Validity of distance-varying filters for individual HRTFs on the horizontal plane *

☆ SALVADOR César[1], SAKAMOTO Shuichi[1], TREVIÑO Jorge[1], and SUZUKI Yôiti[1]

[1]Res. Inst. Elect. Comm. and Grad. Sch. Info. Sci., Tohoku University

1 Introduction

Head-related transfer functions (HRTFs) are a central tool to present three-dimensional (3D) sound. They are linear filters that describe the transmission of sound from a point in space to the eardrums of a listener [1]. The HRTFs are highly individual because they depend on the external anatomical shapes of the listeners. Individual HRTFs and their interaural difference contain auditory cues to perceive the direction [1] and distance [2] of sound sources.

Individual HRTFs should ideally be obtained for the set of all points in space around the listener. Construction of 3D datasets, however, is a complex and time consuming task. In many applications, it is enough to present sounds on a horizontal plane at the height of the ears. A main reason is that the human auditory system resolve sounds more accurately in this plane [3]. Construction of HRTFs throughout the horizontal plane, though, is still a demanding task.

Distance-varying filters (DVFs) [4–6] are acoustic propagators that aim to synthesize the HRTFs on the full horizontal plane. Ideally, they require an initial HRTF dataset obtained for a continuous, circular distribution of sound sources at a single distance (see Fig. 1). DVFs operate in transform domains provided by generalized Fourier transforms that uses orthonormal basis functions defined on the continuous circle.

Recently, the authors have proposed a magnitude-dependent band-limiting threshold (MBT) to design band-limited DVFs that can be applied to initial HRTF datasets obtained for a discrete, circular distribution of sources [5, 6]. The MBT was evaluated on two types of DVFs, respectively formulated in interaural and cylindrical coordinates, and using an average model of a human head. This paper presents an extended evaluation including two individual head models. DVFs are detailed below.

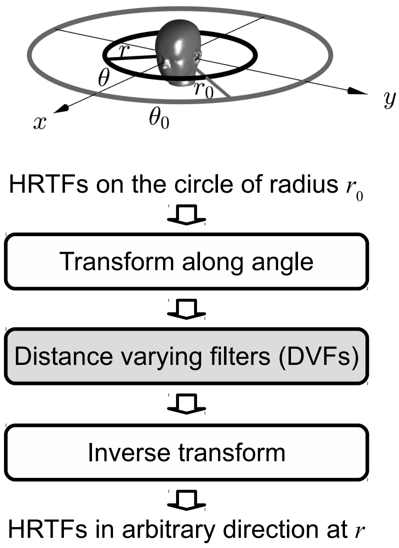


Fig. 1 Synthesis of HRTFs based on DVFs.

2 Distance-varying filters

The formulation of DVFs rely on deriving horizontal-plane solutions to the 3D acoustic wave equation [6, 7]. Deriving horizontal-plane solutions requires to assume invariance along the spatial coordinate outside the horizontal plane. The angular part of the solutions defines the orthonormal basis functions. The radial part of the solutions defines the DVFs. Two types of DVFs are considered in this study. They are respectively formulated in interaural and cylindrical coordinates (see Fig. 2).

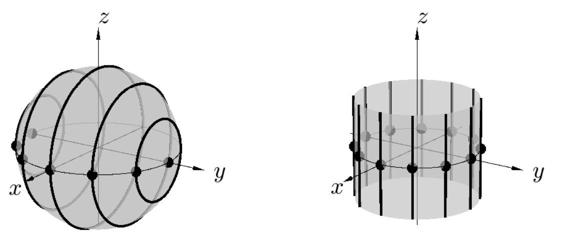
In interaural coordinates, invariance along the polar angle is assumed. This assumption is equivalent to considering a semicircular array of data points that is replicated along vertical circles. Interaural DVFs need to use two filtering stages because signal processing for the semicircles in front and behind the listener must be treated separately. This leads to results on the front and back regions that take

---

* 水平面内における HRTF 距離可変操作法の評価,
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/大学院情報科学)

Table 1 Functions and variables used in (1) and (2).

Coordinates	$\Phi_\ell(\theta)$	$\mathcal{D}_\ell(r_0, r; f)$	$d\mu(\theta_0)$	Θ_1	Θ_2	L_1	L_2
Interaural	$\left(\ell+\frac{1}{2}\right)^{\frac{1}{2}} P_\ell(\sin\theta)$	$\dfrac{r^{-\frac{1}{2}} H_{\ell+\frac{1}{2}}(\frac{2\pi f}{c} r)}{r_0^{-\frac{1}{2}} H_{\ell+\frac{1}{2}}(\frac{2\pi f}{c} r_0)}$	$\cos\theta_0 d\theta_0$	$-\frac{\pi}{2}$	$\frac{\pi}{2}$	0	$\frac{L}{2}-1$
Cylindrical	$\left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \exp(j\ell\theta)$	$\dfrac{H_\ell(\frac{2\pi f}{c} r)}{H_\ell(\frac{2\pi f}{c} r_0)}$	$d\theta_0$	$-\pi$	π	$-\frac{L}{2}$	$\frac{L}{2}-1$



(a) Interaural coordinates. Semicircular array of data points ($x \geq 0$) replicated along vertical circles.

(b) Cylindrical coordinates. Circular array of data points replicated along vertical lines.

Fig. 2 Underlying assumptions to formulate DVFs.

different values along the axis connecting the ears, namely the interaural axis (y-axis).

In cylindrical coordinates, invariance along height is assumed. This assumption is equivalent to considering a circular array of data points that is replicated along infinite vertical lines. Cylindrical DVFs consider the horizontal plane as a whole and, hence, there are no discontinuities in the results.

In practice, DVFs need to operate on initial HRTF datasets obtained for discrete, circular distributions of sources. Properly accounting for a finite number of sources at discrete angles requires to limit the number of orthonormal basis functions involved in the generalized Fourier transforms. This is equivalent to restrict the action of DVFs to a limited angular bandwidth in the transform domains. A frequency-dependent band-limiting threshold [8] is available for the design of band-limited DVFs. This threshold, however, tends to overestimate limits for lower frequencies and closer distances.

Recently, the authors have introduced a magnitude-dependent band-limiting threshold (MBT) [5, 6] to deal with the problem of overestimation at lower frequencies ans closer distances. A concise formulation of HRTF synthesis using band-limited DVFs designed with the MBT is presented below.

By $\mathcal{H}^{\text{init}}$, we denote the initial HRTF dataset for a circular distribution of sources at a distance r_0. By $\mathcal{H}^{\text{synth}}$, we denote the synthesized HRTFs for arbitrary target angle θ and distance $r < r_0$. For a given frequency f, the synthesis of $\mathcal{H}^{\text{synth}}$ from $\mathcal{H}^{\text{init}}$ is defined by the following expression:

$$\mathcal{H}^{\text{synth}}(r, \theta; f) = \sum_{\ell=L_1}^{L_2} \Phi_\ell(\theta) \hat{\mathcal{D}}_\ell(r_0, r; f) \times \int_{\theta_0=\Theta_1}^{\Theta_2} \overline{\Phi_\ell(\theta_0)} \mathcal{H}^{\text{init}}(r_0, \theta_0; f) d\mu(\theta_0). \tag{1}$$

The integral in (1) defines a generalized Fourier transform involving the orthonormal basis functions Φ on azimuthal angle θ_0. The function μ denotes the mathematical measure used for integration; Θ_1 and Θ_2 are the limits of integration. The band-limited transform-domain DVFs that variate the distance from r_0 to r are denoted by $\hat{\mathcal{D}}_\ell$. Finally, the sum over ℓ represents an inverse transform. The number of terms in the sum, delimited by L_1 and L_2, is decided in accordance with the number of data points in the initial HRTF dataset.

Band-limited DVFs are calculated as follows:

$$\hat{\mathcal{D}}_\ell(r, r_0; f) = \begin{cases} s \cdot \mathcal{D}_\ell(r, r_0; f) & \text{if } |\mathcal{D}_\ell| \leq s^{-2}, \\ 0 & \text{else,} \end{cases} \tag{2}$$

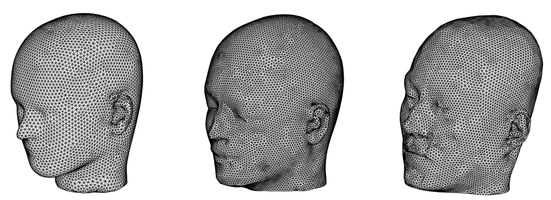
where $\mathcal{D}_\ell$ denotes band-unlimited DVFs, and the scale factor s that defines the MBT, s^{-2}, is calculated according to the following expression:

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

Table 1 details the functions and variables used to define (1) and (2) when HRTF synthesis is performed in interaural and cylindrical coordinates. In this table, P_ℓ is the Legendre polynomial of order ℓ, H_ℓ is the Hankel function of order ℓ, and L is the number of data points in the initial HRTF dataset.

Note that integration in interaural coordinates is performed along a semi circle and, therefore, two separate calculation stages are required for the semicircles in front and behind the listener. This limitation is overcome in cylindrical coordinates because integration is performed along a full circle.

The models above described have been evaluated by the authors using a dummyhead model [5,6]. The next two sections present extended evaluations including two models of real human listeners.



(a) Head A. (b) Head B. (c) Head C.

Fig. 3 Head models.

3 Evaluation method

Evaluations were performed by comparing synthesized datasets $\mathcal{H}^{\text{synth}}$ and target datasets $\mathcal{H}^{\text{target}}$. The head models in Fig. 3, and the algorithm in [9], were used to calculate circular datasets $\mathcal{H}^{\text{init}}$ at a radius $r_0 = 150$ cm, and circular datasets $\mathcal{H}^{\text{target}}$ at radii r ranging from 15 cm to 149 cm in regular intervals of 1 cm. The model A corresponds to the head of the SAMRAI dummy head, whereas the models B and C corresponds to the heads of two real individuals. The datasets $\mathcal{H}^{\text{synth}}$ at target radii r ranging from 15 cm to 149 cm were synthesized from $\mathcal{H}^{\text{init}}$ using (1). All datasets were calculated for the left ears and on for sources on full circles using an angular interval of 1°.

Overall accuracy along frequency was calculated based on the spectral distortion (SD):

$$\text{SD}(\theta)=\left[\frac{1}{f_2-f_1}\int_{f_1}^{f_2}\left[20\log_{10}\left|\frac{\mathcal{H}^{\text{synth}}(\theta,f)}{\mathcal{H}^{\text{target}}(\theta,f)}\right|\right]^2 df\right]^{\frac{1}{2}}, \quad (4)$$

whereas overall accuracy along azimuth was calculated based on the circular correlation (CC):

$$\text{CC}(f) = \frac{\int_{-\pi}^{\pi}\mathcal{H}^{\text{synth}}(\theta,f)\overline{\mathcal{H}^{\text{target}}(\theta,f)}d\theta}{\int_{-\pi}^{\pi}\left|\mathcal{H}^{\text{synth}}(\theta,f)\right|^2 d\theta \times \int_{-\pi}^{\pi}\left|\mathcal{H}^{\text{target}}(\theta,f)\right|^2 d\theta}. \quad (5)$$

4 Results

The spatial resolutions of the head models in Fig. 3 allow for reliable analyses of the results up to around 16 kHz.

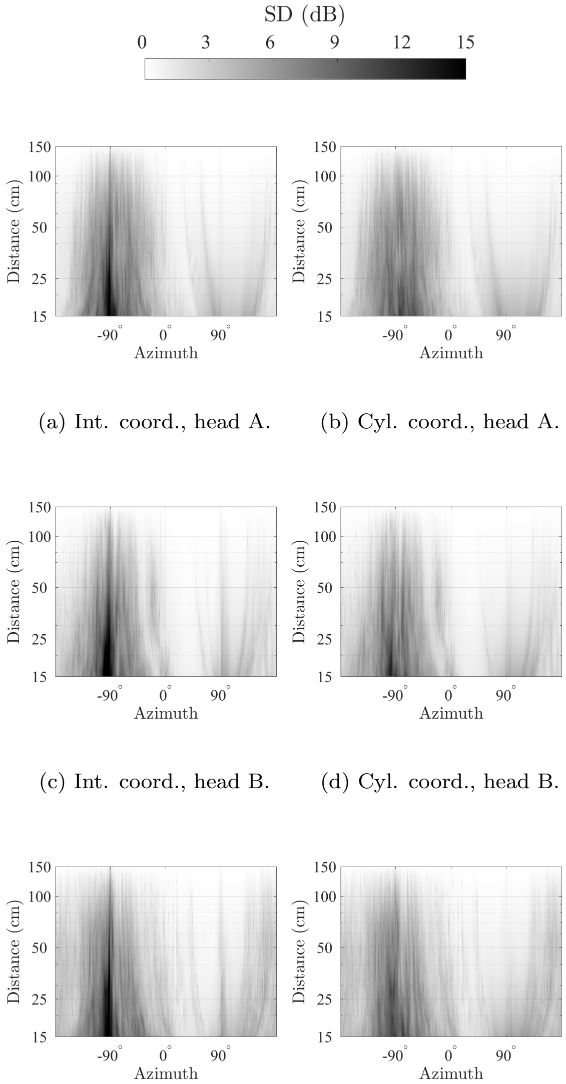
Figure 4 shows the SD results. In general, accuracy decreases with decreasing distance. However, higher accuracies are observed for sources facing the left ear around 90° when compared to the opposite side around −90°. On the opposite, moreover, synthesis in cylindrical coordinates outperforms synthesis in interaural coordinates. The reasons behind this are the discontinuous results along the axis connecting the ears yielded by the method in interaural coordinates. Small differences among the head models can further be observed in other angular regions.

Figure 5 shows the CC results. Accuracy generally decreases with decreasing distance. Individual differences can be observed at closer distances. For example, inaccuracies tend to concentrate below 50 cm and around 10 kHz for the head model A, below 25cm and around 10 kHz for the model head B, and below 25 cm and around 9 kHz for the head C. Small differences between the interaural and cylindrical methods can also be observed at lower frequencies.

5 Conclusion

Evaluations of band-limited DVFs for HRTF synthesis at near distances were presented. Head models of real listeners were considered. Performance of band-limited DVFs along frequency and angle generally decreases with decreasing distance. The analysis of the results suggests that more efforts are required for the accurate synthesis of individual features at closer distances. Efforts for an accurate synthesis should be concentrated on the contralateral side and around 10 kHz. Perceptual tests could provide more insight into the validity of the results.

Acknowledgements A part of this study was supported by the JSPS Grant-in-Aid for Scientific Research no. JP24240016 and no. JP16H01736, and by the A3 Foresight Program for "Ultra-realistic acoustic interactive communication on next-generation Internet". The authors thank M. Otani for admitting us to use his BEM solver [9].

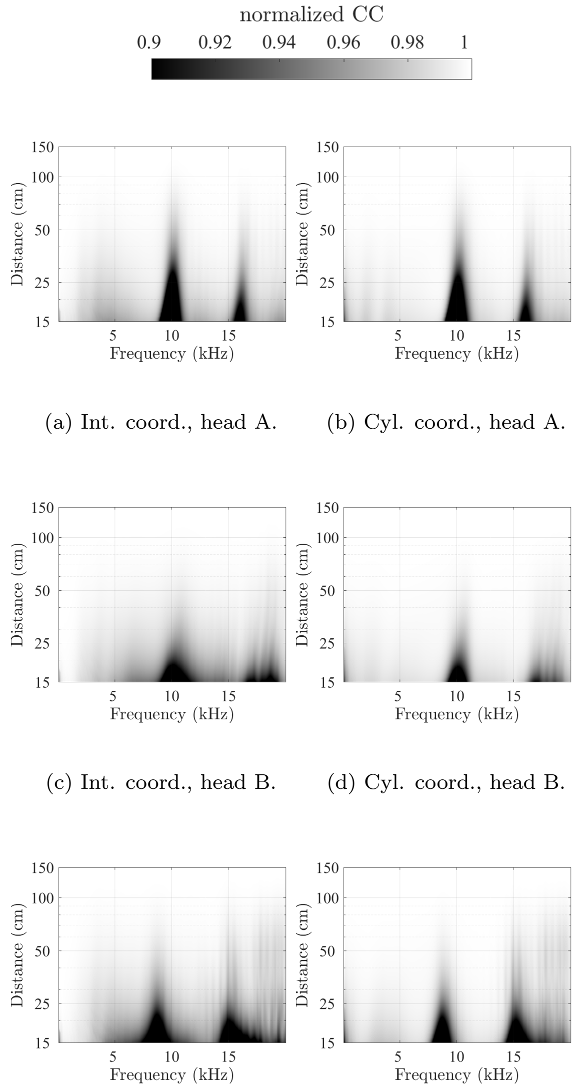


(a) Int. coord., head A. (b) Cyl. coord., head A.

(c) Int. coord., head B. (d) Cyl. coord., head B.

(e) Int. coord., head C. (f) Cyl. coord., head C.

Fig. 4 Spectral distortion (SD) for the left ear.



(a) Int. coord., head A. (b) Cyl. coord., head A.

(c) Int. coord., head B. (d) Cyl. coord., head B.

(e) Int. coord., head C. (f) Cyl. coord., head C.

Fig. 5 Circular correlation (CC) for the left ear.

References

[1] J. Blauert, *Spatial hearing: The psychophysics of human sound localization*, revised ed. MIT Press, 1997.

[2] M. Morimoto *et al., Proc. Congress Acoust. Soc. Jpn.*, 1975.

[3] P. Minnaar *et al.*, *J. Audio Eng. Soc.*, vol. 53, no. 10, pp. 919–929, 2005.

[4] G. Enzner *et al.*, in *Proc. IEEE WASPAA*, Oct. 2011, pp. 157–160.

[5] C. D. Salvador *et al.*, in *Proc. Spring Meeting Acoust. Soc. Jpn.* Yokohama, Japan: Acoustical Society of Japan, Mar. 2016.

[6] C. D. Salvador *et al.*, *Acoust. Sci. Technol.*, vol. 38, no. 1, pp. 1–13, 2017.

[7] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography.* London, UK: Academic Press, 1999.

[8] N. A. Gumerov and R. Duraiswami, *Fast multipole methods for the Helmholtz equation in three dimensions.* Maryland, USA: Elsevier, 2004.

[9] M. Otani and S. Ise, *J. Acoust. Soc. Am.*, vol. 119, no. 5, pp. 2589–2598, May 2006.