2015 International Conference on Intelligent Information Hiding and Multimedia Signal Processing

A compact representation of the head-related transfer function inspired by the wavelet transform on the sphere

Jorge TREVINO*, Shichao HU*, Cesar SALVADOR*, Shuichi SAKAMOTO*, Junfeng LI†, Yôiti SUZUKI*
*Graduate School of Information Sciences and Research Institute of Electrical Communication
Tohoku University, Sendai, Japan 980–8577
Email: {jorge@ais., shichao@ais., salvador@ais., saka@ais., yoh@}riec.tohoku.ac.jp
†Institute of Acoustics
Chinese Academy of Sciences, Beijing, China 100190
Email: lijunfeng@hccl.ioa.ac.cn

***Abstract*—The head-related transfer function (HRTF) characterizes the propagation of sound from its source to the listener's ears. It is commonly used in the research of spatial sound and its applications. The HRTF takes different values depending on the source's position relative to the listener and its frequency. Storing its values directly results in considerably large data sets. Previous research to encode the HRTF using harmonic functions cannot handle local features efficiently. This paper advances a new analysis method to represent raw HRTF data using local functions of the azimuth and elevation angles. The proposal treats features of different scales separately, in a manner similar to the wavelet transform. This allows us to compress HRTF data considerably while accurately preserving its features at all scales. The proposal yields better accuracy than harmonic analysis methods for the finer spatial patterns; this holds even when our method uses more aggressive compression.**

***Keywords*-head-related transfer function; wavelet transform; functions on the sphere; compression;**

I. Introduction

Humans can perceive spatial information from the sound that reaches their ears [1]. A natural example of this is our capacity to estimate the position of sound sources by listening to them. To accomplish this, humans mainly rely on cues related to the acoustic propagation path between the source and their ears. For example, sounds that reach the listener from one side are louder at the ear facing the source; it is harder for sound to travel to the opposite ear since the head acts as an obstacle. In addition to interaural differences like this one, our sense of hearing also takes advantage of the interaction of sound waves with the listener's torso, head and pinnae. The specifics of these interactions are related to the listener's anthropometry; they are also strongly related to the relative position of the sound source and its frequency. Analysis of the cues that these interactions imbue in the sound before it reaches the listener's ears is critical in the study of spatial hearing and the development of 3D sound applications.

The sound localization cues are modeled by the head-related transfer function (HRTF), which characterizes the effects that the listener's body has on the sounds they hear [1], [2]. The HRTF is defined as the sound pressure produced by a single sound source at the listener's ears divided by the pressure it would generate at the position of the head's center when the listener is not there. HRTFs vary among individuals and must be measured or otherwise calculated for each person. High resolution HRTF data sets must store the results of these measurements or calculations for sound sources positioned at many different angles relative to the listener's front and for frequencies comprising the full hearing range of 20 Hz to 20 kHz [3]–[5]. The large amount of data makes working with these data sets difficult. An alternative representation of HRTF data is needed to facilitate its analysis and use.

Previous research efforts attempt to model HRTF data sets using harmonic functions [6]–[8]. In particular, it is convenient to treat the HRTF data for each frequency seperately as a function of the azimuth and elevation angles. In this context, the HRTF can be represented as a weighted sum of spherical harmonic functions; this is the equivalent of the Fourier transform for functions on the sphere. This approach can reduce the amount of information needed to store the coarse features of the HRTF. However, harmonic functions have poor spatial resolution since all of them take large values at most directions. This makes them inadequate to encode the finer details and localized features present in the HRTF.

In this paper, we advance a new method to encode HRTFs at individual frequencies using a collection of local functions. The target HRTF data is expressed as the weighted sum of functions that take significant values only over a small range of directions. The proposal resembles the wavelet transform [9] for functions of the azimuth and elevation angles. Our results show that the original data can be recovered with good accuracy using a small number of these functions. The compression rate and accuracy of the proposal exceeds that of conventional methods based on the spherical harmonics when the target HRTF contains localized features.

978-1-5090-0188-0/15 $31.00 © 2015 IEEE
DOI 10.1109/IIH-MSP.2015.108

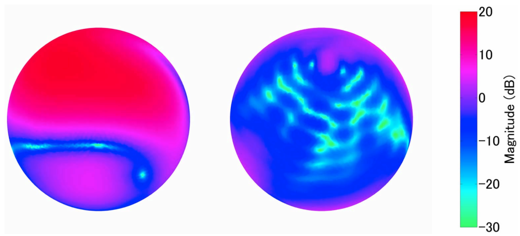

Figure 1. An example of HRTF data. The figure shows the left and right views of the data calculated for the left ear assuming sound sources of 7.4 kHz. A total of 21,162 sound source directions are covered by this data set.

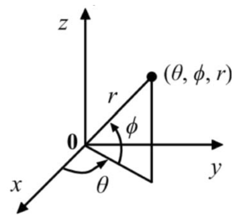

Figure 2. The spherical coordinate system used in this paper. The listener's front lies in the positive direction of the x-axis.

II. Spherical harmonic representation of the HRTF

The head-related transfer function takes values depending on the direction of the sound source relative to the listener's front and its frequency. Its analysis is often simplified by considering either single directions or single frequencies. In this paper, we focus on the latter approach and treat the HRTF as a collection of functions of the azimuth and elevation angles. An example HRTF at 7.4 kHz is shown in Fig. 1.

It is convenient to use a spherical coordinate system when working with functions of the directions. This paper will assume the system shown in Fig. 2. Functions of the azimuth and elevation angles (θ, φ) are commonly decomposed into their harmonic components using the equivalent of the Fourier transform on the sphere. This is also known as the spherical harmonic expansion and is given by the following equation [10]:

$$\mathrm{H}(\theta,\varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathrm{B}_{nm} \mathrm{Y}_{nm}(\theta,\varphi). \quad (1)$$

Here, the function $\mathrm{H}(\theta,\varphi)$ is represented as the weighted sum of spherical harmonic functions $\mathrm{Y}_{nm}(\theta,\varphi)$ with coefficients B_{nm}. The coefficients fully characterize the original function; however, there is an infinite number of them. The spherical harmonic function of order n and degree m can be calculated using the following formula:

$$\mathrm{Y}_{nm}(\theta,\varphi) = \begin{cases} \mathrm{A}_{n|m|}\mathrm{P}_{n|m|}(\sin\varphi)\sin(|m|\theta) & m<0 \\ \frac{1}{\sqrt{2}}\mathrm{A}_{n0}\mathrm{P}_{n0}(\sin\varphi) & m=0 \\ \mathrm{A}_{nm}\mathrm{P}_{nm}(\sin\varphi)\cos(m\theta) & m>0 \end{cases} \quad (2)$$

Figure 3. Reconstruction of HRTF data from a finite number of spherical harmonic expansion coefficients. The original data corresponds to 21,162 directions and a frequency of 7.4 kHz. The order 10 reconstruction uses 121 coefficients, at order 20 it comprises 441 coefficients, for order 25 this rises to 676 coefficients, and at order 30 a total of 961 coefficients are used.

Here, P_{nm} is a Legendre polynomial, while A_{nm} is the normalization factor:

$$\mathrm{A}_{nm} = (-1)^m \sqrt{\frac{2n+1}{2\pi} \cdot \frac{(n-m)!}{(n+m)!}}. \quad (3)$$

The expansion coefficients B_{nm} in Eq. (1) can be calculated as follows [10]:

$$\mathrm{B}_{nm} = \int_{-\pi}^{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{H}(\theta,\varphi)\mathrm{Y}_{nm}(\theta,\varphi)\cos\varphi d\varphi d\theta. \quad (4)$$

Equation (4) can be used to represent the HRTF data for a particular frequency as a set of expansion coefficients. Meanwhile, Eq. (1) can be used to recover the original data from these coefficients. Practical applications, however, must limit the decomposition to a maximum order N, thus considering only the first $(\mathrm{N}+1)^2$ coefficients. Furthermore, the integral in Eq. (4) must be approximated as a numerical quadrature since HRTF data is available only for a finite number of sound source directions.

Figure 3 shows the result of encoding the HRTF data of Fig. 1 using Eq. (4) and decoding it using Eq. (1). The results for a maximum order of 10 (121 coefficients) show that the coarse features of the HRTF are preserved despite the drastic reduction in the amount of data (less than one percent of the original). However, localized structures, such as the deeps in the stripe on the left side, or the patterns seen on the right side can only be recovered when considering very high orders. Increasing the order from 20 (441 coefficients) to 25 (676 coefficients) does little to improve the reconstruction accuracy. The more intrincate patterns in the target data only begin to appear when a maximum order of 30 (961 coefficients) is considered.

III. A new representation of the HRTF using local functions

Representing localized data using the spherical harmonic functions requires very high orders and, consequently, a

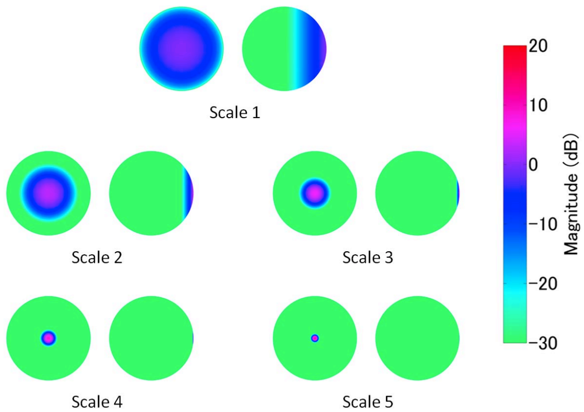

Figure 4. Examples of local functions of different size or scale seen from their centers and sides. The functions are defined as the product of a two-dimensional gaussian and an cosine.

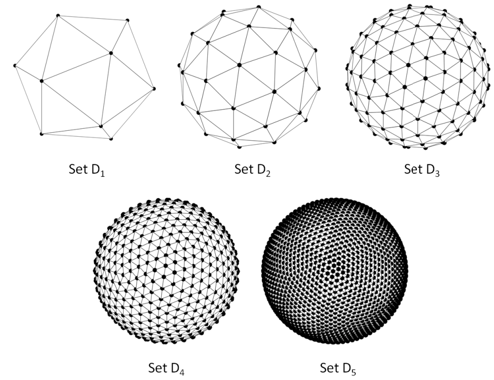

Figure 5. Samplings of the directions used to orientate the analysis functions of scale levels up to five. They follow a dyadic pattern, with each edge of the polyhedra being divided into two edges to generate the next scale level.

large number of coefficients. The reason behind this is the low spatial resolution of the harmonic functions. All spherical harmonics take large values at most directions. The localized features found in HRTF data sets may be encoded more efficiently using local functions.

The problem of characterizing local features in time signals and images is typically handled by using the wavelet transform [9]. The strict formulation of a discrete wavelet transform, needed to represent sampled data, is difficult when the domain of analysis is the sphere. This research introduces a simpler alternative by defining a set of local functions and looking for their linear combination that best fits the target data in the least-squares sense.

A. *Local analysis functions*

The local functions used in our proposal arise from the same ideas behind the Gabor and Morlet wavelets. We apply a gaussian window to the data in order to select a small region in the analysis domain. Harmonic analysis is then carried out on the windowed data. This procedure results in the following analysis functions:

$$\mathrm{W}^{k\sigma}(\theta,\varphi) = e^{-\frac{\lambda_0^2(\theta,\varphi)}{2\sigma^2}} \cos\left[k\lambda_0(\theta,\varphi)\right]. \tag{5}$$

The parameters k and σ set the frequency of the harmonic part and the width of the gaussian window, respectively. The function $\lambda_0(\theta,\varphi)$ stands for the angular distance between the front (zero azimuth and elevation) and the point on the sphere with azimuth θ and elevation φ.

The functions in Eq. (5) must be scaled to match the size of the local features in the target data without altering their shape or energy content. They must also be displaced so that the analysis covers the entire sphere. For given parameters k and σ, the scaled and displaced functions are:

$$\mathrm{W}^{k\sigma}_{s\Omega}(\theta,\varphi) = \sqrt{\mathrm{R}_s}\, e^{-\frac{\mathrm{R}_s^2\lambda_\Omega^2(\theta,\varphi)}{2\sigma^2}} \cos\left[k\mathrm{R}_s\lambda_\Omega(\theta,\varphi)\right]. \tag{6}$$

The subindex s denotes an integer scale level. The resulting functions are more localized for larger values of the scaling factor R_s. The functions $\lambda_\Omega(\theta,\varphi)$ denote the angular distance between the center position Ω and some other point on the sphere. Analysis functions for scales one through five are shown in Fig. 4.

The HRTF data can be represented as the weighted sum of the functions in Eq. (6) for all scales and displacements:

$$\mathrm{H}(\theta,\varphi) = \sum_{s=1}^{\infty}\sum_{\Omega\in\mathrm{D}_s} \mathrm{C}_{s\Omega}\mathrm{W}^{k\sigma}_{s\Omega}(\theta,\varphi). \tag{7}$$

The sets of directions D_s should define regular samplings of the sphere with an average separation between samples that decreases as the scale level s increases. Equation. (7) is the equivalent of Eq. (1); however, there is no exact formula to calculate the coefficients $\mathrm{C}_{s\Omega}$. Nevertheless, appropriate coefficients can be estimated using the least-squares method to invert the linear system that results from truncating Eq. (7) to a maximum scaling factor.

IV. Results and Discussion

The local functions proposed in the previous section can characterize HRTF data. Applying the proposal requires tuning several parameters related to the size of the analysis functions, their frequency content and their positioning over the sphere. The optimal choice of parameters depends on the spatial and frequency resolution required by the target data.

To evaluate the performance of the proposed method, we use it to represent the HRTF data shown in Fig. 1. The oscillation rate k is fixed as 0.5, while the width parameter for the spatial window, σ, is set to a value of 2. The scaling factor is chosen to grow dyadically, according to the formula $\mathrm{R}_s = 2^s$. The analysis functions are positioned according to the sets shown in Fig. 5. These sets provide almost-regular samplings of all directions and their growth closely resembles that of the scaling factor.

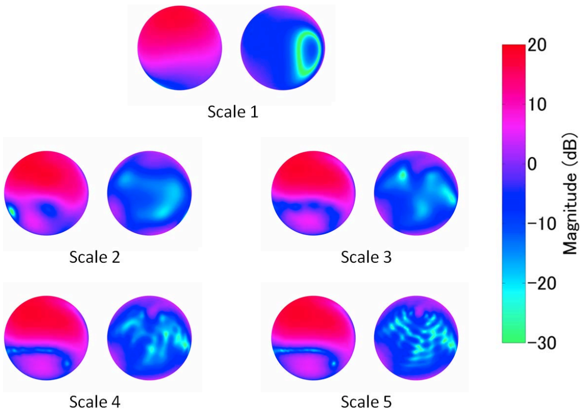

Figure 6. Results for HRTF reconstruction from the expansion coefficients obtained by applying the proposed local analysis functions. Scale levels one to five comprise 12, 54, 216, 858 and 3,420 coefficients in ascending order.

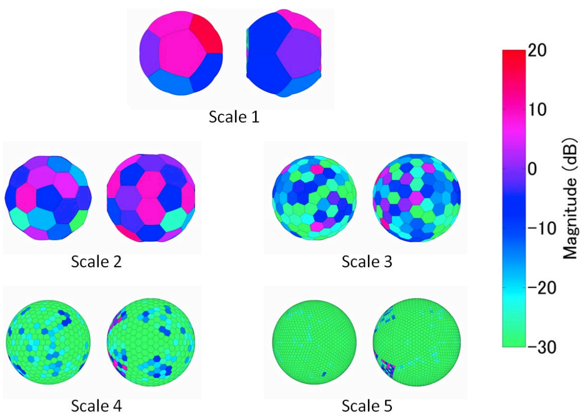

Figure 7. Values for the expansion coefficients corresponding to HRTF data analyzed up to the fifth scale level. Each coefficient is shown as a polygon centered in the direction of its corresponding analysis function.

Figure 6 shows the HRTF data reconstructed from the expansion coefficients obtained using the proposed functions at different scale levels. Meanwhile, Fig. 7 shows the value of the expansion coefficients. Most of the coefficients at the higher scale levels (four and five) are small and have little impact in the reconstruction of the HRTF data. They can be discarded to achieve a more compact representation without a significant loss of accuracy.

Figure 8 shows the result of reconstructing the HRTF data from a limited number of expansion coefficients taken from the five-scale-levels decomposition. The figure shows the results of discarding the smallest coefficients and preserving only as many as used by the spherical harmonic approach at maximum orders of 10 (121), 20 (441), 25 (676) and 30 (961). Comparing these results with Fig. 3 shows that the proposed functions can preserve the finer details along the stripe on the left side and the patterns on the right side better than the spherical harmonics, even when fewer coefficients are used.

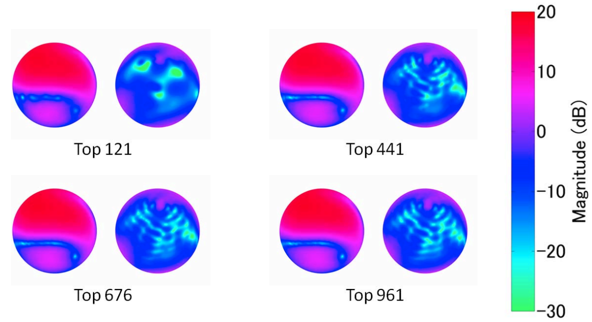

Figure 8. Reconstruction of HRTF data after applying the proposed expansion up to the fifth scale level and then discarding the smallest coefficients. The figure shows the results of keeping the same number of coefficients as used in the spherical harmonic expansion of orders 10 (121 coefficients), 20 (441 coefficients), 25 (676 coefficients) and 30 (961 coefficients).

Acknowledgments

Parts of this research were supported by Grant-in-Aid of JSPS for Scientific Research (no. A24240016) to SY and the A3 Foresight Program for ”Ultra-realistic acoustic interactive communication on next-generation Internet.” The authors wish to acknowledge Makoto Otani’s efforts in developing the program used to calculate the HRTF data.

References

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

[2] H. Møller, M.F. Sørensen, D. Hammershøi and C.B. Jensen, *Head-Related Transfer Functions of Human Subjects*, J. Audio Eng. Soc, Vol. 43 (5), pp. 300–321, 1995.

[3] M. Morimoto, Y. Ando and Z. Maekawa, *On head-related transfer function in distance perception*, Proc. Congr. Acoust. Soc. Jap., pp. 137–138, 1975.

[4] 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. Technol., Vol. 35 (3), pp. 159–165, 2014.

[5] M. Otani and S. Ise, *Fast calculation system specialized for head-related transfer function based on boundary element method*, J. Acoust. Soc. Am., Vol. 119 (5), pp. 2589–2598, 2006.

[6] M.J. Evans, *Analyzing head-related transfer functions measurements using surface spherical harmonics*, J. Acoust. Soc. Am., Vol. 104 (4), pp 2400–2411, 1998.

[7] R. Duraiswami, D.N. Zotkin and N.A. Gumerov, *Interpolation and range extrapolation of HRTFs*, Proc. IEEE ICASSP 2004., pp. 45–48, 2004.

[8] 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., Vol. 98, pp. 72–82, 2012.

[9] I. Daubechies, et al. *Ten lectures on wavelets*, Society for Industrial and Applied Mathematics, 1992.

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