

ROBUST SOUND FIELD REPRODUCTION INTEGRATING MULTI-POINT SOUND FIELD CONTROL AND WAVE FIELD SYNTHESIS

Noriyoshi Kamado, Hiroshi Saruwatari, and Kiyohiro Shikano

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

ABSTRACT

For a reproduced sound field, the competing goals between the listening area and reproduction accuracy in an actual environment is one of the most important problems in sound field reproduction using loudspeakers. In this paper, we propose a new method of balancing these goals with absolute accuracy using an inverse filter of the room acoustics: the null space of a generalized inverse matrix given by a compensation filter of the wave field outside the control points. To develop an expression for the compensation filter, we use the loudspeaker driving function of wave field synthesis (WFS) instead of the filter used in conventional studies. By using WFS, the proposed method overcomes the compensation limitation of auditory distance and azimuth perception outside the control points. The results of computer simulations revealed that the proposed method balances the competing goals and has wide applicability in a spatial domain with high accuracy of reproduction both under free-field conditions and in a simulation model with room reflection.

Index Terms— Sound field reproduction, Multi-point controlled inverse filtering, Wave field synthesis

1. INTRODUCTION

The ultimate objective of sound field reproduction using loudspeakers is to perfectly reproduce the characteristics of natural hearing over the entire spatial and frequency domains. Reproduction methods are classified into two groups: reproduction of multi-point pressure and that of wavefronts. Many of the systems in the first group are based on inverse filtering. The *transaural system* based on the multiple input/output inverse theorem (MINT) [1] is one of the typical systems aimed at reproducing binaural recording to achieve a realistic sensation, where an inverse filter compensates for transfer functions at the user's ears including room reverberation. However, outside the control points (*sweet spot*), the inverse filter does not compensate satisfactorily, and it is known that a method based on the inverse filtering of multi-point controlled reproduction does not work well with user movements. *Multi-point controlled wavefront synthesis* (MCWS) [2] is a technique to compensate for this problem using a number of control points located in the listening area. However, ensuring a wide sweet spot by locating a sensor array on the control points fills the reproduction space with a large set of sensors, which is impractical. With the practical location of the control points, the potential problem in the reproduction area persists in MCWS since the wavefront formed outside the control points is entirely different from that formed in the primary sound field, and sound localization degrades considerably when the user moves from the controlled area.

To mitigate the effect of a listener's movement in such multi-point control-based methods, we have proposed an improved inverse filter design. This method can maintain the perceptual direction of sound outside the sweet spot, while keeping the accuracy at the control points by substituting arbitrary subspace components of MINT with a simple emphasis filter that drives the nearest loudspeaker to

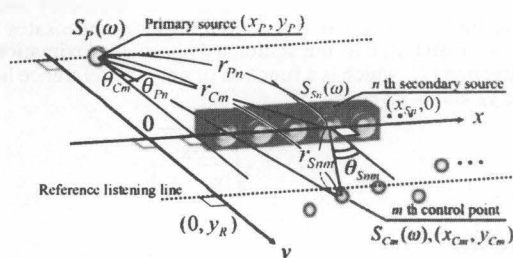


Fig. 1. Configuration of WFS and MCWS.

the primary source [3]. However, this emphasis filter cannot compensate for the perceptual distance, and the reproducible position of the primary source becomes circumscribed only at the loudspeaker position. Therefore, this framework still requires further improvement in accuracy and a flexible method to reproduce the arbitrary primary wavefront regardless of the loudspeaker position.

On another front, in recent years, many of the sound field reproduction systems in the second group, based on wavefront synthesis, have been extensively investigated. *Wave field synthesis* (WFS) [4] is premised on an anechoic reproduction environment and offers a large listening area with a high perceptual reproduction quality for multiple listeners. However, the reproduction accuracy decreases in practice owing to the inherent disadvantages of physical inaccuracies found in the Kirchhoff-Helmholtz integral and room reverberation [5]. Therefore, WFS cannot accurately reproduce the sound field at the sweet spot.

In this paper, we propose a new robust sound field reproduction method that enables high-accuracy reproduction with a wide listening area by integrating our previously proposed multi-point sound field reproduction method and WFS. In the proposed method, the wavefront for the desired spatial cue outside the sweet spot is derived from the approximation of the WFS-synthesized wavefront, but the sweet-spot sound is perfectly preserved. The efficacy of the proposed method is ascertained by objective evaluation through computer-based simulations. The main contribution of this paper is to bridge the theories of multi-point control-based sound field reproduction and WFS, while achieving accurate sweet-spot reproduction and wide-area wavefront generation.

2. CONVENTIONAL METHOD OF SOUND FIELD REPRODUCTION

2.1. WFS

In this section, WFS and MCWS are described theoretically and the equations used for numerical calculations are derived in detail. The geometric configuration and parameters in WFS are depicted in Fig. 1, where $S_p(\omega)$ and $S_{s,n}(\omega)$ denote the spectra of the primary and n th secondary sources, respectively, on the x - y horizontal plane.

The spectrum of the n th secondary source, which synthesizes the

primary spherical wavefront, $S_{S_n}(\omega)$ is expressed as [6]

$$S_{S_n}(\omega) = Q_n^{(WFS)}(\omega) S_P(\omega) = S_P(\omega) \sqrt{\frac{jk}{2\pi}} C(y_R, y_P) \frac{\exp(-jkr_{Pn})}{\sqrt{r_{Pn}}} \frac{\cos \theta_{Pn}}{G(\theta_{Pn}, \omega)} \Delta x, \quad (1)$$

where $Q_n^{(WFS)}(\omega)$ is the wavefront synthesis filter, j is the imaginary unit, k is the wavenumber (ω/c), c is the sound velocity, ω is the angular frequency, Δx is the interelement interval among the secondary sources, r_{Pn} is the distance between the primary source and the n th secondary source, and θ_{Pn} is the angle between the y -axis and the line connecting the n th secondary and primary sources. $G(\theta_{Pn}, \omega)$ is a distance-independent directivity function defined only under far-field conditions. (y_R, y_P) is a function that compensates for the level of mismatch due to the stationary phase approximation along the x -direction [7], which is a function of only the reference listening distance y_R and is given as

$$C(y_R, y_P) = \sqrt{\frac{|y_R|}{|y_R - y_P|}}. \quad (2)$$

2.2. MCWS

The geometric parameters of MCWS are shown in Fig. 1. MCWS controls the spatial spectra at the control points, which are located on the x - y horizontal plane in front of the secondary sources, and generates the desired wavefront. Here, $S_{C_m}(\omega)$ denotes the secondary wavefront spectrum at the m th control-point position. Also, θ_{C_m} is the angle between the y -axis and the line connecting the m th control point and the primary source, $\theta_{S_{nm}}$ is the angle between the y -axis and the line connecting the m th control point and the n th secondary source, r_{C_m} is the spatial distance between the m th control point and the primary source, $r_{S_{nm}}$ is the spatial distance between the m th control point and the n th secondary source, N is the number of secondary sources, and M is the number of control points.

Here, we derive the spectrum of the secondary source $S_{S_n}(\omega)$, which synthesizes the primary spherical wavefront. The transfer function between the n th secondary monopole source and the m th control point, $H_{nm}(\omega)$, is written as

$$Z_{nm}(\omega) = \frac{\exp(-jkr_{S_{nm}})}{r_{S_{nm}}}, \quad (3)$$

From Eq. (3), we define the transfer function matrix

$$\mathbf{Z}(\omega) = \begin{bmatrix} Z_{1,1}(\omega) & Z_{2,1}(\omega) & \cdots & Z_{N,1}(\omega) \\ Z_{1,2}(\omega) & Z_{2,2}(\omega) & \cdots & Z_{N,2}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ Z_{1,M}(\omega) & Z_{2,M}(\omega) & \cdots & Z_{N,M}(\omega) \end{bmatrix}. \quad (4)$$

We write the secondary wavefront spectrum vector at the m th control-point position as

$$S_C(\omega) = \mathbf{Z}(\omega) S_S(\omega), \quad (5)$$

where

$$S_C(\omega) = [S_{C1}(\omega), S_{C2}(\omega), \dots, S_{CM}(\omega)]^T, \quad (6)$$

$$S_S(\omega) = [S_{S1}(\omega), S_{S2}(\omega), \dots, S_{SN}(\omega)]^T, \quad (7)$$

and the superscript T denotes the transpose of the vector/matrix. If the primary wavefront spectrum is equal to the secondary wavefront spectrum at the control-point position, Eq. (5) can be transformed into

$$S_C(\omega) = \mathbf{W}(\omega) S_P(\omega), \quad (8)$$

where

$$\mathbf{W}(\omega) = \begin{bmatrix} \frac{e^{-jkr_{C1}}}{r_{C1}} & \frac{e^{-jkr_{C2}}}{r_{C2}} & \cdots & \frac{e^{-jkr_{CM}}}{r_{CM}} \end{bmatrix}^T. \quad (9)$$

From Eqs. (5) and (8) and the Moore-Penrose (MP) generalized inverse matrix of $\mathbf{Z}(\omega)$, $\mathbf{Z}^+(\omega)$, we obtain the secondary source spectrum vector $S_S(\omega)$ with a wavefront synthesis filter of MCWS

$Q^{(MC)}(\omega)$ in the form,

$$S_S(\omega) = Q^{(MC)}(\omega) S_P(\omega) = \mathbf{Z}^+(\omega) \mathbf{W}(\omega) S_P(\omega). \quad (10)$$

3. PROPOSED METHOD

To improve the robustness against a shift of the user's position in MCWS and MINT, we have proposed an inverse filter design method in which a wavefront radiated from the loudspeaker nearest to the primary source is inserted in the subspace not spanned for the reproduced signal space (sweet-spot-signal space) in the inverse filter matrix [3]. This method can approximately provide a sound field even outside the control points without any degradation of reproduction at the control points (sweet spot). However, this method has a major disadvantage of the localization reproduction ability outside the sweet spot because the reproducible primary source position is quantized within the loudspeaker position, i.e., we cannot compensate for the perceptual distance outside the sweet spot.

In contrast, WFS can reproduce an arbitrary wavefront that compensates for the perceptual distance. However, as described in Sect. 1, WFS cannot accurately generate sound pressure at the control points owing to the existence of room reverberation and the approximation in the theory.

The above-mentioned facts imply that the advantages and disadvantages of our previous method and WFS are complementary; this motivates us to propose an approach combining multi-point sound field reproduction and WFS in this paper. In the proposed method, the wavefront outside the sweet spot is derived from an approximation of the WFS-synthesized wavefront, and we insert it in the subspace in the inverse filter matrix. As the result, we can simultaneously achieve the following: (a) we can reproduce perfect sound pressures in the sweet spot (control points) that are not disturbed by the WFS wavefront, and (b) we can perceive the approximated wavefront reproduced by WFS outside the sweet spot. The detailed algorithm is described below.

Utilizing singular value decomposition, the generalized inverse matrix $\mathbf{Z}^-(\omega)$ of the transfer impedance matrix $\mathbf{Z}(\omega)$ can be denoted as

$$\mathbf{Z}^-(\omega) = \underbrace{\mathbf{V}(\omega)}_{(N \times N)} \underbrace{\begin{bmatrix} \Lambda(\omega) \\ \mathbf{S}(\omega) \end{bmatrix}}_{(N \times M)} \underbrace{\mathbf{U}^H(\omega)}_{(M \times M)}, \quad (11)$$

where the superscript H denotes the complex conjugate transposition of a matrix, $\mathbf{V}(\omega)$ and $\mathbf{U}(\omega)$ are the unitary matrices whose columns are the right and left singular vectors of $\mathbf{Z}(\omega)$, respectively, and $\Lambda(\omega)$ is

$$\Lambda(\omega) = \text{diag}[\lambda_1(\omega), \dots, \lambda_M(\omega)], \quad (12)$$

where λ_m is expressed with the singular values σ_m of $\mathbf{Z}(\omega)$ as

$$\lambda_m(\omega) = \begin{cases} \frac{1}{\sigma_m(\omega)} & (\text{if } \sigma_m(\omega) \neq 0), \\ 0 & (\text{otherwise}). \end{cases} \quad (13)$$

The MP generalized inverse matrix $\mathbf{Z}^+(\omega)$ can be obtained by setting $\mathbf{S}(\omega)$ to be a zero matrix. However, the MP-type inverse filter is specific 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 a controlled area.

Next, to approximate $T(\omega)$, which is the wavefront control filter outside the control points in the subspace (or nullspace) of $\mathbf{Z}^-(\omega)$ with arbitrary components $\mathbf{S}(\omega)$ in Eq. (11), obtain the generalized inverse matrix $\mathbf{Z}^-(\omega)$ closest to $T(\omega)$. We utilize the Frobenius norm as the distance measure and we obtain $\mathbf{Z}^-(\omega)$ to minimize $F(\omega) = \|\mathbf{Z}^-(\omega) - T(\omega)\|_F$. Since the Frobenius norm is not changed by the multiplication of unitary matrices, $F(\omega)$ can be rewritten as

$$F(\omega) = \|\mathbf{V}^H(\omega) (\mathbf{Z}^-(\omega) - T(\omega)) \mathbf{U}(\omega)\|_F = \left\| \begin{bmatrix} \Lambda(\omega) - \mathbf{V}_{\text{span}}^H(\omega) T(\omega) \mathbf{U}(\omega) \\ \mathbf{S}(\omega) - \mathbf{V}_{\text{null}}^H(\omega) T(\omega) \mathbf{U}(\omega) \end{bmatrix} \right\|_F, \quad (14)$$

where $\mathbf{V}_{\text{span}}(\omega)$ is a matrix composed of the first M columns of $\mathbf{V}(\omega)$. Since $\Lambda(\omega)$ is a constant matrix, $F(\omega)$ can be minimized if and only

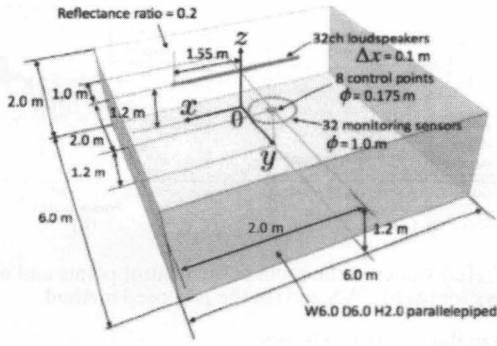


Fig. 2. Configurations of loudspeaker array, control points, monitoring sensors and listening room.

if $S(\omega) - V_{\text{null}}^H(\omega)T(\omega)U(\omega) = 0$; thus, the optimal inverse filter is obtained as follows by setting $S(\omega) = V_{\text{null}}^H(\omega)T(\omega)U(\omega)$ in Eq. (11):

$$\begin{aligned} Z_{\text{opt}}^-(\omega) &= \underset{Z^-(\omega)}{\text{argmin}} F(\omega) \\ &= V(\omega) \left[V_{\text{null}}^H(\omega)T(\omega)U(\omega) \right] U^H(\omega). \end{aligned} \quad (15)$$

Next, we design the filter to guarantee the sound field accuracy outside the control points. As a method of generating the desired wavefront, WFS was introduced in this study. From Eq. (1), the spectrum at the control points can be written in terms of the impedance matrix $Z(\omega)$ as

$$S_C(\omega) = Z(\omega)Q^{(\text{WFS})}(\omega)S_P(\omega), \quad (16)$$

$$Q^{(\text{WFS})}(\omega) = [Q_1^{(\text{WFS})}(\omega), \dots, Q_N^{(\text{WFS})}(\omega)]^T. \quad (17)$$

Equally, the spectrum of MCWS at the same control points can be written as

$$S_C(\omega) = Z(\omega)Q^{(\text{MC})}(\omega)S_P(\omega) = Z(\omega)Z^+(\omega)W(\omega)S_P(\omega). \quad (18)$$

Equation (17) is equivalent to Eq. (18) because WFS and MCWS synthesize identical primary wavefronts, and from the equivalence of these equations, the filter $T(\omega)$ has to satisfy the condition $Q^{(\text{WFS})}(\omega) = T(\omega)W(\omega)$. Therefore, the filter $T(\omega)$ is obtained as

$$T(\omega) = Q^{(\text{WFS})}(\omega)W_{\text{ortho}}(\omega), \quad (19)$$

where $W_{\text{ortho}}(\omega)$ is an orthonormal vector of $W(\omega)$ obtained by singular value decomposition.

4. COMPUTER-BASED SIMULATION AND DISCUSSION

4.1. Simulation conditions and evaluation criteria

To illustrate the properties of the proposed method, the frequency domain and spatial domain descriptions of the synthesized wavefront are used for numerical simulations. The configuration of the simulation system is shown in Fig. 2. The simulation was conducted via 32ch linear-array loudspeakers for reproduction in a free field and listening room situation. The reflected waves under the room condition are regarded as direct waves generated from imaginary sources obtained by *image method* [8]. In this numerical simulation, we use the imaginary sources up to the second order. To show the wide applicability of the proposed method in the spatial domain, we evaluate the sound field reproduction error

$$E_S(x, y, \omega) = \frac{|SWF(x, y, \omega) - PWF(x, y, \omega)|^2}{|SWF(x, y, \omega)|^2}, \quad (20)$$

where $PWF(x, y, \omega)$ denotes the primary wavefront radiated by primary sources and $SWF(x, y, \omega)$ denotes the synthesized secondary wavefront.

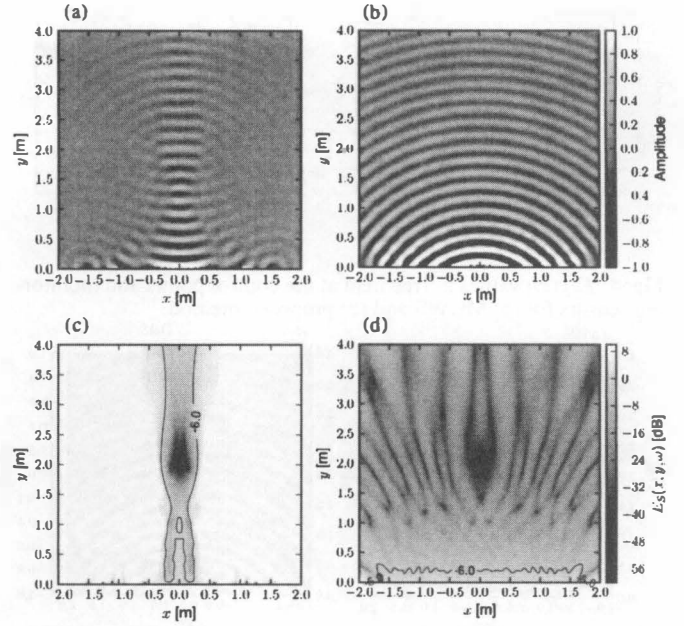


Fig. 3. Wavefronts synthesized at 1600 Hz in free field by (a) MCWS and (b) proposed method. $E_S(\omega)$ values of MCWS and the proposed method are represented in (c) and (d), respectively.

To evaluate the accuracy of the reproduction at the control points, we compared the normalized quadratic reproduction error of the proposed method with that of the MCWS at the control points using [9, 10]

$$E_{LS}(\omega) = \sum_{x_0, y_0} \frac{|SWF(x_0, y_0, \omega) - PWF(x_0, y_0, \omega)|^2}{|SWF(x_0, y_0, \omega)|^2}, \quad (21)$$

where x_0 and y_0 denote the coordinates of the observation point, including the locations of the control points and monitoring sensors. We calculate $E_{LS}(\omega)$ at the monitoring sensor location to evaluate the accuracy of the proposed method outside the sweet spot.

4.2. Simulation results

Figures 3(a) and 3(b) respectively show the wave fields synthesized by MCWS and the proposed method in the free field for a monopole source at position $(x_p, y_p, z_p) = (0.0, -1.0, 1.2)$ [m], where the radiated signal frequency is 1600 Hz. The evaluated wavefront frequency of 1600 Hz is the upper limit of major cues for sound source localization. Figures 3(c) and 3(d) show the values of $E_S(x, y, \omega)$ corresponding to Figs. 3(a) and 3(b), respectively. In Figs. 3(c) and 3(d), the dark areas represent high reproduction accuracy with a small error. In addition, the black solid line indicates the level of $E_S = -6.0$ dB, which is suggested to be the acceptable reproduction region limit for comparison purposes [9, 10]. As can be seen in Figs. 3(a) and 3(c), the reproduction error of MCWS is large because the MP-type inverse filter cannot guarantee the correct wavefront outside the control points. In contrast, in Figs. 3(b) and 3(d), the reproduction error of the proposed method is smaller than that of MCWS, and is generally smallest in the vicinity of the control points.

In Figs. 4(a) and 4(b), the $E_{LS}(\omega)$ values of WFS, MCWS and the proposed method at the control points and monitoring sensors are shown. The frequency of 1700 Hz is indicated in these figures, corresponding to the WFS spatial aliasing frequency. In Fig. 4(a), the $E_{LS}(\omega)$ of the proposed method at the control points is extremely small compared with that of WFS throughout all the frequencies, and is equivalent to that of MCWS. Thus, the reproduction accuracy at the control points is maintained in the proposed method. In addition, in Fig. 4(b), the $E_{LS}(\omega)$ of the proposed method at the monitoring sensors is smaller than that of MCWS below the spatial aliasing

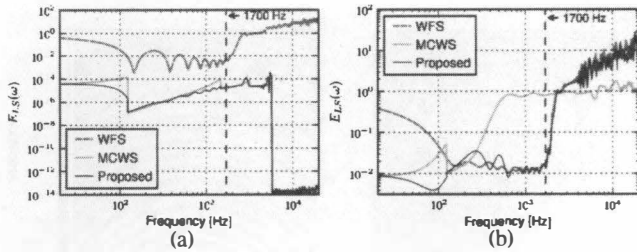


Fig. 4. $E_{LS}(\omega)$ values in free field at the control points and monitoring sensors for (a) MCWS and (b) proposed method.

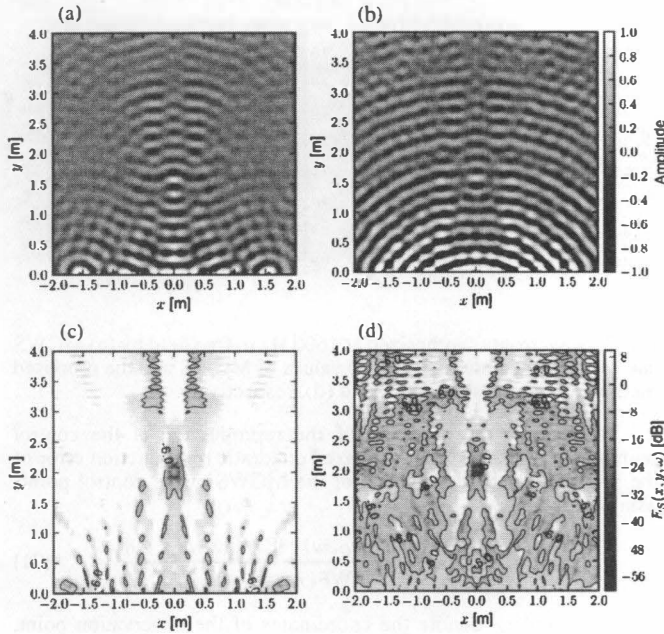


Fig. 5. Wavefronts synthesized at 1600 Hz in virtual room by (a) MCWS and (b) proposed method. $E_S(\omega)$ values of MCWS and the proposed method are represented in (c) and (d), respectively.

frequency, and is almost equivalent to or smaller than that of WFS. Thus, it is possible to compensate for the reproduction accuracy outside the control points using the proposed method. However, in the proposed method, the coloration problem of WFS will occur outside the control points because the error-increasing tendency of the proposed method is similar to that of WFS above the spatial aliasing frequency. In addition, the discontinuity of $E_{LS}(\omega)$ due to the truncation of singular values in SVD is shown in the results of MCWS and the proposed method.

Figures 5(a) and 5(b) respectively show the wave fields synthesized by MCWS and the proposed method in the room for a monopole source at position $(x_p, y_p, z_p) = (0.0, -1.0, 1.2)$ [m]. Figures 5(c) and 5(d) show the $E_S(\omega)$ values for Figs. 5(a) and 5(b), respectively. From Figs. 5(a) and 5(b), we can see that reflected waves in the room disturb the synthesized wavefront. In Figs. 5(c) and 5(d), since the interference of room-reflected waves occurs, the acceptable reproduction regions become narrower than that in free-field simulation results; however, the proposed method maintains this region compared with MCWS. Figures 6(a) and 6(b) show the $E_{LS}(\omega)$ values of WFS, MCWS and the proposed method at the control points and monitoring sensors. In Fig. 6(a), the proposed method also maintains the reproduction accuracy at the control points. However, in Fig. 6(b), $E_{LS}(\omega)$ increases outside the control points for all of the methods, and that of the proposed method fluctuates significantly between the results of MCWS and WFS. Thus, as the room reverberation increases, the reproduction errors outside the control points of the proposed method probably

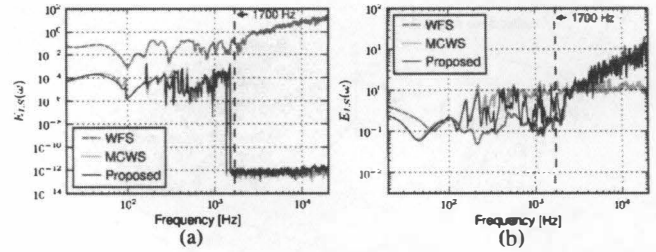


Fig. 6. $E_{LS}(\omega)$ values in the room at the control points and monitoring sensors for (a) MCWS and (b) the proposed method.

approach to the results of MCWS.

5. CONCLUSION

In this paper, we proposed a new method of balancing the listening area and reproduction accuracy with absolute accuracy using an inverse filter of the room acoustics: the null space of the generalized inverse matrix given by a compensation filter of the wave field outside the control points. To develop an expression for the compensation filter, we used the loudspeaker driving function of WFS instead of the filter used in conventional studies. By using WFS, the proposed method overcomes the compensation limitation of auditory distance and azimuth perception outside the control points. The results of computer simulations revealed that the proposed method balances the above goals and has wide applicability in a spatial domain with high accuracy of reproduction both under free-field conditions and in a simulation model with room reflection.

6. REFERENCES

- [1] M. Miyoshi, Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. on ASSP*, vol. 36, no. 2, pp. 683–705, 1988.
- [2] N. Kamado, H. Hokari, S. Shoji, H. Saruwatari, K. Shikano, "Sound field reproduction by wavefront synthesis using directly aligned multi point control," *Proc. 40th AES Conference*, 2010.
- [3] S. Miyabe, M. Shimada, T. Takatani, H. Saruwatari, K. Shikano, "Multi-channel inverse filtering with loudspeaker selection and enhancement for robust sound field reproduction," *Proc. IWAENC 2006*, 2006.
- [4] A. J. Berkhout, D. de Vries, P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, vol. 93, no. 5, pp. 2764–2778, 1993.
- [5] W. de Bruijn, *Application of wave field synthesis in videoconferencing*, Ph.D. thesis, Delft University of Technology, 2004.
- [6] D. de 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.
- [7] E. W. Start, *Direct sound enhancement by wave field synthesis*, Ph.D. thesis, Delft University of Technology, 1997.
- [8] J. B. Allen, D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.
- [9] O. Kirkeby, P. A. Nelson, "Reproduction of plane wave sound fields," *J. Acoust. Soc. Am.*, vol. 94, no. 5, pp. 2992–3000, 1993.
- [10] P. A. Gauthier, A. Berry, W. Woszczyk, "Sound-field reproduction in-room using optimal control techniques: Simulations in the frequency domain," *J. Acoust. Soc. Am.*, vol. 117, no. 2, pp. 662–678, 2006.