

## PAPER

## Distance-varying filters to synthesize head-related transfer functions in the horizontal plane from circular boundary values

César D. Salvador<sup>\*</sup>, Shuichi Sakamoto<sup>†</sup>, Jorge Treviño<sup>‡</sup> and Yôiti Suzuki<sup>§</sup>

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

*(Received 28 March 2016, Accepted for publication 29 June 2016)*

**Abstract:** This paper derives a continuous-space model to describe variations in magnitude of complex head-related transfer functions (HRTFs) along angles and radial distances throughout the horizontal plane. The radial part of this model defines a set of horizontal-plane distance-varying filters (HP-DVFs) that are used to synthesize the HRTFs for arbitrary sound source positions on the horizontal plane from initial HRTFs obtained for positions on a circular boundary at a single distance from the head of a listener. The HP-DVFs are formulated in terms of horizontal-plane solutions to the three-dimensional acoustic wave equation, which are derived by assuming invariance along elevation angles in spherical coordinates. This prevents the free-field inaccurate distance decay observed when assuming invariance along height in cylindrical coordinates. Furthermore, discontinuities along the axis connecting the ears are also overcome, which appear when assuming invariance along the polar angle in interaural coordinates. This paper also presents a magnitude-dependent band-limiting threshold (MBT) for restricting the action of filters to a limited angular bandwidth, which is necessary in practice to enable discrete-space models that consider a finite number of sources distributed on the initial circle. Numerical experiments using a model of a human head show that the overall synthesis accuracy achieved with the proposed MBT outperforms the one achieved with the existing frequency-dependent threshold, especially at low frequencies and close distances to the head.

**Keywords:** Sound source distance, Head-related transfer functions, Hankel functions

**PACS number:** 43.66.Pn, 43.66.Qp, 43.66.Ba [doi:10.1250/ast.38.1]

### 1. INTRODUCTION

Head-related transfer functions (HRTFs) are a central tool for the realistic, binaural presentation of three-dimensional auditory spaces [1,2]. Their aim is to characterize the scattering of sound due to the external anatomy of a listener, mainly determined by the shape of their pinna, head, and body. HRTFs are represented by linear filters relating the position of a sound source and the sound pressure generated by that source at the ears of the listener [3].

HRTFs contain the auditory cues that are essential to perceive the direction of sound sources in anechoic conditions. Auditory cues for the perception of distance are also present in the HRTFs for sources within 1 m of the

listener's head [4–6]. Examples of numerical simulations of circular datasets for left-ear HRTFs at far and near distances are shown in Fig. 1. The trend of ipsilateral spectral features at the far distance to concentrate around the ear's direction ( $\theta = 90^\circ$ ) in the near distance illustrates the so-called acoustic parallax effect [6,7]. The decrease in energy of contralateral spectral features at the near distance illustrates the prominence of the head shadowing effect [5]. Such distance cues are particularly useful when presenting auditory scenes comprised of lateral sounds near the heads of the listeners [8].

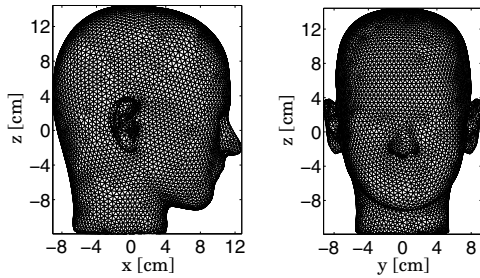
Binaural presentation would ideally require knowing the HRTFs for all points in three-dimensional space. However, measuring or calculating these HRTF datasets is a complex and time consuming task. A physically-motivated synthesis method described in [9–11] makes it possible to calculate the HRTFs for arbitrary positions in space, once an initial HRTF dataset for discrete positions on a sphere surrounding the listener is given. Prior

<sup>\*</sup>e-mail: salvador@ais.riec.tohoku.ac.jp

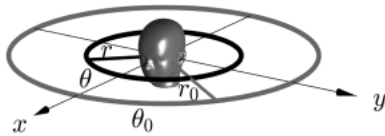
<sup>†</sup>e-mail: saka@ais.riec.tohoku.ac.jp

<sup>‡</sup>e-mail: jorge@ais.riec.tohoku.ac.jp

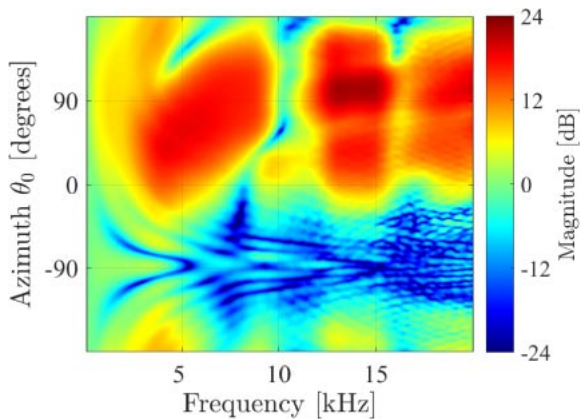
<sup>§</sup>e-mail: yoh@riec.tohoku.ac.jp



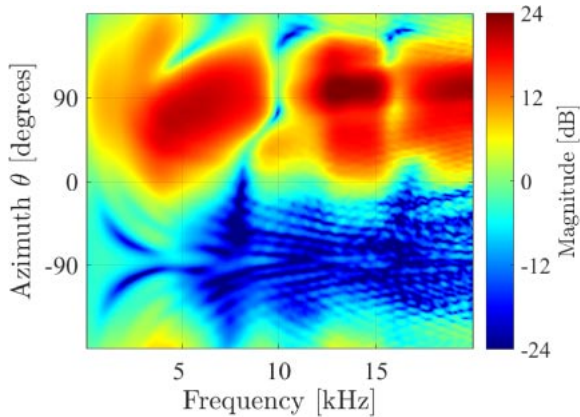
(a) Lateral and front views of the symmetric head model used in this paper for numerical experiments. The grid consists of 14,096 points with an average cell length of 5.1 mm.



(b) Source positions in circular boundaries.



(c) HRTFs at  $r_0 = 150$  cm.



(d) HRTFs at  $r = 25$  cm.

**Fig. 1** Two circular HRTF datasets for the left ear ( $\theta_0 = \theta = 90^\circ$ ) calculated with the boundary element method (BEM) [15]. Azimuthal angles  $\theta_0$  and  $\theta$  are measured from the positive  $x$ -axis, which indicates the front position ( $\theta_0 = \theta = 0^\circ$ ). Magnitudes indicate the difference in sound pressure level at the left ear compared to that which would be observed at the position of the head's center in free-field conditions.

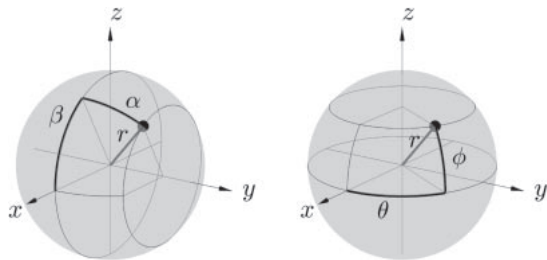
knowledge about the HRTFs at the target positions is not required when applying this model-based method. This is a useful result, since spherical datasets are typically obtained for only one far distance [12]. Nevertheless, optimally distributing sound sources on the sphere is a non-trivial problem [13,14], and obtaining the HRTFs for all required initial positions might not be feasible for many applications.

In many applications, it is enough to present sounds from positions on a horizontal plane at the height of the listener's ears, even though this may not be enough to present complex scenes such as those including reverberations. The reason behind this lies in the fact that real life sound sources are often located on or close to this horizontal plane. Furthermore, the human auditory system can resolve sounds more accurately in this plane [16]. Consequently, horizontal-plane virtual auditory displays would be an important use case. Obtaining HRTFs for a dense set of positions throughout the plane, however, can still be very demanding and time consuming. It would be useful, then, if a method similar to the one presented in [9–11] could be applied to synthesize the HRTFs for arbitrary positions on the horizontal plane, once an initial, circular HRTF dataset is known at a single (typically far) distance.

The synthesis approach in [9–11] relies on the principle of acoustic reciprocity [17] to consider HRTF datasets as sound pressure fields generated by a point source at the ear. This approach originates from three-dimensional solutions to the acoustic wave equation formulated in spherical coordinates, where solutions can be separated into a radial and an angular part [18, Chapter 6]. The angular part defines an orthonormal basis on the sphere, which is used to represent the directional information in the HRTFs that will remain unchanged. The radial part is then used to formulate distance-varying filters in the represented domain. Unfortunately, ambiguous results are obtained when deriving methods for the horizontal plane by simply restricting the spherical representation to a circle [19,20]. This is because, even when considering only a plane, sound still propagates along three dimensions.

Two approaches for horizontal-plane synthesis of HRTFs that take into consideration the spatial propagation of sound have been explored in [21]. Both rely on finding horizontal-plane solutions to the three-dimensional acoustic wave equation by assuming sound fields that are invariant along a spatial coordinate [18, Chapters 4 and 6].

The first approach assumes invariance along height in cylindrical coordinates. This assumption is equivalent to considering vertical, infinite-length line sources of sound. Approximate solutions for the case of finite-length line sources have been proposed in [22] based on the stationary phase method [18, pp. 137–140]. However, these solutions



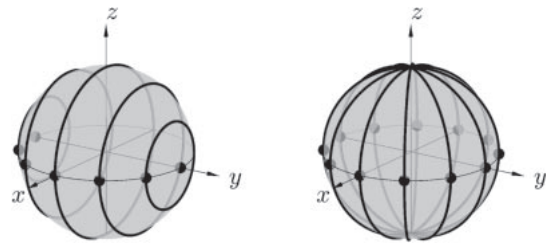
(a) Interaural coordinates. (b) Spherical coordinates.

**Fig. 2** Two ways of describing sound source positions used to characterize HRTF datasets. The front position is along the positive  $x$ -axis. The ears of the listener are along the  $y$ -axis.

are only accurate at a specified distance from the origin. In either way, the main limitation of modeling in cylindrical coordinates is the impossibility of considering the spherical symmetry of sound field propagation due to external sound sources. As a result, the free-field pressure decay along radial distance towards the head's center can not be accurately predicted. The accurate prediction of such distance decay pattern is an important requirement because it matches the distance decay due to point sources of sound, which are the kind of sources used in the definition of HRTFs.

The second approach attempts to account for the spherical, radial convergence towards the head's center by assuming invariance along vertical angles measured from the horizontal plane, which are described by the polar angle  $\beta$  in the interaural coordinate system shown in Fig. 2(a). For a given distance, this assumption is equivalent to considering vertical, circular sources of sound that are parallel to the median plane ( $xz$ -plane), and whose radii decrease as they move away from this plane. This leads to a method that requires the division of the horizontal plane into two regions, in front of and behind the listener, delimited by an axis connecting the ears' canals ( $y$ -axis). In practice, this can be understood as considering a semi-circular array of circular sources (in front or behind), as illustrated in Fig. 3(a) for the frontal case. Nevertheless, this method yields results on the front and back regions that take different values along the interaural axis. Discontinuities on the lateral sides are hence inevitable, precisely where the distance cues are more prominent [8].

In this paper, we properly account for the spherical, radial convergence towards the head's center by assuming invariance along the elevation angle  $\phi$  defined in the spherical coordinate system shown in Fig. 2(b). This assumption is equivalent to considering vertical, semi-circular sources of sound that connect points of equal azimuthal angle  $\theta$  and are terminated by the north and south poles. The vertical semicircles have equal radii; they



(a) Semicircular array of circular sources in interaural coordinates. (b) Circular array of semicircular sources in spherical coordinates.

**Fig. 3** Arrays of sound sources on the horizontal plane.

are also continuously and uniformly distributed around the  $z$ -axis. In practice, this can be interpreted as considering a circular array of semicircular sources, as illustrated in Fig. 3(b). The underlying assumption therefore leads to solutions to the acoustic wave equation that uniformly consider the horizontal plane as a whole. Consequently, there are no singularities or discontinuities in the results. The angular part of the solutions results in a suitable basis for representing directional information on full horizontal circles. The radial part is used to formulate a set of distance-varying filters to be applied in the represented domain. Our solutions, hence, define a method to synthesize HRTFs at arbitrary distances on the horizontal plane, once a circular HRTF dataset is known.

Another issue addressed in this paper deals with angular resolution in practical filter implementation. Continuous, circular distributions of sources are necessarily considered when deriving distance-varying filters from analytic solutions to the acoustic wave equation. Yet HRTF datasets are obtained in practice for discrete distributions of sources. Relating discrete and continuous distributions requires additional assumptions. In particular, following the Nyquist sampling theorem on the circle [23], we consider that continuous distributions have finite angular bandwidths. Implementations therefore require the establishment of thresholds to restrict the action of distance-varying filters to the corresponding angular bandwidth. We examine an existing threshold [24] to limit the angular bandwidth according to frequency and find that this traditional threshold does not ensure accurate results, especially at low frequencies and positions close to the head. To cope with this problem, we propose a threshold to limit the angular bandwidth in accordance with the magnitude decay of spherical wavefronts along distance in free-field.

The remainder of this paper is organized as follows. Section 2 reviews the method in interaural coordinates. Section 3 presents our proposal in spherical coordinates. Section 4 addresses the issue on angular bandwidth limitation. Section 5 presents and discusses the results. Conclusions are stated in Sect. 6.

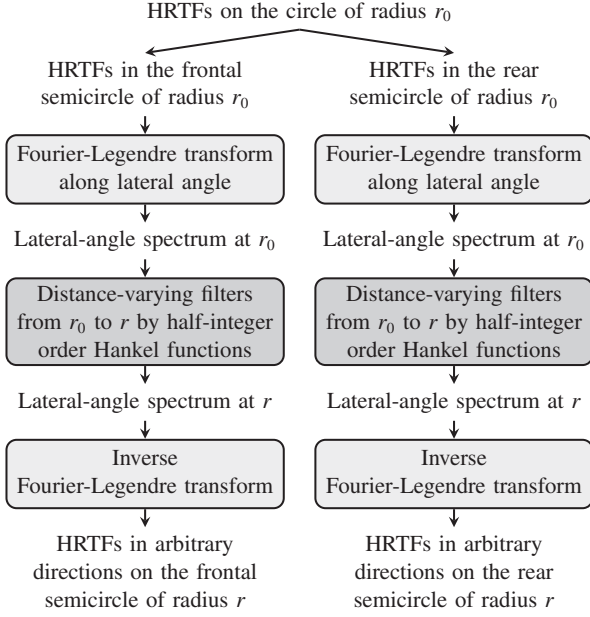


Fig. 4 Block diagram of the synthesis method in interaural coordinates [21].

## 2. SYNTHESIS METHOD IN INTERAURAL COORDINATES

In this section, we discuss the fundamentals of the synthesis method used in [21] when continuous distributions of sources are assumed. This method is formulated on the coordinate system shown in Fig. 2(a), where a point in space is represented by its radial distance  $r$ , lateral angle  $\alpha \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ , and polar angle  $\beta \in [-\pi, \pi]$ . These coordinates are a suitable choice for representing the underlying assumption that the interaural HRTF (the left ear HRTF divided by the right ear HRTF) is constant along polar angle  $\beta$  for the hypothetical case of a spherical head. This method can hence be formulated by solving the acoustic wave equation in interaural coordinates, where sound fields are assumed to be invariant along vertical circles parallel to the median plane ( $xz$ -plane) and centered on the interaural axis ( $y$ -axis). An overview of this method is illustrated in Fig. 4.

### 2.1. Sound Field Propagation in Interaural Coordinates

By  $\Psi^{\text{int}}(r, \alpha, \beta, t) \propto R(r)A(\alpha)B(\beta)\exp(-j\omega t)$ , we denote a time-harmonic sound pressure field with angular frequency  $\omega = 2\pi f$  and independent spatial variables, where  $j = \sqrt{-1}$ . The upper index refers to the interaural coordinate system. Sound fields of this form simplify the three-dimensional wave equation to a sum of three ordinary differential equations such as the ones shown in the following expression [18, p. 183]:

$$\frac{1}{R} \frac{d}{dr} \left[ r^2 \frac{d}{dr} \right] R + \underbrace{\frac{1}{\cos \alpha} \frac{1}{A} \frac{d}{d\alpha} \left[ \cos \alpha \frac{d}{d\alpha} \right] A + \frac{1}{\cos^2 \alpha} \frac{1}{B} \frac{d^2}{d\beta^2} B}_{=-n(n+1)} = -r^2 \frac{\omega^2}{c^2}, \quad (1)$$

where  $c$  is the speed of sound in air.

Solutions to (1) on the horizontal plane ( $xy$ -plane) are obtained by setting a separation constant  $n$  and assuming  $B(\beta)$  is constant. The term involving  $B(\beta)$  thus vanishes. The term involving  $A(\alpha)$  results in a Legendre differential equation, whose solution  $A(\alpha) = P_n(\sin \alpha)$  is the Legendre polynomial of order  $n \in \mathbb{N}$ . The radial term thereof produces a Bessel differential equation; one of its solutions is given in terms of the Hankel function of the second kind and half-integer order  $H_{n+\frac{1}{2}}(\frac{\omega}{c}r)$ . The radial term  $R(r) = r^{-\frac{1}{2}}H_{n+\frac{1}{2}}(\frac{\omega}{c}r)$  is known as the spherical Hankel function.

Sound pressure fields can be defined by the general solution to (1) in the frequency domain according to

$$\Psi^{\text{int}}(r, \alpha, \omega) = \sum_{n=0}^{\infty} \mathcal{C}_n^{\text{int}}(\omega) r^{-\frac{1}{2}} H_{n+\frac{1}{2}}\left(\frac{\omega}{c}r\right) P_n(\sin \alpha). \quad (2)$$

By virtue of the orthonormality property for  $P_n$ , the coefficients  $\mathcal{C}_n^{\text{int}}$  can be calculated using the following expression:

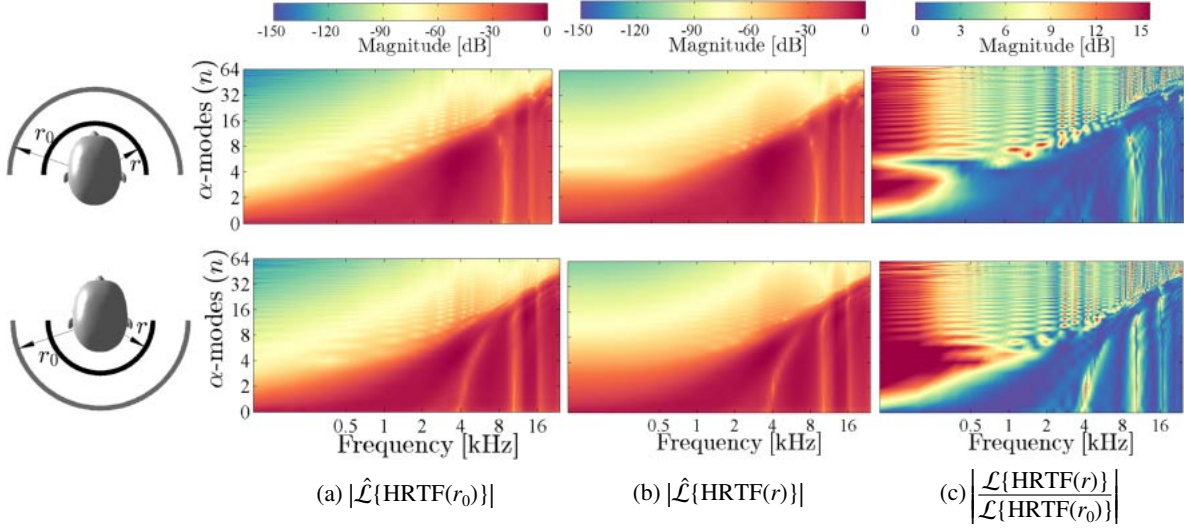
$$\mathcal{C}_n^{\text{int}}(\omega) = \frac{n + \frac{1}{2}}{r^{-\frac{1}{2}} H_{n+\frac{1}{2}}\left(\frac{\omega}{c}r\right)} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \Psi^{\text{int}}(r, \alpha, \omega) P_n(\sin \alpha) \cos \alpha d\alpha. \quad (3)$$

Note that distance and angle variations in (2) are fully represented by  $R(r)$  and  $P_n$ , respectively. Coefficients  $\mathcal{C}_n^{\text{int}}$  hence exclusively depend on  $\omega$  and (3) holds for any other distance.

Given the independence of  $\mathcal{C}_n^{\text{int}}$  with respect to distance, the relation between the sound pressure fields at two different distances  $r$  and  $r_0$  can be formulated as follows:

$$\Psi^{\text{int}}(r, \alpha, \omega) = \sum_{n=0}^{\infty} \left( n + \frac{1}{2} \right)^{\frac{1}{2}} P_n(\sin \alpha) \underbrace{\left[ \frac{r^{-\frac{1}{2}} H_{n+\frac{1}{2}}\left(\frac{\omega}{c}r\right)}{r_0^{-\frac{1}{2}} H_{n+\frac{1}{2}}\left(\frac{\omega}{c}r_0\right)} \right]}_{\text{Filters: } D_n^{\text{int}}(r, r_0, \omega)} \times \underbrace{\left( n + \frac{1}{2} \right)^{\frac{1}{2}} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \Psi^{\text{int}}(r_0, \alpha_0, \omega) P_n(\sin \alpha_0) \cos \alpha_0 d\alpha_0}_{\text{Fourier-Legendre transform along lateral angle: } \mathcal{L}\{\Psi^{\text{int}}\}}. \quad (4)$$

Here, the term on the second line is a Fourier-Legendre transform along  $\alpha_0$ , the term above in brackets defines distance-varying filters in a horizontal half-plane, and the



**Fig. 5** Analysis of distance variation between the datasets shown in Figs. 1(c) and 1(d), from  $r_0 = 150$  cm to  $r = 25$  cm, using the Fourier-Legendre transform  $\mathcal{L}$  defined in (4). Frequencies and lateral-angle modes are depicted along the horizontal and vertical axes, respectively. The colorbars in panels (a) and (b) correspond to magnitudes of normalized Fourier-Legendre transforms  $\hat{\mathcal{L}}$ , while the colorbar in (c) corresponds to magnitudes of ratios between non-normalized Fourier-Legendre transforms  $\mathcal{L}$  at  $r$  and  $r_0$ .

sum of Legendre polynomials on  $\alpha$  is an inverse Fourier-Legendre transform.

The principle of reciprocity [17] implies that HRTF datasets can also be characterized by the sound pressure around the listener due to a point source at each of their ears. Applying (4) to these datasets then defines the synthesis method shown in Fig. 4. Given that lateral angles are in  $[-\frac{\pi}{2}, \frac{\pi}{2}]$ , Fourier-Legendre transforms need to act on separate datasets defined over two semicircles in front and behind the listener. Equation (4) first decomposes the dataset using an integral over  $\alpha_0$  and then reconstructs it at a target angle  $\alpha$ . It should be noted that these two variables are independent.

## 2.2. Fourier-Legendre Analysis of Distance Variation

To investigate the relative variation over distance of circular HRTF datasets, we calculate the ratio between normalized Fourier-Legendre transforms. We analyzed the datasets in Figs. 1(c) and 1(d) to obtain the illustrative results shown in Fig. 5. General remarks are as follows.

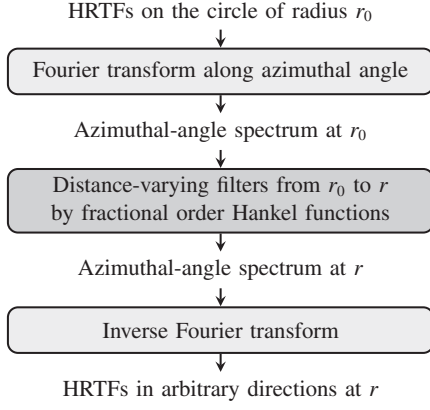
- i) In Figs. 5(a) and 5(b), it can be appreciated that most of the energy in the Fourier-Legendre transform of the datasets is concentrated in the lower lateral-angle modes ( $\alpha$ -modes). The transition between  $\alpha$ -modes of high and low energy increases as frequency increases.
- ii) In the distance variation patterns shown in Fig. 5(c), which were calculated based on the ratio between the Fourier-Legendre transforms illustrated in Figs. 5(a) and 5(b), it is observed that the distance variation patterns corresponding to high-energy  $\alpha$ -modes are

relatively simple as opposed to the ones corresponding to low-energy modes.

- iii) The nature of the Fourier-Legendre transform in (4) implies that frontal and rear regions must be analyzed separately because  $\alpha_0 \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ . These analyses are respectively shown in the top and bottom panels of Fig. 5. Furthermore, distance variations follow different patterns in each of these two regions.

## 3. SYNTHESIS METHOD IN SPHERICAL COORDINATES

In this section, we detail our method to synthesize the HRTFs at arbitrary points on the horizontal plane from HRTF datasets given on a single circular boundary. Continuous distributions of sources are assumed throughout this section. The proposal is formulated on the coordinate system shown in Fig. 2(b), where a point in space is represented by its radial distance  $r$ , azimuthal angle  $\theta \in [-\pi, \pi]$ , and elevation angle  $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ . The underlying idea is to generate spherical datasets by replicating circular datasets along semicircles connecting points of equal azimuth  $\theta$  and terminated by the north and south poles. Semicircles of this type define the meridians. Our method thus originates from an equivalent problem consisting of solving the acoustic wave equation on these coordinates, under the assumption that sound fields are invariant along meridians. Although this assumption can appear one step removed from physical reality when considering the three-dimensional space, it is accepted here for the practical purpose of having a uniform distribution of semicircular sources on a circle, as shown in Fig. 3(b). This



**Fig. 6** Block diagram of the synthesis method in spherical coordinates.

assumption allows us to derive a set of distance-varying filters that takes into consideration the radial symmetry of acoustic propagation towards the head center. An overview of our method is illustrated in Fig. 6.

### 3.1. Sound Field Propagation in Spherical Coordinates

By  $\Psi^{\text{sph}}(r, \theta, \phi, t) \propto \mathfrak{R}(r)\Theta(\theta)\Phi(\phi)\exp(-j\omega t)$ , we denote a sound field varying independently along each one of its coordinates. The upper index refers to the spherical coordinate system. Sound fields of this form also decouple the wave equation into a sum of ordinary differential equations [18, pp. 183]:

$$\frac{1}{\mathfrak{R}} \frac{d}{dr} \left[ r^2 \frac{d}{dr} \right] \mathfrak{R} + \frac{1}{\cos^2 \phi} \underbrace{\frac{1}{\Theta} \frac{d^2}{d\theta^2} \Theta}_{=-m^2} + \frac{1}{\cos \phi} \frac{1}{\Phi} \frac{d}{d\phi} \left[ \cos \phi \frac{d}{d\phi} \right] \Phi = -r^2 \frac{\omega^2}{c^2}. \quad (5)$$

We look for solutions to (5) on a horizontal plane using separation constant  $m$ . We assume that  $\Phi(\phi)$ , the elevation-dependent portion of the solution, is constant. The term involving  $\Phi(\phi)$  thus vanishes and the azimuthal solution oscillates according to  $\Theta(\theta) = \exp(jm\theta)$ , where  $m \in \mathbb{Z}$ .

We deal with the remaining radial term in (5) using auxiliary function  $\mathcal{R}(r) = r^{\frac{1}{2}}\mathfrak{R}(r)$  to obtain the following Bessel differential equation for it:

$$\frac{d^2}{dr^2} \mathcal{R} + \frac{1}{r} \frac{d}{dr} \mathcal{R} + \left[ \frac{\omega^2}{c^2} - \frac{\left( \frac{m^2}{\cos^2 \phi} + \frac{1}{4} \right)}{r^2} \right] \mathcal{R} = 0. \quad (6)$$

For the sake of simplicity, we denote the term inside the parentheses by a parameter  $\mu$  as follows:

$$\mu^2 = \frac{m^2}{\cos^2 \phi} + \frac{1}{4}, \quad (7)$$

where  $\phi \in (-\frac{\pi}{2}, \frac{\pi}{2})$ . Solutions to (6) can now be obtained in terms of the well-known Bessel functions, among which we choose  $\mathcal{R}(r) = H_{\mu}(\frac{\omega}{c}r)$ , the Hankel function of the second kind, and fractional order  $\mu$ . The radial solution finally results in  $\mathfrak{R}(r) = r^{-\frac{1}{2}}H_{\mu}(\frac{\omega}{c}r)$ .

We can thus define the sound pressure field in a horizontal plane in terms of our general solution to (5) in the frequency domain in such a way that

$$\Psi^{\text{sph}}(r, \theta, \omega) = \sum_{m=-\infty}^{\infty} \mathcal{C}_m^{\text{sph}}(\omega) r^{-\frac{1}{2}} H_{\mu} \left( \frac{\omega}{c} r \right) \exp(jm\theta). \quad (8)$$

We interpret this result as an incoming converging wave due to a radiating semicircular arc along a meridian. As a consequence of the orthonormality of complex exponentials, coefficients  $\mathcal{C}_m^{\text{sph}}$  remain defined as follows:

$$\mathcal{C}_m^{\text{sph}}(\omega) = \frac{1}{2\pi r^{-\frac{1}{2}} H_{\mu} \left( \frac{\omega}{c} r \right)} \int_{-\pi}^{\pi} \Psi^{\text{sph}}(r, \theta, \omega) \exp(-jm\theta) d\theta. \quad (9)$$

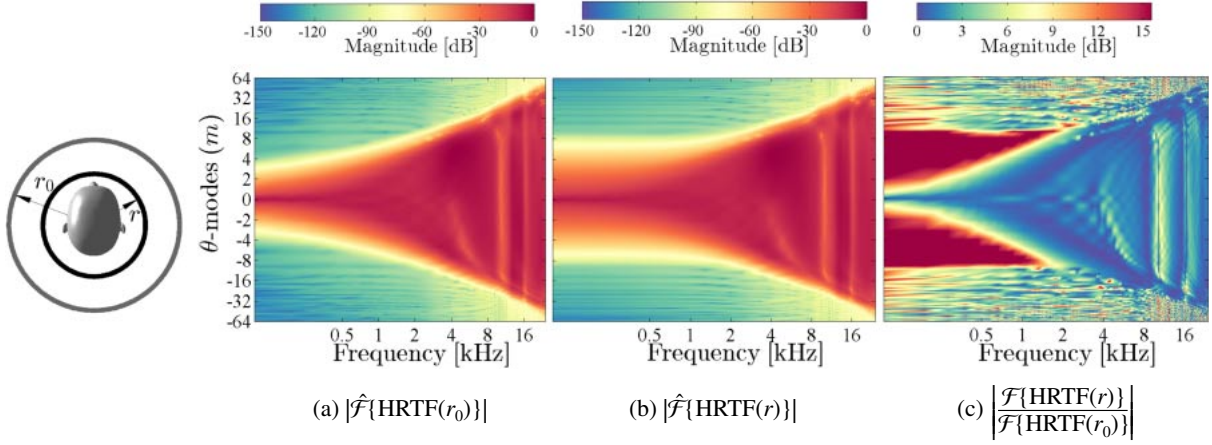
Note that distance and angle variations in (8) are fully described by  $\mathfrak{R}(r)$  and  $\exp(jm\theta)$ , respectively. Coefficients  $\mathcal{C}_m^{\text{sph}}$  hence depend on  $\omega$  only and (9) holds for any distance.

Because coefficients  $\mathcal{C}_m^{\text{sph}}$  are invariant with respect to distance, the sound pressure fields at two different distances  $r$  and  $r_0$  are finally linked according to the following expression:

$$\Psi^{\text{sph}}(r, \theta, \omega) = \sum_{m=-\infty}^{\infty} \left( \frac{1}{2\pi} \right)^{\frac{1}{2}} \exp(jm\theta) \underbrace{\left[ \frac{r^{-\frac{1}{2}} H_{\mu} \left( \frac{\omega}{c} r \right)}{r_0^{-\frac{1}{2}} H_{\mu} \left( \frac{\omega}{c} r_0 \right)} \right]}_{\text{Filters: } D_m^{\text{sph}}(r, r_0, \omega)} \times \underbrace{\left( \frac{1}{2\pi} \right)^{\frac{1}{2}} \int_{-\pi}^{\pi} \Psi^{\text{sph}}(r_0, \theta_0, \omega) \exp(-jm\theta_0) d\theta_0}_{\text{Fourier transform along azimuthal angle: } \mathcal{F}\{\Psi^{\text{sph}}\}}. \quad (10)$$

Here, the term in the second line represents a Fourier transform along  $\theta_0$ . The term in brackets defines the set of horizontal-plane distance-varying filters (HP-DVF) with parameter  $\mu$  defined in (7). Note that  $\mu$  can correspond to horizontal planes at different elevation angles  $\phi \in (-\frac{\pi}{2}, \frac{\pi}{2})$ . For the goals of this paper, we delimit our discussion hereafter to the case  $\phi = 0$ . Finally, the sum of complex exponentials on  $\theta$  describes an inverse Fourier transform.

Applying (10) to HRTF datasets defines the synthesis method shown in Fig. 6. Because azimuthal angles are in  $[-\pi, \pi]$ , Fourier transforms act on datasets defined on complete circular boundaries around the listener. Equation (10) first decomposes the dataset using an integral over  $\theta_0$  and then reconstructs it at a target angle  $\theta$ . Note that these two variables are independent.



**Fig. 7** Analysis of distance variation between the datasets shown in Figs. 1(c) and 1(d), from  $r_0 = 150$  cm to  $r = 25$  cm, using the Fourier transform  $\mathcal{F}$  defined in (10). Frequencies and azimuthal-angle modes are depicted along the horizontal and vertical axes, respectively. The colorbars in panels (a) and (b) correspond to magnitudes of normalized Fourier transforms  $\hat{\mathcal{F}}$ , while the colorbar in (c) corresponds to magnitudes of ratios between non-normalized Fourier transforms  $\mathcal{F}$  at  $r$  and  $r_0$ .

### 3.2. Fourier Analysis of Distance Variation

Similarly to the examples in Sect. 2.2., those shown in Fig. 7 are based on the ratio between the Fourier transforms of the datasets shown in Figs. 1(c) and 1(d). General remarks are as follows.

- i) In Figs. 7(a) and 7(b), it can be observed that most of the energy in the Fourier transform of the datasets is concentrated in the lower azimuthal-angle modes ( $\theta$ -modes). The transition between  $\theta$ -modes of high and low energy increases as frequency increases. The decay of energy along the transition occurs faster than in the Fourier-Legendre domain, as can be observed when comparing with the examples shown in Figs. 5(a) and 5(b).
- ii) In the distance variation patterns shown in Fig. 7(c), which were calculated based on the ratio between the Fourier transforms illustrated in Figs. 7(a) and 7(b), it can be appreciated that the distance variation patterns corresponding to high-energy  $\theta$ -modes of the datasets are relatively simple, in contrast with the ones corresponding to low-energy modes.
- iii) Because Fourier transforms are applied on full circular boundaries at once, the distance variations between two circular datasets are described by a single set of patterns.

## 4. BANDWIDTH-LIMITED DISTANCE-VARYING FILTERS

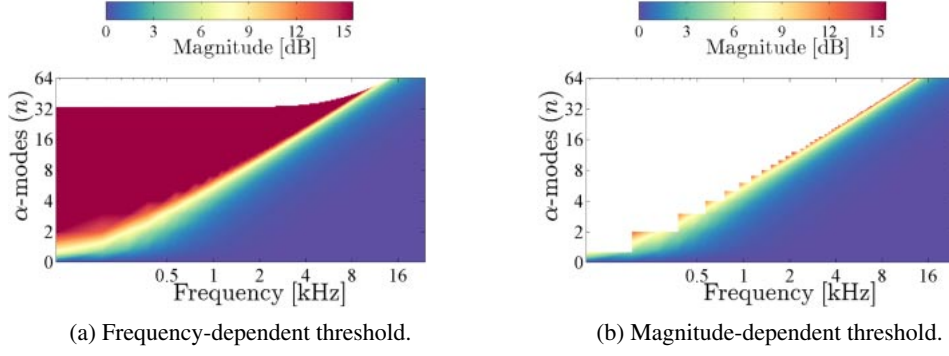
In this section, we address an issue regarding practical implementation of distance-varying filters. Continuous distributions of sources were considered in Sects. 2 and 3 when formulating distance-varying filters. Yet HRTF datasets are obtained for discrete distributions of finite

angular resolution and, hence, representations of datasets on transform domains are accurate in a limited angular bandwidth. Implementations thus require thresholds to restrict the action of distance-varying filters to a limited angular bandwidth.

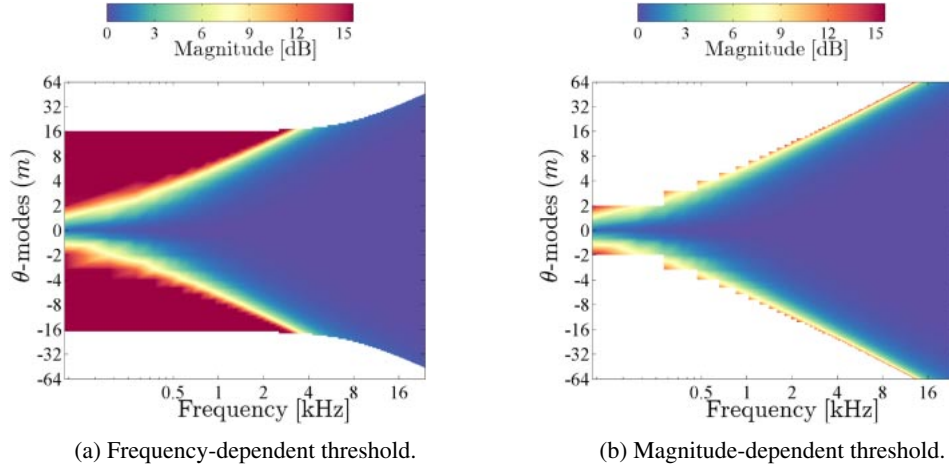
Figures 5 and 7 show that most of the energy in angular spectra of datasets is concentrated in low angular modes up to a frequency-dependent high-to-low energy transition. Small values taken at higher modes would, in practice, lead to a low signal-to-noise ratio that can negatively impact the results; this is in agreement with the ill-conditioned nature of the acoustic propagation problem [18, pp. 291–293]. Furthermore, the behavior of these low-energy components at different distances follows a complex pattern that may be difficult to capture in a general model. For these reasons, we focus our analysis on limiting the angular bandwidth so as to preserve the high-energy modes only.

Limited angular bandwidths are modeled by truncating the sums along modal indices  $n$  in (4) and  $m$  in (10) to a finite number of terms. In [24], a frequency-dependent threshold to decide bounds for modal indices was deduced by using the asymptotic expansion of spherical Hankel functions for large orders. In general, bounds  $L$  for  $n \leq \lfloor L \rfloor$  or  $|m| \leq \lfloor L \rfloor$ , beyond which the distance-varying filters are set to zero, can be defined as follows [24]:

$$L = \left\lceil \left[ \frac{1}{\ln\left(\frac{r}{r_h}\right)} \ln \left( \frac{\left(\frac{r}{r_h}\right)^{\frac{3}{2}}}{\left(\frac{r}{r_h} - 1\right)^{\frac{3}{2}} \epsilon} \right) + 1 \right]^4 \right\rceil$$



**Fig. 8** Distance-varying filters in interaural coordinates  $\hat{D}_n^{\text{int}}(r_0, r)$  from  $r_0 = 150$  cm to  $r = 25$  cm, whose directional bandwidths have been limited according to the frequency-dependent threshold in (11) and proposed magnitude-dependent threshold in (12). Colorbars indicate magnitude gains.



**Fig. 9** Proposed distance-varying filters in spherical coordinates  $\hat{D}_m^{\text{sph}}(r_0, r)$  from  $r_0 = 150$  cm to  $r = 25$  cm, whose directional bandwidths have been limited according to the frequency-dependent threshold in (11) and proposed magnitude-dependent threshold in (12). Colorbars indicate magnitude gains.

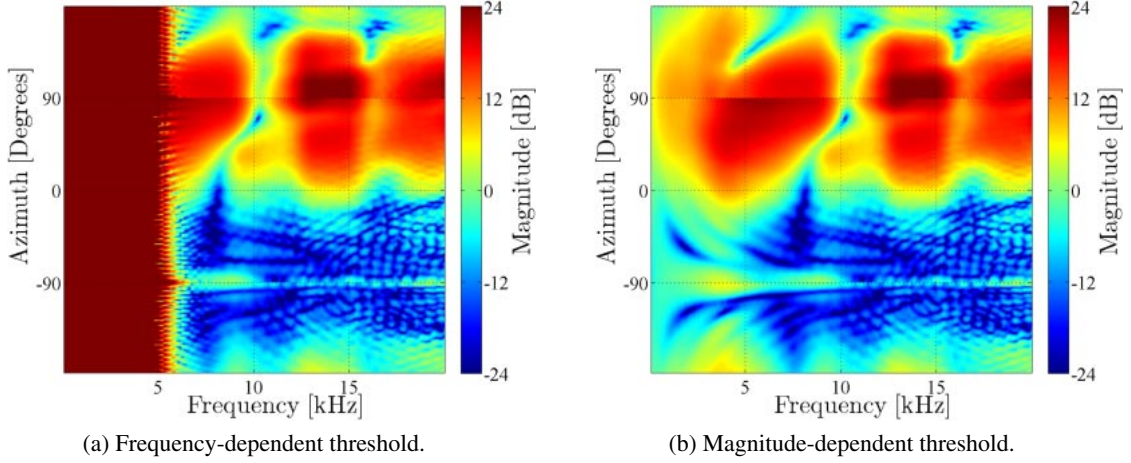
$$+ \left[ \frac{\omega r_h}{c} + \frac{1}{2} \left( \ln \frac{1}{\epsilon^3} + \ln \left( \frac{\omega r_h}{c} \right)^{\frac{1}{2}} \right)^{\frac{2}{3}} \left( \frac{\omega r_h}{c} \right)^{\frac{1}{3}} \right]^{\frac{1}{4}}. \quad (11)$$

Here,  $\epsilon$  is a specified approximation error to represent spatial information inside the smallest radius  $r_h$  containing the head, due to a source at a distance  $r$ . Examples of filters designed using this threshold, for  $\epsilon = 10^{-5}$ ,  $r_h = 15$  cm and  $r = 25$  cm, are shown in Figs. 8(a) and 9(a). Filters designed with this threshold attempt to fully approximate distance-varying patterns, such as the ones shown in Figs. 5(c) and 7(c), by further including information on spectral regions comprising high modes at low frequencies. Nevertheless, resulting filters have the risk of overemphasizing components of low signal-to-noise ratio, as angular spectra of initial datasets at far distances contain low-energy components in these spectral regions.

To cope with this problem, we first observe that the distance-varying filters in (4) and (10) have magnitude responses that are monotonically increasing functions of relative distance, and of  $n$  or  $m$ , respectively. We then propose to restrict their angular bandwidths by clipping their magnitude responses according to the ratio between the surface area of the spheres at  $r_0$  and  $r$ . In addition, to generalize the distance-varying filters to work in both the far and near fields, we scale the filters by a factor  $\frac{r}{r_0}$ . Such scaling is equivalent to the use of Hankel functions normalized by the magnitude of their far-field asymptotic forms [10] or by the magnitude of the near-field radiation of a point source [25]. We formulate a magnitude-dependent band-limiting threshold (MBT) according to the following expression:

$$\hat{D}(r, r_0, \omega) = \begin{cases} \frac{r}{r_0} D(r, r_0, \omega) & \text{if } |D(r, r_0, \omega)| \leq \frac{r_0^2}{r^2}, \\ 0 & \text{else.} \end{cases} \quad (12)$$





**Fig. 10** HRTF datasets at  $r = 25$  cm synthesized with the distance varying filters in interaural coordinates. Directional bandwidths were limited according to the frequency-dependent threshold in (11) and the proposed magnitude-dependent threshold in (12).

Here,  $\hat{D}$  is a distance-varying filter with limited angular bandwidth and  $D$  is one of the filters defined in (4) or (10). Examples of filters designed with this threshold, for the methods in interaural coordinates and spherical coordinates, are respectively shown in Figs. 8(b) and 9(b), where it is observed that the action of the filters is properly restricted to the spectral regions of high-energy components. Magnitude gains at the edges correspond to thresholds  $\frac{v_2^2}{r^2}$  in (12), while staircase patterns are due to using a frequency resolution of 93.75 kHz, which in turn correspond to using 512 samples and a sampling frequency of 48 kHz. Note that no prior knowledge on the target HRTFs is required when using (11) or (12).

## 5. EVALUATION OF THE SYNTHESIS ACCURACY

In this section, we numerically compare the performances of the method in the interaural coordinates defined in (4) and the proposal in the spherical coordinates defined in (10).

### 5.1. Objective Measures of Overall Accuracy

Evaluations were performed by comparing synthesized datasets denoted by  $\hat{\mathcal{H}}$  and target datasets denoted by  $\mathcal{H}$ .

Overall accuracy along frequency was calculated by a logarithmic spectral distance that has been shown to be suitable for predicting audible differences between measured and synthesized HRTFs [26]. We refer to this distance as the spectral distortion (SD) in decibels and write it as follows [26]:

$$SD(\theta) = \left[ \frac{1}{f_2 - f_1} \int_{f_1}^{f_2} \left[ 20 \log_{10} \left| \frac{\hat{\mathcal{H}}(\theta, f)}{\mathcal{H}(\theta, f)} \right| \right]^2 df \right]^{\frac{1}{2}}. \quad (13)$$

Overall accuracy along angles, described for convenience by azimuthal angle  $\theta$ , was calculated based on circular correlations (CC) that provide a good measure for the similarities of the directional patterns of  $\hat{\mathcal{H}}$  and  $\mathcal{H}$ , independently of a possible gain mismatch [11]. Normalized CC is defined as follows [11]:

$$CC(f) = \frac{\int_{-\pi}^{\pi} \hat{\mathcal{H}}(\theta, f) \overline{\mathcal{H}(\theta, f)} d\theta}{\left[ \int_{-\pi}^{\pi} |\hat{\mathcal{H}}(\theta, f)|^2 d\theta \times \int_{-\pi}^{\pi} |\mathcal{H}(\theta, f)|^2 d\theta \right]^{\frac{1}{2}}}, \quad (14)$$

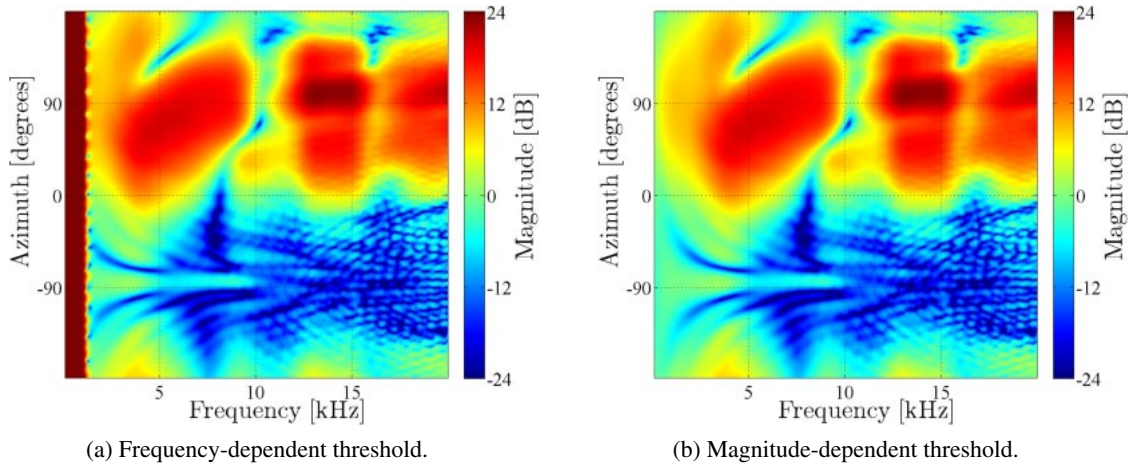
with the overbar denoting the complex conjugate.

### 5.2. Conditions of the Evaluation

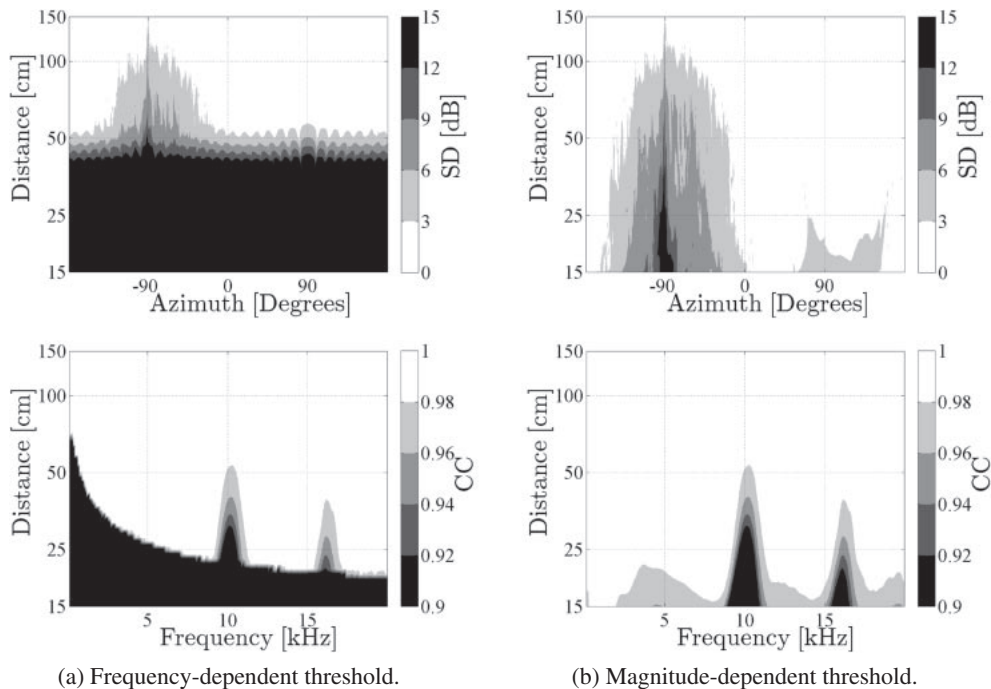
Using the head model described in Fig. 1(a) and the boundary element method (BEM) in [15], we calculated an initial HRTF dataset at a radius  $r_0 = 150$  cm and target datasets at radii  $r$  ranging from 15 to 149 cm in regular spacing intervals of 1 cm. Every dataset was calculated for 360 sound sources equiangularly distributed on the horizontal plane. Such a high angular resolution allowed us to focus the study on the distance effects. We considered frequency bins ranging from  $f_1 = 93.75$  Hz to  $f_2 = 19,875$  Hz, with a sampling frequency of 48 kHz. Because the head model is symmetric, only the left ear was examined.

### 5.3. Results

To illustrate the local performance first, we present examples of HRTF datasets synthesized in interaural and spherical coordinates in Figs. 10 and 11, respectively. Discontinuous results along  $\theta = \pm 90^\circ$  can be observed in Figs. 10(a) and 10(b). This is because Fourier-Legendre transforms are applied along two hemi-circumferences separated by the interaural axis and, unfortunately, differ-



**Fig. 11** HRTF datasets at  $r = 25$  cm synthesized with the proposed distance varying filters in spherical coordinates. Directional bandwidths were limited according to the frequency-dependent threshold in (11) and the proposed magnitude-dependent threshold in (12).



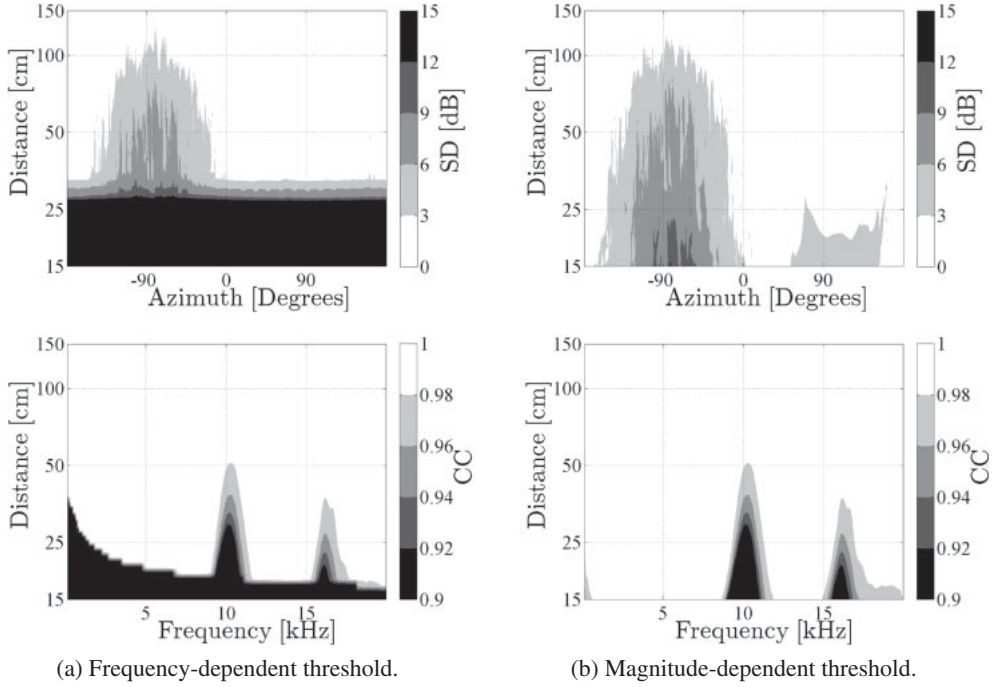
**Fig. 12** Overall accuracies achieved with the distance-varying filters in interaural coordinates. The angular bandwidths of filters were limited according to the frequency-dependent threshold in (11) and proposed magnitude-dependent threshold in (12). Overall accuracies were calculated using the spectral distortion (SD) in (13) and normalized circular correlation (CC) in (14).

ent values are obtained when approaching this axis. Such a drawback is not present in the results obtained with the proposal in spherical coordinates, as can be observed in Figs. 11(a) and 11(b), as Fourier transforms are applied along all azimuth angles at once.

Overall accuracies obtained with the method in interaural coordinates, calculated with (13) and (14), are displayed in Fig. 12. It can be observed that the filters designed with the frequency-dependent threshold yield

poor overall accuracies at low frequencies and close distances. These limitations are overcome when using filters designed with the proposed magnitude-dependent threshold.

Resulting overall accuracies of the proposal in spherical coordinates, calculated with (13) and (14), are displayed in Fig. 13. Notwithstanding the poor overall accuracies at low frequencies and close distances yielded by the frequency-dependent threshold, the overall perform-



**Fig. 13** Overall accuracies achieved with the proposed distance-varying filters in spherical coordinates. The angular bandwidths of filters were limited according to the frequency-dependent threshold in (11) and proposed magnitude-dependent threshold in (12). Overall accuracies were calculated using the spectral distortion (SD) in (13) and normalized circular correlation (CC) in (14).

ance in spherical coordinates under this condition outperforms the method in interaural coordinates. Performance is further improved when using filters designed with the proposed magnitude-dependent threshold nonetheless. In this case, our proposal performed better than the method in interaural coordinates.

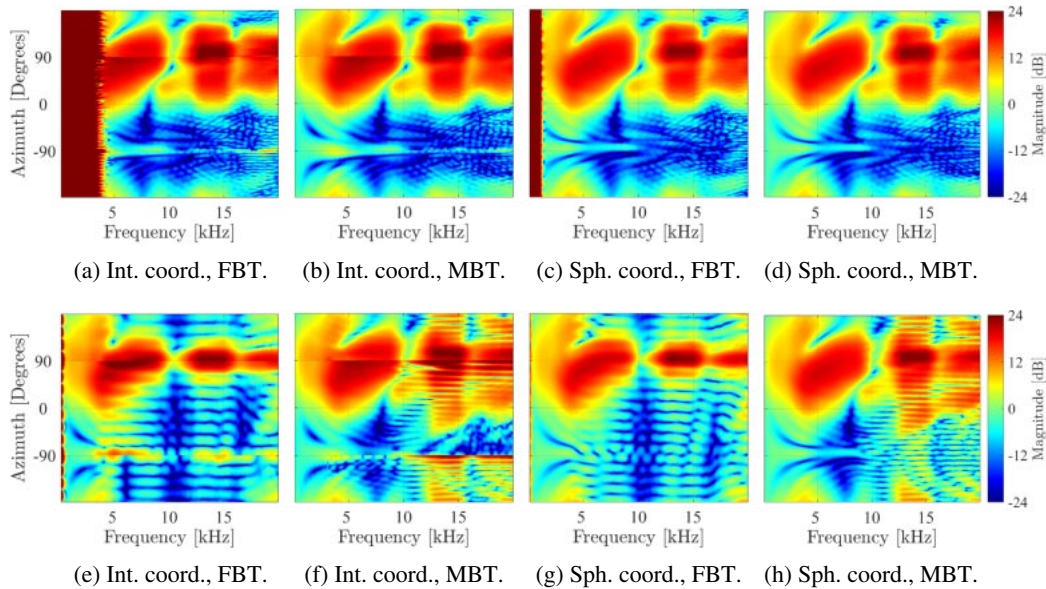
In Fig. 13(b), high SD values are still observed for distance variations around  $\theta = -90^\circ$ . These values would ultimately not have corresponded to significant errors in the synthesized HRTFs, as they result from low levels in the initial and target datasets for the contralateral side. In the same panel, low CC values appear around 10 kHz. The reason behind these inaccuracies is that the initial HRTFs have an energy content that is relatively low in this frequency region, as can be observed in Fig. 1(c). The distance-varying filters rely on a simple model that struggles to handle the head-model features in this frequency region, such as the displacement of salient features towards the interaural axis. In general, the human head model used in this study can be considered to be valid up to an average frequency of 16.6 kHz only. In effect, according to the spatial sampling criterion of four elements per wavelength  $\lambda$  [15], the average cell length  $\frac{\lambda}{4} = 5.1$  mm limited the evaluations up to an average frequency of  $\frac{c}{\lambda} = 16.6$  kHz. Measures of overall accuracy, therefore, provide reliable information for frequencies below this limit only.

#### 5.4. Synthesis Examples from Lower Angular Resolutions

Two initial HRTF datasets at a radius  $r_0 = 150$ , for angular resolutions of  $3^\circ$  (120 sources) and  $10^\circ$  (36 sources), were considered for the synthesis examples shown in Fig. 14. A resolution of  $3^\circ$  allows for modal representations up to  $n = |m| = \frac{120}{2} = 60$ , which correspond to an angular bandwidth still comprising the required initial information for an acceptable synthesis, as can be verified in Figs. 5(a) and 7(a). As a consequence, the results shown in the first row of Fig. 14 only differ slightly from those shown in Figs. 10 and 11. A resolution of  $10^\circ$ , on the other hand, allows for modal representations up to  $n = |m| = \frac{36}{2} = 18$ , and correspondingly an angular bandwidth comprising initial information up to around 8 kHz, as can also be verified in Figs. 5(a) and 7(a). Results shown in the second row of Fig. 14 are therefore expected to be accurate only up to around 8 kHz. Interestingly, it is verified therein that synthesis in spherical coordinates using the MBT outperformed the other methods.

## 6. CONCLUSION

We have herein presented and evaluated a set of horizontal-plane distance-varying filters (HP-DVFs) to synthesize HRTFs at arbitrary distances on the horizontal plane, once a dataset of HRTFs is known on a single circle around the listener.



**Fig. 14** HRTF datasets at  $r = 25$  cm synthesized from initial datasets at  $r_0 = 150$  cm with lower initial angular resolutions:  $3^\circ$  (first row) and  $10^\circ$  (second row). Synthesis was performed using the methods in interaural and spherical coordinates, with the frequency-dependent band-limiting threshold (FBT) in (11) and the magnitude-dependent band-limiting threshold (MBT) in (12).

The HP-DVFs stemmed from the solution of the acoustic wave equation for sound fields that are assumed to be invariant with respect to elevation angles in spherical coordinates. Results were free of the lateral side discontinuities that appear when sound fields are assumed to be invariant along polar angles in interaural coordinates. Moreover, to properly account for HRTF datasets obtained for discrete distributions that limit the angular bandwidth, we introduced a magnitude-dependent band-limiting threshold (MBT) to restrict the action of distance-varying filters to limited angular bandwidths.

We finally evaluated the performance numerically using a model of a human head valid up to 16.6 kHz. Results show that overall accuracies obtained with the proposed HP-DVFs in spherical coordinates outperform the existing distance-varying filters in interaural coordinates. Furthermore, overall accuracies obtained with the proposed MBT outperformed overall accuracies achieved with the traditional frequency-dependent threshold, especially at low frequencies and distances close to the head.

A perceptual evaluation of the HP-DVFs by means of detectability of differences, and localization tests along azimuth and distances, could provide more insight into the validity of the suggested approach.

## ACKNOWLEDGMENTS

The authors wish to thank Makoto Otani for his efforts in developing the BEM solver used to generate the reference HRTF data. This study was supported by a Grant-in-Aid of JSPS for Scientific Research (no. 24240016), the

Foresight Program for “Ultra-realistic acoustic interactive communication on next-generation Internet,” and the Cooperative Research Project Program of RIEC Tohoku University (H24/A14).

## REFERENCES

- [1] H. Møller, “Fundamentals of binaural technology,” *Appl. Acoust.*, **36**, 171–218 (1992).
- [2] M. Morimoto and Y. Ando, “On the simulation of sound localization,” *J. Acoust. Soc. Jpn. (E)*, **1**, 167–174 (1980).
- [3] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, revised ed. (MIT Press, Cambridge, MA, London, 1997).
- [4] M. Morimoto, Y. Ando and Z. Maekawa, “On head-related transfer function in distance perception,” *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 137–138 (1975) (in Japanese).
- [5] D. S. Brungart and W. M. Rabinowitz, “Auditory localization of nearby sources. Head-related transfer functions,” *J. Acoust. Soc. Am.*, **106**, 1465–1479 (1999).
- [6] H.-Y. Kim, Y. Suzuki, S. Takane and T. Sone, “Control of auditory distance perception based on the auditory parallax model,” *Appl. Acoust.*, **62**, 245–270 (2001).
- [7] D. Brungart, “Auditory parallax effects in the HRTF for nearby sources,” *IEEE WASPAA*, pp. 171–174 (1999).
- [8] D. S. Brungart, “Near-field virtual audio displays,” *Presence: Teleop. Virt. Env.*, **11**, 93–106 (2002).
- [9] R. Duraiswami, D. N. Zotkin and N. A. Gumerov, “Interpolation and range extrapolation of HRTFs,” *Proc. IEEE ICASSP*, Vol. 4, pp. 45–48 (2004).
- [10] W. Zhang, T. Abhayapala and R. A. Kennedy, “Insights into head-related transfer function: Spatial dimensionality and continuous representation,” *J. Acoust. Soc. Am.*, **127**, 2347–2357 (2010).
- [11] M. Pollow, K.-V. Nguyen, O. Warusfel, T. Carpentier, M. Müller-Trapet, M. Vorländer and M. Noisternig, “Calculation of head-related transfer functions for arbitrary field points

- using spherical harmonics,” *Acta Acust. united Ac.*, **98**, 72–82 (2012).
- [12] K. Watanabe, Y. Iwaya, Y. Suzuki, S. Takane and S. Sato, “Dataset of head-related transfer functions measured with a circular loudspeaker array,” *Acoust. Sci. & Tech.*, **35**, 159–165 (2014).
- [13] S. Smale, “Mathematical problems for the next century,” *Math. Intell.*, **20**, 7–15 (1998).
- [14] A. Bates, Z. Khalid and R. Kennedy, “Novel sampling scheme on the sphere for head-related transfer function measurements,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, **23**, 1068–1081 (2015).
- [15] 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).
- [16] H. Wallach, “On sound localization,” *J. Acoust. Soc. Am.*, **10**, 270–274 (1939).
- [17] D. N. Zotkin, R. Duraiswami, E. Grassi and N. A. Gumerov, “Fast head-related transfer function measurement via reciprocity,” *J. Acoust. Soc. Am.*, **120**, 2202–2215 (2006).
- [18] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic Press, London, 1999).
- [19] M. Thomas, J. Ahrens and I. Tashev, “A method for converting between cylindrical and spherical harmonic representations of sound fields,” *Proc. IEEE ICASSP*, pp. 4723–4727 (2014).
- [20] J.-M. Lee, J.-W. Choi and Y.-H. Kim, “Reproduction of a higher-order circular harmonic field using a linear array of loudspeakers,” *J. Acoust. Soc. Am.*, **137**, EL227–EL233 (2015).
- [21] G.ENZNER, M. Krawczyk, F.-M. Hoffmann and M. Weinert, “3D reconstruction of HRTF-fields from 1d continuous measurements,” *Proc. IEEE WASPAA*, pp. 157–160 (2011).
- [22] S. Spors and J. Ahrens, “Efficient range extrapolation of head-related impulse responses by wave field synthesis techniques,” *Proc. IEEE ICASSP*, pp. 49–52 (2011).
- [23] T. Ajdler, C. Faller, L. Sbaiz and M. Vetterli, “Sound field analysis along a circle and its applications to HRTF interpolation,” *J. Audio Eng. Soc.*, **56**, 156–175 (2008).
- [24] N. A. Gumerov, A. E. O’Donovan, R. Duraiswami and D. N. Zotkin, “Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation,” *J. Acoust. Soc. Am.*, **127**, 370–386 (2010).
- [25] A. Kan, C. Jin and A. van Schaik, “A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function,” *J. Acoust. Soc. Am.*, **125**, 2233–2242 (2009).
- [26] 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).