

# A new signal processing procedure for stable distance manipulation of circular HRTFs on the horizontal plane \*

☆ SALVADOR César<sup>1</sup>, SAKAMOTO Shuichi<sup>1</sup>, TREVIÑO Jorge<sup>1</sup>, and SUZUKI Yôiti<sup>1</sup>

<sup>1</sup>Res. Inst. Elect. Comm. and Grad. Sch. Info. Sci., Tohoku University

## 1 Introduction

Head-related transfer functions (HRTFs) are a central tool to present 3D sound. They characterize the acoustic path from the position of a sound source to the ears of a listener [1]. HRTFs contain comprehensive auditory cues to perceive the direction of sound sources. Auditory cues for the perception of distance are also present in the HRTFs for sources within ca. 1 m of the listener's head [2–4]. Such distance cues are particularly useful for enhancing the realism when presenting auditory scenes comprised of lateral sounds near the heads of the listeners [3].

HRTF datasets should ideally be obtained for all points in space. Obtaining 3D datasets, however, is a complex and time consuming task. In many applications, it is enough to present sounds on the horizontal plane. A main reason is that the human

auditory system resolve sounds more accurately in this plane [5]. Obtaining HRTFs throughout the horizontal plane, though, is still a demanding task.

In [6], two kinds of ideal distance-varying filters (DVF) were applied to synthesize HRTFs on the full horizontal plane, once an initial circular HRTF dataset for a single distance has been obtained (see Fig. 1). The mathematical details to formulate both kinds of ideal DVFs are described in [7], where continuous distributions of sources on the circle were assumed. Ideal DVFs are classified according to the coordinate system used for their analytical formulation: the interaural coordinates in Fig. 2(a) and the cylindrical coordinates in Fig. 2(b).

Interaural DVFs needs to use two filtering stages because signal processing for the semicircles in front and behind the listener must be treated separately. This leads to 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 [3]. Cylindrical DVFs, on the other hand, considers the horizontal plane as a whole and, hence, there are no discontinuities in the results. However, they produce inaccurate predictions of pressure de-

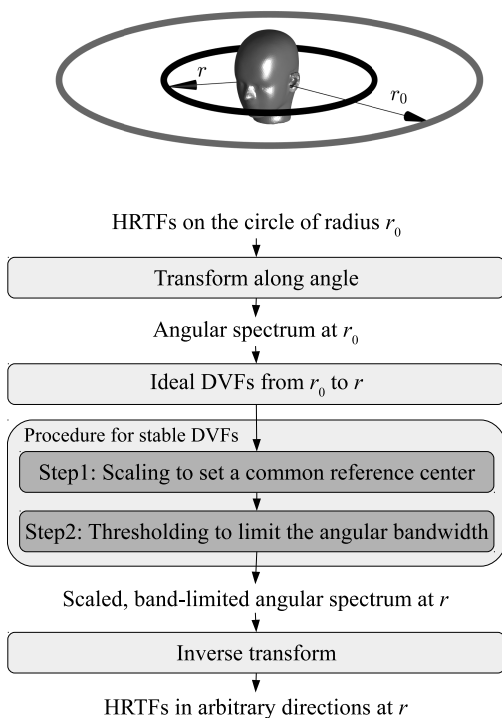
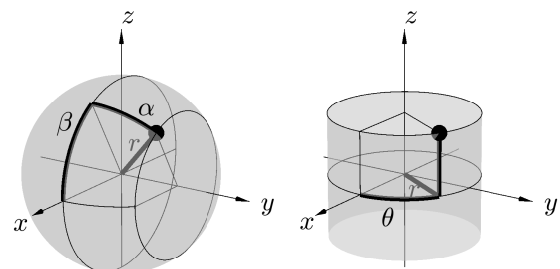


Fig. 1 Overview of synthesis methods.



(a) Interaural.

(b) Cylindrical.

Fig. 2 Coordinate systems.

\*水平面における距離を考慮した頭部伝達関数合成法, サルバドルセザル, 坂本修一, トレビーニョホルヘ, 鈴木陽一 (東北大・通研/院情科)

cay along radial distance. According to [6], the reason behind these inaccuracies is the consideration of infinite line sources instead of point sources when modeling sound propagation.

This article introduces a two-step procedure for stable distance manipulation of HRTFs using ideal DVFs. The first step copes with the problem of inaccurate distance decay observed in cylindrical coordinates. We postulate that these inaccuracies are mainly due to the absence of a common reference center rather than to the fact of considering infinite line sources. Given that a vertical, infinite line source is symmetric with respect to any point in the horizontal plane, the change of pressure along radial distance should differ at most by a scaling factor when compared with the change of pressure due to a point source. When obtaining initial HRTFs, on the other hand, the pressure at the head's center when the listener is not there is used as a reference signal to define a reference center. We propose to link these concepts for deriving factors to scale the ideal DVFs so as to define a reference center that matches the one of the initial HRTFs.

The second step of our procedure is related to the assumption of continuous distributions of sources on the circle that underlies the formulation of ideal DVFs. In practice, DVFs need to operate on initial HRTF datasets obtained for sources at discrete angular positions on a circle around the listener. Properly accounting for angular sampling requires to restrict the action of ideal DVFs to limited angular bandwidths. A band-limiting frequency-dependent threshold has been derived in [9]. This threshold, however, tends to overestimate limits and does not ensure accurate synthesis at distances close to the head. We propose an alternative, new approach for choosing thresholds based on the fact that magnitudes of ideal DVFs increase monotonically with increasing relative distance. We therefore formulate a magnitude-dependent threshold according to the ratio between areas of concentric surfaces traversed by sound at the initial and desired distances.

## 2 Ideal DVFs

We review in this section two HRTF synthesis methods that are based on the ideal DVFs explored in [6]. Both methods rely on finding horizontal-plane solutions to the three-dimensional acoustic

wave equation when assuming sound fields that are invariant along a spatial coordinate [7]. Continuous distributions of sources are therefore assumed throughout this section as well as in section 3.1. A general overview of the methods is shown in Fig. (1).

### 2.1 Interaural DVFs

The first method relies on invariance along polar angle  $\beta \in [-\pi, \pi]$  in the interaural coordinate system shown in Fig. 2(a), where  $r$  denotes radial distance and  $\alpha \in [-\frac{\pi}{2}, \frac{\pi}{2}]$  denotes lateral angle. Interaural 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  $\beta$ .

By  $\mathcal{H}^{\text{int}}$ , we denote the HRTFs for a distribution of sources on the horizontal plane ( $xy$ -plane). The upper index refers to the interaural coordinates. For a given angular frequency  $\omega = 2\pi f$ , the relation between  $\mathcal{H}^{\text{int}}$  at two different distances  $r$  and  $r_0$  is formulated according to [6, 7]:

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

Here, the term on the second line is a Fourier-Legendre transform along  $\alpha_0$ . The orthogonal basis functions are the Legendre polynomials  $P_n$  of order  $n$ , which are the angular part of the general solution to the wave equation. The term above in brackets defines the interaural DVFs in terms of the spherical Hankel functions  $h_n$  of order  $n$ , which are in turn the radial part of the solution. The sum of Legendre polynomials on  $\alpha$  represents an inverse Fourier-Legendre transform. Note that lateral angles  $\alpha_0$  are in  $[-\frac{\pi}{2}, \frac{\pi}{2}]$  and, therefore, Fourier-Legendre transforms need to act on separate datasets defined over two semicircles in front and behind the listener.

### 2.2 Cylindrical DVFs

The second method relies on invariance along height  $z$  in the cylindrical coordinate system shown in Fig. 2(b), where  $r$  denotes radial distance and  $\theta \in [-\pi, \pi]$  azimuthal angle. Assuming height-invariance is equivalent to considering infinite line sources of sound.

By  $\mathcal{H}^{\text{cyl}}$ , we denote the HRTFs for a distribution of sources on the horizontal plane. The upper index refers to the cylindrical coordinates. The relation between  $\mathcal{H}^{\text{int}}$  at two different distances  $r$  and  $r_0$  is formulated according to [6, 7]:

$$\mathcal{H}^{\text{cyl}}(r, \theta, \omega) = \sum_{m=-\infty}^{\infty} \left( \frac{1}{2\pi} \right)^{\frac{1}{2}} \exp(jm\theta) \underbrace{\left[ \frac{H_m(\frac{\omega}{c}r)}{H_m(\frac{\omega}{c}r_0)} \right]}_{\text{Cylindrical DVFs}} \times \underbrace{\left( \frac{1}{2\pi} \right)^{\frac{1}{2}} \int_{-\pi}^{\pi} \mathcal{H}^{\text{cyl}}(r_0, \theta_0, \omega) \exp(-jm\theta_0) d\theta_0}_{\text{Fourier transform along azimuthal angle}}. \quad (2)$$

Here, the term in the second line clearly represents a Fourier transform along  $\theta_0$ , where complex exponentials are in turn the angular part of the solution to the wave equation. The term in brackets defines the cylindrical DVFs in terms of the Hankel functions  $H_m$  of order  $m$ , which are the radial part of the solution. The sum of complex exponentials on  $\theta$  describes an inverse Fourier transform. Note that azimuthal angles  $\theta_0$  are in  $[-\pi, \pi]$  and, hence, Fourier transforms operate on complete circular boundaries.

### 3 Procedure for stable DVFs

In this section, we detail the two-step procedure to implement the ideal DVFs in (1) and (2).

#### 3.1 Scaling to set a reference center

The first step is mainly intended to cope with the problem of inaccurate distance decay observed in cylindrical coordinates. The principle used here, however, also applies to the method in interaural coordinates. In general, we propose to establish a common reference center for both the ideal DVFs and the initial HRTF dataset by normalizing the components of the ideal DVFs according to the radiation of an infinite-line (or point) source whose wave fronts converge to the reference center in empty space. This is equivalent to dividing each cylindrical (or spherical) Hankel function by its far-field asymptotic form [8]. Replacing the Hankel functions in (1) and (2) by their normalized forms results in ideal DVFs scaled by the following factor:

$$s = \begin{cases} \frac{r}{r_0} & \text{interaural coordinates,} \\ \left( \frac{r}{r_0} \right)^{\frac{1}{2}} & \text{cylindrical coordinates.} \end{cases} \quad (3)$$

Ideal DVFs multiplied by this scaling factor are thus generalized to operate in both the far and near fields.

### 3.2 Thresholding for angular band-limiting

The second step is intended to restrict the angular spectra of ideal DVFs so that they operate only on a limited angular bandwidth. This is necessary in practice to account for datasets of HRTFs obtained for discrete distributions of sound sources.

Limited angular bandwidths are modeled by truncating the sums along  $n$  in (1) or  $m$  in (2) up to a maximum order. In [9], a frequency-dependent threshold to decide maximum orders was deduced by using the large-order asymptotic form of Hankel functions. Bandwidth limits yielded by this threshold, however, are overestimated and present the risk of emphasizing high-order components that typically have a low signal-to-noise ratio [7].

We have observed that ideal DVFs have magnitude responses that are monotonically increasing functions of order and relative distance. Based on this observation, we propose an alternative approach for restricting the angular bandwidth by clipping the magnitude responses of ideal DVFs according to the ratio between the surface area of two hypothetical concentric spheres (or cylinders) at  $r_0$  and  $r$ .

In summary, we formulate both the scaling operation and the magnitude-dependent band-limiting threshold using the following expression:

$$\hat{D}(r, r_0, \omega) = \begin{cases} s \cdot D(r, r_0, \omega) & \text{if } |D| \leq s^{-2}, \\ 0 & \text{else.} \end{cases} \quad (4)$$

Here,  $\hat{D}$  defines our scaled band-limited distance-varying filters (SB-DVFs),  $D$  denotes the cylindrical or interaural DVFs respectively defined in (1) or (2), and  $s$  is the scaling factor defined in (3).

### 4 Evaluation of overall accuracy

In this section, we numerically compare the performances of the procedure described in [9] and our procedure defined in (4). We use both procedures to implement the ideal DVFs in (1) and (2). Using a human head model and the algorithm in [11], we have 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 intervals of 1 cm.

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 based on the spectral distortion (SD).

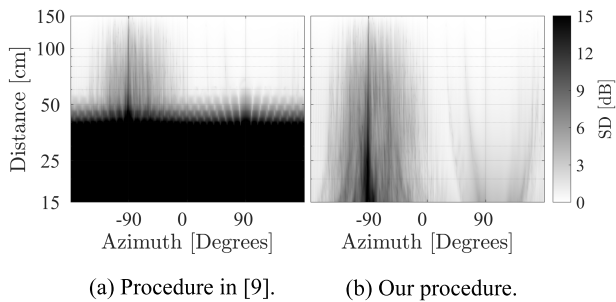


Fig. 3 SD for interaural DVFs.

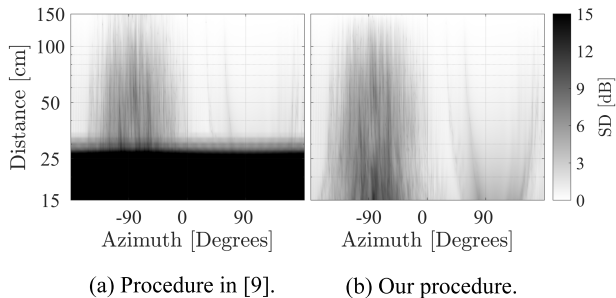


Fig. 4 SD for cylindrical DVFs.

This metric has been shown to be suitable for predicting audible differences between measured and synthesized HRTFs [10]. The SD in decibels is defined according to [10]:

$$SD(\theta) = \left[ \frac{1}{f_2 - f_1} \int_{f_1}^{f_2} \left[ 20 \log_{10} \left| \frac{\hat{\mathcal{H}}(\theta, f)}{\mathcal{H}(\theta, f)} \right| \right]^2 df \right]^{\frac{1}{2}}. \quad (5)$$

It can be observed in Figs. 3(a) and 4(a) that DVFs designed with the procedure in [9] yield poor overall accuracies at close distances. These limitations are overcome when designing DVFs according to our procedure in (4), as can be observed in Figs. 3(b) and 4(b). In Fig. 4(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 initial and target datasets obtained for sources on the opposite side of the evaluated ear, which present low energy values since they are acoustically shadowed by the head.

## 5 Conclusion

We have herein presented a new procedure for the realization of ideal distance-varying filters (DVFs), which are used to synthesize head-related transfer functions (HRTFs) on the full horizontal plane, once

an initial circular HRTF dataset for a single distance has been obtained. Our procedure comprises two steps: i) scaling to set a reference center for the ideal DVFs so that this center matches the one used to obtain the initial dataset of HRTFs, and ii) magnitude-dependent thresholding for limiting the angular bandwidth of ideal DVFs, which is required in practice to account for discrete distributions of sources on the initial circle. Our procedure therefore defines scaled band-limited distance-varying filters (SB-DVFs). We performed numerical experiments using a model of a human head. We found that, when applying our two-step procedure to implement ideal DVFs, the overall accuracies outperformed the ones obtained when using conventional procedures.

**Acknowledgements** A part of this study was supported by a JSPS Grant-in-Aid for Scientific Research (no. 24240016) and the Foresight Program for “Ultra-realistic acoustic interactive communication on next-generation Internet”. The authors thank Makoto Otani for admitting us to use his BEM solver to calculate HRTF datasets, and Junfeng Li for fruitful discussion.

## References

- [1] J. Blauert, *Spatial Hearing*, rev. ed., 1997.
- [2] M. Morimoto *et al.*, *Proc. Congress Acoust. Soc. Jpn.*, 1975.
- [3] D. Brungart, *et al.*, *J. Acoust. Soc. Am.*, Vol. 106, No. 3, pp. 1465–1479, Sep. 1999.
- [4] H.-Y. Kim, *et al.* *Appl. Acoust.*, Vol. 62, No. 3, pp. 145–270, Mar. 2001.
- [5] P. Minnaar, *J. Audio Eng. Soc.*, Vol. 53, N0. 10, pp. 919–929, 2005.
- [6] G. Enzner, *Proc. IEEE WASPAA*, pp. 157–160, Sep. 2011.
- [7] E.G. Williams, *Fourier Acoustics*, 1999.
- [8] W. Zhang, *et al.*, *J. Acoust. Soc. Am.*, Vol. 127, No. 4, pp. 2347–2357, Sep. 2010.
- [9] N. Gumerov, *Fast multipole methods for the Helmholtz equation in three dimensions*, 2004.
- [10] K.-S. Lee and S.-P. Lee, *IEEE. Trans. Audio, Speech, Language Process.*, Vol. 19, No. 6, pp. 1780–1790, Aug. 2011.
- [11] M. Otani, S. Ise, *J. Acoust. Soc. Am.*, Vol. 119, pp. 2589–2598, May 2006.