Acoust. Sci. & Tech. **26**, 3 (2005)

PAPER

A study on switching of the transfer functions focusing on sound quality

Akihiro Kudo*, Haruhide Hokari and Shoji Shimada

Nagaoka University of Technology,
1603–1 Kamitomioka-machi, Nagaoka, 940–2188 Japan

(*Received 6 May 2004, Accepted for publication 15 November 2004*)

Abstract: Many papers have described moving sound image schemes that use loudspeakers or headphones. Since most of these schemes switch the spatial transfer function being used, wave discontinuity occurs at the moment of switching, which degrades the sound quality. While the characteristics of the wave discontinuity depend on the moving sound image scheme used, no paper appears to have considered the relationship between the wave discontinuity and the scheme used. To rectify this omission, this paper examines three approaches: the simple switching approach, the overlap-add approach, and the fade-in·fade-out approach. The sound degradation caused by the wave discontinuity is assessed, and an objective measure, spectrum distortion width, is introduced to quantify the wave discontinuity. Subjective assessments, carried out using Scheffe's comparison tests, verify that the overlap-add approach with modified hamming window and the fade-in·fade-out approach were better than the other methods.

Keywords: Moving sound image, Switching transfer functions, Bandwidth equation, Wave discontinuity, Sound quality degradation

PACS number: 43.66.Pn, 43.66.Ki [DOI: 10.1250/ast.26.267]

1. INTRODUCTION

A perception of a moving sound image can be achieved by switching transfer functions which include spatial transfer characteristics. Many papers have targeted this schemes assuming the use of loudspeakers or headphones.

The simplest method using loudspeakers is to physically move the active loudspeaker itself. However, switching the input signal of static loudspeakers is more often used to reproduce moving sound through auditory stimulation [1], since the physical approach is impractical. Some schemes are described below.

Perrott *et al.* [1] studied the conditions under which subjects perceived a continuous moving sound image on the horizontal plane, using two loudspeakers. The stimulus test signal to drive each loudspeaker, was broad band noise with 50 ms duration time, which included the rest time for driving the adjacent loudspeaker. Two listening subjective assessment tests were done. First, the center of the two loudspeakers was located in the front of the subject, and the opening angle between two loudspeakers was set at 6°, 40°, and 160°. Second, the direction angle of the center position was set at 0°, 45°, and 90°, on the condition that the opening angle was held steady at 40°. The results indicated that 1) sound in continuous motion was easily perceived as the opening angle was small, and 2) the effect didn't depend on the sound source direction.

Taking the case of loudspeakers located on the upper transverse plane at regular interval, Nakajima *et al.* [2] studied the interval angle between loudspeakers and the duration of the switching time, as if the sound in motion were continuous. The stimulus signal had cross-fading characteristics with 15 ms duration; the loudspeakers were driven sequentially. The distance from the center of the subject's head to the loudspeakers' mounting circle was 1.5 m, and the loudspeaker interval angles were 22.5°, 36°, 45°, 54°, when the number of loudspeakers were 9, 6, 5, and 4, respectively. The stimulus test signals were the pure tone (250 Hz, 500 Hz, 1,000 Hz, 2,000 Hz, 4,000 Hz) and critical band noises (center frequency: 250 Hz, 450 Hz, 1,000 Hz, 2,150 Hz, 4,000 Hz). They concluded as follows: 1) moving sound images of the critical band noises were perceived as being more continuous than the pure tones, 2) interval angle between loudspeakers was in inverse proportion to the switching time duration, 3) deviation of the perception of the continuity of the moving sound image was large between subjects.

Kinoshita *et al.* [3] also studied the conditions necessary for the continuous perception of moving sound images in the horizontal and median planes, using three

*e-mail: kudo@audio.nagaokaut.ac.jp

loudspeakers driven sequentially. The interval angle between loudspeakers was 30°, and the distance from the center position of the subject's head to each loudspeaker was 1.5 m. The stimulus signal had a cross-fading characteristic with 10 ms duration time. Three experiments were carried out. In order to assess the condition of continuous perception, the first experiment examined the relationship between maximum switching time duration of the stimulus signal and four stimulus signals on the horizontal plane. White noise (200–7,000 Hz), pink noise (200–7,000 Hz), 1/3 octave band noise (center frequency: 1,000 Hz), and female voice were used as the stimulus signals. The second experiment examined the dependency of the minimum cross-fading duration time on the shapes of the cross-fading window. The final experiment studied the impact of the switching time duration of the stimulus signals on the direction of movement and the direction angle of loudspeaker center position on the horizontal and median plane. The opening angles were 0°, 45°, 90°, 135° and 180° on the horizontal plane, and 0°, 45°, 90°, 135° and 180° on the median plane. From the results, it was concluded that 1) the maximum switching time duration was minimum when the white noise and the pink noise were used, the time was 73 ms, the time of 1/3 octave band noise was 267 ms, and that of the female voice was 580 ms; 2) continuous perception of the moving sound image depended on the instantaneous sound pressure level fluctuation and the kind of stimulus signal; 3) it was necessary to hold the constant energy of stimulus signals steady in cross-fading to permit slow-moving sounds to be perceived; and 4) the switching time is only slightly dependent on moving angle and the direction angle.

Mizushima *et al.* [4] examined a continuity of sound image movement using two loudspeakers in horizontal plane. Several rest times were inserted into the stimulus signals which drive loudspeakers. They considered rest times (1 ms, 2 ms, 4 ms and 8 ms), opening angle (20°, 40° and 60°) and duration time of stimulus signal (5 ms, 10 ms, 20 ms, 40 ms, 80 ms). A white noise (500–8,000 Hz) was used in this experiment. The results indicated that 1) moving sound image was easy to be perceived continuously as the duration time was long and the rest time was short, especially the rest time must be 2 ms or less. 2) duration time of stimulus signal which includes the rest time needs to exceed the auditory integration time 100–150 ms.

As described above, moving sound image schemes that use loudspeakers have been studied under the assumption that the sound image movement was to be continuously perceived.

On the other hand, most of the moving sound image schemes that use headphones assume that the spatial transfer functions are switched [5–9]. While a few papers stated only that "the stimulus signal was sequentially convolved with the head-related impulse responses," no paper has provided a detailed explanation of the convolution processing. This processing may differ for every researcher because the details remained unclear.

An example of the moving sound image schemes that use headphones was presented by Matsumoto *et al.* Matsumoto *et al.* [10] picked up the conventional cross-fading method and the overlap-add method which is their proposed method. The sound in motion created by a loudspeaker was used as the basis of comparison, and the two methods were compared by objective analysis using a spectrogram. The results demonstrated that the spectrogram of moving sound produced by the overlap-add method was closer to the spectrogram of sound movement created by loudspeakers than that of the conventional method.

Because the moving sound image schemes using headphones are based on switching spatial transfer functions and the moving sound image schemes using loudspeakers are achieved by switching stimulus signals to drive the loudspeakers sequentially, both schemes share the basic concept of "switching."

We note, however, that a wave discontinuity occurs at the moment of switching [10], since the schemes using headphones switch between different spatial transfer functions. On this point, the schemes that use headphones are essentially different from those that use loudspeakers. The wave discontinuity is perceived as a click noise. Because the wave discontinuity is yielded whenever the sound image moves, the quality of the moving sound image is degraded significantly. It is clearly very important to study this sound quality degradation, because users readily perceive this sound quality degradation when using headphones.

Even though the characteristics of the wave discontinuity depend on the moving sound image schemes used, no paper appears to have considered the relationship between the wave discontinuity and the schemes used.

This paper studies the impact of several moving sound image schemes that use headphones on the click noises yielded by the wave discontinuities as a basic study for achieving moving sound images that equal the quality of the images created with loudspeakers.

As the moving angle of the sound image is changed, objective evaluations assess the wave discontinuity using an objective measure, spectrum distortion width. Subjective evaluations examine the perception of the click noises using Scheffe's paired comparison tests.

2. PRINCIPLE OF MOVING SOUND IMAGE

Figure 1 shows the relationship of the transfer functions used for realizing a moving sound image. In Fig. 1,

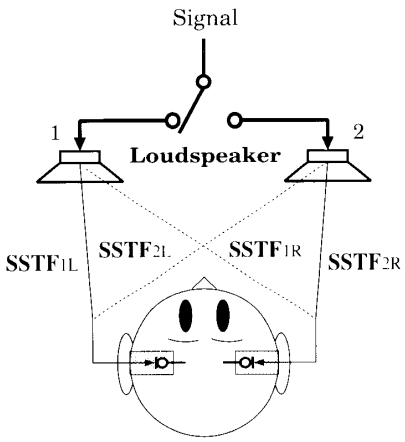

Fig. 1 Relationship of the transfer functions.

subscript L (R) denotes left (right) ear. The Spatial Sound Transfer Function (SSTF) models the conversion of the input signal of the loudspeaker into the output signal of the miniature microphone at the entrance of the ear canal of the listener. A moving sound image is achieved if the loudspeaker that radiates the sound wave is switched.

The wave discontinuities are yielded by the time difference and the amplitude difference between the impulse responses of switching spatial transfer functions, and the wave discontinuities increase as these differences increase, as describe in Section 4.1. These differences are caused by the difference in loudspeaker-ear distance between the two loudspeakers. That is, the wave discontinuity increases with the distance difference. Moreover, a geometric analysis showed that the distance difference is largest when the center of the two loudspeakers is located at or near the front of the subject under the condition that the opening angle between the two loudspeakers is fixed. The details are described in the Appendix. It can be concluded that the switching of the spatial transfer functions toward the frontal direction yields the most easily perceived wave discontinuity. Hence, this study focuses on switching toward the frontal direction.

Thurlow *et al.* [11] observed the head movements of subjects in localizing sound sources. They classified the head movements into three movements: tip, pivot and rotate, and reported that rotate was more common than the other two movements. Uematsu *et al.* [12] examined the impacts of the head movements on sound localization. They focused on two head movements, left-right and up-down, and reported that the left-right movements contributed more strongly to sound localization than up-down movements. These papers indicate that the head rotation on the horizontal plane is most important for sound localization. Since the head rotation and the sound source movement on the circumference are related in that the relative arrangement of subject and loudspeaker is identical

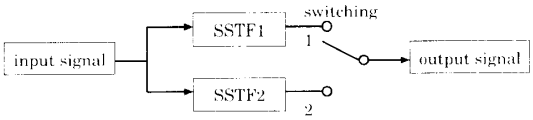

Fig. 2 Concept of switching transfer functions.

in space, the sound source movement also contributes to sound localization. Our final goal is to realize correct moving sound image localization using headphones. Therefore, this paper studied the left-right movements since they strongly contribute to sound localization.

3. SWITCHING TRANSFER FUNCTION METHODS

The concept of switching transfer functions is shown in Fig. 2. The blocks "SSTF 1,2" indicate transfer functions that represent different sound source positions. A moving sound image is simply achieved by switching the transfer functions of both ear sides.

This section separates the switching transfer function schemes into three groups: the simple switching approach, the overlap-add approach, and the fade-in-fade-out approach. We create an algorithm for each group.

3.1. Simple Switching Method (Method A)

Figure 3 illustrates the simple switching method. The input signal is convolved with $ssir_1$ and $ssir_2$ which were measured at different sound source positions, where ssir denotes the impulse response of SSTF. The moving sound image is created by concatenating the signals in alternate rectangular windows frame by frame. This method has the characteristic that a wave discontinuity occurs at each signal concatenation point.

3.2. Overlap-Add Method

As an example of the overlap-add approach, we selected the normal overlap-add method of [10,13] and the overlap-add method with modified hamming window.

- Normal overlap-add method (method B)

Figure 4 illustrates the normal overlap-add method. The input signal is windowed frame by frame. The signal is

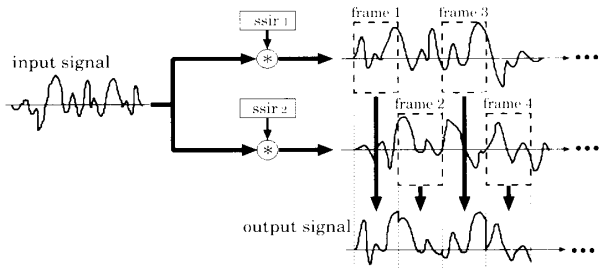

Fig. 3 Schematic diagram of the simple switching method.

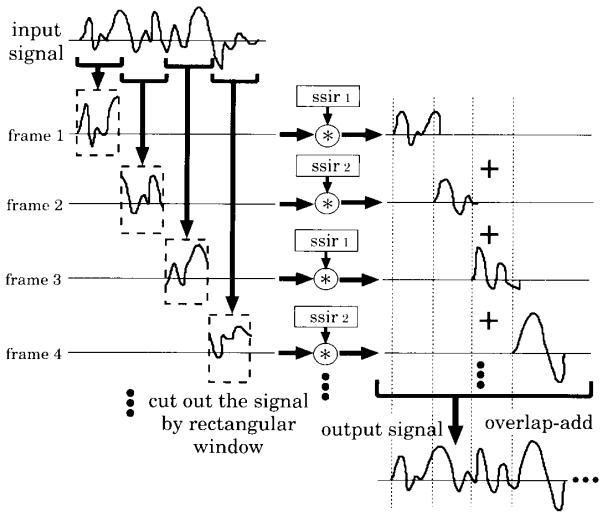

Fig. 4 Schematic diagram of the normal overlap-add method.

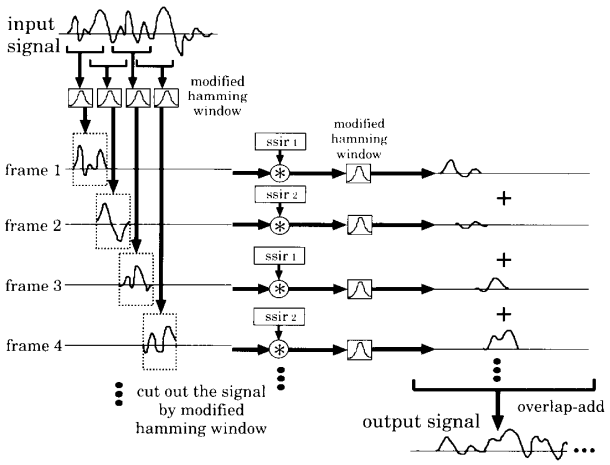

Fig. 5 Schematic diagram of the overlap-add method with modified hamming window.

convolved with each ssir, which were measured at different sound source positions. The convolved signals are overlap-added to create the moving sound image. This method creates wave discontinuity in the overlap areas.

- Overlap-add method with modified hamming window (method C)

Figure 5 illustrates the overlap-add method with modified hamming window. The input signal is weighted with a modified hamming window [14] on a frame-wise basis, and the signal is convolved with each ssir, which were measured at different positions. The convolved signal is weighted again with the modified hamming window. The signals are overlap-added to create the moving sound image.

3.3. Fade-In·Fade-Out Method

Figure 6 illustrates the fade-in·fade-out method. The input signal is windowed so that the frames overlap. In the

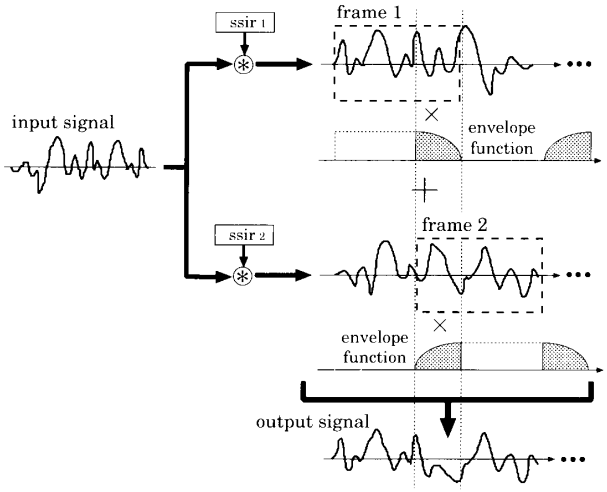

Fig. 6 Schematic diagram of the fade-in·fade-out method.

overlap areas, the signals are weighted by envelope functions $f(t)$, $g(t)$ which satisfy [3]

$$f^2(t) + g^2(t) = 1 \tag{1}$$

3.3.1. Shapes of the envelope functions

While there are many envelope functions that satisfy Eq. (1), we adopt the following envelope function pairs,

- $\sqrt{t}$ Window method (method D)

$$\begin{aligned} f(t) &= \sqrt{t} \\ g(t) &= \sqrt{1-t} \end{aligned} \tag{2}$$

- Cosine window method (method E)

$$\begin{aligned} f(t) &= \cos(\pi t) \\ g(t) &= \sin(\pi t) \end{aligned} \tag{3}$$

- Fourier series window method (method F)

Moreover, we take the following unknown envelope function pair as a Fourier series expression.

$$\begin{aligned} f(t) &= \sum_{k=0}^{N-1} a_k \cos(k\pi t) \\ g(t) &= f(1-t) \end{aligned} \tag{4}$$

We assume that the envelope function pair satisfies the boundary conditions and the sum-square-constant conditions. Those conditions are shown as

Boundary conditions:

$$\begin{aligned} f(0) &= 1 \\ f(\pm 1) &= 0 \end{aligned} \tag{5}$$

Sum-square-constant conditions:

$$\begin{aligned} f(1/2) &= \sqrt{2}/2 \\ f^2(1/4) + g^2(1/4) &= 1 \end{aligned} \tag{6}$$

The Fourier series is decided by Eq. (4), Eq. (5), and Eq. (6); $N = 4$, because there are four linearly independent

equations.

$$
\begin{aligned}
a_0 &= \frac{1+\sqrt{2}}{4} \\
a_1 &= \frac{1}{4}\left(1+\sqrt{\frac{5-2\sqrt{2}}{2}}\right) \\
a_2 &= \frac{1-\sqrt{2}}{4} \\
a_3 &= \frac{1}{4}\left(1-\sqrt{\frac{5-2\sqrt{2}}{2}}\right)
\end{aligned}
\tag{7}
$$

$$
\begin{aligned}
f(t) &= \left(1+\sqrt{2}\right)/4 \\
&\quad + (1/4)\left(1+\sqrt{(5-2\sqrt{2})/2}\right)\cos(\pi t) \\
&\quad + \left\{(1-\sqrt{2})/4\right\}\cos(2\pi t) \\
&\quad + (1/4)\left(1-\sqrt{(5-2\sqrt{2})/2}\right)\cos(3\pi t) \\
g(t) &= \left(1+\sqrt{2}\right)/4 \\
&\quad - (1/4)\left(1+\sqrt{(5-2\sqrt{2})/2}\right)\cos(\pi t) \\
&\quad + \left\{(1-\sqrt{2})/4\right\}\cos(2\pi t) \\
&\quad - (1/4)\left(1-\sqrt{(5-2\sqrt{2})/2}\right)\cos(3\pi t)
\end{aligned}
\tag{8}
$$

Figure 7 shows the envelope functions $f(t)$ of each method. The Fourier series window method varies more smoothly than the $\sqrt{t}$ Window method or the Cosine window method.

4. OBJECTIVE EVALUATION

4.1. Causes of Wave Discontinuity

Figure 8 shows an example of moving sound image localization. The instantaneous amplitude of the input signal changes smoothly, and the impulse responses ssir_1, ssir_2 are ideal. T denotes the time difference between the impulse responses, and A_1 and A_2 denote the amplitudes of ssir_1 and ssir_2, respectively.

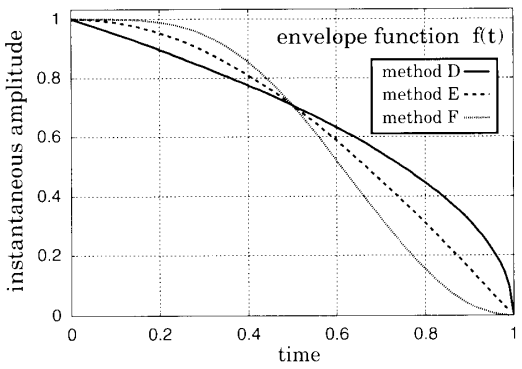

Fig. 7 The curves of envelope function $f(t)$.

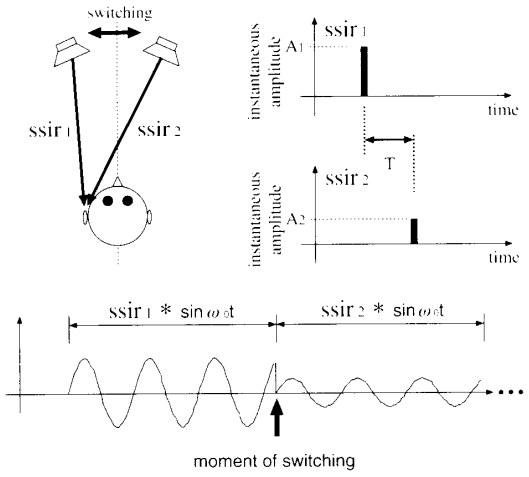

Fig. 8 Causes of wave discontinuity.

From this figure, a wave discontinuity is created by amplitude difference $A_1 - A_2$ and time difference T.

Furthermore, the frequency spectrum expands at the moment of switching. This implies that the wave discontinuity can be evaluated from the width of the frequency spectrum.

4.2. Spectrum Distortion Width σ_ω

Section 4.1. indicated that the width of the frequency spectrum can be used to evaluate wave discontinuities.

For this we adopt the bandwidth equation described by Cohen [15]. This bandwidth equation represents "efficient bandwidth" in the frequency domain, and can evaluate the amplitude difference and time difference yielded by the switching of transfer functions.

The bandwidth equation is transformed from continuous time to discrete time as follows.

$$
\sigma_\omega^2 = \sum_{k=0}^{N-1}(k - \langle k\rangle)^2 |S(k)|^2 \tag{9}
$$

k is a discrete frequency, N is a discrete bandwidth, $\langle k\rangle$ is a mean frequency [15]. $|S(k)|$ is the frequency amplitude characteristic, its squared sum is normalized. The mean frequency is defined by

$$
\langle k\rangle = \sum_{k=0}^{N-1} k \cdot |S(k)|^2 \tag{10}
$$

The mean frequency represents the centroid of the energy distribution in the frequency domain.

In this study, the output of the bandwidth equation is identified as the Spectrum Distortion Width (SDW).

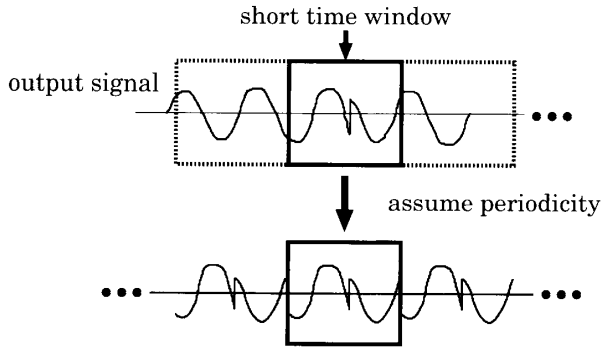

Fig. 9 Wave discontinuity created by assuming periodicity of the output signal.

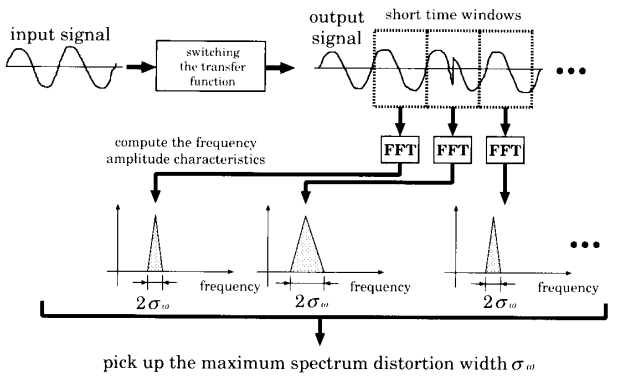

Fig. 10 Schematic diagram of the objective evaluation measure.

4.3. Problems in Computing Spectrum Distortion Width

To compute the SDW, first, the output signal is segmented using short time windows. However, there are two problems in computing SDW.

4.3.1. Time position dependency of the short time window

SDW depends on the time position of the short time windows, because the windowed waveform varies as the time positions of the windows are shifted.

4.3.2. Wave discontinuity caused by assuming signal periodicity

Because the windowed signals are subjected to FFT (Fast Fourier Transform) to compute the SDW, it is assumed that the signal is periodic in the time domain. However, the signal within the window that includes the wave discontinuity is not periodic as is shown in Fig. 9. Consequently, the wave discontinuity cannot be evaluated.

The above argument suggests that SDW values are rather arbitrary. The influence of the time position dependency is reduced if the input signal is periodic and the length of the short time window equals an integral multiple of the input signal period. One idea is to use a pure tone as the input signal; a pure tone satisfies the condition that the length of the short time window is an integer multiple of the input signal period.

The length of the short time window is set as a power-of-two for FFT as is shown in Eq. (11).

$$2^m/f_s = r/f_p \tag{11}$$

2^m is the length of the short time window, f_s denotes the sampling frequency, r is an integer, and f_p denotes the frequency of the pure tone.

In addition, the SDW is computed for several short time window lengths.

4.4. Procedure of Computing the Maximum Spectrum Distortion Width

Our procedure for computing the SDW is shown in Fig. 10. It provides an objective evaluation measure since we use the maximum value of the SDW, hereafter we call it Maximum Spectrum Distortion Width σ_ω max (MSDW).

4.5. Analysis Conditions

A moving sound image which alternates between two front positions is created.

The moving angle is altered, because the characteristics created by switching transfer functions depend on the moving angle. The computational conditions are listed in Table 1.

Table 1 Analysis conditions.

Method	A	B	C	D, E, F
Frame length	2,048 taps			
Frame shift	2,048 taps		256 taps	2,048 taps
Switching time	8,092 taps (170 ms)			
Fade-in· fade-out time				2,048 taps (43 ms)
Pure tone length	48,000 taps (1s)			
Pure tone frequency	$f_p = (48000/256) \cdot r$ [Hz] $r = 1$–80			
Moving angle	10° (355°–5°), 20° (350°–10°) 30° (345°–15°), 40° (340°–20°) 50° (335°–25°), 60° (330°–30°)			
Short time window	rectangular			
Short time window length	256 taps, 512 taps 1,024 taps, 2048 taps			
Short time window frame shift	half of frame shift			
Sampling frequency	48 kHz			
HATS	SAMRAI (KOUKEN)			
Distance between sound source and subject	1.5 m			

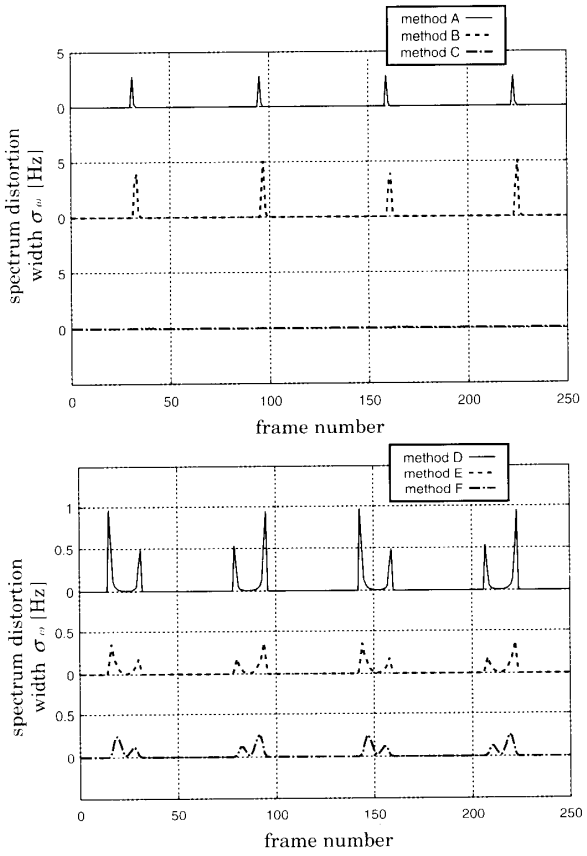

Fig. 11 Calculated SDW values, where the moving angle is 10 (355 –5), the pure tone frequency is 750 Hz and the length of the short time window is 256 taps.

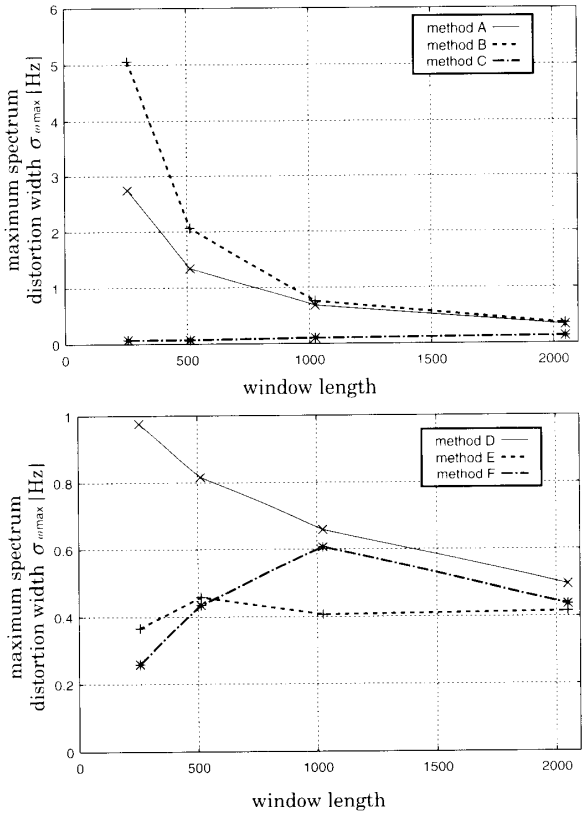

Fig. 12 Relationship between MSDW and the short time window length; the moving angle is 10 (355 – 5), the pure tone frequency is 750 Hz.

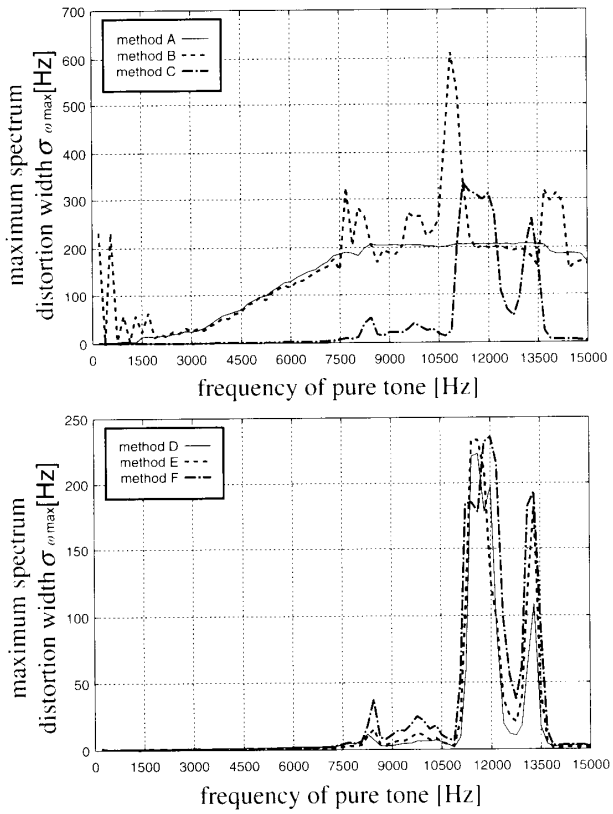

Fig. 13 Relationship between MSDW and pure tone frequency: the moving angle is 10 (355 –5).

4.6. Results

4.6.1. Calculated SDW values

Some calculation examples are shown in Fig. 11. The results indicate that method A and method B yield more wave discontinuity than the other methods.

4.6.2. Relationship between MSDW and short time window length

Figure 12 plots MSDW versus short time window length for the 6 methods. The results indicate that 1) the MSDW falls as the short time window length increases, 2) the MSDW values yielded by the methods are quite different if the short time window length is short, especially in the case of 256 taps.

256 taps are adopted as the short-time window length, because this setting emphasizes the differences between the methods.

4.6.3. Relationship between MSDW and pure tone frequency

Figures 13 and 14 plot MSDW values versus pure tone frequency. The results show that 1) methods A and B yield larger MSDW than the other methods, 2) method B has larger MSDW than method A, 3) the fade-in-fade-out method has smaller MSDW than method C.

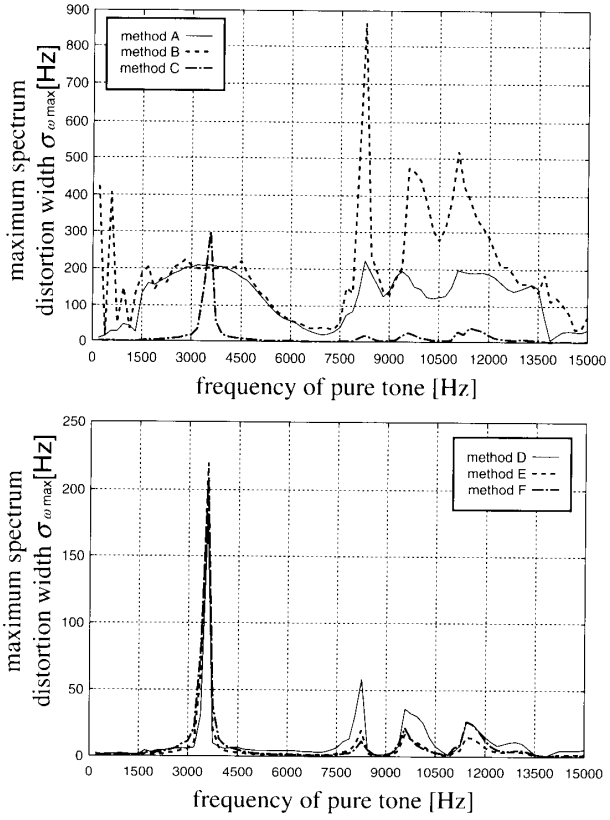

Fig. 14 Relationship between MSDW and pure tone frequency: the moving angle is 30° (345°–15°).

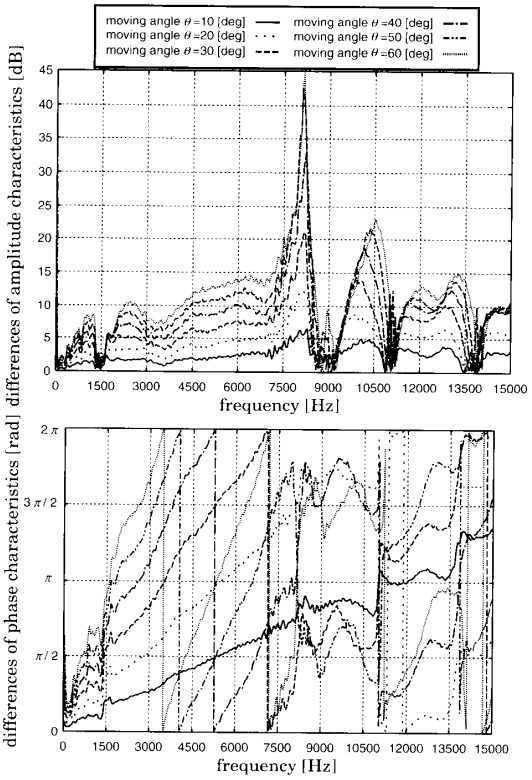

Fig. 15 AD and PD values where the PD values are limited to 2π.

4.6.4. Wave discontinuity yielded by the fade-in·fade-out method

This section discusses the wave discontinuity yielded by the fade-in·fade-out method. Our assumption is that wave discontinuity is related to the amplitude difference and the phase difference created by the switching of transfer functions. The amplitude difference (*AD*) and the phase difference (*PD*) are given as follows.

$$AD = \left| 20 \log_{10} \left| \frac{\mathrm{SSTF}_1(k)}{\mathrm{SSTF}_2(k)} \right| \right| \tag{12}$$

$$PD = |\theta_1(k) - \theta_2(k)| \tag{13}$$

where k is discrete frequency, $\mathrm{SSTF}_1(k)$, $\mathrm{SSTF}_2(k)$ are switching transfer functions, which are represented by

$$\mathrm{SSTF}_1(k) = |\mathrm{SSTF}_1(k)| \cdot e^{j\theta_1(k)}$$

$$\mathrm{SSTF}_2(k) = |\mathrm{SSTF}_2(k)| \cdot e^{j\theta_2(k)}$$

Figure 15 plots the AD and PD values versus the pure tone frequency. This figure indicates that AD is very large at 7,875 Hz but very small at 1,500 Hz, where the influence of PD is strong. These two pure tone frequencies (1,500 Hz and 7,875 Hz) were used in calculating the impact of moving angle on the MSDW of methods D, E, and F, see Fig. 16.

The results indicate that 1) at 1,500 Hz, the differences between the three methods are small until 30° (345°–15°), method F yields the worst MSDW over 30° (345°–15°), 2) at 7,875 Hz, method D (F) yields the worst (smallest)

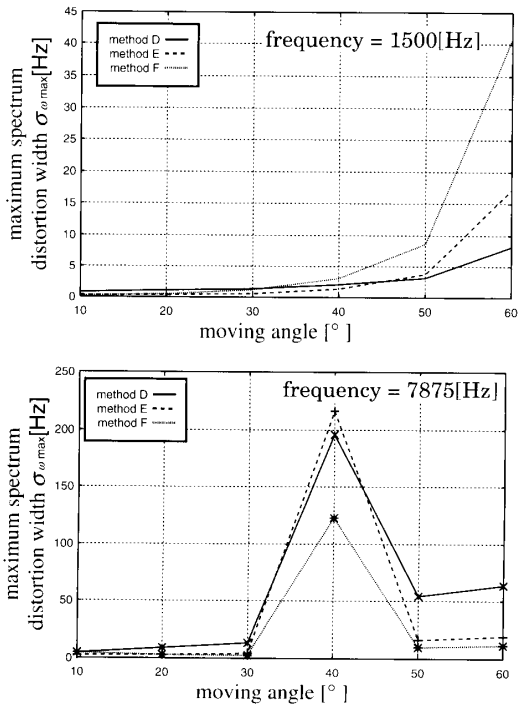

Fig. 16 MSDW versus the moving angle.

MSDW, 3) the MSDW range at 7,875 Hz is wider than that at 1,500 Hz. That is, MSDW is determined more by AD than PD. Therefore, it is concluded that method F is the best fade-in·fade-out method.

5. SUBJECTIVE EVALUATION

5.1. Measurement of the SSTF

The SSTF of several subjects were measured in a test room: a rectangular soundproofed chamber that reduced outside noise by at least 50 dB. Its reverberation time was about 0.1 s. The procedure used to measure the SSTF is described below:

Each subject sat on a seat equipped with a headrest in the test room. Next, an M-sequence signal (100 Hz–15 kHz) was radiated from each loudspeaker. The sound source azimuth directions were 5°, 15°, 30°, 330°, 345°, 355°, and the source distance from the subject to the face of each loudspeaker was 1.45 m. The loudspeakers (Soundevice SD-0.6 models) were level with the subject's ears. The sound pressure level of the signal was 65 dB at the entrance of the subject's ear canal. Miniature microphones (RION UC-92 types) were inserted about 5 mm into the entrance of the subject's ear canal. The outputs of the microphones were converted into a 16-bit linear pulse code with a 48 kHz sampling frequency by an A/D converter (SDS μ DASBOX 16A). Finally, the impulse responses of the SSTF were calculated using the Hadamard conversion method. These SSTF were corrected to match the characteristics of the headphones (SONY MDR-ED238) and loudspeakers [16].

5.2. Subjects

Eight males aged between 22 and 24 years participated in the experiment; none had any history of hearing problems of any kind.

5.3. Stimuli

The stimuli are described in Table 2, where the sound pressure level was adjusted at the entrance of the ear canal of a HATS (KOKEN SAMRAI).

The sound source was transformed into the output signal by the switching transfer function methods. As in the objective evaluation, the output signal was made to oscillate between two positions. All parameters are shown in Table 3.

Table 2 Stimuli parameters.

Male voice	Duration time	about 5.2 s
	Sound pressure level	60–65 dB
Female voice	Duration time	about 4.4 s
	Sound pressure level	60–66 dB
Musical sound	Duration time	about 5.2 s
	Sound pressure level	63–66 dB

Table 3 Parameters of the output signals.

Method	A	B	C	D, E, F
Frame length	2,048 taps			
Frame shift	2,048 taps		256 taps	2,048 taps
Switching time	24,576 taps (about 0.5 s)			
Fade-in·fade-out time				2,048 taps (43 ms)

The output signals were digitally synthesized on a computer (Sun SparcStation 10) with a sampling frequency of 48 kHz and 16-bit quantization. The signals were converted from digital to analog by a D/A converter (SDS μ DASBOX 16A), and lowpass filtered (cut-off frequency = 20 kHz).

5.4. Procedure

Each subject sat on a seat in the test room wearing the headphones (SONY MDR-ED238) as shown in Fig. 17. A test signal and two evaluation signals, which were created by the switching transfer function methods, were reproduced via the headphones (see Fig. 18).

The subject's task was to compare the click noises in the two test signals using a five grade comparison scale, see Table 4. These assessment procedures are based on Scheffe's pair comparison method.

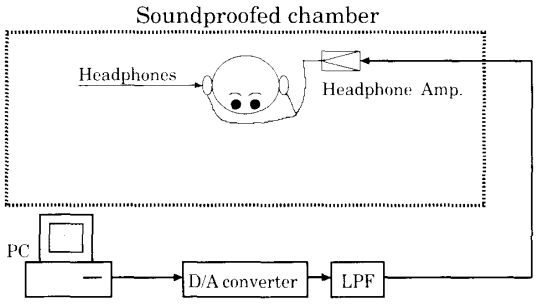

Fig. 17 Experimental system.

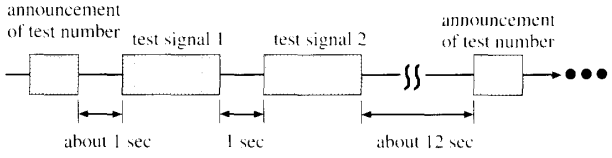

Fig. 18 Layout of the test signals. The test number preceded the evaluation pair.

Table 4 Five-grade comparison scale used in Scheffe's comparison test. The evaluation presented the test sound first (see Fig. 18).

Grade	Comparison of impairment
2	Degradation is not worrisome at all
1	Degradation is not worrisome
0	Degradation is same level
−1	Degradation is worrisome
−2	Degradation is very worrisome

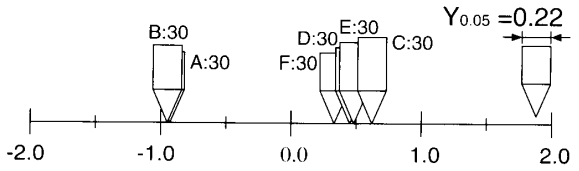

Fig. 19 Results of the preliminary test: male voice.

5.5. Preliminary Test and Its Result

Scheffe's comparison method required ${}_{18}P_2 = 306$ judgments per stimulus for comparing all pairs, since there are six methods and the moving angles are 10° (355°–5°), 30° (345°–15°) and 60° (330°–30°). However, we endeavored to simplify the number of judgments in order to reduce subject's burden. To this end, a preliminary test was performed before the main test. In the objective evaluation, the MSDWs of the fade-in·fade-out methods were very small. If there is no significant difference between the fade-in·fade-out methods, we can take one method from the three fade-in·fade-out methods and simplify the number of judgments. The preliminary test was conducted with 30° (345°–15°) moving angle. Five subjects participated in this test.

The result of a male voice is shown in Fig. 19. Using a female voice and a musical sound yielded quite similar tendencies.

In this figure, X:Y denotes method X and moving angle Y. For example, A:30 denotes that the switching transfer function method was method A and the moving angle was 30° (345°–15°). The position of the arrows corresponds to the average of the score. The width of the arrows represents the 95% confidence interval, and the overlap of arrows means that there is no significant difference between them.

We concluded that method F was the best fade-in·fade-out methods in Section 4.6.4, but there were no significant differences between the fade-in·fade-out methods. This means that the MSDWs were too small and the click noises yielded by the fade-in·fade-out methods were not perceivable. Because there were no significant differences between the methods, method F is adopted as the fade-in·fade-out method in the main test.

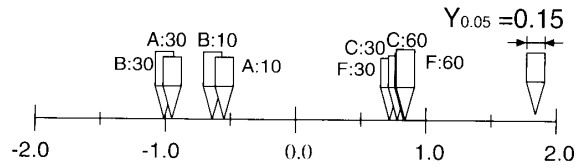

Fig. 20 Results of the main test: male voice.

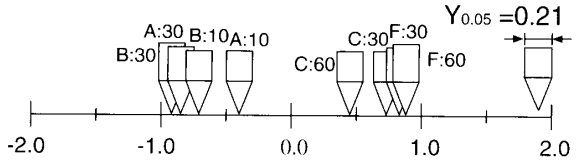

Fig. 21 Results of the main test: female voice.

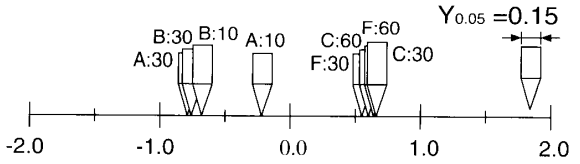

Fig. 22 Results of the main test: music.

5.6. Results of Main Test

The results of the main test are shown in Figs. 20–22. The interpretation results of the data are similar to those of the preliminary test.

The results in Figs. 20–22, indicate that 1) as for methods C and F, there is little sound quality degradation, and there is no significant difference between them, 2) as for methods A and B, the sound quality degrades only slightly as the moving angle decreases.

6. DISCUSSION

To examine the validity of the objective evaluation, this section considers the correspondence of objective evaluation and subjective evaluation via the moving angles and the switching transfer function methods.

6.1. Moving Angle

The objective evaluation obtained different tendencies of the MSDWs in each moving angle, so we separately considered the MSDWs at the boundary of 7.5 kHz.

In the objective evaluation, the MSDWs of 30° (345°–15°) moving angle were larger than that of 10° (355°–5°) moving angle for all methods, at 7.5 kHz or less. The MSDWs were nearly identical for methods A and B, except for specific frequencies, and the MSDWs of 10° (355°–5°) moving angle are larger than that of the 30° (345°–15°) moving angle for methods C, D, E, and F, at 7.5 kHz or more.

In the subjective evaluation, the click noises were perceived less easily as the moving angle decreased for methods A and B. Consequentially, good correspondence between objective and subjective evaluations was obtained, except for the musical sound due to its broader bandwidth. While the female voice had a bandwidth of 8 kHz, the musical sound has a broader bandwidth, and the differences of the MSDWs between methods A and B is small at those frequencies. On the other hand, there is no significant difference for methods C and F, except for method C with 60° (330° –30°) moving angle. The click noises couldn't be perceived, since the MSDWs was very small at each moving angle.

6.2. Switching Transfer Function Methods

From the objective evaluation we can rank the methods in order of decreasing MSDWs as, method B, method A, method C, then methods D, E and F.

The subjective evaluation showed that methods A and B yielded larger click noises than the other methods, and there was a significant difference between methods A and B with 10° (355° –5°) moving angle. In contrast, there was no significant difference between them with 30° (345° –15°) moving angle. The subjects could not discriminate the difference in click noises between methods A and B, since the MSDWs were too large at 30° (345° –15°) moving angle.

There was no significant difference between methods C and F, except for the female voice. This means that the subjects could not perceive the click noises, because the difference of the MSDWs between methods was very small.

6.3. Overall Correspondence of Subjective and Objective Evaluation

The objective evaluation indicated that MSDW increased with the moving angle for all methods. The subjective evaluation, on the other hand, matched this characteristic only for methods A and B.

7. CONCLUSION

This paper studied the impact of the moving sound image scheme used on the wave discontinuity created when switching the scheme to achieve moving sound images that, as the final goal, equal the quality of the images created with loudspeakers.

First, the switching transfer function schemes were grouped into the simple switching approach, the overlap-add approach, and the fade-in·fade-out approach. We created an algorithm for each approach.

For the overlap-add approach, the normal overlap-add method and the overlap-add method with modified hamming window were examined. For the fade-in·fade-out approach, the $\sqrt{t}$ Window method, the Cosine window method, and the Fourier series window method were examined.

Second, the sound quality degradation caused by wave discontinuity was assessed; spectrum distortion width was used to quantify the wave discontinuity. Scheffe's paired comparison tests were carried out as subjective assessments. The results are as follows.

Objective evaluation

The overlap-add method with modified hamming window (method C) and the fade-in·fade-out method (methods D, E, and F) exhibited only slight levels of wave discontinuity. The simple switching method (method A) had less wave discontinuity than the overlap-add method (method B). Among the fade-in·fade-out methods (methods D, E, and F), the Fourier series window method (method F) created less wave discontinuity than the others.

Subjective evaluation

The overlap-add method with modified hamming window (method C) and the Fourier series window method (method F) had less sound quality degradation than the other methods. Furthermore, there was no significant difference between them. In the simple switching method (method A) and the normal overlap-add method (method B), the sound quality degraded only slightly as the moving angle was increased.

REFERENCES

[1] D. R. Perrott and T. Z. Strybel, "Some observations regarding motion without direction," in *Binaural and Spatial Hearing in Real And Virtual Environments*, R. H. Gilkey and T. R. Anderson, Eds. (Lawrence Erlbaum Associates Publishers, Mahwah, N.J., 1997), Chap. 14, pp. 275–294.

[2] T. Nakajima, K. Tamaribuchi and S. Saito, "Perception of the motional image induced from distributed sound sources excited sequentially," *Trans. Tech. Comm. Psychol. Physiol. Acoust.*, H-90-6 (1990).

[3] I. Kinoshita and S. Aoki, "Effects of source signals on the perception of continuity for sequentially reproduced sound," *Trans. Tech. Comm. Psychol. Physiol. Acoust.*, H-93-2 (1993).

[4] K. Mizushima, S. Nakanishi and M. Morimoto, "Continuity of sound image caused by successive presentation of stimulus in the horizontal plane," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 437–438 (1992).

[5] F. L. Wightman and D. J. Kistler, "Resolution of front-back ambiguity in spatial hearing by listener and source movement," *J. Acoust. Soc. Am.*, **105**, 2841–2853 (1999).

[6] T. Nishino, S. Kajita, K. Takeda and F. Itakura, "Interpolation of the head related transfer function on the horizontal plane," *J. Acoust. Soc. Jpn. (J)*, **55**, 91–99 (1999).

[7] H. Uematsu, M. Kato and M. Kashino, "The influence of sound source movement on the extracranial localization," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 403–404 (2000).

[8] D. Kimura and Y. Suzuki, "A consideration about the effect of head movement on the sound localization," *IEICE Tech. Rep.*, EA2001-44, pp. 57–64 (2001).

[9] D. Katsumi, Y. Watanabe and H. Hamada, "The effects of HRTF on elevational localization by virtual sound imaging,"

Proc. Spring Meet. Acoust. Soc. Jpn., pp. 581–582 (2002).
[10] M. Matsumoto, M. Tohyama and H. Yanagawa, "A method of interpolating binaural impulse responses for moving sound images," *Acoust. Sci. & Tech.*, **24**, 284–292 (2003).
[11] W. R. Thurlow, J. W. Mangels and P. S. Runge, "Head movements during sound localization," *J. Acoust. Soc. Am.*, **42**, 489–493 (1967).
[12] M. Kato, H. Uematsu, M. Kashino and T. Hirahara, "The effects of head motion on human sound localization," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 505–506 (2001).
[13] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, N.J., 1975).
[14] D. W. Griffin and J. S. Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Trans. Acoust. Speech Signal Process.*, **ASSP-32**, 236–242 (1984).
[15] L. Cohen, *Time-Frequency Analysis* (Prentice-Hall, Englewood Cliffs, N.J., 1995).
[16] S. Yano, H. Hokari, S. Shimada and H. Irisawa, "A study on the transfer function of sound localization using binaural earphones," *AES the 104th Convention*, Amsterdam, no. 4658 (1998).

APPENDIX SUPPLEMENT ON THE DISTANCE DIFFERENCE

Figure A.1 shows the geometric relationship between the subject and loudspeakers. r is the radius of the subject's head, R is the distance from the center of the head to a loudspeaker, D_1 and D_2 are distances from the right side ear to the two loudspeakers, respectively, $\Delta\theta$ is the moving angle. The figure indicates the stimulus signal created when the active loudspeaker was moved from A_1 to A_2 with $\Delta\theta$ angle. θ is the azimuth of loudspeaker A_1. Now consider the distance difference $D_1 - D_2$ that depends on θ; for this we introduce function $p(\theta)$ normalized by $\sqrt{R^2 + r^2}$

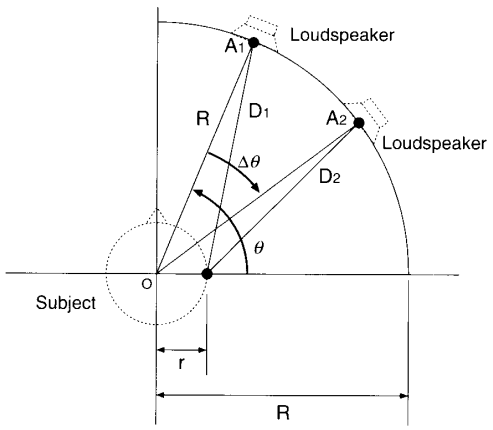

Fig. A.1 Geometric relation between subject and loudspeaker movement.

$$p(\theta) = (D_1 - D_2)/\sqrt{R^2 + r^2} = \sqrt{1 - \frac{2Rr}{R^2 + r^2}\cos\theta} - \sqrt{1 - \frac{2Rr}{R^2 + r^2}\cos(\theta - \Delta\theta)} \quad \text{(A·1)}$$

Where the distances D_1 and D_2 are as follows

$$D_1 = \sqrt{R^2 + r^2 - 2Rr\cos\theta}$$
$$D_2 = \sqrt{R^2 + r^2 - 2Rr\cos(\theta - \Delta\theta)}$$

If $\Delta\theta$ is a fractional moving angle, we get

$$\begin{aligned} p(\theta) &= \sqrt{1 - a\cos\theta} - \sqrt{1 - a\cos(\theta - \Delta\theta)} \\ &= \sqrt{1 - a\cos\theta} - \sqrt{1 - a(\cos\theta\cos\Delta\theta + \sin\theta\sin\Delta\theta)} \\ &\approx \sqrt{1 - a\cos\theta} - \sqrt{1 - a(\cos\theta + \Delta\theta\sin\theta)} \\ &= \sqrt{1 - a\cos\theta}\left(1 - \sqrt{1 - (a\Delta\theta\sin\theta)/(1 - a\cos\theta)}\right) \\ &= \sqrt{1 - a\cos\theta}\left(1 - \sqrt{1 - q(\theta)}\right) \end{aligned} \quad \text{(A·2)}$$

Where

$$a = 2Rr/(R^2 + r^2)$$
$$q(\theta) = a\Delta\theta\sin\theta/(1 - a\cos\theta), \quad |q(\theta)| \le 1$$

From Eq. (A·2), the distance difference $p(\theta)$ depends on $q(\theta)$. If the $q(\theta)$ takes a maximum value, the distance difference $p(\theta)$ is maximized. Therefore, consider a differential coefficient of $q(\theta)$,

$$q'(\theta) = a\Delta\theta(\cos\theta - a)/(1 - a\cos\theta)^2 \quad \text{(A·3)}$$

That is, the distance difference has a maximum value at $0 \le \theta \le \pi/2$, where θ equals $\cos^{-1}(a)$. In the case of $R = 1.45$ m and $r = 0.085$ m, we calculate

$$\theta = \cos^{-1}\{2Rr/(R^2 + r^2)\} = \cos^{-1}(0.1168) = 83^\circ \quad \text{(A·4)}$$

Therefore, it is concluded that the distance difference has a maximum value when the center of the two loudspeakers is located at or near the front.