Reproduction of Varied Sound Image Localization for Real Source in Stereo Audio System.

Satoshi Okuro* and Yoshinobu Kajikawa*
* Faculty of Engineering Science, Kansai University
Yamate-cho, Suita-shi, Osaka, 564-8680, Japan
E-mail: kaji@kansai-u.ac.jp Tel: +81-6-6368-1121

***Abstract*—In this paper, we propose a sound reproduction system which can realize varied sound image localization in stereo audio systems. The proposed system can suppress unnatural variations of sound image localization with listener's movement and maintain the absolute position of sound image so that a real source exists in the corresponding position. Generally, human being perceives the direction of sound image on horizontal plane according to Interaural Level Difference (ILD) and Interaural Time Difference (ITD) between signals arriving at both ears. Accordingly, unnatural variation of sound image localization accompanying listener's movement is due to the differences of ILD and ITD between the stereo audio system and the real source. The proposed system therefore compensates ILD and ITD using digital filters. Some subjective assessment tests with ten subjects demonstrate that fixed sound image can be realized in the proposed system when listener moves away by giving appropriate signal level ratios.**

I. Introduction

Recently, sound field reproduction systems have been actively studied. Sound field control techniques with loudspeakers can be classified into two methods which control narrow points and large areas, respectively.

The former method reproduces any desired sound at any desired point (e.g. listener's ears). To do this, the method of controlling points needs to remove the influence of transfer functions between loudspeakers and control points and crosstalk paths, i.e. the so-called crosstalk canceller. In this method, acoustic transfer functions from loudspeakers to any desired points must be given. Hence, if a listner moves, the system performance will be degraded because acoustic transfer functions will change.

On the other hand, the latter method typically uses a palanar or linear loudspeakers to reconstruct sound field. This method referred to as wave field synthesis (WFS) can control wide areas. However a great number of loudspeakers are needed according to the sampling theorem of the space if the sound field is controlled over the audio bandwidth of 20 kHz [1]-[4]. Such an approach is impractical.

To solve this problem, the methods with a few loudespeakers which are allocated appropriately are reserched recently[6]. Such methods can greatly reproduce any sound fields at sweet spot. However, in this method, the system performance will be degraded when a listner moves away from the sweet spot because the reproduction of sound fields is specified by a particular location. Hence, sound image positions must be fixed regardless of user's movements in sound field reproduction systems. Figure 1 shows an example of common stereo systems where sound image positions must be fixed regardless of user's movements. In Fig. 1, when the same signal is produced from the two loudspeakers, the listener at P_c perceives sound image at P_r. However, if listener moves to P_a, the sound image moves to P_f. This problem does not arise in real sound sources. In this paper, we propose a sound reproduction method generating sound image at any fixed position regardless of user's movements. Moreover, we discuss the system performance and parameter values required for the proposed method through subjective assessment test results.

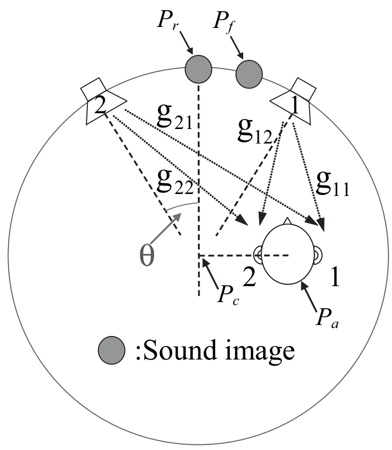

Fig. 1. Ordinary stereo audio system

II. Proposed Method for fixed sound image position

We consider only horizontal plane where listner's ears and loudspeakers exist. Moreover, we assume that reflections and reverberations does not exist, and the listener's position can be obtained by RF tag or triangulation techniques with video camera.

A. Basic Principle of Sound Image Perception

Human perception of sound image mainly depends on interaural level difference (ILD) and interaural time difference (ITD)[7]. ILD forms the basis of the "intensity-difference theory" of directional hearing. If sound pressure level at one side is higher than that at the other side, Human perceives sound image at higher amplitude side. ITD is the difference in arrival time of a sound between two ears. If the time difference becomes large, then the sound image moves to the side in which the sound arrives first. This phenomenan is known as Haas effect. Hence, the proposed method controls sound image fixed at any position by digital filters.

B. Filter Design Process for Fixed Sound Image

It is assumed that a listener perceives sound image at P_r in any of these cases where the listener stands at P_c and P_a as shown in Fig. 1. Figure 2 shows the block diagram of the proposed method. The filter design process for fixed sound image are as follows:

1) Obtain the position P_a where a listener stands.
2) Estimate a sound image position where the listeners standing at P_c would perceive. Generally, sound image positions are controled by the signal level ratio L_{ref} of right channel to left channel. Hence, the azimuthal angle θ_1 of the sound image in Fig. 3 can be estimated based on the sine law as

$$\theta_1 = sin^{-1}(\frac{1-L_{ref}}{1+L_{ref}}sin\theta), \tag{1}$$

where θ is the azimuthal angle of loudspeakers at the sweet spot.
3) Calculate θ_2 which is the azimuthal angle of the sound image at position P_r, and then calculate the signal level ratio L_c to perceive sound image at position P_r.
4) Obtain the distances of transfer paths $G_{11}(k)$, which is the transfer function between the right ear and right loudspeaker, and $G_{22}(k)$, which is the transfer function between the left ear and left loudspeaker. Next, select the longer transfer path and then calculate the distance decay ratio $|\beta_{mn}|$ and the phase difference $e^{j(\Delta\Phi_{mn}\times k)}$. Here, m indicates the loudspeaker number and n indicates the ear number (right is "1" and left is "2"). Moreover, the distance decay ratio $|\beta_{mn}|$ is independent of the frequency k and $\Delta\phi_{mn}$ represents the group delay derived from the distance difference in the free sound field.
5) Design the digital filters for fixed sound image. The purpose of the proposed method compensates the distance difference in the both channel and gives a new level ratio L_c to perceive the sound image at P_r. The digital filters can be designed as

$$\begin{bmatrix} H_1(k) \\ H_2(k) \end{bmatrix} = \begin{bmatrix} G_{11}^{-1}(k) \\ \sqrt{\frac{L_c}{L_{ref}}}G_{22}^{-1}(k) \end{bmatrix}. \tag{2}$$

In this equation, $G_{11}(k)$ and $G_{22}(k)$ are unknown. Hence, $G_{11}(k)$ and $G_{22}(k)$ are replaced with $|\beta_{mn}|$ and phase delay ratio $e^{j(\Delta\Phi_{mn}\times k)}$ in the step 4, that is, Eq. (2) is rewritten as

$$\begin{bmatrix} H_1(k) \\ H_2(k) \end{bmatrix} = \begin{bmatrix} \frac{1}{|\beta_{11}|}e^{-j\frac{2\pi k}{N}\Delta\phi_{11}} \\ \sqrt{\frac{L_c}{L_{ref}}}\frac{1}{|\beta_{22}|}e^{-j\frac{2\pi k}{N}\Delta\phi_{22}} \end{bmatrix}, \tag{3}$$

where, $|\beta_{11}|\, e^{-j(\Delta\phi_{11}\times k)}$is equal to 1, if the distance of $G_{11}(k)$ is longer than the distance of $G_{22}(k)$.

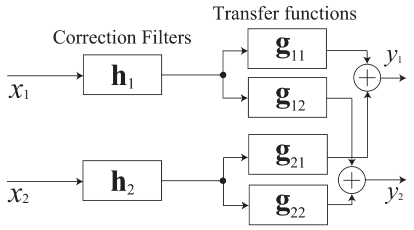

Fig. 2. Block diagram of proposed system

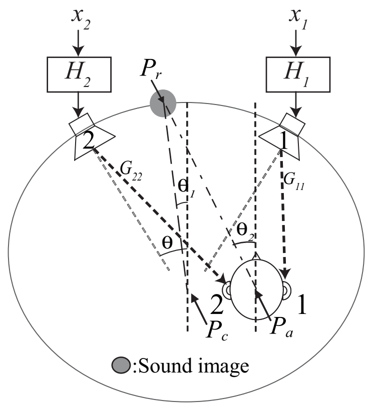

Fig. 3. Stereo audio system when listener moved.

In this filter design process, L_c is the important parameter to fix sound image at any position. Therefore, we disucuss how to obtain appropriate signal level ratio L_c through some subjective assessment tests in the next chapter.

III. Disucuss the signal level ratio L_c with subjective assessment

A. Experimental setup

Figures 4 and 5 show the GUI interface and the arrangement for subjective assessment, respectively. Table I shows the experimental conditions.

In Fig. 4, the signal level ratio L_c is controled by moving the slider on the upper side. L_c is equal to -20 dB if the slider is moved at the leftmost position, and L_c is equal to 20 dB iif the slider is moved at the rightmost position. In addition, the step size of the movements of the slider is 0.5 dB. This implies the number of steps of slider is 81. Boxes in the right-of-center are used to indicate sound image position provided by the subjective assessment test. The box number represents the corresponding sound image position in Fig. 5. Hence, if a black circle is indicated in the 1st box, the subject adjusts the slider so as to perceive the sound image at the corresponding

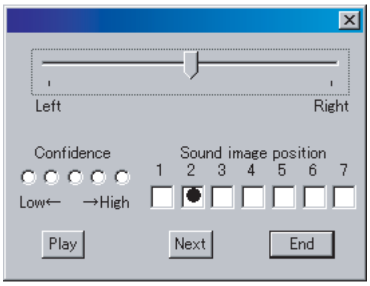

Fig. 4. System interface for subjective assessment.

TABLE I
EXPERIMENTAL CONDITIONS.

Noise signal	White noise
Sampling frequency	44.1 kHz
Loudspeaker	FUJITSU TEN ECLIPSE TD510
Loudspeaker set up angle	± 45 rad
Number of subject	10
Presented sound level	70 dB

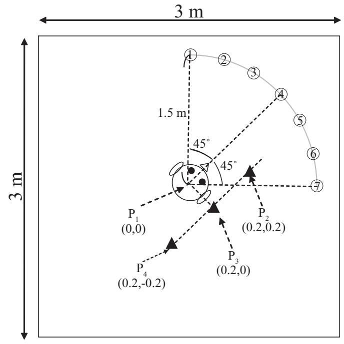

Fig. 5. Speaker arrangement

Table 2 Subjective evaluation results at position P_1.

Sound Image position	Averaged level ratio	Averaged confidence	Variance	Standard error
1	-18.23	4.53	2.06	0.48
2	-11.50	4.00	4.28	0.69
3	-5.86	3.96	6.58	0.86
4	0.09	4.36	6.41	0.84
5	6.59	4.07	7.17	0.89
6	11.90	4.14	0.89	0.32
7	18.78	4.58	1.27	0.38

Table 3 Subjective evaluation results at position P_2.

Sound Image position	Averaged level ratio	Averaged confidence	Variance	Standard error
1	-19.06	4.63	1.53	0.41
2	-12.22	4.03	1.35	0.39
3	-7.55	4.03	2.05	0.48
4	-2.78	4.19	5.26	0.76
5	3.44	4.24	3.82	0.65
6	9.49	4.54	3.83	0.65
7	18.71	4.87	2.83	0.56

Table 4 Subjective evaluation results at position P_3.

Sound Image position	Averaged level ratio	Averaged confidence	Variance	Standard error
1	-19.05	4.40	1.21	0.37
2	-12.64	4.33	4.53	0.71
3	-7.29	4.23	5.21	0.76
4	-2.12	4.43	4.50	0.71
5	3.47	4.47	3.62	0.63
6	10.79	4.47	4.56	0.71
7	19.19	4.50	1.00	0.33

Table 5 Subjective evaluation results at position P_4.

Sound Image position	Averaged level ratio	Averaged confidence	Variance	Standard error
1	-18.98	4.57	1.44	0.40
2	-12.55	3.97	2.57	0.53
3	-7.82	3.87	3.56	0.63
4	-2.90	4.11	7.11	0.89
5	3.24	3.91	6.06	0.82
6	8.63	4.04	3.49	0.62
7	17.93	4.80	4.01	0.67

position. The sound image is presented in random order three times, but the same position is not presented successively. After adjusting the slider position to fit the sound image to the position indicated in boxes, the subject evaluates the image width and the confidence of the image position in five levels with the radio boxes in the left-of-center. The GUI system reproduces the white noise whose level is adjusted based on the slider position when the subject clicks "Play" button. Using the GUI system shown in Fig. 4, we determine the appropriate signal level ratio L_c

In Fig. 5, seven loudspeakers are arranged at each position ①-⑦ and are labeled by the corresponding number. The loudspeakers at positions ① and ⑦ are used for reproducing the sound signal and the other loudspeakers are dummy.

Subjects sat on the arm-chair at each position P_1–P_4 in Fig. 5 and were instructed to face forward and set their head on head rest during subjective assessment. Subjective assessment tests were carried out in a sound-attenuated chamber (3 m × 3 m × 2 m). The number of subjects was 10 including 9 males and 1 female.

B. Procedure of subjective assessment

1) Sit on the arm chair located at any instructed position.
2) Check the sound image position which is presented on the right boxes in the GUI.
3) Adjust the slider so as to perceive the sound image at the presented position on the GUI while listening the presented sound.
4) Evaluate the sound image width and the confidence of the sound image position in five levels with the radio boxes on the left-of-center.
5) Click the "Next" bottom and return to the step 2.
6) Click the "End" bottom after 7 (sound image positions) × 3 = 21 evaluations.

C. Results of subjective assessment

Experimental results at positions P_1–P_4 are shown in Tables 1–4. It can be seen from these tables, that the averaged confidences at positions P_2–P_4 are the almost same as sweet spot P_1. Hence, the proposed method set to appropriate level ration L_c provides the fixed sound image at any position even if the listener moves away from the sweet spot. However, Table

4 shows that the averaged confidence at position (d) is lower than those at other positions. This is because of the directivity of loudspeakers.

Next, we examine the appropriate signal level ratio L_c from experimental results. In the proposed method, a new listener's position becomes a new sweet spot acoustically. Hence, the normalized azimuthal angle of the sound image θ_i is defined as So we introduce nomalized angle of sound image position$\hat{\theta}_i$ as Eq. 4.

$$\hat{\theta}_i(m) = \begin{cases} -\theta_i(m)/\theta_i(1) & \theta_i(m) < 0 \\ \theta_i(m)/\theta_i(7) & otherwise \end{cases}, \quad (4)$$

where i indicates the listener position P_i, and m indicates the sound image position, and $\theta_i(m)$ indicates the angle of sound image position. Table 5 shows the angles and nomalized angles of sound image position for each listening position. Moreover, Figs. 6 and 7 show the azimuthal and normalized azimuthal angles versus averaged signal level ration, respectively. In these figures, the error bars represent the standard errors and the red line shows the approximate line obtained by least square method. These figures does not include the results at sound image positions 1 and 7 because the signal level ratios at these positions are saturated. It can be seen from Figs. 6 and 7 that the normalized azimuthal angle has stronger correlation to the signal level ratio L_c than the azimuthal angle. Moreover, the approximate line is within 95% confidence intervals. Hence, the proposed method can fix the sound image using Eq. (5) even if the listener moves away from the sweet spot.

$$L = 17 * \hat{\theta}_i(m) + 0.3. \quad (5)$$

IV. Conclusion

In this paper, we proposed the method for providing the fixed sound image even if the listener moves away from the sweet spot in stereo audio systems. Moreover, we examined the appropriate signal level ratio through subjective assessment test. Experimental results demonstrated that the appropriate signal level ratio yields fixed sound image at any positions. However, if the listener approaches to loudspeakers, then the sound image cannot be reproduced at appropriate positions. This is because of the directivity of loudspeakers. In the future, we will apply the obtained signal level ratio to the proposed sound reproduction system and explore a novel method for providing multiple fixed sound images.

Table 6 Azimuthal and normalized azimuthal angles.

	P_5		P_2	
Sound image position	Azimuthal angle	Normalized Azimuthal angle	Azimuthal angle	Normalized Azimuthal angle
1	-45.00	-1.00	-45.00	-1.00
2	-30.00	-0.67	-32.66	-0.73
3	-15.00	-0.33	-20.23	-0.45
4	0.00	0.00	-7.59	-0.17
5	15.00	0.33	5.36	0.16
6	30.00	0.67	18.75	0.57
7	45.00	1.00	32.73	1.00
	P_3		P_4	
Sound image position	Azimuthal angle	Normalized Azimuthal angle	Azimuthal angle	Normalized Azimuthal angle
1	-50.60	-1.00	-57.27	-1.00
2	-37.05	-0.73	-42.55	-0.74
3	-23.13	-0.46	-26.94	-0.47
4	-8.75	-0.17	-10.30	-0.18
5	6.20	0.16	7.37	0.16
6	21.79	0.57	25.92	0.58
7	38.04	1.00	45.00	1.00

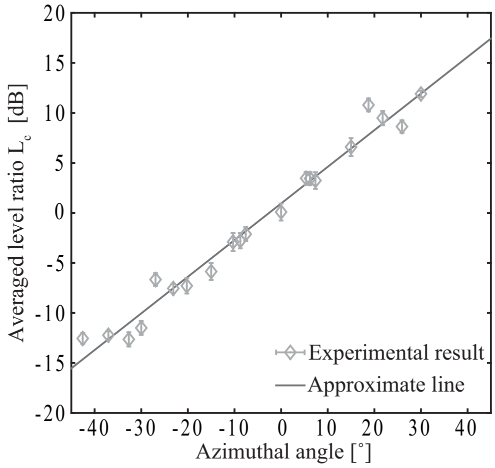

Fig. 6. Azimuthal angle versus averaged level ratio

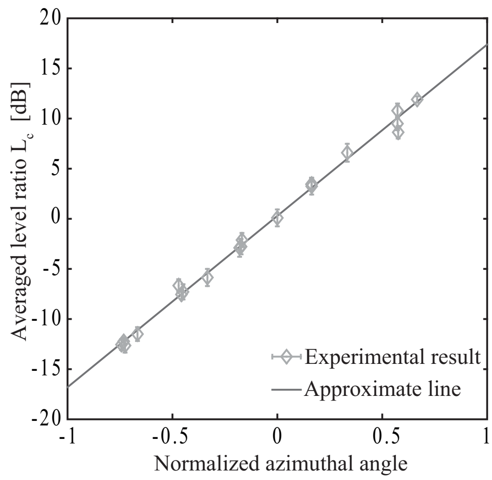

Fig. 7. Normalized azimuthal angle versus averaged level ratio

References

[1] A.J. Berkhout, D. D. Vries, and P. Vougel "Acoustic control by wave field synthesis," J. Acoust. Soc. Amer., vol. 93, no. 5, pp. 2764–2778, 1993.

[2] E. Verheijen, "Sound field reproduction by wave field synthesis," Ph.D. dissertation, Delft Univ. of Technol., Delft, The Netherlands, 1997.

[3] D. D. Vries, "Sound reinforcement by wavefield synthesis: Adaptation of the synthesis operator to the loudspeaker directivity characteristics," J. Audio Eng. Soc., vol. 44, no. 12, pp. 1120–1131, 1996.

[4] D. D. Vries, Wave field synthesis, ser. AES Monograph. New York:Audio Eng. Soc., 2009.

[5] K. Hiyama, S. Komiyama, and K. Hamasaki, "The minimum number of loudspeakers and its arrangement for reproducing the spatial Impression of diffuse sound field," 113th AES convention , Oct. 2002.

[6] Y. Makita, "On the directional localization of sound in the stereophonic sound field," Europiano Broadcasting Union Technical Review, Par–tech, vol. 73, pp. 102–108, 1962.

[7] J. Blauert, Spatial Hearing, The MIT Press, Cambridge 1986.