Moving sound image representation method

Mitsuo Matsumoto, Susumu Yamanaka*, Mikio Toyama*, Hiroaki Nomura***

**Kogakuin University, ** East Asia University*

ABSTRACT

This paper introduces an algorithm simulating a moving sound image as a time-variant acoustic event and reports on the results of an investigation of methods for interpolation of binaural responses. As moving sound images require numerous binaural responses, the use of interpolation methods is necessary. We investigated two interpolation methods; one is a linear waveform interpolation method considering of initial time delay difference and the other is a linear interpolation method using the minimum-phase component of a response and an inter-aural time difference model. The methods are evaluated by using a spectrogram of binaural signals simulated by the algorithm. The results showed that the linear interpolation considering the initial time-delay difference resulted in highly precise interpolation and that the inter-aural time difference was maintained in both the methods.

1. INTRODUCTION

A moving sound image is a time-variant acoustic event. Time-variant acoustic events can be simulated by a number of successive time-invariant events. For example, cross-fading technique has been applied to simulate a moving sound image by using binaural responses. The cross-fading technique does not reflect actual physical situation. Because during sound image direction transition, two sound images are heard at once. In this paper, we introduce a convolution method to simulate a moving sound image as a time-variant acoustic event by using successively changed binaural responses.

From practical point of view, since measuring a number of successively changed binaural responses is not a realistic task, the use of interpolation methods is necessary. Previously proposed interpolation methods [1], [2], [3] are estimated by spectrum distortion [1], SD and signal to deviation ratio [2], [3] SDR from time-invariant point of view. In the case of moving sound images as a time-variant event, estimating interpolation precision by these evaluation methods is somewhat inappropriate. We report on an interpolation of binaural-response- interpolation methods that can be used to deal with moving sound images; one is a linear interpolation method that considers an initial time delay difference and the other is a linear interpolation method that uses the minimum-phase component and a binaural model. We estimate the accuracy of the interpolation methods by using the convolution algorithm from the time-variant point of view. Furthermore, inter-aural cross correlation, IACC also estimate the accuracy of the methods.

2. MOVING SOUND IMAGE

Suppose that a subject is listening to a moving sound source that moves from P_0 to P_4 through P_1, P_2, and P_3 in Fig. 1. We considered m samples of a source signal $x(n)$ to be reproduced at P_0, P_1, P_2, P_3, and P_4, respectively. Simultaneously, the binaural response between the sound source and the subject changes as the sound source moves as shown in Fig. 1. Signals $y(n)$ at the subject's ear is calculated by adding the tail of $x_0(n)*h_0(n)$, which is shown as the overlapped hatched areas in Fig. 1, to $x_1(n)*h_1(n)$, etc.

We evaluated the two interpolation methods by how well they predicted the signals at the subject's ears in accordance with the algorithm described above.

3 INTERPOLATION METHODS

In this paper we investigate interpolation methods, a linear

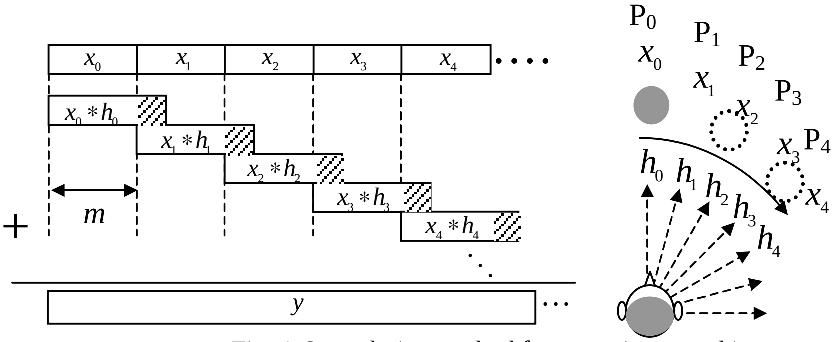


Fig. 1 Convolution method for a moving sound image

waveform interpolation method considering initial time delay differences and a linear interpolation method that uses minimum-phase component of a binaural response.

3.1 A linear interpolation considering initial time delay difference

Figure 2 illustrates a linear interpolation method. In the figure, $h_i(n)$ and $h_k(n)$ denote respectively binaural impulse responses of the i_{th} and k_{th} directional sound sources. As seen in Fig. (a), initial time delays, for example d_i samples for $h_i(n)$ depend on the source directions. To equalize the initial time delays, let $h_k(n)$ time-shift by the time delay difference $d_i - d_k$ that is determined by the cross-correlation between $h_i(n)$ and $h_k(n)$. Figure 2(b) shows the responses equalized regarding the time delay difference. Intermediate response $h'_j(n)$ is interpolated according to the equation

$$h'_j(n) = \frac{1}{a+b}\{bh'_i(n) + ah'_k(n)\} (eq.1),$$

where a denotes the angle between the i_{th} and j_{th} sound source directions and b denotes the angle between those of the j_{th} and k_{th}.

And the initial time delay of the response d_j is determined by d_i and d_k in the same manner.

$$d_j = \frac{1}{a+b}\{bd_i + ad_k\}.$$

Then, let $h'_j(n)$ time-shift by $d_j - d_i$.

In this method, the binaural response to be interpolated depends on only the adjacent binaural responses. What this means is that the interpolated response is determined not by using binaural information, for example inter-aural time difference, but by using monaural information.

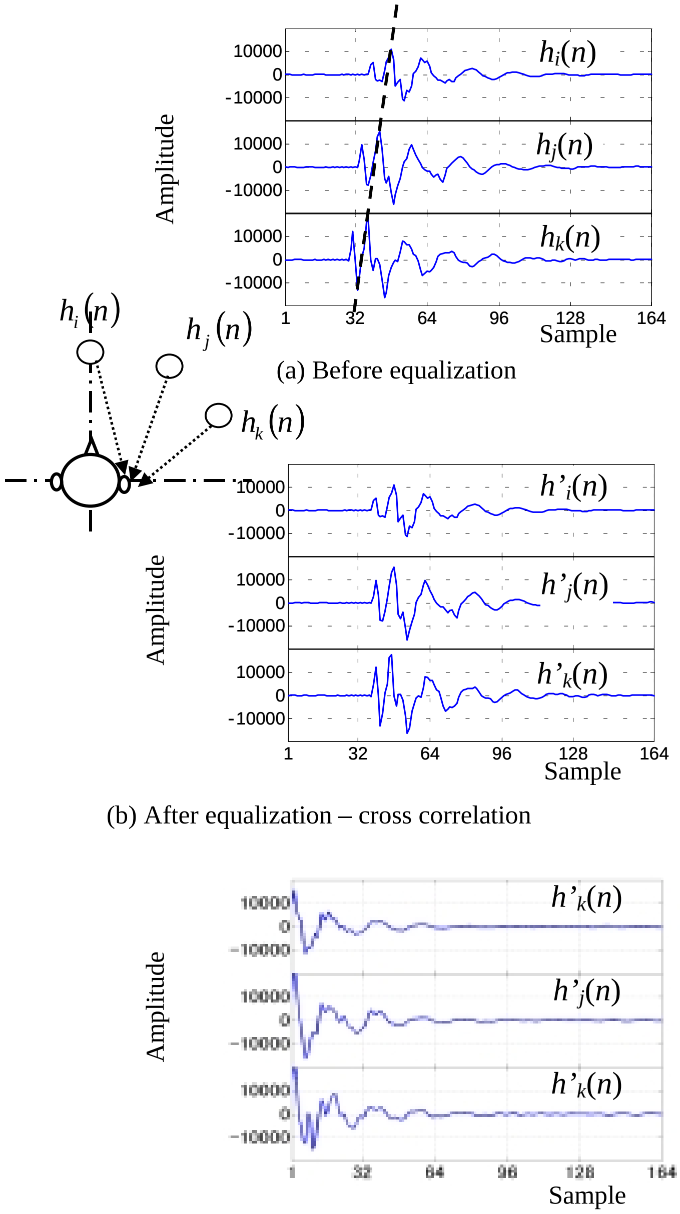
Furthermore, an initial time delay difference is determined by the cross-correlation. Since the cross-correlation relates to power spectrums of the two responses, regarding "time of arrival" issue, the cross-correlation does not give the initial time delay difference accurately. However, as the power spectrum might reflect hearing impression, we use the cross-correlation to determine the time difference.

3.2 A linear method using minimum-phase component

In this method, the minimum-phase component of a binaural response is extracted to equalize initial time delay [2]. Figures 3(a) and 3(b) show impulse responses before and after equalization. Interpolation is in the same manner according to the Eq. 1. Inter-aural time difference, the ITD in the direction is determined by the equation listed below.

$$ITD = \frac{d}{c}\sin\theta$$

where c, d and θ denote the velocity of sound, the distance between the two ears and direction of the sound source. Then, the ITD is applied to the pair of the binaural responses for the source direction. In this paper, we assume that c is 340 m/s and d is 22 cm respectively.



(a) Before equalization

(b) After equalization – cross correlation

(c) After equalization – minimum-phase

Fig. 2. Initial time delay equalization

These two methods are close with regard to initial time delay equalizing. However, how to reproduce ITDs of the two methods is different. In the first method, although an ITD is not considered explicitly, the ITD is maintained implicitly

while interpolating. Conversely, an ITD on the basis of a binaural model is explicitly determined in the second method.
In this research, we used dummy head responses downloaded from a web site affiliated by the MIT Media [4] as measured binaural responses that changed in 5-degree steps. We also made responses in 1-degree steps using a linear interpolation method without initial time-delay equalizing and used these as the "original" responses.

In this paper, we attempted to interpolate 5- and 10-degree impulse responses from 0- and 15-degree responses in the same way as the interpolation is done at 15-degree angular intervals.

4. RESULTS

From the statistical estimation viewpoint, impulse responses interpolated by the linear waveform interpolation using the cross-correlation and the linear method using the minimum-phase component were estimated by SDR[2]. Figure 3 shows the SDRs of the two responses interpolated for the right ear. The X- and Y-axes indicate the sound source directions and SDRs respectively. The upper straight lines are the SDRs of the linear interpolation using the cross-correlation and the lower ones are those of the linear method using minimum-phase. As seen in this figure, the interpolation precision of the linear interpolation using the cross-correlation is higher than that of the linear method using the minimum-phase.
Figure 4(a) shows the spectrogram of a signal at the subject's left ear in the case of the "original" response. The horizontal and vertical axes represent, respectively, time and frequency. Figure 4(b) and Fig. 4(c) show the spectrograms of the two interpolation methods. Through Fig. 4(a) to Fig. 4(c), spectrograms of right ear that is ipsilateral of the sound source are similar. However, three spectrograms of left ear, through Fig. 4(a) to Fig. 4(c), which is contra lateral of the source are different from each other.
Regarding the IACCs shown in Fig. 5, the downward arcs of the linear method using the cross-correlation in Fig. 5(b) is closer to that of the "original" and conversely, the arc in 5(c) is somewhat different from that of the "original". However, the outline of the arc in 5(c) is similar to those of the "original" and the linear method using the cross-correlation. The figure shows that the two interpolation methods maintain inter-aural time differences.

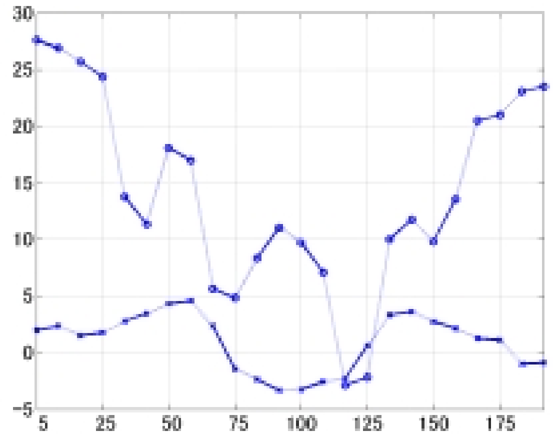


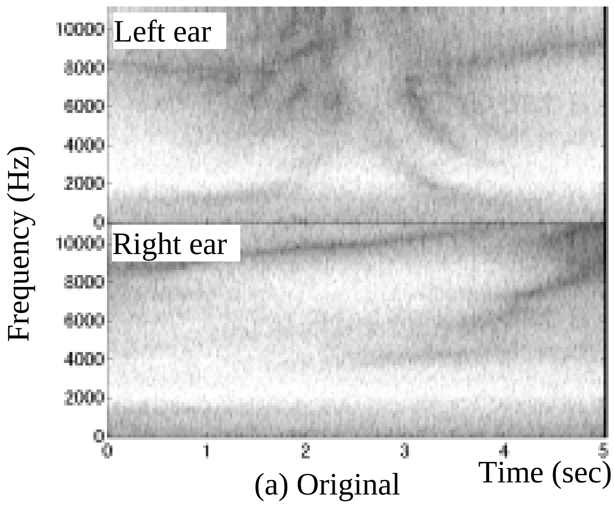
Fig. 3 SDRs of the two methods

5. SUMMARY

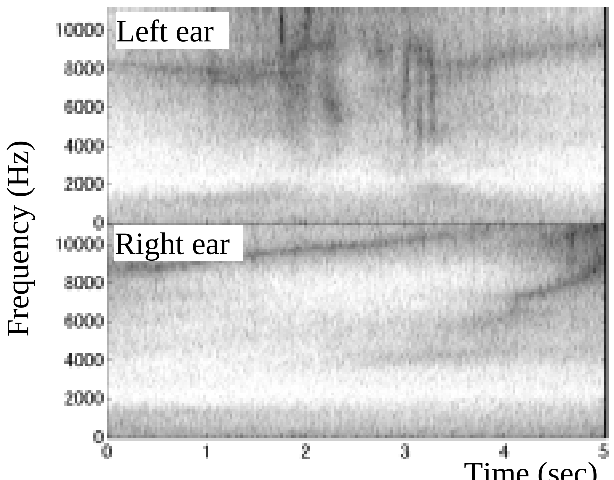
We reported the development an algorithm to simulate a binaural signal reproduced by a moving sound source that we used to evaluate the two interpolation methods.
We investigated two methods of binaural interpolation for moving sound images. The first is a linear waveform interpolation method that considers the initial time-delay difference and the second is the linear method that uses the minimum-phase component of the binaural response.
Results show that both the two methods maintain inter-aural time difference.
The linear waveform interpolation method is better suited than the linear method using the minimum-phase component to the task of interpolating moving sound images regarding the IACC. "Time of arrival" issue in the first method and hearing impression will be investigated as a further study. This research was partly underwritten by the support system in place for R&D activities in the field of info-communications operated by the auspices of the Telecommunications Advancement Organization of Japan, and the International Communications Foundation.

REFERENCES

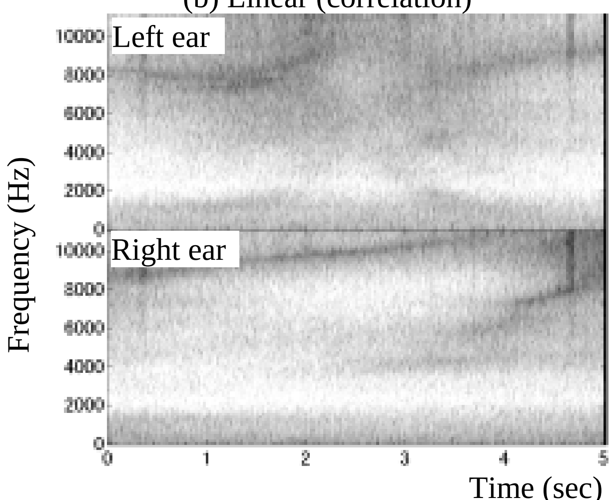
[1] T. Nishino *et al.*, J. Acoust. Soc. Japan. 57, pp. 685-692 (2001) (in Japanese)
[2] M. Uchiyama *et al.*, Proc. I.O.A., **20** (5), pp.231-238 (1998)
[3] M. Toyama *et al.*, IEEE Workshop MMSP, pp. 221-226 (1999)
[4] MIT web site: http://sound.media.mit.edu/KEMAR.html



(a) Original

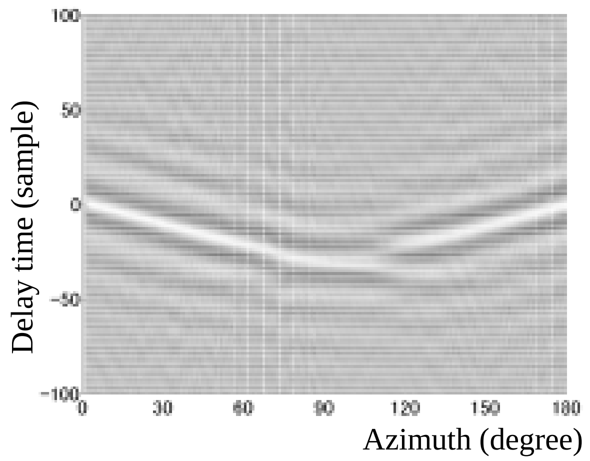


(b) Linear (correlation)

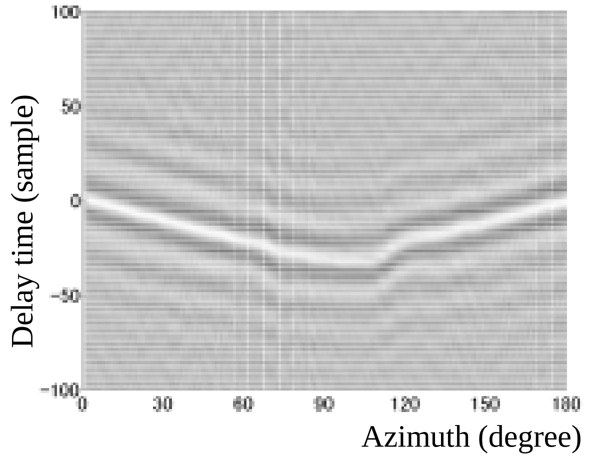


(c) Linear (min-phase)

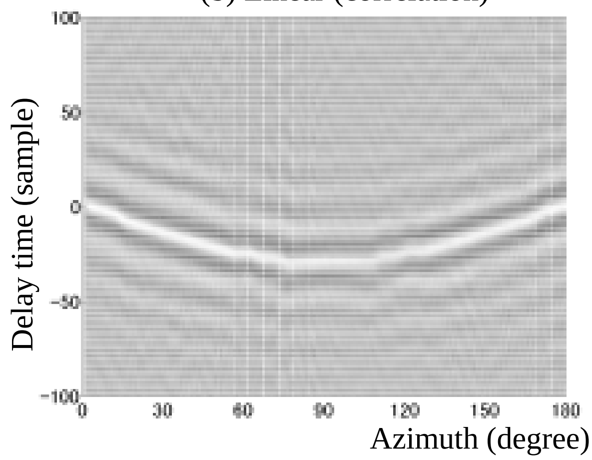
Fig 4. Spectrograms of a moving sound image



(a) Original



(b) Linear (correlation)



(c) Linear (min-phase)

Fig 5. Inter aural cross correlation (IACC) of a moving sound image