PAPER

Sound localization with multi-loudspeakers by usage of a coincident microphone array

Jun Aoki, Haruhide Hokari* and Shoji Shimada

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

Abstract: We examine multi-channel microphone arrangements to achieve precise and stable sound image localization in the horizontal plane when multi-loudspeakers are used. In this paper, six different coincident microphone arrays, cardioid microphones with different directions, are tested. We derive equations to model the system and define a system evaluation measure. The sound localization assessment shows that our equations approximately agree with the assessment results, and that the system evaluation measure must suit the microphone arrangement used. These results confirm that while the perception of lateral localization is difficult, three of the six arrays provide good sound localization. Last, we clarify that the coincident microphone array can also provide stable sound localization in multi-channel recording.

Keywords: Coincident microphone array, Perception of sound image localization, Cardioid microphone

PACS number: 43.38.Md, 43.38.Vk, 43.66.Qp   [DOI: 10.1250/ast.24.250]

1.   INTRODUCTION

Many multi-channel reproduction systems have been researched. In particular, many papers have discussed multi-channel loudspeaker arrangements for sound image reproduction systems that use the head-related transfer function (HRTF) and for HDTV systems [1,2]. Several loudspeaker arrangements for multi-channel stereophonic sound system have been published recently, see Recommendation ITU-R BS. 775-1 [3]. These arrangements are compatible with one another and so have been widely applied in areas such as DVD, HDTV, and digital film sound. It is certain that their application will involve the use of multi-channel reproduction systems (i.e. multiple loudspeakers). This means that multi-channel microphone arrangements must be optimized.

Over the last few years, several papers have examined microphone arrangements for multi-channel sound recording. To create truly effective multi-channel sound recording system, various recording factors such as directional stability, spatial impression, depth, and ambient atmosphere must be considered. We have focused on directional stability, an important factor in sound (image) localization, and are examining multi-channel microphone arrangements

to achieve precise and stable sound image localization in the horizontal plane to support the use of multi-loudspeakers. While several multi-channel microphone arrangements such as Fukada-Tree and OCT-Surround [4] have been proposed, they emphasize spatial impression and depth as well as directional stability. Their aim slightly differs from ours because they demand an extremely stable frontal image. Furthermore, since these arrangements use spaced microphones, their recording signal outputs have not only level difference but also phase difference. It has been reported that the direction of the wavefront created in two-channel ($2/0 : X/Y$ represents loudspeaker arrangement where X is the number of front loudspeakers, Y is the number of back loudspeakers.) stereo varies with the frequency of the sound source when signals that have phase difference are recreated by loudspeakers [5]. However, the coincident microphone array has in phase outputs if the distance between the sound source and the array is sufficiently long. Also for 2/0 stereo, reports indicate that in phase signals can accurately regenerate real sound sources [6]. Furthermore, it is well known that a coincident pair of microphones can provide more stable sound localization than a spaced pair of microphones in two-channel recording (see, for example, [7]). These facts imply that existing multi-channel microphone arrangements do not well regenerate real sound sources and that a

*e-mail: hokari@vos.nagaokaut.ac.jp

coincident microphone array can provide stable sound localization in multi-channel recording too. However, more investigation is needed to confirm these ideas.

This paper addresses the perception of sound localization. We examine several coincident cardioid arrays and derive equations to model the system. Sound sources are recorded by the arrays and recorded signals are reproduced by multi-loudspeakers; we define a system evaluation measure. We weigh our equations and the system evaluation measure against the sound localization assessment results. The loudspeaker arrangements in our study are based on 3/2 stereo, which is a recommended reference loudspeaker arrangement for multi-channel stereophonic sound systems according to Recommendation ITU-R BS. 775-1.

2. DERIVATION OF EQUATIONS

While the direct approach is to find the optimum microphone arrangement by conducting actual trials, the time and effort involved in examining all possible arrangements makes this impractical. This problem can be easily resolved by deriving theoretical equations that can model the system. Our approach is to extend the equations used to model the reproduction side to cover the recording side as well.

2.1. Reproduction Side Equations

2.1.1. Equations at low frequencies

Leakey [8,9] assumed that, at low frequencies, sound localization mainly depends on the interaural time difference (ITD) and that, in 2/0 stereo as shown in Fig. 1(a), if the ITD produced by the two loudspeaker signals equals the ITD produced by the real source, the ITD produced by the two loudspeaker signals creates a sound image on the
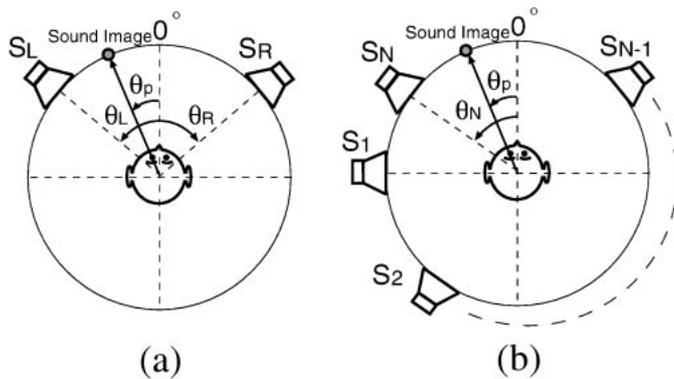
Fig. 1 Sound image reproduction systems. The interaural time difference (ITD) produced by the loudspeaker signals creates a sound image on the direction of θ_p. (a) Two-channel reproduction system. 2 loudspeakers are equidistant from the listener. 2/0 stereo represents this loudspeaker arrangement. (b) Multichannel reproduction system. N loudspeakers are equidistant from the listener.

direction of the real source. Leakey derived the following equation:

$$\sin\theta_p = \frac{L\sin\theta_L + R\sin\theta_R}{L + R} \tag{1}$$

where θ_L and θ_R are the azimuth angles and L and R are the signal amplitudes of the left and right loudspeakers $S_{L,R}$, respectively; θ_p represents the perceived angle of the sound image.

In Fig. 1(b), Bernfeld [10] extended Eq. (1) to cover multi-channel loudspeakers; he derived the following equation:

$$\sin\theta_p = \frac{\sum_{i=1}^{N} A_i \sin\theta_i}{\sum_{i=1}^{N} A_i} \tag{2}$$

where $\theta_i(i = 1, 2, \cdots, N)$ is the azimuth angle and $A_i(i = 1, 2, \cdots, N)$ is the signal amplitude of loudspeaker $S_i(i = 1, 2, \cdots, N)$.

2.1.2. Equations at high frequencies

At high frequencies, Leakey [8,9] also emphasized the ITD of the slowly varying envelope function of the sound waveform and derived the following equation:

$$\sin\theta_p = \frac{L^2\sin\theta_L + R^2\sin\theta_R}{L^2 + R^2} \tag{3}$$

According to Takahashi *et al.* [11], Eq. (3) is applicable to wide-band signals as well as high-frequency signals because this equation well agrees with sound localization assessment results gained with white noise ($\sim$20 kHz) in asymmetric loudspeaker arrangement to the median plane.

Taking this report as our base, we extended Eq. (3) by applying Leakey's and Bernfeld's theories to derive the following equation:

$$\sin\theta_p = \frac{\sum_{i=1}^{N} A_i^2 \sin\theta_i}{\sum_{i=1}^{N} A_i^2} \tag{4}$$

The following assumptions are implicit in Eqs. (1)–(4).

- The loudspeakers of the reproduction system are equidistant from the listener.
- The listener faces the front (i.e., 0° direction) and the head is immobile.
- The distance between the loudspeakers and the center of the head is sufficiently long compared to the distance between ears, i.e., arriving sound waves from the loudspeakers at the ears can be regarded as plane waves.
- Loudspeaker signals are in phase but have different amplitude or polarity.

2.2. Equations Covering Both Sides

In 2/0 stereo, Clark *et al.* [12] also emphasized the interaural phase difference (IPD) at low frequencies and derived the same equation as Eq. (1) for the reproduction side. Their analyses made the same assumptions described in Section 2.1. Furthermore, they extended Eq. (1) by replacing the signal amplitudes of the two loudspeakers L and R with the polar equations of a pair of figure-8 microphones respectively, and defined a theoretical equation that also covered the recording side. By the way, a pair of figure-8 microphones, arrayed at a lateral angle of $90°$, forms a coincident array (this array is well known as *the Blumlein array*). Their idea takes advantage of the assumption of Eq. (1), that the loudspeaker signals are in phase, as well as the assumption that the characteristics of a coincident microphone array mean that the coincident microphone outputs are in phase (i.e., the idea holds only under these assumptions).

We apply the above idea and extend Eqs. (1)–(4) by replacing the signal amplitudes of the respective loud-speakers $A_i(i = 1, 2, \cdots, N)$ with the polar equations of cardioid microphones as shown in Eq. (5); the resulting extended equations are theoretical equations that cover both sides. Needless to say, cardioid microphones can form a coincident array.

$$A_i = 0.5 + 0.5\cos(\theta_{M_i} - \theta_r) \qquad (i = 1, 2, \cdots, N) \quad (5)$$

where $\theta_{M_i}(i = 1, 2, \cdots, N)$ is the azimuth angle of microphone $M_i(i = 1, 2, \cdots, N)$ (i.e., direction of maximum sensitivity), and θ_r represents the recorded angle of the real source.

We mainly use two theoretical equations, calculated by Eqs. (2), (5) and Eqs. (4), (5), and call them *low* and *high frequency equation*, respectively.

3. SYSTEM EVALUATION MEASURE

The main purpose of this study is find out multi-channel microphone arrangements that make $\theta_r = \theta_p$ as closely as possible. We introduce a *system evaluation measure* (SEM) to assess arrangement performance.

First, the unsigned error e_t or e_e between the *desired azimuth angle* θ_d, given by $\theta_r = \theta_p$, and the *theoretical azimuth angle* θ_t, given by the low and high frequency equations, or *experimental azimuth angle* θ_e, gained by sound localization assessment results, is defined by

$$\begin{cases} e_t = |\theta_d - \theta_t|, & \textit{if use } \theta_t \\ e_e = |\theta_d - \theta_e|, & \textit{if use } \theta_e \end{cases} \quad (6)$$

where $0° \le e_t(, e_e) \le 180°$; attention must be paid to the sign when calculating $e_t(, e_e)$, as shown by Fig. 2.

It follows that SEM_t and SEM_e are defined as follows;
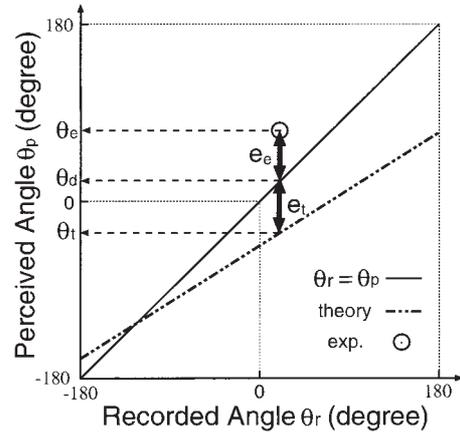
Fig. 2 Example of calculating e_t and e_e. $\theta_r = \theta_p$ is ideal line, *theory* is a sample line given by a theoretical equation and *exp.* is a sample data extracted from sound localization assessment results.

$$\begin{cases} SEM_t = \sqrt{\dfrac{1}{D}\sum_D e_t{}^2} \\[4ex] SEM_e = \dfrac{\displaystyle\sum_M \sqrt{\dfrac{1}{D}\sum_D e_e{}^2}}{M} \end{cases} \quad (7)$$

where D is the number of directions, M is the number of listeners, and $0° \le SEM_t(, SEM_e) \le 180°$. SEM_t and SEM_e represent theoretical and experimental data, respectively.

SEM_t and SEM_e imply the standard deviation between θ_d and θ_t or θ_e over all horizontal directions, so values close to $0°$ indicate that the system (combination of microphone and loudspeaker arrangements) has better performance.

4. RECORDING

4.1. Microphone Arrangements

Six coincident cardioid arrays (p1–p6), see Fig. 3 were examined. The direction of the arrowhead represents the direction of maximum sensitivity of the microphone, so
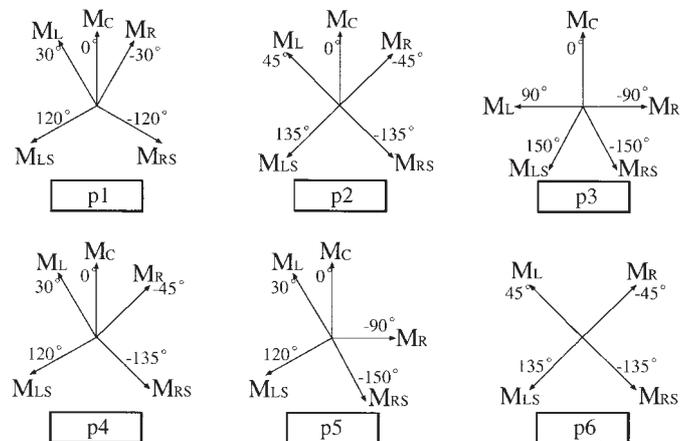
Fig. 3 Coincident cardioid arrays examined.

$\theta_{M_i}(i = 1, 2, \cdots, N)$ of Eq. (5) equals this direction.

These arrangements were selected for the following reasons.

a. Five-Channel (loudspeaker arrangement: 3/2 stereo)

Symmetric arrangement to the median plane: p1, p2, p3

p1 is configured so that the azimuth angles of the microphones equal those of the loudspeakers.

p2 is based on p6 (see Paragraph **b**.). p3 is based on INA5 [13] which aims to provide a recording angle of 360°.

Asymmetric arrangement to the median plane: p4, p5

p4 and p5 are adopted in order to examine the impact of asymmetric arrangements.

b. Four-Channel (loudspeaker arrangement: 2/2 stereo)

Quadraphonic arrangement: p6

p6 is known to be suitable for four-channel stereo since it offers good stereo location [14]. We adopted it in order to examine the effect of the center-channel. Other reasons for adopting p6 are (i) clarification of the cause of sound image elevation found in our previous work [15] and (ii) confirmation of the expectation, four-channel systems give slightly better localization than five-channel systems, raised by the theoretical equations and system evaluation measure (see Section 6).

4.2. Recording Conditions

4.2.1. Recording signals

The recording signals used were three band-limited noise samples (200–600 Hz, 2–15 kHz, 200 Hz–15 kHz) created by limiting white-noise of 20 Hz–20 kHz [16] using LPF and HPF (−135 dB/oct). The signal bands were determined so as to satisfy the concepts of the equations and the characteristics of the microphones and loudspeakers used.

4.2.2. Recording systems

An example of the appearance of the recording system is shown in Fig. 4. Each coincident array, formed by placing the cardioid microphones on the same vertical axis, was placed in an anechoic chamber and surrounded by 24 loudspeakers located at 15° intervals; the loudspeakers output the three band-limited noise samples (200–600 Hz, 2–15 kHz, 200 Hz–15 kHz) to be captured by the arrays. Further details about the recording conditions are given below.

- Distance between loudspeakers and microphones: 2 m
- Height of loudspeakers and microphones: 1.15 m
- Model of loudspeakers: Soundevice MODEL SD-0.6
- Model of microphones: audio-technica ATM15a (cardioid pattern)
- Band-Limit: 15 kHz
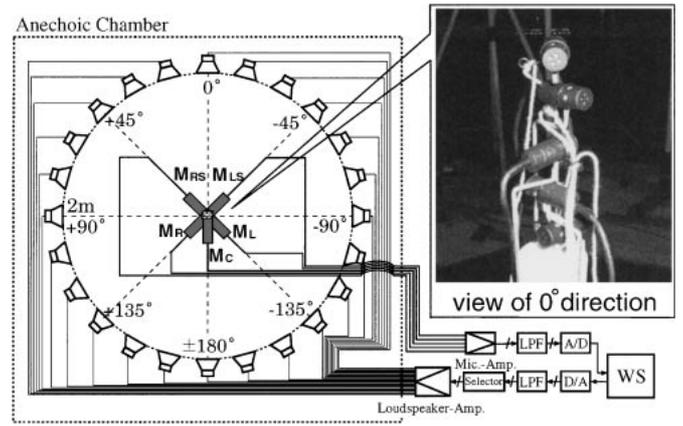
Fig. 4 Recording system for p2.
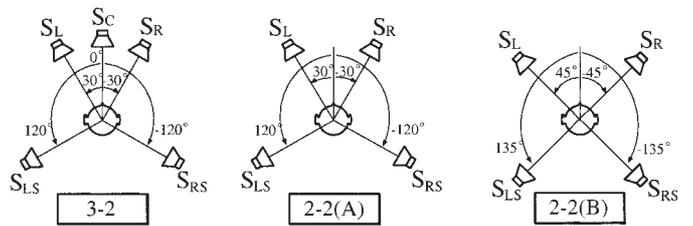
Fig. 5 Loudspeaker arrangements.

5. SOUND LOCALIZATION ASSESSMENT

5.1. Loudspeaker Arrangements

We examined three loudspeaker arrangements for sound localization assessment: ((3-2, 2-2(A), 2-2(B)), see Fig. 5. 2-2(A) and 2-2(B) are taken from Furumi *et al.* [1] while 2-2(A) is equivalent to 2/2 stereo as described in Recommendation ITU-R BS. 775-1.

2-2(A) was adopted in order to examine the effect of the center-channel. According to Furumi *et al.*, 2-2(B) is a suitable arrangement for multi-channel systems that use the HRTF; we examined this arrangement to confirm the effect of not using HRTF.

5.2. Tests

We conducted seven tests (TYPE 1–7), each of which used a different microphone and loudspeaker arrangement, as shown in Table 1.

Table 1 Type of tests.

Mic \ Sp	3-2	2-2(A)	2-2(B)
p1	TYPE 1	—	—
p2	TYPE 2	—	—
p3	TYPE 3	—	—
p4	TYPE 4	—	—
p5	TYPE 5	—	—
p6	—	TYPE 6	TYPE 7

In the tests of TYPE 1–5, recorded signals of the respective microphones $M_{L,C,R,LS,RS}$ were cut into a suitable length and became the input signals of the respective loudspeakers $S_{L,C,R,LS,RS}$ (see Figs. 3, 5 and Section 5.3.1). In the tests of TYPE 6,7, recorded signals of the respective microphones $M_{L,R,LS,RS}$ became the input signals of the respective loudspeakers $S_{L,R,LS,RS}$ in a similar way.

5.3. Test Conditions
5.3.1. Test signals
The test signals (i.e., stimuli) were made by combining three recorded signal segments, as shown in Fig. 6. This arrangement (duration time and repetition number) is based on a report by Yamaji *et al.* [17].
5.3.2. Test systems
The test system is shown in Fig. 7. In the tests, the loudspeakers were hidden by an acoustically transparent curtain from the subject, and markers (1–48) were placed at 7.5° intervals for the subject to refer to: the subject sat in a seat with a headrest and the subject's head was fixed against the headrest. In one trial, 24 (number of recorded signal: 15° intervals) stimuli per signal were presented in random order and the subject was directed to write the marker number of the perceived direction of the sound image on a sheet, ignoring the height of the sound image, spread of the sound image, sound color, etc. All tests (TYPE 1–7) were performed only once (one trial). Therefore, the listening number of times with the same signal per subject was only one. Further details about the test conditions are given below.

- Distance between loudspeakers and center of the subject's head: 2 m
- Subjects: 6 men ranging in age from 22 to 24
- Sound pressure level: 60 dB(A)

The height of loudspeakers, model of loudspeakers, and band-limit are the same as those used for recording (see Section 4.2.2).

Based upon the results of tests of TYPE 1–3 (see Section 5.4.3), the tests of TYPE 6 and 7 were performed using only the 200 Hz–15 kHz signal. These tests were performed immediately after the test of TYPE 5 and the subject answered a question that asked if the subject perceived any change (higher or lower) in the height of the sound image in order to determine if the center-channel influenced the sound image.

5.4. Test Results
The results are shown in Figs. 8–14. Circle size indicates the number of subjects perceiving that sound image direction (i.e., if the circle is large many subjects perceived the same direction). The linear curve of $\theta_r = \theta_p$ and localization curves of the low and high frequency equations, calculated by Eqs. (2), (4) and (5), are also plotted.
5.4.1. TYPE 1–3
In Fig. 8, it is noticed that the result of 200–600 Hz represents many instances of front-back confusion, or vice versa, for the low frequency equation; the result of 2–15 kHz inidicates data spread. Contrary to these results, the result of 200 Hz–15 kHz indicates stable localization with
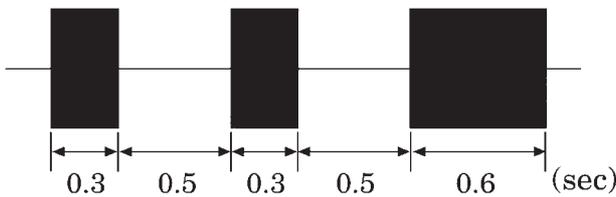
Fig. 6   Test signal.
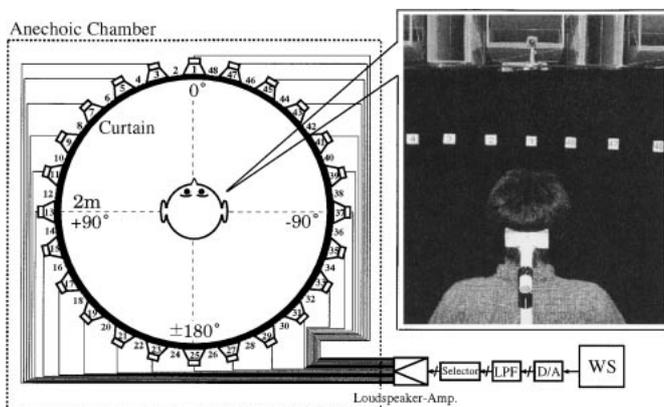
Fig. 7   Test system.
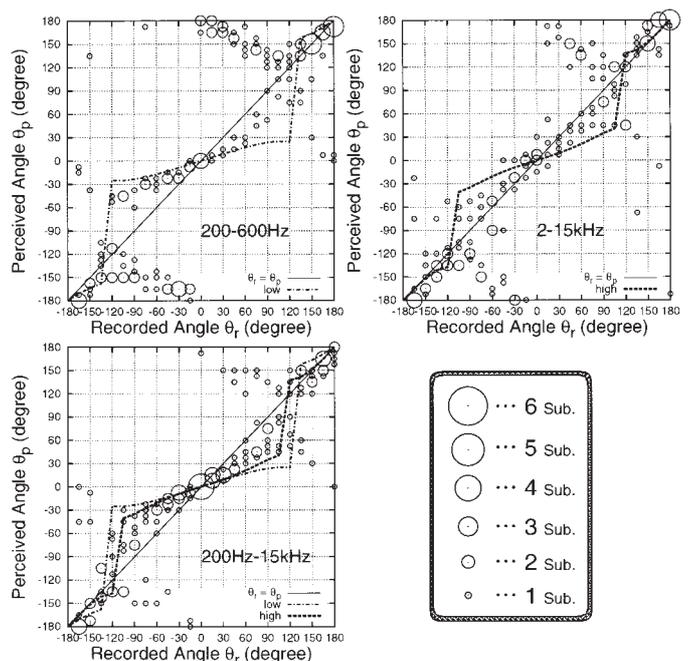
Fig. 8   Results of TYPE 1.
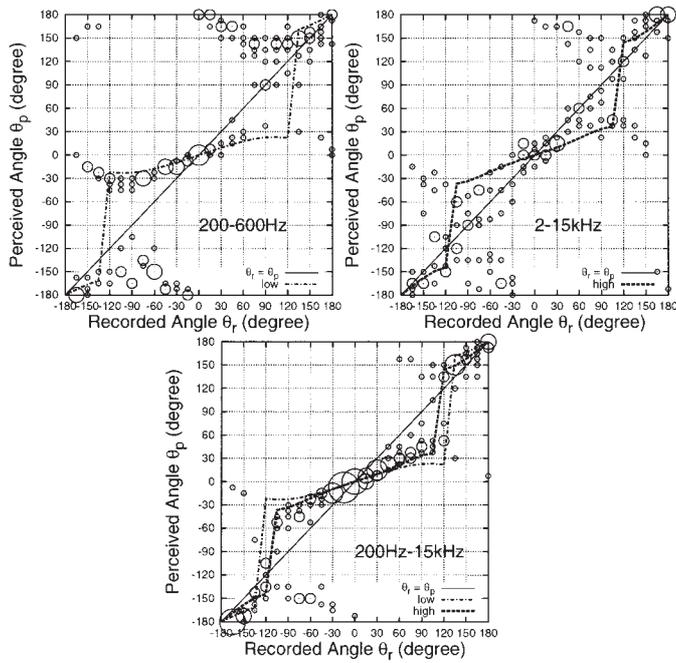
Fig. 9 Results of TYPE 2.
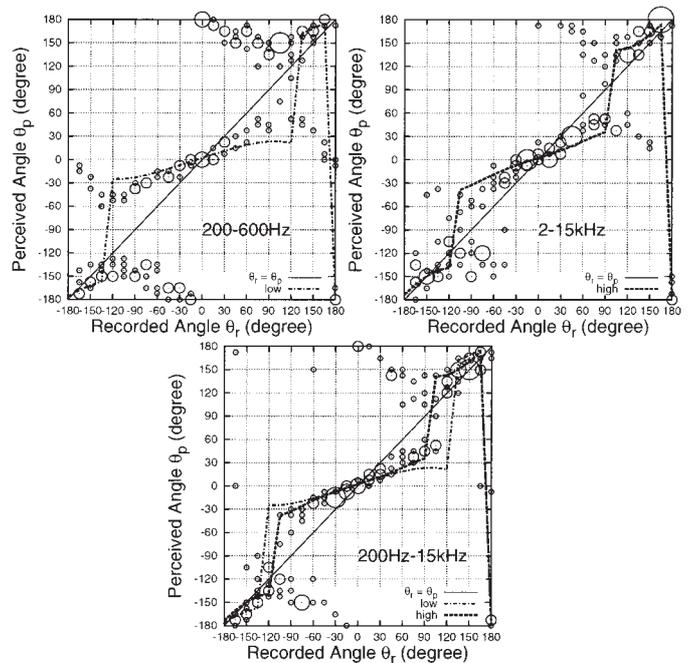
Fig. 11 Results of TYPE 4.
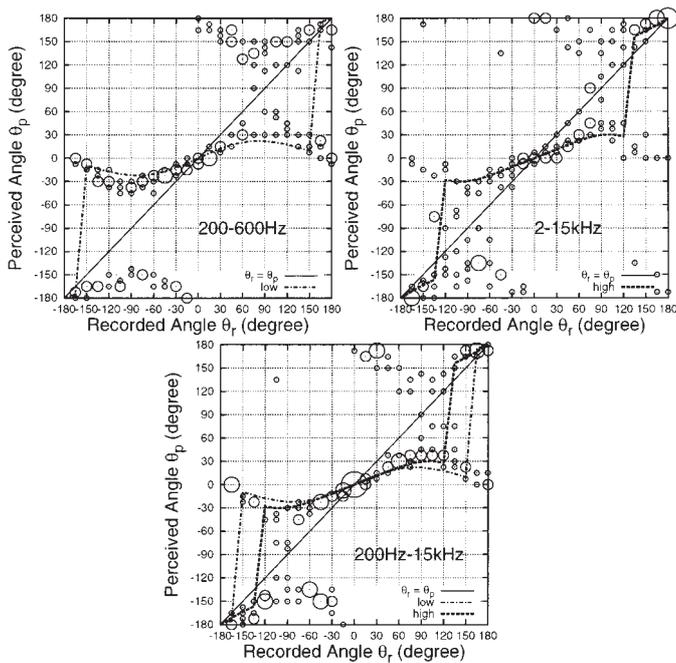
Fig. 10 Results of TYPE 3.
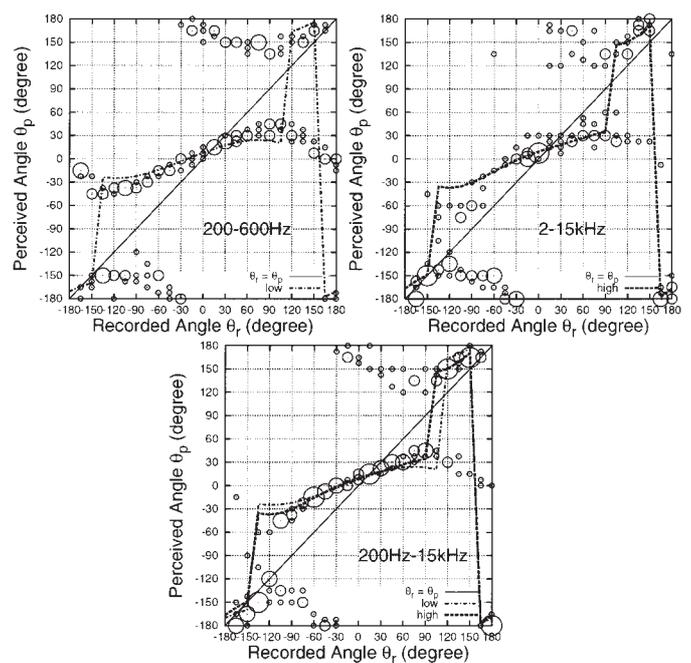
Fig. 12 Results of TYPE 5.

few instances of confusion.

These results confirm that the theoretical equations introduced in this paper approximately agree with all results. The result of 200 Hz–15 kHz is closer to the high frequency equation than the low frequency equation. This result agrees well with the result of Takahashi *et al.* [11].

In Figs. 9 and 10, the results of the three signals show a similar tendency to the TYPE 1 results. With regard to the result of 200 Hz–15 kHz, it is noted that TYPE 2 yields

slightly more stable localization than TYPE 1 and that the result of TYPE 3 indicates concentrated localization to the front regardless of the recorded angle.

5.4.2. TYPE 4,5

Figures 11 and 12 show that the localization is asymmetric with regard to the median plane due to the asymmetric microphone arrangements (p4 and p5). These results imply that the microphones must be located symmetrically. Figures 8, 9, and 11 show that a slight

change in the microphone angle influences the sound image only slightly because the results of TYPE 1, 2 and 4 are not greatly different.

5.4.3.　TYPE 6,7

From Figs. 8–10, it is found that the 200–600 Hz and 2–15 kHz signals allowed front-back, or vice versa, error and so unstable localization. Therefore, since it is very difficult to distinguish the difference from the other TYPEs, these tests were performed using only the 200 Hz–15 kHz signal.

Figures 9, 13, and 14 show that the results of TYPE 6
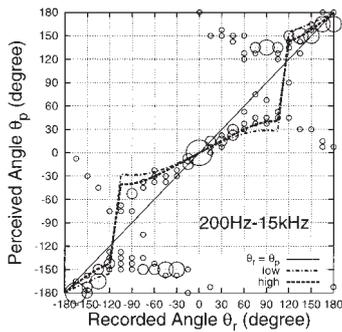
Fig. 13　Results of TYPE 6.
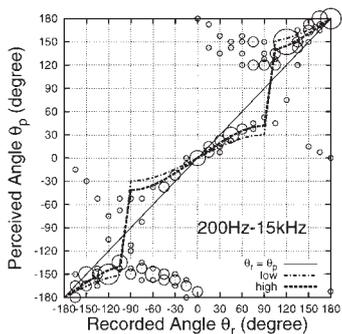
Fig. 14　Results of TYPE 7.
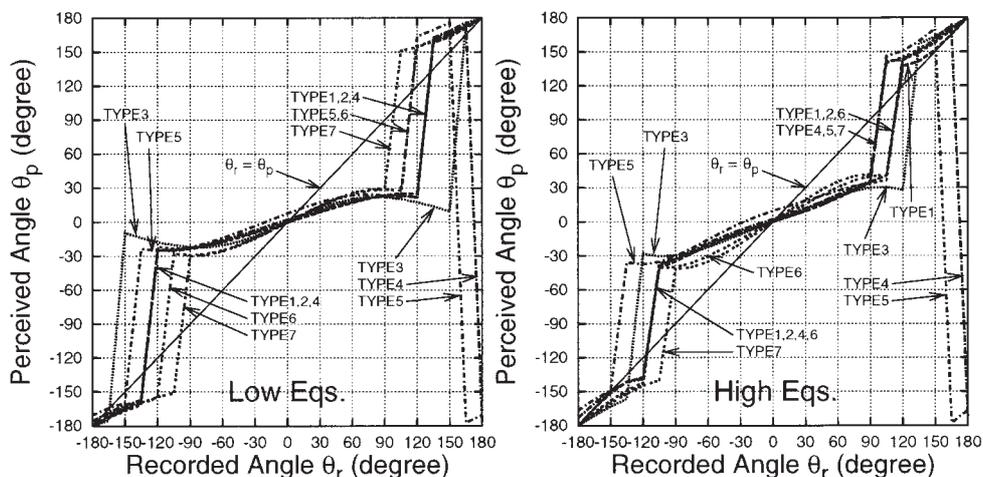
and 7 demonstrate many instances of front-back confusion and poor front stability compared to the result of TYPE 2. The subjects reported that the height of the sound image was either higher or lower compared to the test of TYPE 5. These results indicate that the center-channel stabilizes the front localization and pins the perceived height of the sound image, which agrees with the results given in previous studies (see, for example, [2]).

6.　DISCUSSION

Figure 15 plots, for the theoretical equations for various TYPEs, the relation of θ_p versus θ_r. Figure 16 shows SEM, calculated using Eq. (7) with $D = 24$, $M = 6$, as a function of TYPE.

A consideration of Figs. 15, 16 (SEM_t), and Takahashi *et al.*'s report [11] yields the following points:

1) In all TYPEs, the localization range when using the 2–15 kHz and 200 Hz–15 kHz signals is wider than that achieved with the 200–600 Hz signal, i.e., experimental azimuth angle more closely approaches the desired azimuth angle.

2) TYPE 1 and 2 give well-balanced localization compared to TYPE 3–5. The localization offered by TYPE 3 is concentrated towards the front. TYPE 4 and 5 give asymmetric localization with respect to the median plane.

3) TYPE 6 and 7 with four channels give slightly better localization than all TYPEs with five channels.

4) The localization offered by TYPE 1, 2, and 4, which have slightly different microphone arrangements and slightly different microphone angle, are virtually the same.

5) Lateral localization is poor regardless of the TYPE.

The above points and an examination of Figs. 8–14 and 16 (SEM_e) yield the following conclusions:

1) The expected result was achieved even though the

Fig. 15　Theoretical equations for various TYPEs. θ_p versus θ_r.
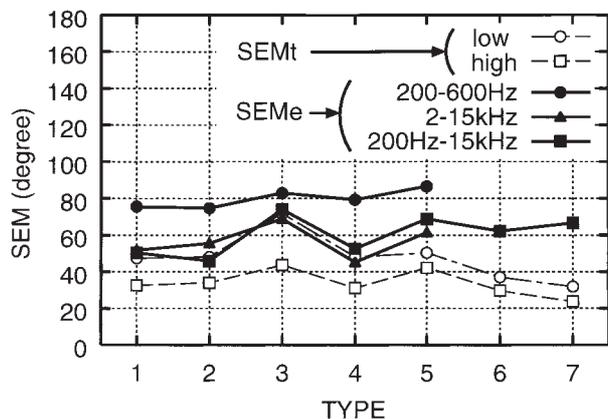
Fig. 16 SEM as a function of TYPE.

characteristics of the signals, 200–600 Hz and 2–15 kHz gave rise to front-back confusion and data spread, respectively. This supports the finding that the SEMs of the 2–15 kHz and 200–15 kHz signals are small compared to those of 200–600 Hz.

2) The expected result was achieved. This means that the better localization offered by TYPE 1 and 2 confirms the finding that the SEMs of TYPE 1 and 2 are small compared to those of TYPE 3 and 5.

3) Contrary to expectation, TYPE 6 and 7 yielded worse localization than TYPE 2. This supports the finding that the SEMs of TYPE 6 and 7 were larger than those of TYPE 2 (1 and 4) due to their high rate of front-back confusion. Further, it was reported that subjects perceived the height of the sound image was slightly raised or lowered. It is estimated that these results are caused by the absence or presence of the center-channel, i.e., the directional stability and the height of the sound image depend on the absence or presence of the center-channel.

4) The expected result was achieved. This supports the finding that the SEMs of TYPE 1, 2, and 4 are much the same.

5) The expected result was achieved. However, because lateral localization perception was slightly improved with the 2–15 kHz and 200 Hz–15 kHz signals, there is a possibility of developing a method that can control lateral localization.

These facts confirm that our theoretical equations are available to find the optimum system and that SEM can evaluate system performance in terms of localization stability and precision. Moreover, the fact that most of six subjects noted the same perceived direction of the sound image in only one trial indicates that the coincident microphone array can also provide stable sound localization in multi-channel recording.

7. CONCLUSIONS

This paper examined six coincident microphone (cardioid pattern) arrays to achieve precise and stable sound image localization in the horizontal plane when multi-loudspeakers are used.

We derived equations to model the system and defined a system evaluation measure. Extensive sound localization trials were conducted to assess the system, our equations, and the system evaluation measure. The following points were clarified.

- The theoretical equations and system evaluation measure introduced in this paper are valid.
- While lateral localization is difficult to achieve, TYPE 1, 2 and 4 systems provide slightly better localization, i.e., microphone arrangements p1, 2 and 4 are better.
- The coincident microphone array can also provide stable sound localization in multi-channel recording.

ACKNOWLEDGEMENT

REFERENCES

[1] Y. Furumi, H. Hokari and S. Shimada, "A study on sound image reproduction with multi-channel transaural system," *IEICE Tech. Rep.*, EA99-53, pp. 9–16 (1999).

[2] K. Kurozumi, S. Komiyama, K. Ohgushi, K. Tsujimoto, A. Morita and J. Ujihara, "Sound system suitable for high definition television," *J. ITE*, **42**, 579–587 (1988).

[3] ITU-R BS. 775-1, "Multichannel stereophonic sound system with and without accompanying picture," Geneva (1992–1994).

[4] K. Hamasaki, "Multichannel sound recording for digital broadcasting," *J. Acoust. Soc. Jpn.* (*J*), **57**, 610–616 (2001).

[5] Y. Makita, "On the directional localisation of sound in the stereophonic sound field," *E.B.U Rev. pt. A*, **73**, 102–108 (1962).

[6] K. Nakabayashi, "A method of analyzing the quadraphonic sound field and its application," *J. Acoust. Soc. Jpn.* (*J*), **33**, 116–127 (1977).

[7] S. Lipshitz, "Stereo microphone techniques — Are the purists wrong?," *J. Audio Eng. Soc.*, **34**, 716–744 (1986).

[8] D. M. Leakey, "Some measurements on the effects of interchannel intensity and time differences in two channel sound systems," *J. Acoust. Soc. Am.*, **31**, 977–986 (1959).

[9] D. M. Leakey, "Stereophonic sound systems," *Wireless World*, **66**, 154–160 (1960).

[10] B. Bernfeld, "Simple equations for multichannel stereophonic sound localization," *J. Audio Eng. Soc.*, **23**, 553–557 (1975).

[11] T. Takahashi, H. Hokari and S. Shimada, "A perception of sound image localization on asymmetric arranged loudspeakers," *IEICE Tech. Rep.*, EA96-55, pp. 25–32 (1996).

[12] H. A. M. Clark, G. F. Dutton and P. B. Vanderlyn, "The 'stereosonic' recording and reproducing system," *J. Audio Eng. Soc.*, **6**, 102–117 (1958).

[13] G. Theile, "Natural 5.1 music recording based on psycho-acoustic principles," *AES 19th Int. Conf. Proc.*, pp. 201–229 (2001).

[14] M. A. Gerzon, "Recording techniques for multichannel

stereo," *Brit. Kinematography Sound & Telev.*, **53**, 274–279 (1971).

[15] J. Aoki, H. Hokari and S. Shimada, "A predictive equation for the direction of sound image localization considered sound pickup," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 585–586 (2002).

[16] *AUDIO TEST CD-1 91 TEST SIGNALS FOR HOME AND LABORATORY USE* (Jpn. Audio Soc.).

[17] T. Yamaji, H. Hokari and S. Shimada, "Stimuli for sound localization test —In case of loudspeaker listening—," *IEICE Tech. Rep.*, EA2001-115, pp. 63–67 (2002).