Acoust. Sci. & Tech. **24**, 5 (2003)

PAPER

A method of interpolating binaural impulse responses for moving sound images

Mitsuo Matsumoto[1,*], Mikio Tohyama[1,†] and Hirofumi Yanagawa[2,‡]

[1]*Kogakuin University,*
2665–1, Nakano-machi, Hachioji, 192–0015 Japan
[2]*Chiba Institute of Technology,*
2–17–1, Tsudanuma, Narashino, 275–0016 Japan

(*Received 16 December 2002, Accepted for publication 3 June 2003*)

Abstract: Previously introduced method of interpolating binaural implse responses and algorithm to simulate a moving sound image were evaluated objectively. The method interpolates the responses taking into account the arrival time difference due to changes in the direction of a moving sound source. For the angular interval of 15°, the average of the SDR values of our method, 23 dB was larger than that of the simple method, 9.9 dB. The variances of the *SDR* values showed our method interpolated the responses more independently of the azimuths of the sound source than the simple method. The responses interpolated using our method changed smoothly as the source direction changed. We have evaluated the algorithm by comparing a moving sound image simulated using the algorithm with an actual moving sound image recorded using a rotating dummy head and with a moving sound image simulated using a conventional method. The spectrogram of the binaural signal of the moving sound image, and no ripples were seen.

Keywords: Moving sound image, Interpolation, Binaural impulse response, HRTF, Time-variant convolution

PACS number: 43.66.Qp [DOI: 10.1250/ast.24.284]

1. INTRODUCTION

Virtual-reality systems require realistic simulation of moving sound images. While a moving sound image can be simulated using a series of binaural impulse responses, measuring a number of such responses is time consuming. A method is needed to interpolate these responses, and various methods have been proposed from the static point of view [1–4]. Kistler *et al.* used principle components analysis to model head-related transfer function (HRTF) magnitude functions. They introduced a minimum-phase function for the phase characteristics of HRTF. Furthermore, they approximated HRTFs using a linear combination of five basic magnitude functions and minimum-phase ones. This "approximation" can be considered "interpolation." A linear combination of two minimum-phase functions does not always result in a minimum-phase one.

For moving sound images, the distance between the sound source and listener changes as the source moves. The direction of the sound in relation to the listener also changes. We thus need to consider arrival time when interpolating the binaural impulse responses. Nishino *et al.* also used a magnitude function and a minimum-phase function to model HRTFs [3]. They interpolated an HRTF using two modeled HRTFs. They replaced the initial time delay of the HRTF with the modeled delay that had the highest covariance with the actually measured HRTF. They did not present, however, a basis for linear interpolation of the delays. Hartung, Braasch, and Sterbing used linear interpolation using three HRTFs on a hemisphere in the frequency domain [5]. Though the phase differences of two of the three HRTFs were taken into account, the three HRTFs interpolated the HRTF linearly. Cross fading, a well-known method for moving a sound image, also interpolates the responses linearly. Our group proposed a method for interpolating binaural impulse responses considering arrival time difference [6]. Furthermore, we proposed an algorithm of time-variant convolution for moving a sound image [6].

*Current affiliation: Chiba Institute of Technology, e-mail: matsu@yana.net.it-chiba.ac.jp
†Current affiliation: Global Information and Telecommunication Institute, Waseda University, e-mail: m_tohyama@waseda.jp
‡e-mail: yanagawa@net.it-chiba.ac.jp

We have evaluated our interpolation method in comparison with a conventional linear method numerically. We also evaluated our algorithm for moving a sound image by comparing a moving sound image simulated using the algorithm with an actual sound image recorded using a rotating dummy head and with a moving sound image simulated using the conventional cross-fading technique. After reviewing the conventional method and our algorithm for moving a sound image, we will describe our objective evaluation.

2. METHODS FOR INTERPOLATING BINAURAL IMPULSE RESPONSES

2.1. Arrival Time Difference

Figure 1 illustrates how the arrival time changes as a sound source moves. The sound source moves along an arc, starting directly in front of the listener and continuing 90° to the right. The arc is centered on the listener. Since the left ear is off center, the distance between the left ear and the sound source changes as the source moves. This change in distance is reflected in the arrival time difference between the two binaural impulse responses. In this paper, we define a binaural impulse response as an impulse response between a sound source and an ear of a listener.

2.2. Experimental Set-up

We measured the binaural impulse responses of a B&K 4100 dummy head mounted on a turntable in an anechoic chamber by using time-stretched pulses. The experimental set-up is shown in Fig. 2. The dummy head rotated on the axis that went through the center of the head. The loudspeaker, i.e., the sound source, appeared to move on an arc centered on the dummy head. The frequency characteristics of the loudspeaker were not equalized given that the loudspeaker appeared to move around the listener. Thirty-seven responses at a sampling frequency of 44.1 kHz from 0° (front) to 180° (back) via right 5° steps were measured. Each response was 512-samples long. Binaural responses were used after 8-times over-sampling to detect the arrival difference more accurately.

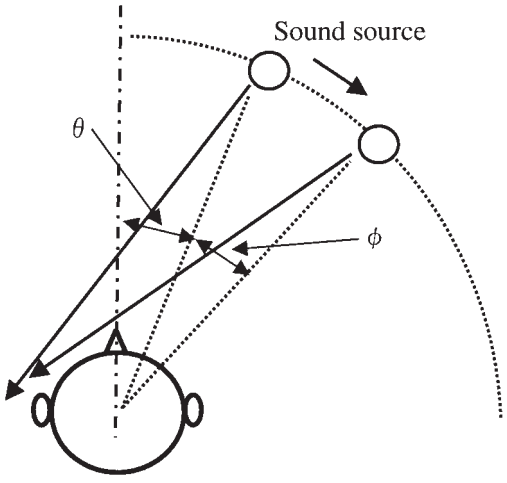

Fig. 1 Configuration of listener and moving sound source.

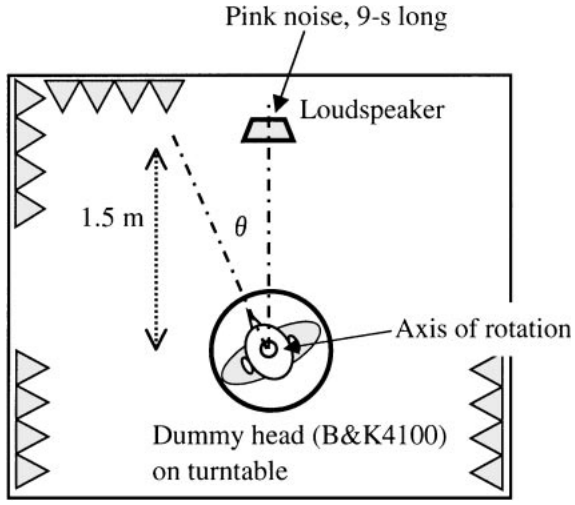

Fig. 2 Experimental set-up.

2.3. Arrival Time Difference Detection

We detected the arrival time difference at the left ear between the two binaural impulse responses by using the Jeffress model [7]. Figure 3 shows the difference between the arrival times when the source was at the azimuth, θ, and at $\theta + \phi$.

The x and y axes denote the azimuth of the sound source and the arrival time difference (number of samples after 8-times over-sampling), respectively. For example, the arrival time difference between the responses at $\theta = 0°$ and $\theta + \phi = 5°$ ($\phi = 5°$) was 8 samples long (22.7 microsecond), shown by the leftmost "$*$" in the figure. The time difference between the responses at $\theta = 0°$ and $\theta + \phi = 10°$ ($\phi = 10°$) was 16 samples long (45.3 microsecond), shown by the leftmost "+". The time difference between the responses at $\theta = 0°$ and $\theta + \phi = 15°$ ($\phi = 15°$) was 24 samples long (68.0 microsecond), shown by the leftmost "$\circ$". These results show that the arrival time difference between binaural impulse responses θ and $\theta + \phi$ is proportional to ϕ. Therefore, we can linearly interpolate the arrival time difference of binaural impulse responses by using the arrival time difference of the two responses.

Figure 4 shows the binaural impulse responses for azimuths of 0°, 5°, 10°, and 15° for the left ear after the arrival times were aligned. The four responses were quite close, except for their amplitudes. These results suggest that linear interpolation of the amplitude of the responses after arrival time alignment by using the difference in the azimuths, ϕ, may be allowed.

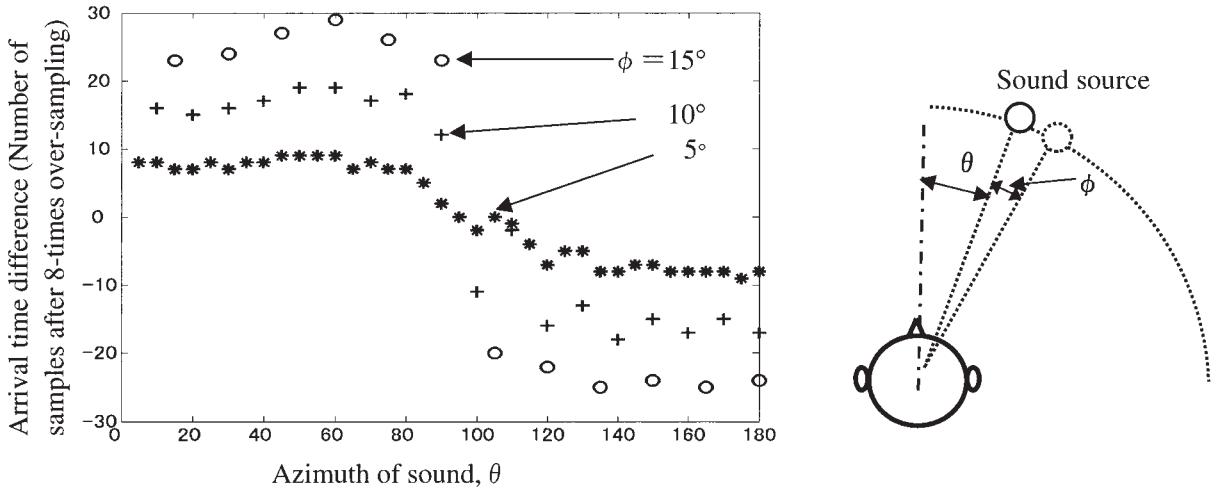

Fig. 3 Arrival-time difference of responses.

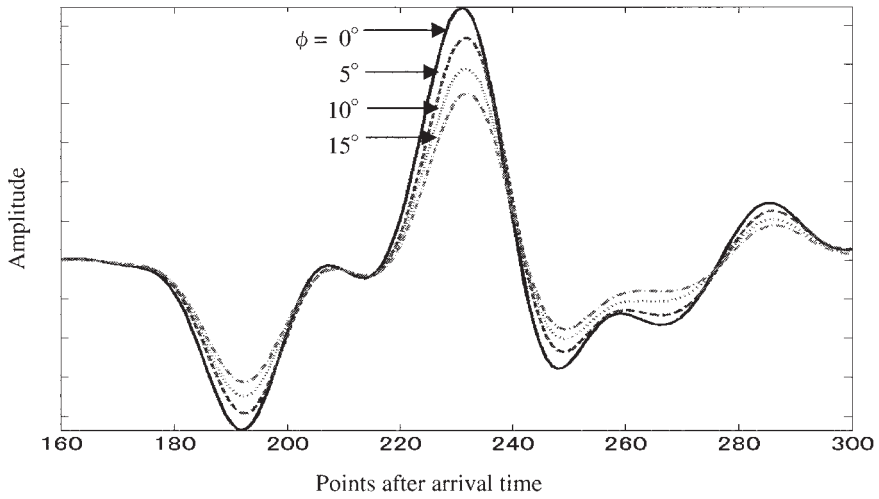

Fig. 4 arrival time aligned responses for left ear.

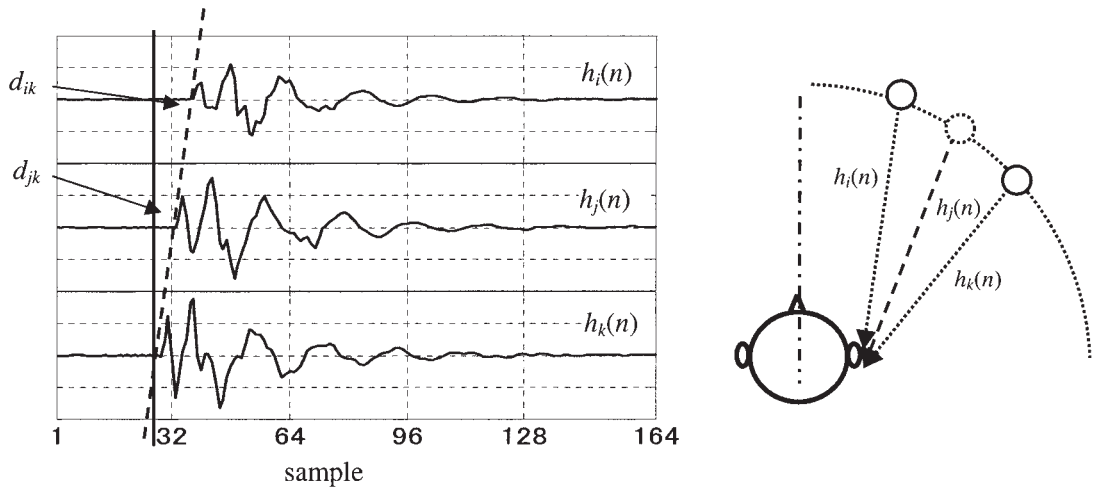

Fig. 5 Our method for linearly interpolating binaural impulse response.

2.4. Linear Interpolation Taking into Account Arrival Time Difference

Figure 5 illustrates our method of linearly interpolating binaural impulse responses taking into account the arrival time differences. The $h_i(n)$ and $h_k(n)$ denote, respectively, the binaural impulse responses of the left ear for the directions of the ith and kth sound sources, and d_{ik} denotes the arrival time difference between them. The binaural

impulse response to be interpolated is denoted as $h_j(n)$, and the arrival time difference between the binaural impulse responses for the ith and jth azimuths is denoted as d_{jk}. The $h_j(n)$ is interpolated along the oblique dotted line given by

$$h_j(n + d_{jk}) = \frac{1}{a+b}\{bh_i(n + d_{ik}) + ah_k(n)\}, \qquad (1)$$

where a denotes the angle between the sound source directions of ith and jth and b denotes the angle between those of jth and kth. The arrival time difference of a response is given by

$$d_{jk} = \frac{b}{a+b} d_{ik}. \qquad (2)$$

Interpolation of the responses at one ear is independent of that at the other. No binaural models are used in this method.

2.5. Evaluation of Interpolation Accuracy

We evaluated interpolation accuracy by using the signal to deviation ratio:

$$SDR = 10 \log \frac{\sum_{n=0}^{N-1} h^2(n)}{\sum_{n=0}^{N-1} \{h(n) - \tilde{h}(n)\}^2}, \qquad (3)$$

where $h(n)$ denotes a measured binaural impulse response, $\tilde{h}(n)$ denotes an interpolated binaural impulse response, and N indicates the number of response samples. As a counterpart to our method, we also investigated the interpolation accuracy of another algorithm. This algorithm (hereafter "simple method"), which is used in the cross-fading technique described later, does not take into account the arrival time difference between binaural impulse responses.

For both methods, we investigated interpolation accuracy for three angular intervals. "Angular interval" denotes the angular difference in the azimuths of the sound source directions to be used for interpolation. First, the binaural impulse responses for every 10° are used to interpolate the intermediate response for every 5°. For example, the responses for 0° and 10° are used to interpolate the response for 5° and so on. Next, the responses for every 15° are used to interpolate the intermediate responses for 5° and 10°. Then, the responses for every 45° are used to interpolate the intermediate responses for every 5° from 5° to 40° and so on.

Figures 6 (a) to (d) shows the *SDR* values of the simple method and those of our method for the left ear (shaded side) and right ear (exposed side). The x and y axes denote the azimuth of the sound source and the *SDR* value, respectively.

Table 1 Averages and variances of *SDR* values.

Left ear (shaded side)

	Simple method		Our method	
Angular interval	Average (dB)	Variance	Average (dB)	Variance
10°	7.3	4.6	16.5	29.0
15°	3.7	4.3	13.9	33.4
45°	−1.6	8.7	9.2	27.8

Right ear (exposed side)

	Simple method		Our method	
Angular interval	Average (dB)	Variance	Average (dB)	Variance
10°	15.3	82.4	24.8	18.1
15°	9.9	56.7	23.0	13.9
45°	1.3	23.4	7.0	55.7

Table 1 shows the averages and variances of the *SDR* values for the two interpolation methods. For angular intervals of 10° and 15°, the average values for our method were much larger than those for the simple method. The variances for the right ear for the simple method were larger than those for our method. Though the variances for the left ear for the simple method were very small, the average *SDR* values were small. This means that our method interpolated the binaural impulse responses more independently of the azimuths of the sound source than the simple method.

For the left ear (shaded side), our method (Fig. 6(c)) had a lower *SDR* value at an azimuth of 100° for an angular interval of 10° due to misdetection of the arrival time difference between the binaural impulse responses for 100° and those for 110° (Fig. 6(c)).

The four folded solid lines at an angular interval of 45° in Fig. 6(c) show that the *SDR* values of our method were higher than those of the simple method, represented by the four U-shaped curves in Fig. 6(a).

For the right ear (exposed side), for the azimuths from 0° to 85°, the *SDR* values with our method were higher than those with the simple method. Conversely, at the azimuths from 95° to 175°, the *SDR* values with the simple method were close to those with our method due to misdetection of the arrival-time differences. This shows that if the angular interval is 45° or more, our method cannot accurately interpolate binaural impulse responses at azimuths from 95° to 175°.

2.6. Application of Interpolation

To simulate a sound image that moves smoothly, we used the binaural impulse responses at 0° and 15° and interpolated the responses from 1° to 14° using both the simple method and our method. The changes in responses are shown in Figs. 7(a) and (b), respectively. The x (width) and y (depth) axes denote the sample and the azimuth of the

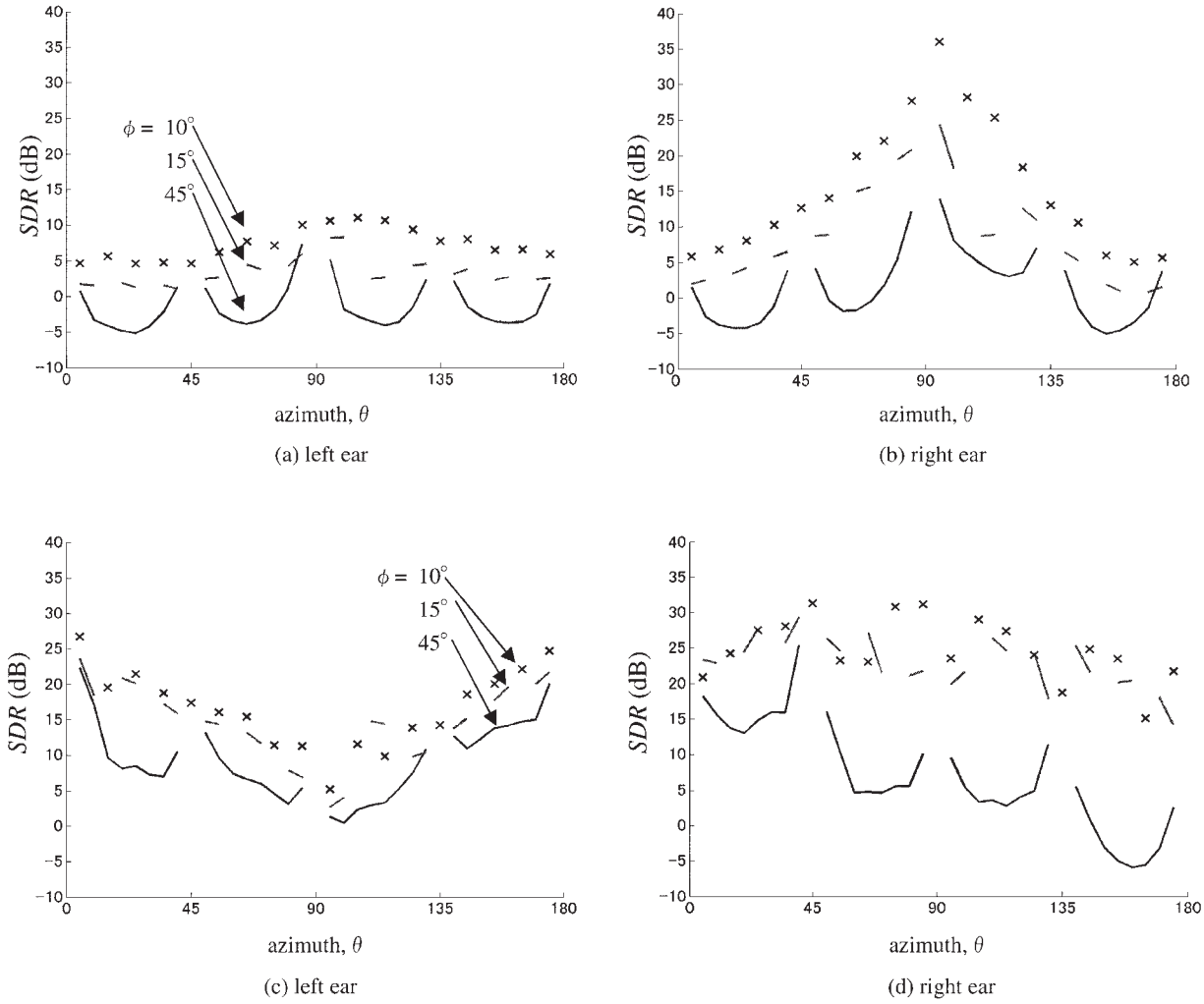

Fig. 6 (a) Signal to deviation ratio, *SDR*, simple method, left ear. (b) Signal to deviation ratio, *SDR*, simple method, right ear. (c) Signal to deviation ratio, *SDR*, our method, left ear. (d) Signal to deviation ratio, *SDR*, our method, right ear.

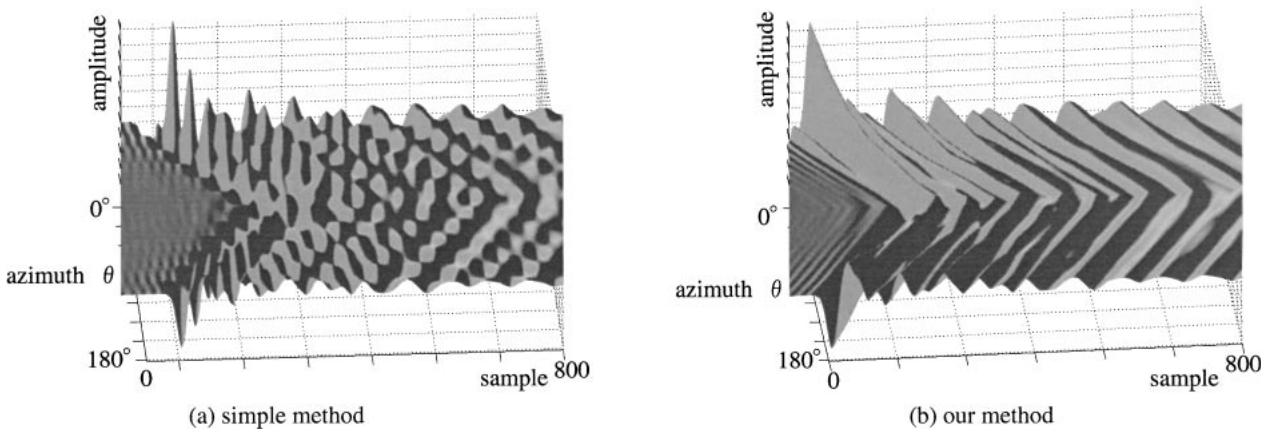

Fig. 7 Changes in impulse responses by azimuth angle.

sound source, respectively. The z axis shows the amplitude of the interpolated and actually measured responses. Many ripples are seen in Fig. 7(a), while none are seen in Fig. 7(b), indicating that a sound image simulated using our method does not include ripples of the amplitude.

3. MOVING SOUND IMAGES SIMULATED USING CONVENTIONAL METHOD AND OUR ALGORITHM

3.1. Conventional Method

The cross-fading technique is a conventional method for moving sound images. It can be represented as

$$\begin{aligned} y(n) &= \alpha x(n) * h_0(n) + (1-\alpha)x(n) * h_1(n) \\ &= x(n) * (\alpha h_0(n) + (1-\alpha)h_1(n)), \end{aligned} \tag{4}$$

where $x(n)$ and $y(n)$ denote the sound source samples and binaural signal samples during direction transition, respectively, $h_0(n)$ and $h_1(n)$ represent the binaural impulse responses at one ear in two directions of the sound source, and α denotes the ratio of $h_0(n)$ to $h_1(n)$.

This technique can be interpreted in two ways. The top line of the Eq. (4) indicates that when a sound image is being moved, sounds from two different sound sources simulate the sound image. This is not a realistic way to represent the actual sound image of the sound source. The bottom line denotes the simple method of linearly interpolating a binaural impulse response, and the interpolated response changes over time. The "$*$" represents time-variant convolution. The bottom line also denotes that no method other than simple linear interpolation can be used in the cross-fading technique.

3.2. Our Algorithm

Assume that sound source $x(n)$ moves and is reproduced as samples x_0, x_1, and x_2, as shown in Fig. 8. The matrix in Fig. 8 shows the calculated binaural signal at an ear.

Suppose that the sound source moves from $x_0(n)$ to

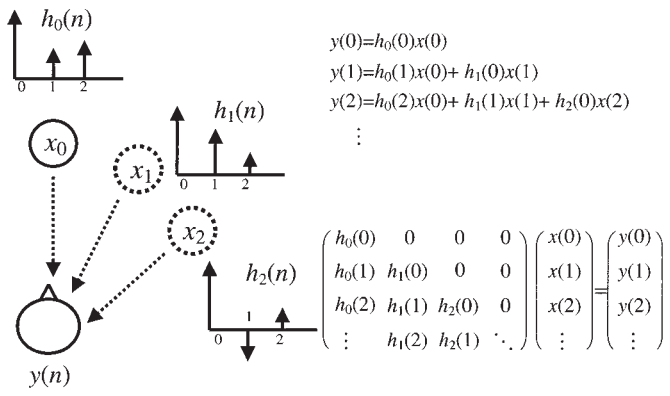

Fig. 8 Our algorithm sample by sample.

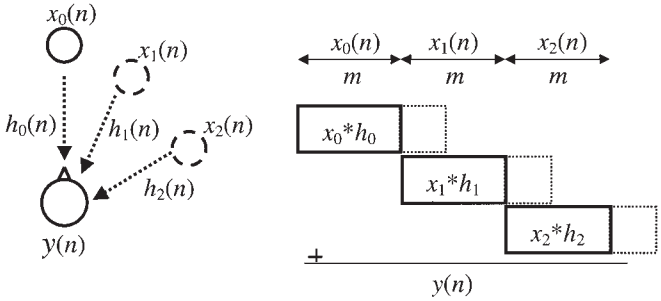

Fig. 9 Our algorithm every m samples.

$x_2(n)$ through $x_1(n)$, as shown by the diagram on the left in Fig. 9. We consider all m samples of source signal $x(n)$ to be reproduced at P_0, P_1, and P_2. The binaural impulse responses between the sound source and the listener change simultaneously as the sound source moves, as shown by the diagram on the right in Fig. 9. Adding the tails of $x_0(n) * h_0(n)$, which are shown as the dotted rectangles in Fig. 9, to $x_1(n) * h_1(n)$, gives signal $y(n)$ at the listener's ear.

This algorithm is very similar to the sectioned convolution add method. However, since impulse responses $h_0(n)$, $h_1(n)$ and $h_2(n)$ differ, there are discontinuities at each dotted square. Therefore, to move a sound image smoothly, we must use numerous slightly different binaural impulse responses. Since this algorithm is independent of the method used to interpolate the responses, we can use various interpolation methods.

3.3. Evaluation of Algorithms

To evaluate our algorithm using our method of interpolation, we compared a moving sound image simulated using our algorithm with an actual moving sound image and one simulated using the cross fading technique as a conventional method. The one simulated using cross fading was produced by time-variant convolution of the binaural impulse responses depicted in Fig. 7(a) and pink noise. The one simulated using our algorithm was produced by time-variant convolution of the responses depicted in Fig. 7(b) and the same pink noise.

The actual moving sound image was recorded while rotating the turntable rather than by moving the loudspeaker, as shown in Fig. 2. The turntable took nine seconds to go halfway round. The loudspeaker reproduced pink noise as the moving sound source.

3.4. Results

Figures 10(a) to (f) show spectrograms of the binaural signals at the left and right ears. Each frame of the spectrograms is 512 samples long and has 256-sample overlapping. The x axes represent the time transition from 0 to 9 seconds, corresponding to the change in azimuth

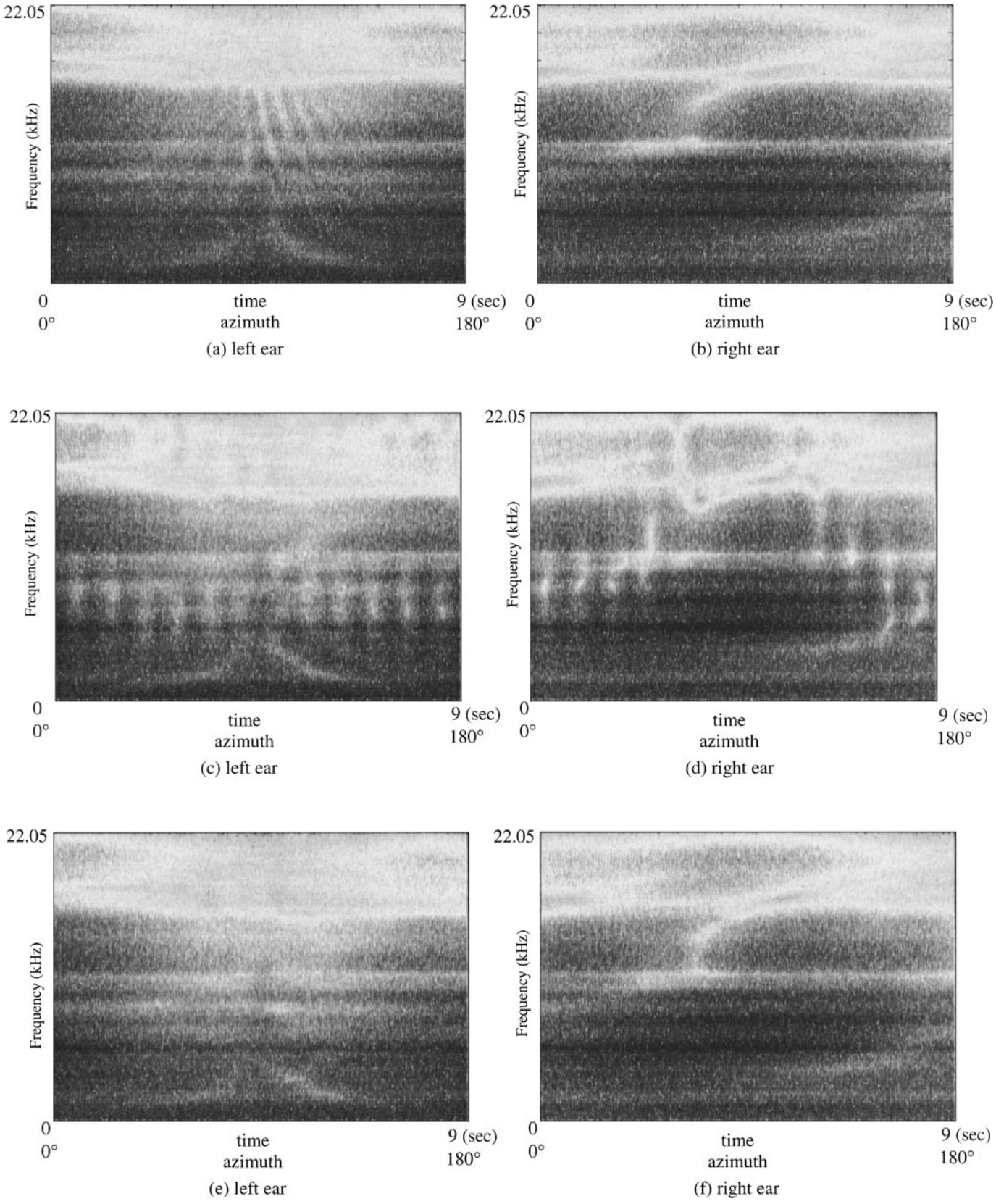

Fig. 10 (a) Spectrogram of binaural signal for actual moving sound, left ear. (b) Spectrogram of binaural signal for actual moving sound, right ear. (c) Spectrogram of binaural signal for sound image simulated using crossfading, left ear. (d) Spectrogram of binaural signal for sound image simulated using crossfading, right ear. (e) Spectrogram of binaural signal for sound image simulated using our algorithm, left ear. (f) Spectrogram of binaural signal for sound image simulated using our algorithm, right ear.

from 0° (front) to 180° (back). The *y* axes represent the frequency up to 22.05 kHz. The gradations show the amplitude of the frequency characteristics.

Figures 10(a) and (b) show the spectrograms of the binaural signals for the actual image of pink noise. At the beginning of the time transition, since the sound source was located in front of the dummy head, the binaural signals at both ears had higher frequency components. As the time transitioned, the sound source moved from the front to the right side of the head, so the left ear became shaded from the signal, resulting in a reduction of the higher frequency level. On the other hand, the right ear became more exposed to the source, and the higher frequency level increased. The downward radial ripples at the center of the time transition in Fig. 10(a) are assumed to be due to diffraction by the shoulders and torso of the dummy head.

Figures 10(c) and (d) show the spectrograms of the binaural signals for the images produced using cross fading. In contrast to the spectrograms in (a) and (b), these spectrograms have ripples in the amplitude-frequency characteristics, seen as vertical stripes.

Figures 10(e) and (f) show the spectrograms of the binaural signals for the images produced using our algorithm. They do not have any ripples in the amplitude-frequency characteristics and are very close to those of the actual sound image. At the center of the time transition, radial ripples are not seen in (e) due to the angular interval of 15°.

Figure 11 shows the time transition of the amplitude level of the 8.8-kHz frequency component of the actual sound image shown in Fig. 10(a). It also shows that of the sound image simulated using cross fading shown in Fig. 10(c) and that of the sound image simulated using the proposed algorithm shown in Fig. 10(e). The *x* and *y* axes denote the time transition from 0 to 9 seconds and amplitude level, respectively. The time transition of the amplitude-frequency characteristics at the frequency of the actual moving sound image decreased gradually as time passed. The amplitude-frequency characteristics of the sound image simulated using the proposed algorithm changed much like those of the actual sound image. In contrast, the characteristics of the sound image simulated using cross fading decreased and ripples appeared as time passed.

Suppose that $A_{\mathrm{act}}\,(f_k, m)$ in the Eq. (5) denotes the time transition of the amplitude-frequency level of the binaural signal of the actual sound image at a frequency of f_k. The M in the equation denotes the number of the frames, and m is the frame index. Across (f_k, m) and Aproposed (f_k, m) denote those for a sound image simulated using cross fading and our algorithm. Furthermore, $d_1(f_k)$ and $d_2(f_k)$ are calculated as follows.

$$\begin{aligned} d_1(f_k) &= \sum_{m=0}^{M-1} |A_{\mathrm{act}}(f_k, m) - A_{\mathrm{cross}}(f_k, m)|^2 \\ d_2(f_k) &= \sum_{m=0}^{M-1} |A_{\mathrm{act}}(f_k, m) - A_{\mathrm{our}}(f_k, m)|^2 \end{aligned} \tag{5}$$

The $d_1(f_k)$ represents the error in the frequency domain between the actual sound image and the sound image simulated using cross fading. The $d_2(f_k)$ represents that between the actual image and the sound image simulated using the proposed algorithm.

We introduce error ratio *ER*:

$$ER(f_k) = 10 \log \frac{d_2(f_k)}{d_1(f_k)}.$$

As shown in Fig. 12, from 3 to 11 kHz, $d_2(f_k)$ was much

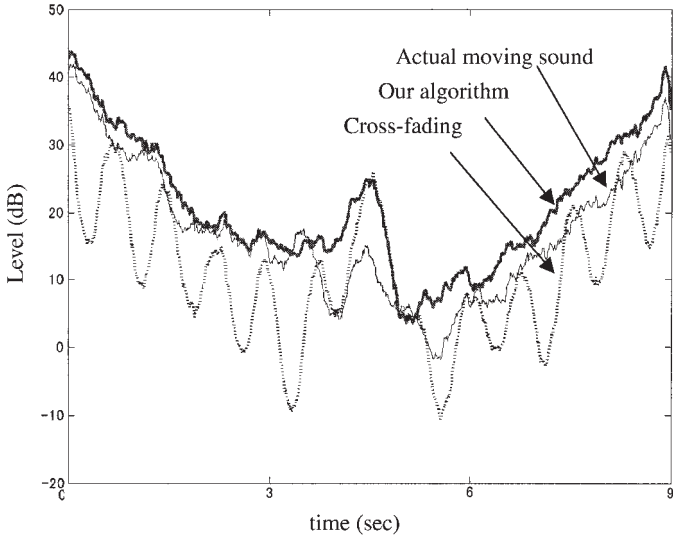

Fig. 11 Time transition of amplitude-frequency lebel characteristics.

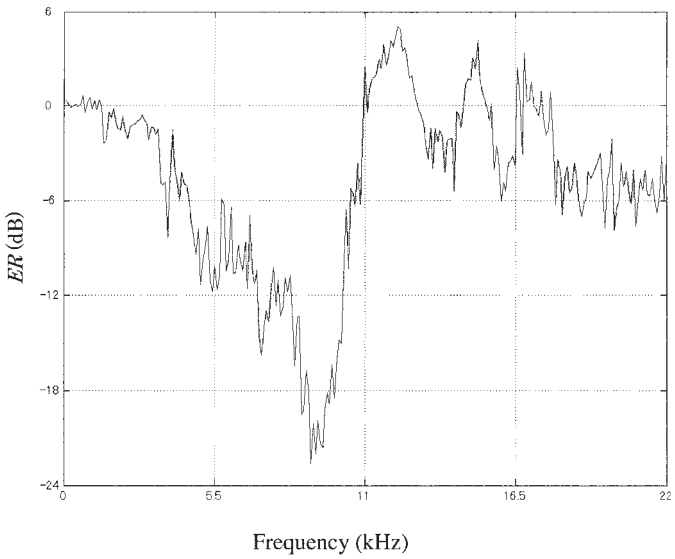

Fig. 12 Error ratio, *ER*.

smaller than $d_1(f_k)$. This means that the time-transition of the amplitude-frequency characteristics of a sound image simulated using the proposed algorithm are much closer to those of the actual image than those of a sound image simulated using cross fading. For the frequency range from 11 to 16.5 kHz, the sound image simulated using cross fading was closer to that of the actual one than that simulated using the proposed algorithm.

4. SUMMARY

We have described a method for interpolating binaural impulse responses and an algorithm for simulating a moving sound image. The method takes into account the arrival time difference due to changes in the direction of a moving sound source. Evaluation of the interpolation accuracy using the signal deviation ratio showed that the binaural impulse responses interpolated using this method are closer to ones actually measured than those using a simple linear interpolation method that does not consider the arrival-time difference. The responses interpolated using our method changed smoothly as the sound source direction changed.

We also described an algorithm for simulating a moving sound source. We recorded an actual moving sound image by using a dummy head mounted on a rotating turntable. We compared the image simulated using the algorithm with the actual one and one simulated using the cross-fading technique. The spectrogram of the binaural signal of the sound image simulated using our algorithm was closer to that of the actual image and contained no ripples.

ACKNOWLEDGEMENTS

We thank Mr. Satoru Gotoh for his assistance with the experimental set-up and signal processing and Mr. Kiyoaki Terada for his comments on an early draft of the manuscript.

REFERENCES

[1] M. Uchiyama and M. Tohyama, "Sound image control for Internet," *Proc. Inst. Acoustics*, **20**, 231–238 (1998).
[2] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *J. Acoust. Soc. Am.*, **91**, 1637–1647 (1992).
[3] T. Nishino, S. Kajita, K. Takeda and F. Itakura, "Interpolation of head related transfer functions of azimuth and elevation," *J. Acoust. Soc. Jpn.* (*J*), **57**, 685–692 (2001).
[4] M. Toyama, M. Uchiyama and H. Nomura, "Head related transfer function representation of directional sound for spatial acoustic events modeling," *IEEE Workshop MMSP*, pp. 221–226 (1999).
[5] K. Hartung, J. Braasch and S. J. Sterbing, "Comparison of different methods for the interpolation of head-related transfer functions," *AES 16th International Conf.*, pp. 319–329 (1999).
[6] M. Matsumoto, S. Yamanaka, M. Tohyama and H. Nomura, "Moving sound image representation method," *EUSIPCO 2002*, no. 485 (2002).
[7] L. A. Jeffress, "A place theory of sound localization," *J. Comp. Physiol. Psychol.*, 35–39 (1948).