INTERPOLATION OF HEAD-RELATED TRANSFER FUNCTIONS BY SPATIAL LINEAR PREDICTION

Ryouichi Nishimura, Hiroaki Kato, and Naomi Inoue

National Institute of Information and Communications Technology (NICT) / Advanced Telecommunications Research Institute International (ATR) 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan

ABSTRACT

Head-related transfer functions (HRTFs) are essential for creating a virtual sound source when sound waves are transmitted through headphones. The measurement of HRTFs is a complicated and time-consuming task. Therefore, the interpolation of HRTFs is crucial for virtual auditory display systems in which both listeners and sound objects are likely to move in a virtual auditory space. In this study, HRTFs are measured with a high spatial resolution in order to develop an effective interpolation method. An analysis of HRTFs indicates that the functions exhibit periodicity in amplitude along the azimuthal angle. The optimum filter coefficients required to interpolate HRTFs from several functions measured along multiple azimuthal directions are derived. Computer simulations indicate that in comparison to conventional methods, the proposed method yields less estimation error in the interpolation of HRTFs. Listening tests indicate that the proposed method can provide better perception of virtual sound.

Index Terms— Audio systems, Acoustic signal processing, Head-Related Transfer Function, Interpolation, Spatial linear prediction

1. INTRODUCTION

Head-related transfer functions (HRTFs), or head-related impulse responses (HRIRs), which are representations of HRTFs in the time domain, are essential factors for virtual auditory display systems to add spatial information to sound when it is presented through headphones or earphones. However, the measurement of HRTFs along all potential directions is a complex and considerably time-consuming task. Moreover, HRTFs differ among individuals, and the measurement of HRTFs necessitates the listener to remain still for a long time. Therefore, interpolation is required. Assuming HRTFs are measured densely for all directions beforehand, interpolation will be useful in implementing the HRTFs into an interactive virtual auditory display system because of the memory and computational limitations.

Many interpolation methods for the measurement of HRTFs have been developed in the last decade. T. Djelani et al. proposed the separation of the initial delay from the HRIR and the interpolation of the initial delay and the remaining part, separately [1]. M. A. Blommer and G. H. Wakefield represented HRIRs in a form of a pole-zero model [2]. Watanabe et al. proposed the interpolation of poles and zeros on the basis of this model [3]. These methods employed only HRIRs of the neighboring measured points to calculate the HRTF within the measured points. J. Chen et al. first executed the Karuhnen-Lóeve expansion of HRIRs for all measured points and interpolated coefficients for reconstruction [4]. This method employed only the coefficients of the neighboring points in the interpolation, although the Karuhnen-Lóeve expansion was applied for HRIRs of all the measured points. T. Ajdler et al. investigated the effects of the microphone position, which is usually not the center of the circle of positions of the speaker in the HRTF measurement, by using a circularly deployed microphone array. They found that the Fourier coefficient of each frequency bin in the HRTF varies according to a sinc function as a speaker moves circularly around the head due to the difference in the time of arrival of the signals [5]. On the basis of their observations, they proposed the interpolation of HRTFs by using a sinc function.

In this study, the periodicity of HRTFs along the azimuthal angle is investigated by measuring HRTFs with a high spatial resolution. Subsequently, the optimum filter coefficients for spatial linear prediction of HRTFs are derived. The performance of the proposed method is evaluated through both computer simulations and listening tests.

2. HEAD-RELATED TRANSFER FUNCTION

2.1. Measurement

HRIRs were measured for every 1°-azimuth in order to analyze their spatial frequency characteristics. The measurement was conducted in an anechoic chamber in which an HRTF measurement system was mounted beneath its ceiling, as schematically depicted in Fig. 1. The arched arm of the system could rotate horizontally and the support of the

^aThis research was partly supported by Grant-in-Aid for Scientific Research by Japan Society for the Promotion of Science (JSPS).



Fig. 1. Setup for the HRTF measurement



Fig. 2. Power spectrum of measured HRTFs for the left ear of HATS (Gray scale indicates the relative level in dB)

speaker could move along the arm. The azimuth and elevation of the speaker were automatically controlled by using a personal computer. An optimized Aoki's time-stretched pulse (OATSP) signal of length 16384 points was used at a sampling rate of 48 kHz and with a resolution of 16 bits. The HRIRs were measured using microphones (Sound Professionals SP-TFB-2) placed at the entrance of both ears of a head and torso simulator (HATS: B&K 4128) and at a sampling rate of 48 kHz and with a resolution of 16 bits. The OATSP signal was played 40 times for each speaker position and then it was averaged to improve the signal-to-noise ratio. The power spectrum of the measured HRTFs for the left ear is shown in Fig. 2. The azimuthal angle of 0° corresponds to the speaker position



Fig. 3. Spatial frequency characteristics of HRTFs

of the frontal direction of the HATS, and it increases clockwise. A similar but symmetric pattern is obtained for the right ear. It must be noted that data for the points between 161° and 164° are missing because the arched arm could not move along these angles due to mechanical limitations.

2.2. Analysis

The power spectrum of the measured HRTFs shown in Fig. 2 were further analyzed through discrete Fourier transformation (DFT) along the azimuthal direction, that is, the column direction in Fig. 2, in order to investigate the spatial frequency characteristics of the HRTFs. The spatial frequency characteristics of the HRTFs are depicted in Fig. 3. The abscissa corresponds to each frequency bin of the HRTFs. The ordinate corresponds to the periodic cycle of the power spectrum level along the azimuthal direction. For example, a cycle period of five means that the wavelength approaching the periodicity corresponds to five degrees. Before applying DFT, the part with the missing data was removed in order to avoid blurring of the periodicity of the HRTFs due to the discontinuity in the data. It is noteworthy that the periodicity, which is shown as white portions in the figure, is remarkable; it is approximately up to 15 kHz, and the cycle period is approximately 6-7°. Considering the Nyquist theorem, an azimuthal resolution of less than 3° is required to measure HRTFs without being affected by the aliasing due to this periodicity.

This periodicity along the azimuthal direction can be attributed to the interference of diffracted sound around the head, because it markedly appears at the contralateral side of the speaker. Therefore, it is expected that a similar periodicity can be observed for heads with similar sizes. In fact, these results hold true for the HRTFs of other people, which are measured using a dummy head created on the basis of precise calculations of the physical features of the people from their MRI data.

$$\boldsymbol{d}_{i}(\omega) \equiv [c(\omega,\theta_{0}+i) \ c(\omega,\theta_{1}+i) \ \dots \ c(\omega,\theta_{K-1}+i)]^{\mathrm{T}} \qquad i=1,2,\dots,k-1$$
(1)

3. FORMULATION

It can be expected that the development of a better interpolation method is possible by using HRTFs measured at multiple positions because the periodicity of HRTFs exists along the azimuthal direction, as described in the previous section. Although the periodicity was determined for the power spectrum of HRTFs, in this section, it is formulated for the complex spectrum of HRTFs.

Consider the interpolation of HRTF along a certain direction in a horizontal plane among the HRTFs measured at a constant interval of k° in the plane. For simplicity, let k denote a factor of 360 so that the number of measured points should become K = 360/k. The problem is to find the optimum filter coefficients w, which satisfy simultaneous equations expressed in Eqs. (1)-(3), when HRTFs with a spatial resolution of 1° are required. In Eqs. (1) and (2), $\theta_i(j) =$ $(0, 1, \ldots, K - 1)$ denotes the angles for which HRTFs are assumed to be measured, and $c(\omega, \theta)$ represents a complex Fourier coefficient for frequency ω of the HRTF for azimuthal angle θ . The transpose and complex conjugate transpose of a matrix is denoted by ^T and ^{*}, respectively. Assume that for symmetry the taplength of w, denoted by N in Eq. (2), is an even number. The optimum solution of Eq. (3) in terms of the minimum mean squared error is represented as

$$\boldsymbol{w}_i(\omega) = (\boldsymbol{C}^*(\omega)\boldsymbol{C}(\omega))^{-1}\boldsymbol{C}^*(\omega)\boldsymbol{d}_i(\omega). \tag{4}$$

Although HRTF measurement with a high spatial resolution is required once in this method, it can be carried out by using a dummy head instead of a real human head. The obtained coefficients are used to interpolate the HRTFs of other people, whose HRTFs are measured more coarsely.

From Eq. (4), it is inferred that the interpolation of the proposed method is executed on each frequency bin independently. C. I. Cheng and G. H. Wakefield suggested the interpolation of HRTFs for each frequency bin [6]. However, their method employed only the magnitude of HRTFs, and therefore special measures was adopted for dealing with their phase components. On the other hand, the proposed method represented the HRTFs in a single complex expression and moreover integrated the HRTFs of multiple positions on the basis of spatial linear prediction.

4. EVALUATION

4.1. Computer simulation

The optimum coefficients were derived assuming that the HRTFs for every 1° azimuth were to be interpolated from those measured for every 5°. The HRTFs of HATS (B&K4128) investigated in Sec. 2.1 were used in the derivation. The obtained coefficients were then applied to HRTFs of other dummy heads, which were created on the basis of the MRI data of real people, to obtain interpolated HRTFs.

The performance of the proposed method was compared with those of two conventional methods:

1) Linear interpolation

The interpolated HRIR for angle θ from those measured at θ_a and θ_b is obtained as

$$h(n,\theta) = \alpha h(n,\theta_a) + (1-\alpha)h(n,\theta_b) \qquad a \neq b, \quad (5)$$

where

$$\alpha = \frac{\theta_b - \theta}{\theta_b - \theta_a}.$$
 (6)

2) Linear interpolation with arrival time correction [7] The initial delay is removed in order to align the HRIRs. The removed initial delay and the remaining part are then interpolated separately. Signals are first up-sampled by eight and down-sampled after interpolation to mitigate the problem of discrete time delay.

The objective performance was evaluated in terms of the mean squared error (MSE) defined as

$$MSE = \frac{1}{2} \sum_{j=L,R} \frac{1}{M} \sum_{\theta} 10 \log \frac{||h_j(n,\theta) - \hat{h_j}(n,\theta)||^2}{||h_j(n,\theta)||^2},$$
(7)

where M (= 284 in the present test) denotes the number of directions for which HRIRs are interpolated, $h(n, \theta)$ represents a measured impulse response for angle θ , the subscript indicates the ear measured for, and $\hat{h}(n, \theta)$ represents an interpolated impulse response.

The results are shown in Fig. 4. The horizontal positions of each condition are intentionally shifted for visibility. It is apparent that the proposed method is superior to conventional methods when the taplength of the filter is greater than two.



Fig. 4. Estimation error of interpolated HRTFs

4.2. Listening test

Unofficial listening tests were conducted to study whether artificial unnatural noise is perceived when a virtual sound source moves because the proposed method consider the continuity neither along frequency nor along the azimuthal angle. Filters with taplengths of two and six employed in the proposed method were compared with the simple linear interpolation and the linear interpolation techniques with arrival time correction. A piece of jazz music was convolved with HRTFs along all the directions (360°); the HRTFs were obtained by interpolating HRTFs of every 15° by means of one of the abovementioned four interpolation methods. Continuous variation in the HRTFs with a short time period could generate an image of sound source rotating around the head. Listeners listened to the stimuli through headphones and were asked to judge the direction along which the sound image was rotating and to rate the convergent of the sound on a scale of 1 to 5. Five listeners participated in the tests and answered twice for each condition. The sound image was rotated clockwise and anti-clockwise equally but in a random order. The correct answer rate for each method is shown as a box plot in Fig. 5. The boxes have lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the boxes. The mean scales for the rating of convergence are shown in the figure.

It is apparent that the linear interpolation (Method 1) and the linear interpolation with arrival time correction (Method 2) are inferior to the proposed method (Method 3 and 4).

5. CONCLUSION

HRTF interpolation based on spatial linear prediction is proposed. The method produces good results in both objective and subjective tests as compared with basic conventional methods. HRIRs for directions other than the horizontal plane can be obtained by a simple extension of the proposed



Fig. 5. Correct answer rate of judging the rotating direction of a virtual sound source and mean rating of the sound convergence. (1: Linear, 2: Linear with correction, 3: Proposed with taplength of two 4: Proposed with taplength of six)

method; this should be investigated further in the near future. Moreover, further interpolation of w_j would be required to estimate the HRTFs with higher spatial resolutions than that used for the derivation of w_j .

6. REFERENCES

- [1] T. Djelani, C. Pörschmann, J. Sahrhage, and J. Blauert, "An interactive virtual-environment generator for psychoacoustic research II: Collection of head-related impulse responses and evaluation of auditory localization," *Acta Acustica united with Acustica*, vol. 86, no. 6, pp. 1046–1053, November 2000.
- [2] M. A. Blommer and G. H. Wakefield, "Pole-zero approximation for head-related transfer functions using a logarithmic error criterion," *IEEE Trans. on Speech and Audio Processing*, vol. 5, no. 3, pp. 278–287, May 1997.
- [3] K. Watanabe, S. Takane, and Y. Suzuki, "A novel interpolation method of HRTFs based on the common-acoustical-pole and zero model," *Acta Acustica united with Acustica*, vol. 91, no. 6, pp. 958–966, November/December 2005.
- [4] J. Chen, B. D. V. Veen, and K. E. Hecox, "A spatial feature extraction and regularization model for the head-related transfer function," *J. Acoust. Soc. Am.*, vol. 97, no. 1, pp. 439–452, January 1995.
- [5] 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.*, vol. 56, no. 3, pp. 156–175, March 2008.
- [6] C. I. Cheng and G. H. Wakefield, "Spatial frequency response surfaces: An alternative visualization tool for head-related transfer functions (HRTFs)," in *ICASSP*'99, March 1999, vol. 2, pp. 961–964.
- [7] M. Matsumoto, S. Yamanaka, M. Tohyama, and H. Nomura, "Effect of arrival time correction on the accuracy of binaural impulse response interpolation," *J. Audio Eng. Soc.*, vol. 52, no. 1/2, pp. 56–61, January/February 2004.