2008 International Workshop on Nonlinear Circuits and Signal Processing
NCSP'08, Gold Coast, Australia, March 6-8, 2008

SOUND QUALITY IMPROVEMENT IN BINAURAL REPRODUCTION BASED ON SOURCE-ORIENTED TIME-VARYING INVERSE FILTER

Yuuta Yuyama†, Shigeki Miyabe[1] †, Hiroshi Saruwatari†, Kiyohiro Shikano†

†Nara Institute of Science and Technology
Nara 630-0192, Japan
Phone:+81-743-72-5287
Email: {yuuta-y, shige-m, sawatari, shikano}@is.naist.jp

ABSTRACT

We have proposed a sound field reproduction system that can retain representation of sound source's directions of arrival even outside control points where a precise sound image is reproduced. However, the problem of the privious method is that the sound quality outside the control points is low. In this paper, we address the problem and the reason of the problem is analyzed from the viewpoint of matrix inversion. Then we propose a new method to improve the sound quality. The subjective evaluation shows that the proposed method improves the sound quality of the previous method.

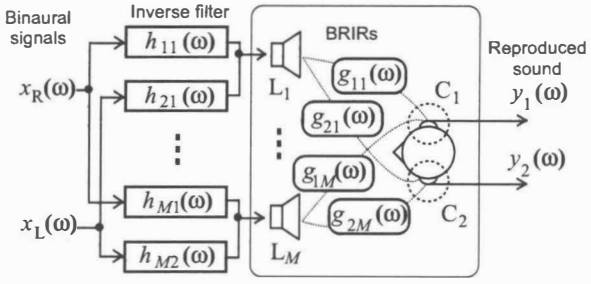

Figure 1: Configuration of a transaural system with two control points and M loudspeakers.

1. INTRODUCTION

Sound field reproduction using loudspeakers can be classified into two groups; whether or not to compensate impulse responses from the loudspeakers to the listener's ears [1]. Those without the compensation are based on a simple idea to pan the signals among multiple loudspeakers and diffuse them widely. Although this approach has an advantage in that the reproduced area is wide, its reproduction accuracy is limited. For the sake of accurate reproduction, *transaural system* reproduces a replica of sound in the human ears [2, 3]. Recorded signal at the human ear (*binaural recording*) can be reproduced at the user's ears by compensating the impulse responses of the user's ears including the reverberation, so-called binaural room impulse responses (BRIRs). Although BRIRs are in general non-minimum phase system, multiple input/output inverse theorem (MINT; [4]) proves that their inverse filters can be designed using more loudspeakers than the control points, i.e., user'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 user's movement. Although some crosstalk cancellers, which compensate anechoic head related transfer functions but BRIRs, successfully expand the sweet spots toward front and back [5], 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 [6], which interrupts listening. To mitigate the effect of the listener's movement, we have proposed an inverse filter design to weight intensity on a loudspeaker in the direction closest to the source's DOA [7]. By optimizing arbitrary components of MINT, directional cues are presented even outside the sweet spot while accuracy at control points is maintained to be in equivalent accuracy as that of the conventional MINT. However, in this method, the enhancement loudspeaker output sound is distorted.

In this paper, we address the proposed method's problem where the sound quality deterioration is pointed out. To solve the problem, we propose a new filter design method of the inverse filter to flatten the frequency characteristics. Efficacy of the proposed method is ascertained in the subjective evaluation.

2. SOURCE-ORIENTED TIME-VARYING INVERSE FILTER

2.1. Enhancement of Loudspeaker with Subspace of Generalized Inverse Matrix

[1] The author is a Research Fellow of the Japan Society for the Promotion of Science

In the transaural system, we must reproduce binaural recording at fixed control points that are arranged at the listener's ears. Such a reproduction can be achieved by designing the inverse filter of the BRIRs. By using more loudspeakers than control points, it is proven that the strict inverse system of the BRIRs with non-minimum phases can be designed [4]. Here we deal with the problem to control sound field around two control points C_n $(n = 1, 2)$ at both ears of the listener using the M (> 2) loudspeakers L_m $(m = 1, ..., M)$. We measure all the transfer functions $g_{nm}(\omega)$ from M loudspeakers L_m $(m = 1, \ldots, M)$ to the control points C_n $(n = 1, 2)$. We show the configuration of such the transaural system in Fig. 1.

We designate the binaural recording in the frequency domain as a two-dimensional vector $\boldsymbol{x}(\omega) = [x_1(\omega), x_2(\omega)]^{\mathrm{T}}$, where ω is angular frequency and $\{\cdot\}^{\mathrm{T}}$ denotes matrix transposition. We measure all the transfer functions $g_{nm}(\omega)$ from L_m to C_n for $m = 1, \ldots, M$, $n = 1, 2$. 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 $\boldsymbol{y}(\omega) = \boldsymbol{G}(\omega)\boldsymbol{H}(\omega)\boldsymbol{x}(\omega) = \boldsymbol{x}(\omega)$, and the binaural recording is reproduced.

The inverse filter $\boldsymbol{H}(\omega)$ can be designed with the generalized inverse matrix $\boldsymbol{G}^{-}(\omega)$ of the transfer functions matrix $\boldsymbol{G}(\omega)$. To obtain generalized inverse matrix, the singular value decomposition (SVD) is applied to the transfer functions matrix $\boldsymbol{G}(\omega)$.

$$\boldsymbol{G}(\omega) = \boldsymbol{U}(\omega) \underbrace{[\boldsymbol{\Gamma}(\omega), \boldsymbol{O}_{2,M-2}(\omega)]}_{2\times M} \boldsymbol{V}^{\mathrm{H}}(\omega), \tag{2}$$

$$\boldsymbol{\Gamma}(\omega) = \mathrm{diag}[\mu_1(\omega), \mu_2(\omega)], \tag{3}$$

where, $\mu_k(\omega)(k = 1, 2)$ denote singular values, diag[·] shows a diagonal matrix composed of the arguments. $\{\cdot\}^{\mathrm{H}}$ denotes the conjugate transposition, $\boldsymbol{O}_{2,M-2}(\omega)$ is a zero $2 \times (M-2)$ matrix. $\boldsymbol{U}(\omega)$ and $\boldsymbol{V}(\omega)$ are unitary matrices with the left eigenvectors and the right eigenvectors corresponding to the eigenvalues. Then generalized inverse matrix of $\boldsymbol{G}(\omega)$, denoted by $\boldsymbol{G}^{-}(\omega)$, can be written as

$$\boldsymbol{G}^{-}(\omega) = \boldsymbol{V}(\omega) \underbrace{\begin{bmatrix} \boldsymbol{\Lambda}(\omega) \\ \boldsymbol{S}_{M-2,2}(\omega) \end{bmatrix}}_{M\times 2} \boldsymbol{U}^{\mathrm{H}}(\omega), \tag{4}$$

$$\boldsymbol{\Lambda}(\omega) = \mathrm{diag}[\lambda_1(\omega), \lambda_2(\omega)], \tag{5}$$

$$\lambda_k(\omega) = \begin{cases} \frac{1}{\mu_k(\omega)} & (\text{if } \mu_k(\omega) \neq 0) \\ 0 & (\text{otherwise}). \end{cases} \tag{6}$$

where $\boldsymbol{S}(\omega)$ is an arbitrary subspace exists in the generalized inverse matrix of a non-square matrix. To fix the ambiguity of the generalized inverse matrix, the Moore-Penrose (MP) generalized inverse matrix $\boldsymbol{G}^{+}(\omega)$ is generally used [8].

$$\boldsymbol{G}^{+}(\omega) = \boldsymbol{V}(\omega) \underbrace{\begin{bmatrix} \boldsymbol{\Lambda}(\omega) \\ \boldsymbol{O}_{M-2,2}(\omega) \end{bmatrix}}_{M\times 2} \boldsymbol{U}^{\mathrm{H}}(\omega). \tag{7}$$

In contrast, our previous method optimizes the nullspace of the generalized inverse matrix. At first, we design a multichannel filter as a *target filter*. Target filter $\boldsymbol{L}(\omega)$ has weighted gain in the channel of loudspeakers whose direction of arrivals (DOAs) are close to that of a source. To make the optimal inverse filter $\boldsymbol{H}(\omega)$, we minimize the distance between the inverse filter and the time-varying target filter $\boldsymbol{L}(\omega)$.

$$\boldsymbol{H}(\omega) = \underset{\boldsymbol{G}^{-}(\omega)}{\mathrm{argmin}} \left\| \boldsymbol{G}^{-}(\omega) - \boldsymbol{L}(\omega) \right\|_{\mathrm{Fr}}. \tag{8}$$

where, $\|\cdot\|_{\mathrm{Fr}}$ denotes Frobenius norm given as $\|[a_{mn}]_{mn}\|_{\mathrm{Fr}} = \sqrt{\sum_m \sum_n |a_{mn}|^2}$. From Eq.(4), square of Frobenius norm between $\boldsymbol{G}^{-}(\omega)$ and $\boldsymbol{T}(\omega)$, defined by $F(\omega)$, can be written as

$$\begin{aligned} F(\omega) &= \left\| \boldsymbol{G}^{-}(\omega) - \boldsymbol{L}(\omega) \right\|_{\mathrm{Fr}}^2 \\ &= \left\| \boldsymbol{V}(\omega) \begin{bmatrix} \boldsymbol{\Lambda}(\omega) \\ \boldsymbol{S}(\omega) \end{bmatrix} \boldsymbol{U}^{\mathrm{H}}(\omega) - \boldsymbol{L}(\omega) \right\|_{\mathrm{Fr}}^2, \end{aligned} \tag{9}$$

Since the Frobenius norm is not changed by multiplication of unitary matrices, $F(\omega)$ can be rewritten as

$$\begin{aligned} F(\omega) &= \left\| \boldsymbol{V}^{\mathrm{H}}(\omega) \left(\boldsymbol{G}^{-}(\omega) - \boldsymbol{L}(\omega) \right) \boldsymbol{U}(\omega) \right\|_{\mathrm{Fr}}^2 \\ &= \left\| \begin{bmatrix} \boldsymbol{\Lambda}(\omega) \\ \boldsymbol{S}(\omega) \end{bmatrix} - \boldsymbol{V}(\omega)^{\mathrm{H}} \boldsymbol{L}(\omega) \boldsymbol{U}(\omega) \right\|_{\mathrm{Fr}}^2 \\ &= \left\| \begin{bmatrix} \boldsymbol{\Lambda} - \boldsymbol{V}_{\mathrm{span}}^{\mathrm{H}}(\omega) \boldsymbol{L}(\omega) \boldsymbol{U}(\omega) \\ \boldsymbol{S}(\omega) - \boldsymbol{V}_{\mathrm{null}}^{\mathrm{H}}(\omega) \boldsymbol{L}(\omega) \boldsymbol{U}(\omega) \end{bmatrix} \right\|_{\mathrm{Fr}}^2 \\ &= \left\| \boldsymbol{\Lambda}(\omega) - \boldsymbol{V}_{\mathrm{span}}^{\mathrm{H}}(\omega) \boldsymbol{L}(\omega) \boldsymbol{U}(\omega) \right\|_{\mathrm{Fr}}^2 \\ &\quad + \left\| \boldsymbol{S}(\omega) - \boldsymbol{V}_{\mathrm{null}}^{\mathrm{H}}(\omega) \boldsymbol{L}(\omega) \boldsymbol{U}(\omega) \right\|_{\mathrm{Fr}}^2, \end{aligned} \tag{10}$$

where, $\boldsymbol{V}_{\mathrm{span}}(\omega)$ is a truncated matrix of $\boldsymbol{V}(\omega)$ and is composed of eigenvectors that span the row space of $\boldsymbol{G}(\omega)$ as

$$\boldsymbol{V}_{\mathrm{span}}(\omega) = [\boldsymbol{v}_1(\omega), \boldsymbol{v}_2(\omega)], \tag{11}$$

Similarly, $\boldsymbol{V}_{\mathrm{null}}(\omega)$ is truncated matrix of $\boldsymbol{V}(\omega)$ and is composed of unit vectors that span the null space of $\boldsymbol{G}(\omega)$ as

$$\boldsymbol{V}_{\mathrm{null}}(\omega) = [\boldsymbol{v}_3(\omega), \ldots, \boldsymbol{v}_M(\omega)], \tag{12}$$

In Eq. (10), $\left\| \boldsymbol{\Lambda}(\omega) - \boldsymbol{V}_{\mathrm{span}}^{\mathrm{H}}(\omega) \boldsymbol{L}(\omega) \boldsymbol{U}(\omega) \right\|_{\mathrm{Fr}}^2$ cannot be changed because $\boldsymbol{\Lambda}(\omega)$ is fixed to satisfy the generalized inverse matrix

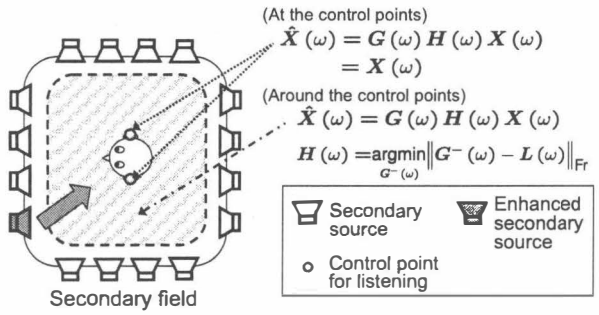

Figure 2: Sound field reproduction system with conventional method [7].

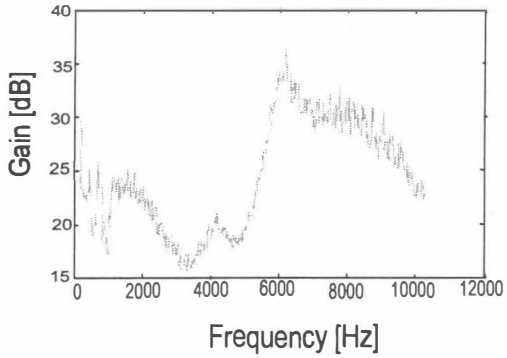

Figure 3: Gain of the inverse filter with MP.

of $\boldsymbol{G}(\omega)$. On the other hand, $\boldsymbol{S}(\omega)$ is arbitrary and the term $\left\|\boldsymbol{S}(\omega) - \boldsymbol{V}_{\text{null}}^{\text{H}}(\omega)\boldsymbol{L}(\omega)\boldsymbol{U}(\omega)\right\|_{\text{Fr}}^2$ can be minimized to zero by a substitution

$$\boldsymbol{S}(\omega) = \boldsymbol{V}_{\text{null}}^{\text{H}}(\omega)\boldsymbol{L}(\omega)\boldsymbol{U}(\omega), \tag{13}$$

then $F(\omega)$ is minimized. Therefore, substituting Eq. (13) in Eq. (4), the optimal solution can be obtained as

$$\boldsymbol{H}(\omega) = \boldsymbol{V}(\omega)\begin{bmatrix} \boldsymbol{\Lambda}(\omega) \\ \boldsymbol{V}_{\text{null}}^{\text{H}}(\omega)\,\boldsymbol{L}(\omega)\,\boldsymbol{U}(\omega) \end{bmatrix}\boldsymbol{U}^{\text{H}}(\omega). \tag{14}$$

Thereby, the user can feel the sound image toward the enhanced source even outside the control points. Simultaneously, the accurate reproduction at the control points can be achieved as well as the inverse filter with MP. Then, we use this as an inverse filter [7]. When the sources DOA change, we must change the channel of enhancement loudspeakers.

2.2. Design of Target Filter

We design the target filter to satisfy the following requirements:

- The difference of delay
- The difference of gain.

As for the first point, we synchronize the peak of the target filter and the inverse filter with MP. At first we obtain the time delay τ when the impulse response of the inverse filter has the largest amplitude in the time domain. Then we give the target filter's linear phases with the delay of τ. As for gain, we arrange it to be the same as the MP generalized inverse filter; thus

$$\boldsymbol{L}_{mn}(\omega) = \begin{cases} \frac{\|\boldsymbol{G}^{+}(\omega)\|_{\text{Fr}}}{\sqrt{2}} \cdot e^{-j\omega\tau} & (\text{if } m = k) \\ 0 & (\text{otherwise}), \end{cases} \tag{15}$$

where k is the number of loudspeakers. We show the configuration of the sound field reproduction system with conventional method in Fig. 2.

3. IMPROVEMENT OF SOUND QUALITY

In this section, we describe the problem of the conventional method, where the deterioration of the sound quality is pointed out. Then we propose a new method to solve the sound quality problem.

3.1. Problem Analysis of Target Filter

In the conventional design of target filter, the target filter output sound is distorted, and pre-echo is noticeable. In the method, we arrange the gain of the target filter to be the same as the MP generalized inverse filter. However, gain of the MP generalized inverse filter is determined for compensation of the characteristics only at the control points and then the appropriate characteristic outside the control points is not guaranteed. Also, the Frobenius norm of the MP generalized inverse matrix is highly related to the condition number of the MP generalized inverse matrix. We show the Frobenius norm of the MP generalized inverse matrix in Fig. 3. The frequency characteristics are diverse over frequency bins. This results in pre-echo which is caused by dip in frequency characteristics. Also the gain is too large in low frequency range, and the sound quality degradation arises.

3.2. Improvement

To flatten the frequency characteristics, we fix the gain of the target filter through all the frequency bins. As the fixed gain, we utilize the average of the Frobenius norm of the MP generalized inverse matrix in all the frequency bins. Thereby, the target filter output sound can realize close to a primary sound;

$$\boldsymbol{L}_{mn}(\omega) = \begin{cases} \rho \cdot e^{-j\omega\tau} & (\text{if } m = k) \\ 0 & (\text{otherwise}), \end{cases} \tag{16}$$

$$\rho = \sqrt{\frac{1}{\pi}\int_0^{\pi} \frac{\|\boldsymbol{G}^{+}(\omega)\|_{\text{Fr}}^2}{\sqrt{2}}\, d\omega}. \tag{17}$$

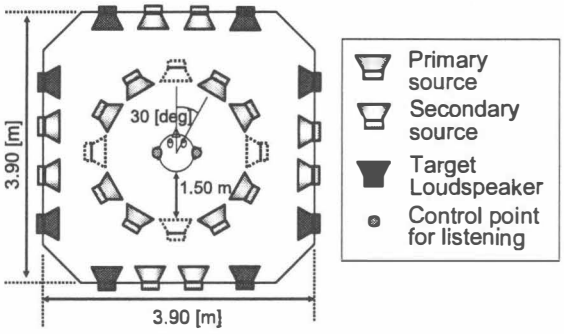

Figure 4: Experimental conditions.

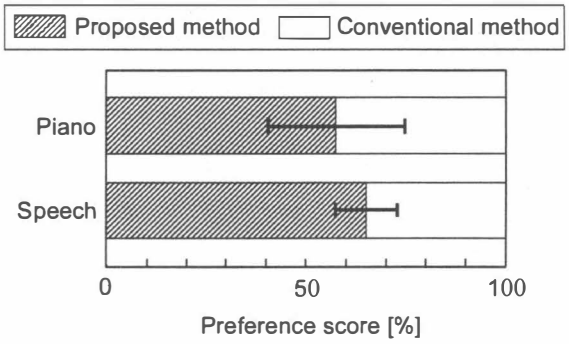

Figure 5: Experimental results. The error bar shows 95% confidence intervals.

4. EXPERIMENT

4.1. Conditions

We compared the sound quality outside the control points in subjective evaluation. The experiment was conducted via sixteen loudspeakers for the reproduction, in a room 3.9 m × 3.9 m with a reverberation time of 160 ms (Fig. 4). We used sources of piano and speech with sampling frequency of 48 kHz. The position of the sound sources are set at 1.5 m apart from the user and their directions are (±30°, ±60°, ±120°, ±150°) clockwise, where the direction in front of the user is set to be 0°. The loudspeakers for reproduction were set in the same directions as the sound sources with different distance from the user. The length of the measured impulse response is 12000 points in 48 kHz sampling, and the FFT length is 131072 points. The frequency range of the control is 150–10000 Hz. We made 48 patterns of signals to be reproduced in simulations. The quality of the proposed method and the conventional method are compared in the XAB test. The length of each stimulus is 15 seconds. The subjects consist of four males and a female in their twenties.

4.2. Result

We show the experimental results in Fig. 5. The scores of the conventional method and the proposed method in piano sound were 57.5% and 42.5%. Although the proposed method can slightly outperform, in the mean values the difference between the two methods in piano sound is not remarkable with statistical confidence. On the other hand, the scores of the conventional method and the proposed method in speech were 65.0% and 35.0%. The difference between the two methods in speech is obvious. Therefore, it is ascertained that the proposed method improves the sound quality of the conventional method.

5. CONCLUSION

We proposed a sound quality improvement method in binaural reproduction based on source-oriented time-varying inverse filter. The result of subjective experiments showed the efficacy of the proposed method.

REFERENCES

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

[2] M. R. Schroeder, and B. S. Atal, "Computer simulation of sound transmission in rooms," *IEEE Conv. Rec.,* vol.7, pp.150–155, 1963.

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

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

[5] 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.

[6] 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.

[7] 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.

[8] 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.