HYBRID STRUCTURE OF INVERSE FILTERING AND DOA-PARAMETERIZED WAVEFRONT SYNTHESIS

Yuuta Yuyama, Shigeki Miyabe, Hiroshi Saruwatari, and Kiyohiro Shikano*

Nara Institute of Science and Technology
8916-5 Takayama-cho, Ikoma-shi, Nara, Japan
{yuuta-y, shige-m, sawatari, shikano}@is.naist.jp

ABSTRACT

Weakness against a user's position shifting is one of the most important problems of binaural reproduction using loudspeakers. In this paper we propose a new method of inverse filtering of room acoustics with high robustness against a user's position shifting by presenting a wavefront estimated from the binaural recording. To analyze and synthesize the wavefront, we introduce a new physical modelling of the superimposition of the wavefronts weighted by multiple sound sources. By following the fluctuation of the wavefront with a time-varying filter, the analysis and synthesis overcome the limitation of the number of sound sources. Utilizing arbitrary components of the generalized inverse matrix, wavefront approximation does not degrade the accuracy of reproduction at the controlled area of the inverse filter.

Index Terms— Spatial audio, binaural recording, direction of arrival, matrix inversion, acoustic fields.

1. INTRODUCTION

Sound reproduction using loudspeakers can be classified into two groups; reproduction of wavefront or that of binaural recording [1]. Many of the small systems in the first group are based on a simple idea to pan multichannel signal, which is robust against a user's positioning and is widely distributed as stereophonic or 5.1 surround. However, to obtain highly realistic sensation, the size of the system is likely to become too large [2, 3].

Reproduction of binaural recording can achieve realistic sensation with relatively small system called *transaural system* [4], where an inverse filter compensates the impulse responses at the user's ears including the reverberation, so-called binaural room impulse responses (BRIRs). Although BRIRs are nonminimum phase systems in general, multiple input/output inverse theorem (MINT) [5] proves that their inverse filters can be designed using more loudspeakers than the control points, i.e., the listener's ears. However, the compensation is not satisfied by the inverse filter outside the control points (*sweet spot*), and it is known that transaural reproduction is weak against a user's movement. The reproduction with MINT is specified to accuracy in the control points, and reproduction outside the control points is not considered. With such specified control, the wavefront outside the control points is formed as an entirely different one in the primary field, and the sound localization degrades considerably when the user moves from the controlled area. Although some crosstalk cancellers, which compensate anechoic head related transfer functions but BRIRs, successfully expand the sweet spots toward front and back [6], their reproduction accuracy is not as good as those based on MINT. To adapt the inverse filter, microphones are required to be set at the user's ears [7], which interrupts listening.

To mitigate the effect of the listener's movement, we have proposed an inverse filter design to maintain wavefront outside the sweet spot [8]. By optimizing arbitrary components of MINT, direction of the primary source is presented even outside the sweet spot with

*Reseach Fellow of the Japan Society for the Promotion of Science.

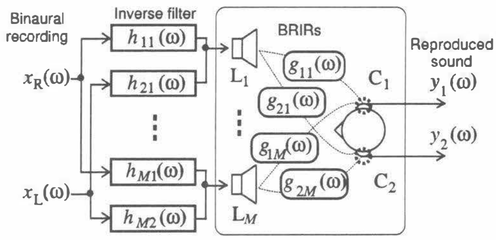

Fig. 1. Configuration of a transaural system with two control points and M loudspeakers.

accuracy at the control points unchanged. However, this method is merely a theoretical framework and can only deal with binaural recording of a single sound source. In this paper, utilizing our previous work, we propose a new time-varying inverse filtering method for binaural recording of multiple sources, which achieves transaural reproduction with high robustness against a user's movement. The wavefront is analyzed with the binaural recording to be reproduced, and the analyzed wavefront is approximated outside the sweet spot. To analyze the wavefront formed by the multiple sources, we introduce a new physical model inspired by binaural cue coding (BCC) [10]. BCC is a multichannel audio coding method by analyzing time- and level-difference among the audio channels consisting of an arbitrary number of sources. Since the wavefront formed by the multiple sources is directly followed by a time-varying filter, the proposed method can reproduce arbitrary number of sources. Efficacy of the proposed method is ascertained in both objective and subjective evaluations.

2. CONVENTIONAL SOUND FIELD REPRODUCTION USING INVERSE FILTER

The transaural system achieves high realistic sensation of source localization by reproducing binaural recording at the listener's ears. Such a reproduction can be achieved by designing the inverse filter of the BRIRs. By using more loudspeakers than the control points, it is proven that the strict inverse system of the BRIRs with nonminimum phases can be designed [5]. Here we deal with the problem of controlling the sound field around two control points C_1 and C_2 at the two ears of the listener using the M (> 2) loudspeakers L_m ($m = 1, \ldots, M$). We show the configuration of such the transaural system in Fig. 1.

We denote the binaural recording in the frequency domain as a two-dimensional vector $\boldsymbol{x}(\omega) = [x_1(\omega), x_2(\omega)]^T$, where ω is angular frequency and $\{\cdot\}^T$ denotes matrix transposition. We measure all the transfer functions $g_{nm}(\omega)$ from L_m to C_n for $m = 1, \ldots, M$, $n = 1, \ldots, 2$. We define a $2 \times M$ matrix $\boldsymbol{G}(\omega) = [g_{nm}(\omega)]_{nm}$, where $[a]_{nm}$ is a matrix that has an element a in the n-th row and m-th column. Then we design an $M \times 2$ inverse filter matrix $\boldsymbol{H}(\omega) = [h_{mn}(\omega)]_{mn}$ to satisfy the following condition

$$\boldsymbol{G}(\omega)\boldsymbol{H}(\omega) = \boldsymbol{I}, \tag{1}$$

where $\boldsymbol{I}$ denotes an identity matrix. By outputting the filtered binaural recording $\boldsymbol{H}(\omega)\boldsymbol{x}(\omega)$ from the loudspeakers, the reproduced signals $\boldsymbol{y}(\omega)$ at the user's ears satisfy the condition $\boldsymbol{y}(\omega) =$

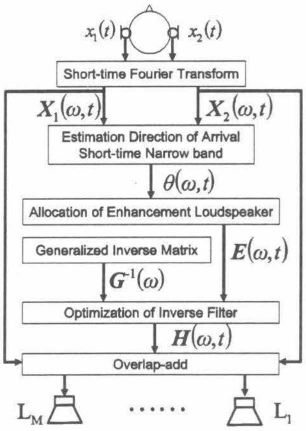

Fig. 2. Configuration of the proposed system.

$G(\omega)H(\omega)x(\omega) = x(\omega)$, and the binaural recording is reproduced.

The condition in Eq. (1) shows that $H(\omega)$ can be designed with the generalized inverse matrix $G^{-}(\omega)$ of $G(\omega)$. To fix ambiguity of $G^{-}(\omega)$, the Moore-Penrose (MP) generalized inverse matrix $G^{+}(\omega)$ is generally used [9]. Utilizing singular value decomposition, the generalized inverse matrix can be denoted as

$$G^{-}(\omega) = V(\omega) \underbrace{\begin{bmatrix} \Lambda(\omega) \\ S(\omega) \end{bmatrix}}_{M\times 2} U^{\mathrm{H}}(\omega), \tag{2}$$

$$\Lambda(\omega) = \mathrm{diag}\left[\lambda_1(\omega), \lambda_2(\omega)\right], \tag{3}$$

$$\lambda_k(\omega) = \begin{cases} \frac{1}{\mu_k(\omega)} & (\text{if } \mu_k(\omega) \neq 0), \\ 0 & (\text{otherwise}), \end{cases} \tag{4}$$

where $\mu_k(\omega)\,(k = 1, 2)$ denote singular values, $U(\omega)$ and $V(\omega)$ are unitary matrices with the left eigenvectors and the right eigenvectors corresponding to the eigenvalues, respectively, $\{\cdot\}^{\mathrm{H}}$ denotes the conjugate transposition, $S(\omega)$ is an arbitrary $(M-2)\times 2$ matrix, and diag$[\cdot]$ shows a diagonal matrix composed of the arguments. The MP generalized inverse matrix $G^{+}(\omega)$ can be obtained by setting $S(\omega)$ to be a zero matrix. Finally we use $G^{+}(\omega)$ as the inverse filter $H(\omega) = G^{+}(\omega)$. However, the MP-based inverse filter is specified to the reproduction at the control points and the reproduction cannot be guaranteed outside the control points. Thus the sound localization degrades considerably when the user moves from the controlled area.

3. OUR PREVIOUS WORKS: INVERSE FILTERING WITH SECONDARY SOURCE SELECTION AND ENHANCEMENT

3.1. Summary

To improve the robustness against the user's position shifting, we have proposed an inverse filter design method to approximate the wavefront formed by a sound source outside the sweet spot without degradation of reproduction at the control points [8]. Such an inverse filter is obtained by embedding a filter to form the wavefront in the arbitrary elements in the nullspace of the generalized inverse matrix shown in Eq. (2). This method can deal with only the binaural recording generated by a single source in a known direction, and this method is an important basis of the proposed method.

3.2. Algorithm

First we design the filter to approximate the wavefront. As a method to form the wavefront, we output sound from only a single loudspeaker located in the source direction. For the parallel use of the filter with inverse filtering, the latency and the gain should be conformed with those of the inverse filter. The filter $T(\omega)$ to satisfy the condition is obtained as

$$T_{mn}(\omega) = \begin{cases} \frac{\|G^{+}(\omega)\|_{\mathrm{Fr}}}{\sqrt{2}} \cdot e^{-j\omega\tau} & (\text{if } m = k), \\ 0 & (\text{otherwise}), \end{cases} \tag{5}$$

where τ is the latency where the MP-based inverse filter has the largest peak in the time domain, k is the number of loudspeaker, and $\|\cdot\|_{\mathrm{Fr}}$ denotes the Frobenius norm given as $\|[a_{mn}]_{mn}\|_{\mathrm{Fr}} = \sqrt{\sum_m \sum_n |a_{mn}|^2}$. We call the multichannel filter $T(\omega)$ designed above the *direction emphasis filter*.

Next, to approximate $T(\omega)$ in the subspace (or nullspace) of $G^{-}(\omega)$ with arbitrary components $S(\omega)$ in Eq. (2), we obtain the generalized inverse matrix $G^{-}(\omega)$ closest to $T(\omega)$. We utilize the Frobenius norm as the distance measure and we obtain $G^{-}(\omega)$ to minimize $F(\omega) = \|G^{-}(\omega) - T(\omega)\|^2_{\mathrm{Fr}}$. Since the Frobenius norm is not changed by multiplication of unitary matrices, $F(\omega)$ can be rewritten as

$$\begin{aligned} F(\omega) &= \left\|V^{\mathrm{H}}(\omega)\left(G^{-}(\omega) - T(\omega)\right)U(\omega)\right\|^2_{\mathrm{Fr}} \\ &= \left\|\begin{bmatrix} \Lambda - V^{\mathrm{H}}_{\mathrm{span}}(\omega)T(\omega)U(\omega) \\ S(\omega) - V^{\mathrm{H}}_{\mathrm{null}}(\omega)T(\omega)U(\omega) \end{bmatrix}\right\|^2_{\mathrm{Fr}} \\ &= \left\|\Lambda(\omega) - V^{\mathrm{H}}_{\mathrm{span}}(\omega)T(\omega)U(\omega)\right\|^2_{\mathrm{Fr}} \\ &\quad + \left\|S(\omega) - V^{\mathrm{H}}_{\mathrm{null}}(\omega)T(\omega)U(\omega)\right\|^2_{\mathrm{Fr}}, \end{aligned} \tag{6}$$

where $V_{\mathrm{span}}(\omega)$ is a matrix composed of the first two column of $V(\omega)$, and $V_{\mathrm{null}}(\omega)$ is a matrix composed of the rest of the components of $V(\omega)$. Since $\Lambda(\omega)$ is indispensable for the generalized inverse matrix, the optimal inverse filter is obtained by the substitution $S(\omega) = V^{\mathrm{H}}_{\mathrm{null}}(\omega)T(\omega)U(\omega)$ as

$$H(\omega) = V(\omega)\begin{bmatrix} \Lambda(\omega) \\ V^{\mathrm{H}}_{\mathrm{null}}(\omega)\,T(\omega)\,U(\omega) \end{bmatrix} U^{\mathrm{H}}(\omega). \tag{7}$$

3.3. Problem Under Existence of Multiple Sound Sources

If the binaural recording is generated by a single source, the design and the filtering should be conducted for binaural recording of each source. However, the separated binaural recording of each source is not available generally in practical situation. If the number of the sources is only two, the sources can be separated by beamforming. However, it is difficult to obtain high-quality source separation of more than three sources from only the two-channel observation of the binaural recording.

4. PROPOSED METHOD

4.1. Motivation and Strategy

To apply the above transaural reproduction accompanied with wavefront approximation to the binaural recording of an arbitrary number of sources, we propose a new time-varying inverse filtering method. To analyze and synthesize wavefront generated by the multiple sources, we introduce a new physical model inspired by BCC [10]. By analyzing parameters concerned with perception of source localization in each narrow subband of each analysis frame, BCC sufficiently encodes the multichannel signals composed of an arbitrary number of sources with low bit rate. Similarly to BCC, we analyze the wavefront as DOA and synthesize the wavefront as in the time-frequency domain. Since source separation is not used in this strategy, we can deal with an arbitrary number of sources.

Here we describe the detail of the proposed physical model. The wavefront generated by a single source in a fixed position is constant in spite of the unsteady behavior of the source signal. The wavefront under existence of multiple sources is a superimposition of the multiple wavefronts generated by each of the sources. Since the superimposition of each wavefront is weighted by the unsteady behavior of each source signal, the superimposed wavefront fluctuates in each narrow subband in each short time duration separately. All we have to reproduce is such time-varying wavefront but the static one of

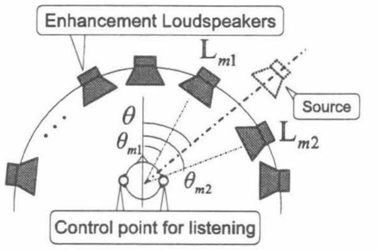

Fig. 3. Configuration of intensity panning.

each of the sources. Thus, we analyze and synthesize the wavefront following the fluctuation with the time-varying filter.

Here we briefly summarize the proposed method. First we estimate the wavefront's direction of arrival (DOA) in each subband of each short-time analysis frame. According to the estimated DOA, panning is determined in each narrow band of each frame. Finally, the inverse filter is optimized in each of the frames.

4.2. Algorithm

[STEP 1] Short-time narrow-band DOA estimation

As a method to achieve the DOA estimation from the binaural recording in each subband in each short-time analysis frame, we utilize a binaural DOA estimation method based on interaural level difference (ILD) and interaural time difference (ITD) [11].

First we apply short-time Fourier analysis to the binaural recording. We denote the binaural recording in the time-frequency domain by $X(\omega, t) = [X_1(\omega, t), X_2(\omega, t)]^{\mathrm{T}}$. The estimated DOA $\theta_L(\omega, t)$ with ILD and the estimated DOA candidate $\theta_{T,\gamma}(\omega, t)$ with ITD are obtained as

$$\theta_L(\omega, t) = \arcsin\left(\frac{20\log|X_2(\omega, t)/X_1(\omega, t)|}{\alpha(\omega)}\right), \tag{8}$$

$$\theta_{T,\gamma}(\omega, t) = \Pi\left(\frac{c(\arg(X_2(\omega, t)/X_1(\omega, t) + 2\pi\gamma))}{r\beta(\omega)}\right)$$
$$\text{with } \Pi(x) = 0.50018x + 0.009897x^3 + 0.00093x^5, \tag{9}$$

where γ is an indefinite integer, c [m s] is the wave propagation speed, r is the head radius, and $\alpha(\omega)$ and $\beta(\omega)$ are scaling coefficients learned from the head related transfer function database. To utilize the property that ITD is reliable in the low frequency while ILD is reliable in the high frequency, we estimate the DOA as

$$\theta(\omega, t) = \begin{cases} \theta_{T,\gamma}(\omega, t) \quad \gamma = 0 & (\text{if } \omega < \kappa_1) \\ \theta_H(\omega, t) & (\text{if } \kappa_1 \leqq \omega \leqq \kappa_2) \\ \theta_L(\omega, t) & (\text{if } \kappa_2 < \omega), \end{cases} \tag{10}$$

$$\theta_H(\omega, t) = \mathrm{argmin}_{\theta_{T,\delta}(\omega,t)}|\theta_L(\omega, t) - \theta_{T,\gamma}(\omega, t)|. \tag{11}$$

In our implementation, we set κ_1 and κ_2 associated with 1000 Hz and 2000 Hz, respectively.

[STEP 2] Time-frequency domain amplitude panning filter

First we describe the filter to define the intensity panning based on sine law [12] among the loudspeakers in each narrow band of each frame. We show the configuration of intensity panning in Fig. 3. We select the two loudspeakers $\mathrm{L}_{m_1(\omega,t)}$ and $\mathrm{L}_{m_2(\omega,t)}$ in the two closest directions $\theta_{m_1(\omega,t)}(\omega, t), \theta_{m_2(\omega,t)}(\omega, t)$ to $\theta(\omega, t)$. Then the amplitude balance $b_{m_1(\omega,t)}(\omega, t), b_{m_2(\omega,t)}(\omega, t)$ between $\mathrm{L}_{m_1(\omega,t)}$ and $\mathrm{L}_{m_2(\omega,t)}$ can be obtained from the following rule:

$$\frac{b_{m_1(\omega,t)}(\omega, t) - b_{m_2(\omega,t)}(\omega, t)}{b_{m_1(\omega,t)}(\omega, t) + b_{m_2(\omega,t)}(\omega, t)} = \frac{\sin\left(\theta(\omega, t) - \left(\dfrac{\theta_{m_1(\omega,t)}(\omega, t) + \theta_{m_2(\omega,t)}(\omega, t)}{2}\right)\right)}{\sin\left(\dfrac{\theta_{m_1(\omega,t)}(\omega, t) - \theta_{m_2(\omega,t)}(\omega, t)}{2}\right)},$$
$$b_{m_1(\omega,t)}(\omega, t)^2 + b_{m_2(\omega,t)}(\omega, t)^2 = 1. \tag{12}$$

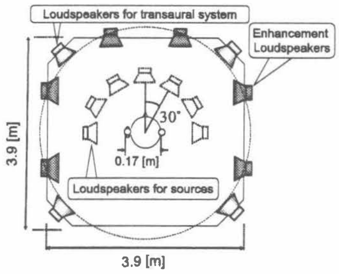

Fig. 4. Experimental conditions.

[STEP 3] Maintenance of gain and delay

Here we describe the maintenance of latency and gain. The latency is determined in the same manner of the conventional direction emphasis filter. As for gain, we equalize the gain in the whole of the reproduction frequency range. The equalized gain is determined by the gain of the MP-based inverse filter normalized by the whole the frequency range. The m-th-row-n-th-column component $E_{mn}(\omega, t)$ of the time-varying direction emphasis filter $\boldsymbol{E}(\omega, t)$ is given by as

$$E_{mn}(\omega, t) = \begin{cases} b_m(\omega, t)\,\sigma_n(\omega, t)\,\rho e^{-j\omega\tau} & (\text{if } m = m_i(\omega, t) \text{ for } i = 1, 2) \\ 0 & (\text{otherwise}), \end{cases} \tag{13}$$

where $\sigma_n(\omega, t)$ is a coefficient to prevent the dip in spectrum caused by phase-out effect, given by

$$\sigma_n(\omega, t) = \frac{e^{j\arg(X_1(\omega,t)+X_2(\omega,t))}}{\sqrt{|X_1(\omega, t)|^2 + |X_2(\omega, t)|^2}} \cdot X_n^*(\omega, t), \tag{14}$$

where $*$ denotes conjugate, and ρ is the total gain of the inverse filter in the full bandwidth of the reproduction shown as

$$\rho = \frac{1}{\sqrt{2}}\sqrt{\frac{1}{2\pi}\int_{-\pi}^{\pi}\|\boldsymbol{G}^+(\omega)\|_{\mathrm{Fr}}^2\, d\omega}. \tag{15}$$

[STEP 4] Embedding of panning to nullspace

In the final, we embed the time-varying direction emphasis filter $\boldsymbol{E}(\omega, t)$ in the nullspace similarly to Eq. (7) as

$$\boldsymbol{H}(\omega, t) = \boldsymbol{V}(\omega)\begin{bmatrix} \boldsymbol{\Lambda}(\omega) \\ \boldsymbol{V}_{\mathrm{null}}^{\mathrm{H}}(\omega)\,\boldsymbol{E}(\omega, t)\,\boldsymbol{U}(\omega) \end{bmatrix}\boldsymbol{U}^{\mathrm{H}}(\omega). \tag{16}$$

The loudspeaker output is obtained by the overlap-and-add method of the processed binaural recording $\boldsymbol{H}(\omega, t)\boldsymbol{X}(\omega, t)$.

5. EXPERIMENT AND DISCUSSIONS

5.1. Experimental Conditions

The experiment was conducted via ten loudspeakers for reproduction, in a room 3.9 m × 3.9 m with a reverberation time of 160 ms. The length of the measured impulse response is 9600 points in 48 kHz sampling, and the inverse filter is designed using zero-padded impulse response with 16384 samples. The frequency range of the control is 150–6000 Hz. In the short-time Fourier analysis of the binaural recordings, we used a Hann window with 8192 samples.

5.2. Evaluation of Reproduction Performance at Control Points

To show that the proposed method does not degrade the reproduction at the control points, we compared the accuracy of the proposed method with that of the MP-based inverse filter. We used binaural recordings of music as the signal to be reproduced. We show the result in Table 1. The degradation is not at a problematic level and the quality is sufficient for high-fidelity listening.

5.3. Evaluation of Sound Quality Apart from the Sweet Spot

We compared the quality of sound outside the controlled area in subjective evaluation. We made in computer simulation the signals at the ears of the user 30 cm apart in front of the control points, and the sound was played back with headphones. The music is an ensemble of more than three instruments. We made four stimuli changing

Table 1. Reproduction performance at control points

Method	SNR [dB]
Conventional method with MP	50.9
Proposed method	49.7

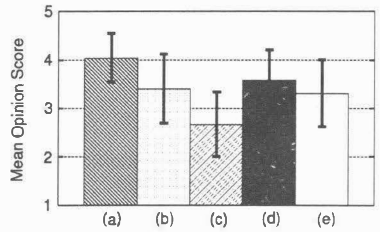

Fig. 5. Mean opinion score of sound quality out of sweet spot. The error bar shows 95% confidence intervals.

the combination of the sources. We show the results in Fig. 5. The methods compared are (a) the original binaural recording, (b) simple playback of the binaural recording with two loudspeakers, (c) the time-varying direction emphasis filter $E(\omega, t)$ described in Sect. 4.2, (d) the conventional MP-based inverse filter, and (e) the proposed method. The subjects consisted of seven males.

As a result, the sound quality of the proposed method has no significant degradation compared with the conventional method and the stereo playback. The time-varying direction emphasis filter degrades the sound quality because of the musical noise caused by the time-varying filter. However, the effect of the degradation caused by the time-varying direction emphasis filter is not problematic in the proposed method because the degradation is masked by the inverse filter.

5.4. Evaluation of Localization Ability Out of Sweet Spot

We compared the sound localization ability outside the sweet spot in the subjective evaluation. The reproduction is conducted in the real environment of the room shown in Fig. 4. As the source signals, we used performances of piano, flute and drums. The positions of the primary sources are three combinations of directions $(-90°, -30°, 30°)$, $(-60°, 0°, 60°)$ and $(-30°, 30°, 90°)$ clockwise in front of the control points. We presented 72 stimuli in a random order, and the subjects were forced to choose the answer from the directions $-90°, -60°, -30°, 0°, 30°, 60°$ and $90°$. The length of each stimulus was 5 seconds. The subjects consisted of six males and a female. We show the results in Figs. 6 and 7.

Comparing (a) and (b), the stereo playback of the binaural recording could not present directions wider than 30°. Although the conventional MP-based inverse filter has high accuracy around the front, the accuracy is low in the lateral directions. In contrast, the proposed method has no significant error and showed a 22.4% higher rate of correct answer than the conventional MP-based inverse filter.

6. CONCLUSIONS

We proposed a new sound field reproduction method robust against a user's position shift. By following the fluctuation of the wavefront with a time-varying inverse filter, the proposed method successfully presents the DOA of the sources with the binaural recording consists of an arbitrary number of sound sources without degrading the accuracy at the sweet spot. The results of subjective and objective experiments ascertained the efficacy of the proposed method. The remaining problem is DOA estimation of wavefront that arrives from behind the listener.

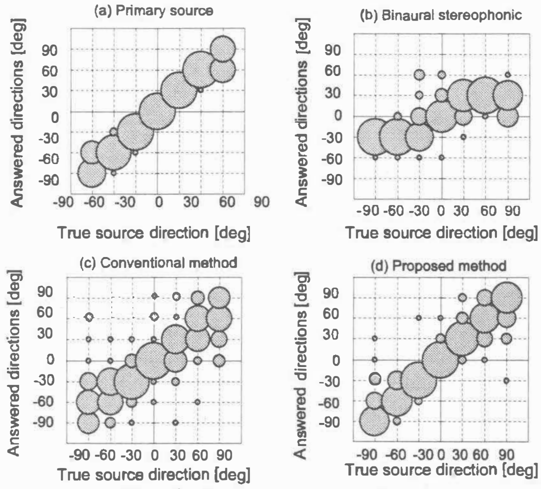

Fig. 6. The answered directions.

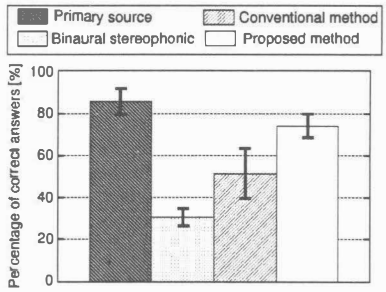

Fig. 7. Rates of correct answers. The error bar shows 95% confidence intervals.

7. REFERENCES

[1] J. Blauert, *Spatial Hearing*, MIT Press, Cambridge, MA, 1983.

[2] S. Spors, H. Buchner, R. Rabenstein, and W. Herbordt, "Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering," *J. Acoust. Soc. Am.*, vol.122, no.1, pp.354–369, 2007.

[3] S. Ise, "A principle of sound field control based on the kirchhoff-helmholtz integral equation and the theory of inverse systems," *Acustica*, vol.85, pp.78–87, 1999.

[4] J. Bauck and D. H. Cooper, "Generalized transaural stereo and applications," *J. Audio Eng. Soc.*, vol.44, no.9, pp.683-705, 1996.

[5] M. Miyoshi, and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. ASSP*, vol.36, no.2, pp.145–152, 1988.

[6] P. A. Nelson, O. Kirkeby, T. Takeuchi, and H. Hamada, "Sound fields for the production of virtual acoustic images," *J. Sound Vib.*, vol.204, no.2, pp.386–396, 1997.

[7] P. A. Nelson, H. Hamada, and S. J. Elliott, "Adaptive inverse filters for stereophonic sound reproduction," *IEEE Trans. Signal Process.*, vol.40, no.7, pp.1621–1632, 1992.

[8] S. Miyabe, M. Shimada, T. Takatani, H. Saruwatari, and K. Shikano, "Multi-Channel Inverse Filtering with Loudspeaker Selection and Enhancement for Robust Sound Field Reproduction," *Proc. IWAENC 2006*, #32, 2006.

[9] Y. Tatekura, S. Urata, H, Saruwatari and K. Shikano, "On-line relaxation algorithm applicable to acoustic fluctuation for inverse filter in multichannel sound reproduction system," *IEICE Trans. Fundam.*, vol. E88-A, no. 7, pp. 1747–1756, 2005.

[10] C. Faller and F. Baumgarte, "Binaural Cue Coding-Part II:Schemes and Applications," *IEEE Trans. Speech And Audio Processing*, vol.11, no.6, pp.520–531, 2003.

[11] J. Mouba, S. Marchand, "A source localization/ separation/ respatialization system based on unsupervised classification of interaural cues," *Proc. 9th Int. DAFx'06*, pp.233–238, 2006.

[12] B. B. Bauer, "Phasor analysis of some stereophonic phenomena," *J. Acoust. Soc. Am.,*, vol.33, pp.1536–1539, 1961.