, the forward transform calculates the scaling coefficients $\boldsymbol{S}$ and wavelet coefficients $\boldsymbol{W}$ of a target function $\boldsymbol{H}$, while the backward step synthesizes the target function from the obtained coefficients. The forward and backward transform (both to the end level) are denoted as $\mathbb{W}$ and $\mathbb{W}^{-1}$, respectively. The coefficients $\boldsymbol{S}$ and $\boldsymbol{W}$ compose the set of analysis coefficients $\boldsymbol{C} = \{\boldsymbol{S}, \boldsymbol{W}\}$ in the spherical wavelet decomposition, which are calculated as $\boldsymbol{C} = \mathbb{W}\{\boldsymbol{H}\}$. The target function can be synthesized with the obtained coefficients, namely $\boldsymbol{H} = \mathbb{W}^{-1}\{\boldsymbol{C}\}$.

The lifting scheme does not construct wavelets directly. However, by implementing the inverse transform with the coefficient set as 1 for a single position and 0 for the rest, the corresponding wavelet or scaling function can be obtained. In such a manner, the analysis function of level ℓ and

position $\vec{m}_{\ell,p}$ or $\vec{v}_{\ell,p}$ in the spherical wavelet analysis can be obtained with the following equation:

$$Y_{\ell,p} = \mathbb{W}^{-1}_{\ell,p}\{\boldsymbol{C}\}. \tag{1}$$

where, $\mathbb{W}^{-1}_{\ell,p}$ implements the backward transform $\mathbb{W}^{-1}$ from the root to the leaf level, where the expansion coefficient $c_{\ell,p} \in \boldsymbol{C}$ is set as 1, and the rest of the coefficients are set as 0. The scaling function of scale level $\ell = 1$

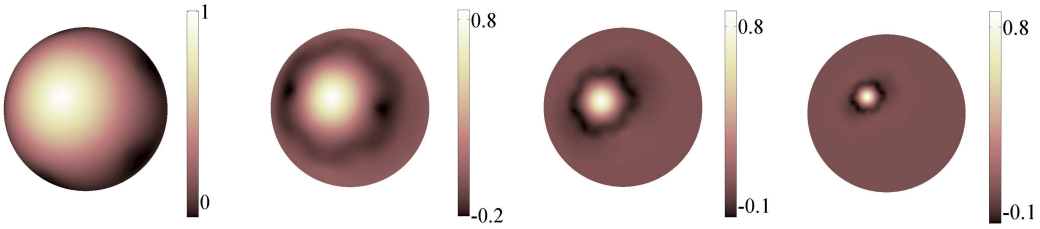

Figure 5: Analysis functions in Eq. 1 for the spherical wavelet transform (the leaf level is set to scale level 6) based on the lifting scheme with the butterfly subdivision: scaling function of scale level 1 (left-most), and wavelets of scale level 1, 2, and 3 from second to the right. The finest resolution corresponds to level 6.

and spherical wavelets from scale $\ell = 1$ to 3 are shown in Fig. 5. The center positions for locating the scaling function and wavelets on the sphere correspond to the mesh vertex in Fig. 2. The low-scale wavelets have greater support, while the high-scales have more compact support, which accounts for the low spatial frequency and high spatial frequency, respectively.

3. Application of spherical wavelets to HRTF representation

This paper focuses on the modeling of the magnitude part of the HRTF, which is a function on a sphere that depends on the direction of the sound source. The sound source direction, at a certain distance in spherical coordinates, is specified by its azimuthal angle $\theta \in [-180°, 180°]$ and elevation angle $\phi \in [-90°, 90°]$, in which $(0°, 0°)$ and $(90°, 0°)$ represent the front and left direction, respectively. The decomposition of the target HRTF magnitudes $H(\theta, \phi)$ can be described using the following matrix form:

$$H(\theta,\phi) = \sum_{\vec{v}_{1,i} \in V_1} S(\vec{v}_{1,i})\varphi_{1,i}(\theta,\phi) + \sum_{\ell=1}^{L-1} \sum_{\vec{m}_{\ell,j} \in M_\ell} W(\vec{m}_{\ell,j}) \cdot \psi_{\ell,j}(\theta,\phi). \tag{2}$$

Here, $\varphi_{1,i}(\theta,\phi)$ and $\psi_{\ell,j}(\theta,\phi)$ are the scaling function and spherical wavelet corresponding to $\vec{v}_{1,i}$ and $\vec{m}_{\ell,j}$, respectively; $S(\vec{v}_{1,i})$ and $W(\vec{m}_{\ell,j})$ are the expansion coefficients in the decomposition. It should be noted that, in order to implement this method, the sampling positions (θ,ϕ) of the target HRTF dataset should match a certain kind of sampling scheme such as the icosahedron subdivision in the spherical wavelet analysis. This can be handled by properly interpolating the HRTF at the needed positions or generating the HRTF at the corresponding positions with a calculation method such as the boundary element method (BEM) [11]. An example of a scaling function $\varphi_{1,i}(\theta,\phi)$ of scale level $\ell = 1$ and spherical wavelet $\psi_{\ell,j}(\theta,\phi)$ of scale level $\ell = 1,2,3$, and the corresponding meshes for locating these analysis functions are shown in Fig. 5 and Fig. 2, respectively. In practice, ℓ is truncated to a level to approximate the target function. The expansion coefficients can be calculated using the lifting scheme introduced in Section 2.

4. Evaluation of the proposed method

In this section, numerical experiments are conducted to validate the effectiveness of the method described in Section 3.

4.1. Objective measurement

The goal of this study is to approximate the original HRTF magnitudes using a set of analysis functions. To measure the approximation error in the spatial domain, the mean normalized error is used, which is defined as

$$E_{mnl} = \frac{1}{N}\sum_{m=1}^{N}\frac{|H_{synth}(\theta_m,\phi_m) - H_{target}(\theta_m,\phi_m)|}{|H_{target}(\theta_m,\phi_m)|}, \tag{3}$$

where $H_{synth}(\theta_m,\phi_m)$ and $H_{target}(\theta_m,\phi_m)$ are the reconstructed HRTF magnitude and target HRTF magnitude at direction (θ_m,ϕ_m), respectively; and N is the total number of HRTF samples under study.

4.2. HRTF database

HRTF datasets are generally obtained through measurements at predefined grid positions or are calculated through simulation methods at arbitrary positions. For the introduced spherical wavelet transform, the sampling distribution of the target dataset matches the vertex of the icosahedron subdivisions. Besides the use of icosahedral grids, the proposed method can also be

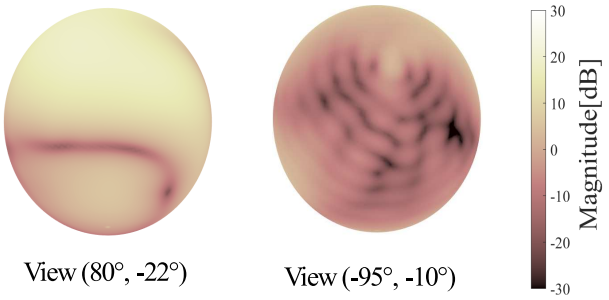

Figure 6: Original HRTFs (left ear) at 7.4 kHz. The panels from left to right are the HRTF with views of (80°, −22°) and (−95°, −10°), respectively.

applied in other hierarchical sampling schemes that are constructed by subdividing the edges of Platonic solids where suitable weighting parameters for the lifting scheme are necessary. The target dataset used for the simulation in the present study is calculated using the BEM for the SAMRAI (Koken) dummy head [11]. 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 directions according to the vertex of the icosahedron subdivision V_6. The number of the HRTF samplings along the direction is 10242 at each frequency bin. This number of samplings is high enough to recover HRTFs at all directions within the audible frequency range [24], and allows for visualizing the fine details of HRTF spatial patterns. The target HRTF at 7.4 kHz for the left ear is shown in Fig. 6.

4.3. HRTF representation using the proposed method

By using the spherical wavelets, the target HRTF magnitudes whose sampling points match the icosahedron subdivision of scale level $\ell = 6$ (10242 points in total) are decomposed to a root level $\ell = 1$. The spatial variations of the HRTF magnitudes can be smoothed by truncating ℓ to a certain level in the reconstruction. Fig. 7 shows that the original HRTF with a total sampling number of 10242 is approximated up to scale 1 (number of coefficients = 42), 2 (number of coefficients = 162), 3 (number of coefficients = 642), and 4 (number of coefficients = 2562), which yield $E_{nml} = 0.260$, 0.105, 0.048 and 0.009 respectively. The expansion coefficients from scale level 1 to 5 are shown in Fig. 8. The target HRTF magnitude is equal to the scaling function at scale level 6. This result suggests that the higher scale spherical wavelets catch finer details, while the low scale spherical wavelets approximate of the

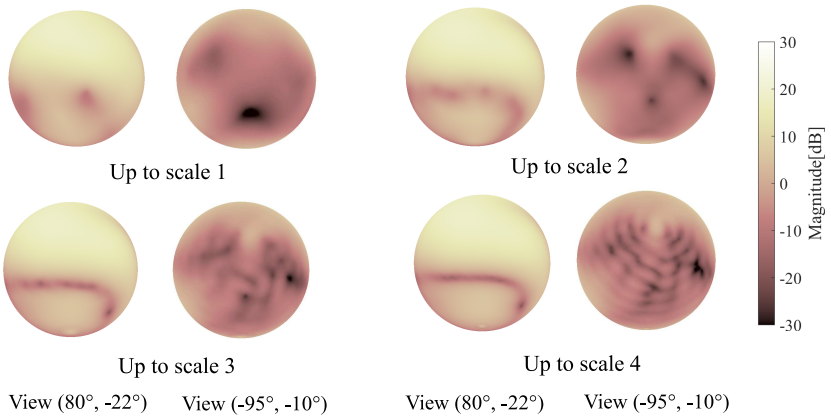

Figure 7: Approximated HRTF (left ear) in Eq. 2 with spherical wavelets up to different scales.

coarse structure of the HRTF. Thus, the original target HRTF with a high sampling amount can be represented by the expansion coefficients. A perfect reconstruction of the target HRTF can be realized by implementing the backward transform using all 5 scale's expansion coefficients.

One of the motivations of this study is to visualize and describe the HRTF local features using the expansion coefficients. In the previous work of the authors' group, the coefficients of a set of local analysis function could describe the local features to some degree [31].

By observing the expansion coefficients in Fig. 8, some interesting guesses may be made. The lowest scale (scale 1) coefficients may describe the spatial filtering effects excluding the pinna effects; the significant differences between the coefficient values of the ipsilateral and contralateral sides clearly reveal the head shadow effect. For the expansion coefficients of scale 5 , most of their values are nearly 0. It suggests that the spatial spectrum at this scale plays little role in the composition of the target HRTF. The coefficient values of scale 4 are also close to 0 except at some small local regions that corresponds to distributions of the sharp spatial variations of the original HRTFs. For the coefficients of scales higher than 1, most of their values are small and the significant values are only taken at some local regions. Since the pinna effects are more likely involved in the fast-changing spatial details [37], these higher scale coefficients seem to have a closer relevance to the pinna effects, which highly depend on the direction. The distribution of the expansion coefficients at different scales not only classify the acoustic

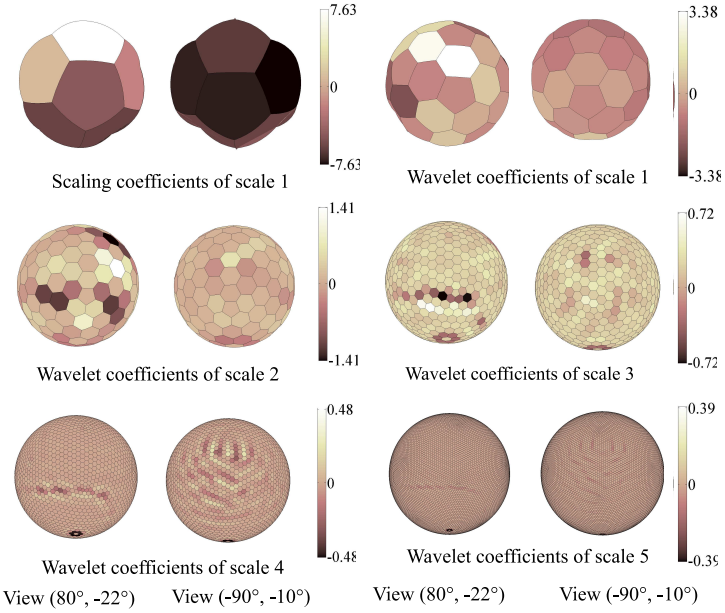

Figure 8: Expansion coefficient values of spherical wavelets at each scale in the decomposition of the target HRTF at 7.4kHz (Fig. 6) from the scale level $\ell = 1$ to the scale level $\ell = 5$.

effects of different anatomical parts, but also exhibits in which direction those filtering effects take place. This is a significant advantage over spherical harmonic decomposition[1]. Therefore, the proposed method provides a tool for visualization and insights of the spatial filtering effects inside the HRTF data. Furthermore, the distribution of the expansion coefficients should also have the individual information across different subject, this may also open a means for HRTF individualization.

4.4. Efficient representation of the HRTF using the proposal

Fig. 8 shows that the expansion coefficients are close to 0 in most directions at the higher levels. This indicates that the rapidly changing spatial

[1]The required number of samples S for modeling with spherical harmonics up to order N with a well-conditioned sampling scheme is denoted as $S = (N+1)^2$. For modeling with spherical wavelets up to the level L, the required number of samples is denoted as $S = 10 \times 4^L + 2$. Therefore, the relationship between modeling with spherical harmonics up to N (spatial frequency) and modeling with spherical wavelets up to L can be given as $(N+1)^2 = 10 \times 4^L + 2$.

variations mainly occur in some small and local regions. An assumption can be made that those low–value coefficients play little role in the reconstruction and could be removed to more efficiently represent the HRTF. To perform the compression, the analysis functions of each scale are first normalized to the same energy. The expansion coefficients with the absolute values above a certain threshold after the normalization are preserved to reconstruct the HRTF. In this manner, the HRTF can be represented by a small number of coefficients.

The following part of this section evaluates this efficient representation of the HRTF in all directions and compares it with the spherical harmonic method when using same number of analysis functions. The comparison is further extended to evaluations of HRTFs in some local regions on the sphere.

4.4.1. Comparison with spherical harmonics in all directions

The spatial representation of the HRTF in all directions is conducted with the same target HRTF as above. To compare the performance with spherical harmonics, the approximation errors are evaluated using the same number of analysis functions between the two methods. Namely, the approximation errors using spherical wavelets (SWs) with the number of coefficients of 121, 441, 676, 961 for reconstruction are compared to the errors when using spherical harmonics (SHs) up to order 10, 20, 25, and 30, respectively, which represent the HRTF with different spatial accuracy. Please note that in this study, we use real spherical harmonic representations with icosahedron samplings, which are well-conditioned for the spherical harmonic decomposition, because the orthonormality error decreases [38]. Fig. 9 plots the comparison results. It shows that our proposed method is slightly better or at least comparable with the conventional method based on real-valued spherical harmonics in terms of the approximation error when representing the HRTF magnitudes in all directions with a same number of analysis functions in the above conditions.

4.4.2. Comparisons with spherical harmonics at local regions

An expected advantage of the proposed method over the spherical harmonic method is to better represent the HRTF local features. Next, comparisons with the spherical harmonic method are conducted with the HRTF represented at some local regions. To perform this local representation, only the significant–value spherical wavelets corresponding to the target local region are selected for the reconstruction; this selection is done in two steps.

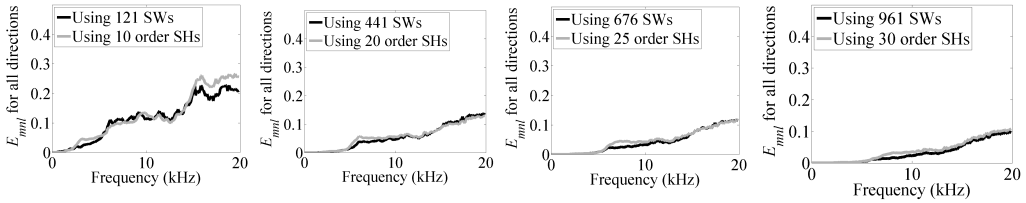

Figure 9: Comparisons of E_{mnl} for the target HRTF in all directions betweeen the approximation using spherical wavelets (SWs) while keeping 121, 441, 676, and 961 highest coefficients for reconstruction and spherical harmonics (SHs) up to order 10, 20, 25, and 30, respectively.

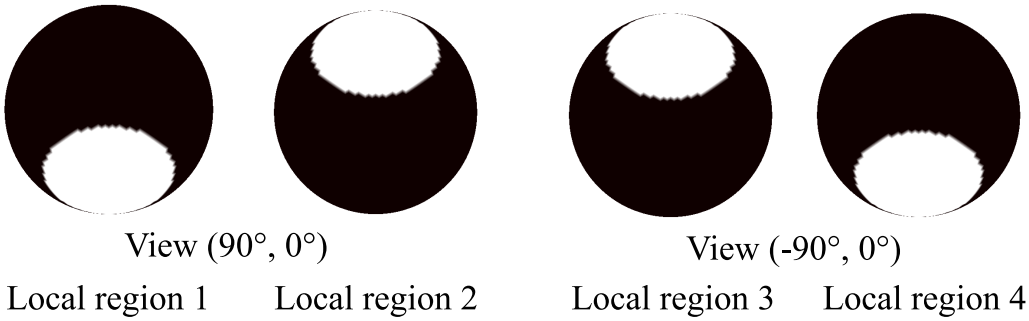

Figure 10: Four local regions (in white) selected for the evaluation of the local representation.

The first selecting step depends on the local area under evaluation. Because the low–scale spherical wavelets contribute to the coarse structure and have an influence over a larger area on the sphere, all analysis functions of scale 1 are preserved in the reconstruction. For the higher–scale spherical wavelets, only those close to the area under evaluation are used. More specifically, the spherical wavelets whose radius of influence intersects the area under evaluation are selected for this local representation. The radius of influence of a spherical wavelet is defined as the distance between its center position (where it has its maximum value) and the position where its amplitude decreases to the minimum value. Second, among the selected wavelets in the first step, only the spherical wavelets with significant expansion coefficient values are preserved.

The target HRTFs at the audible frequencies are reconstructed inside four spherical caps that are centered at (90°,−48°), (90°,48°), (−90°,48°), and (−90°,−48°), respectively, all with a size of 1.40 steradians, and denoted

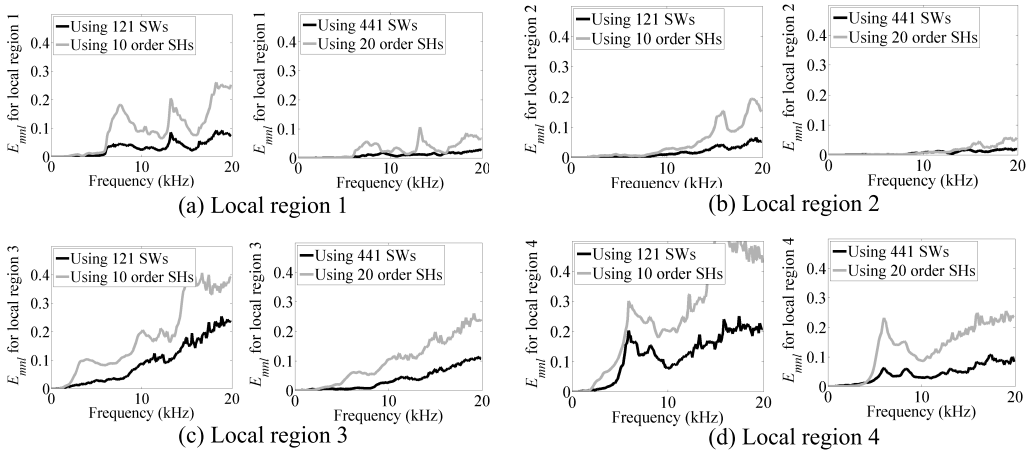

Figure 11: Comparisons of E_{mnl} for representing target HRTFs inside four local regions between the approximation using the same number of spherical wavelets (SWs) and spherical harmonics (SHs).

as local region 1, 2, 3, and 4, respectively. These four regions together cover most directions in space, as shown in Fig. 10. For these four local regions, the E_{mnl} values are compared between the approximation using spherical wavelets with 121 and 441 highest coefficients for reconstruction, and real-valued spherical harmonics up order 10 (number of harmonics=121) and order 20 (number of harmonics=441), respectively. The results shown in Fig. 11 suggest that when using these same number of analysis functions, our proposed method yields smaller approximation errors than for the spherical harmonic method. Therefore, the efficiency of modeling HRTF local features using spherical wavelets is validated.

5. Conclusions

In this study, we represented the HRTF spatial variations using the lifting scheme based spherical wavelets, which are widely used in the field of computer graphics. Numerical experiments showed that when using the same number (121 and 441) of analysis functions, approximation of the HRTF at the evaluated local regions based on spherical wavelets yields smaller errors than for the spherical harmonic method. In addition, the expansion coefficients of the spherical wavelets could well correspond to the direction dependent HRTF local features. This provides a tool for visualizing and analyzing the acoustic filtering effects inside the target HRTF data. Future work

will consider evaluations of several individual HRTFs to study the individual information that benefits from the proposed method.

6. Acknowledgments

This study was partly supported by JSPS KAKENHI Grant Numbers JP24240016, JP16H01736, the A3 Foresight Program for "Ultra-realistic acoustic interactive communication on next-generation Internet." The BEM solver used was developed by Dr. Makoto Otani.

References

[1] V. Mellert, K. Siebrasse, S. Mehrgardt, Determination of the transfer function of the external ear by an impulse response measurement, The Journal of the Acoustical Society of America 56 (6) (1974) 1913–1915.

[2] E. Shaw, Transformation of sound pressure level from the free field to the eardrum in the horizontal plane, The Journal of the Acoustical Society of America 56 (6) (1974) 1848–1861.

[3] J. Blauert, Spatial hearing: the psychophysics of human sound localization, MIT press, 1997.

[4] M. Morimoto, Y. Ando, On the simulation of sound localization., Journal of the Acoustical Society of Japan (E) 1 (3) (1980) 167–174.

[5] F. L. Wightman, D. J. Kistler, Headphone simulation of free-field listening. i: Stimulus synthesis, The Journal of the Acoustical Society of America 85 (2) (1989) 858–867.

[6] Y. Suzuki, D. Brungart, Y. Iwaya, K. Iida, D. Cabrera, H. Kato, Principles and applications of spatial hearing, World Scientific, 2011.

[7] C. D. Salvador, S. Sakamoto, J. Treviño, Y. Suzuki, Design theory for binaural synthesis: Combining microphone array recordings and head-related transfer function datasets, Acoustical Science and Technology 38 (2) (2017) 51–62.

[8] Y. Suzuki, F. Asano, H.-Y. Kim, T. Sone, An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses, The Journal of the Acoustical Society of America 97 (2) (1995) 1119–1123.

[9] P. Majdak, P. Balazs, B. Laback, Multiple exponential sweep method for fast measurement of head-related transfer functions, Journal of the Audio Engineering Society 55 (7/8) (2007) 623–637.

[10] B. F. Katz, Boundary element method calculation of individual head-related transfer function. i. rigid model calculation, The Journal of the Acoustical Society of America 110 (5) (2001) 2440–2448.

[11] M. Otani, S. Ise, Fast calculation system specialized for head-related transfer function based on boundary element method, The Journal of the Acoustical Society of America 119 (5) (2006) 2589–2598.

[12] W. Kreuzer, P. Majdak, Z. Chen, Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range, The Journal of the Acoustical Society of America 126 (3) (2009) 1280–1290.

[13] W. G. Gardner, K. D. Martin, Hrtf measurements of a kemar, The Journal of the Acoustical Society of America 97 (6) (1995) 3907–3908.

[14] V. R. Algazi, R. O. Duda, D. M. Thompson, C. Avendano, The cipic hrtf database, in: Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the, IEEE, 2001, pp. 99–102.

[15] K. Watanabe, Y. Iwaya, Y. Suzuki, S. Takane, S. Sato, Dataset of head-related transfer functions measured with a circular loudspeaker array, Acoustical science and technology 35 (3) (2014) 159–165.

[16] H. Wierstorf, M. Geier, S. Spors, A free database of head related impulse response measurements in the horizontal plane with multiple distances, in: Audio Engineering Society Convention 130, Audio Engineering Society, 2011.

[17] C. T. Jin, P. Guillon, N. Epain, R. Zolfaghari, A. Van Schaik, A. I. Tew, C. Hetherington, J. Thorpe, Creating the sydney york morphological and acoustic recordings of ears database, IEEE Transactions on Multimedia 16 (1) (2014) 37–46.

[18] P. Majdak, M. Noisternig, Aes69-2015: Aes standard for file exchange-spatial acoustic data file format, Audio Engineering Society.(March 2015).

[19] C. D. Salvador, S. Sakamoto, J. Treviño, Y. Suzuki, Dataset of near-distance head-related transfer functions calculated using the boundary element method, to be presented in Audio Engineering Society International Conference on Spatial Reproduction -Aesthetics and Science-, Tokyo, Japan (Aug., 2018).

[20] F. Asano, Y. Suzuki, T. Sone, Role of spectral cues in median plane localization, The Journal of the Acoustical Society of America 88 (1) (1990) 159–168.

[21] Y. Haneda, S. Makino, Y. Kaneda, Common acoustical pole and zero modeling of room transfer functions, IEEE Trans. Speech Audio Process. 2 (2) (1994) 320–328.

[22] M. J. Evans, J. A. Angus, A. I. Tew, Analyzing head-related transfer function measurements using surface spherical harmonics, The Journal of the Acoustical Society of America 104 (4) (1998) 2400–2411.

[23] B.-S. Xie, Recovery of individual head-related transfer functions from a small set of measurementsa), The Journal of the Acoustical Society of America 132 (1) (2012) 282–294.

[24] W. Zhang, T. D. Abhayapala, R. A. Kennedy, R. Duraiswami, Insights into head-related transfer function: Spatial dimensionality and continuous representation, The Journal of the Acoustical Society of America 127 (4) (2010) 2347–2357.

[25] R. Duraiswaini, D. N. Zotkin, N. A. Gumerov, Interpolation and range extrapolation of hrtfs [head related transfer functions], in: Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on, Vol. 4, IEEE, 2004, pp. iv–45.

[26] G. D. Romigh, D. S. Brungart, R. M. Stern, B. D. Simpson, Efficient real spherical harmonic representation of head-related transfer functions, Selected Topics in Signal Processing, IEEE Journal of 9 (5) (2015) 921–930.

[27] M. Pollow, K.-V. Nguyen, O. Warusfel, T. Carpentier, M. Müller-Trapet, M. Vorländer, M. Noisternig, Calculation of head-related transfer functions for arbitrary field points using spherical harmonics decomposition, Acta acustica united with Acustica 98 (1) (2012) 72–82.

[28] A. W. Mills, On the minimum audible angle, The Journal of the Acoustical Society of America 30 (4) (1958) 237–246.

[29] P. Minnaar, J. Plogsties, F. Christensen, Directional resolution of head-related transfer functions required in binaural synthesis, Journal of the Audio Engineering Society 53 (10) (2005) 919–929.

[30] J. Trevino, S. Hu, C. Salvador, S. Sakamoto, J. Li, Y.-O. Suzuki, A compact representation of the head-related transfer function inspired by the wavelet transform on the sphere, in: Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2015 International Conference on, IEEE, 2015, pp. 372–375.

[31] S. Hu, J. Trevino, C. Salvador, S. Sakamoto, J. Li, Y. Suzuki, A local representation of the head-related transfer function, The Journal of the Acoustical Society of America 140 (3) (2016) EL285–EL290.

[32] A. P. Bates, Z. Khalid, R. A. Kennedy, On the use of slepian functions for the reconstruction of the head-related transfer function on the sphere, in: Signal Processing and Communication Systems (ICSPCS), 2015 9th International Conference on, IEEE, 2015, pp. 1–7.

[33] S. Mallat, A wavelet tour of signal processing, Academic press, 1999.

[34] J. D. McEwen, M. P. Hobson, D. J. Mortlock, A. N. Lasenby, Fast directional continuous spherical wavelet transform algorithms, IEEE transactions on Signal Processing 55 (2) (2007) 520–529.

[35] Z. Khalid, R. A. Kennedy, S. Durrani, P. Sadeghi, Y. Wiaux, J. D. McEwen, Fast directional spatially localized spherical harmonic transform, IEEE Transactions on Signal Processing 61 (9) (2013) 2192–2203.

[36] P. Schröder, W. Sweldens, Spherical wavelets: Efficiently representing functions on the sphere, in: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques, ACM, 1995, pp. 161–172.

[37] S. Sekimoto, R. Ogasawara, Y. Iwaya, Y. Suzuki, S. Takane, Numerical investigation of effects of head sizes and ear positions on head-related transfer functions, in: Proc. of The Japan-China Joint Conference of Acoustics 2007, Vol. 6, 2007.

[38] C. D. Salvador, S. Sakamoto, J. Treviño, Y. Suzuki, Boundary matching filters for spherical microphone and loudspeaker arrays, IEEE/ACM Transactions on Audio, Speech, and Language Processing 26 (3) (2018) 461–474.