DESIGN OF TRANSFORM FILTER FOR REPRODUCING ARBITRARILY SHIFTED SOUND FIELD USING PHASE-SHIFT OF SPATIO-TEMPORAL FREQUENCY

Shoichi Koyama, Ken'ichi Furuya, Yusuke Hiwasaki, and Yoichi Haneda

NTT Cyber Space Laboratories, NTT Corporation 3-9-11, Midori-Cho, Musashino-Shi, Tokyo 180-8585, Japan

ABSTRACT

For real-time sound field transmission systems from a far-end to a near-end, the driving signals of a loudspeaker array at the near-end need to be calculated by using only received signals obtained by a microphone array at the far-end. Additionally, having the capability to control the location of the sound field to be reproduced in order to adjust it to the visual images is advantageous. The goal of this study was to develop a method to transform received signals of a microphone array into driving signals of a loudspeaker array in order to reproduce arbitrary shifted sound fields. We analytically derive a transform filter in the spatio-temporal frequency domain. The location of the sound field to be reproduced is controllable only by phase-shift of the transform filter. The proposed method was found to be computationally efficient compared to the conventional method based on a least squares algorithm, and numerical simulation results indicated that reproduction accuracies were almost the same in both methods.

Index Terms— Sound field reproduction, wave field synthesis, spatial Fourier transform, acoustical holography

1. INTRODUCTION

Sound field reproduction methods using many loudspeakers are intended to achieve more realistic audio systems. For real-time recording and reproducing systems such as telecommunication systems, it is preferable to calculate driving signals of a loudspeaker array at the near-end from only received signals of a microphone array at the far-end. We have proposed a method for transforming received signals of planar and linear microphone arrays into driving signals of planar and linear loudspeaker arrays to reproduce sound fields by applying a transform filter in the spatio-temporal frequency domain [1, 2]. By using this method, the reconstructed sound field corresponds to the received sound field at the position of the loudspeaker array. In future applications of sound field reproduction, it would be useful to be able to control the corresponding position of these sound fields, for instance, to adjust them to the visual images. The goal of this study was to derive a transform filter to shift the location of sound field to be reproduced back and forth and around.

Many sound field reproduction methods using planar or linear secondary source distributions, i.e., loudspeaker arrays,

have been proposed. However, most conventional methods require many other parameters than just the received signals of the microphone array to calculate the driving signals. One well-known conventional method is wave field synthesis (WFS) [3, 4]. The traditional formulation of WFS is based on stationary phase approximation, and requires the positions of the primary sources to be reproduced. The formulations presented in [5] and [6] require planar or linear distribution of the sound pressure gradient and the spatial frequency of the entire target sound field, respectively.

Methods based on a least squares algorithm (LS method) are applicable when only a planar or linear sound pressure distribution is known [7, 8]. In this case, the inverse transfer matrix between loudspeakers and control points is numerically calculated, and the sound pressures are controlled at the control points to correspond to those obtained by the microphone array [8]. Therefore, the location of the sound field to be reproduced can be controlled by setting the positions of control points. However, this requires recalculating the inverse transfer matrix between the loudspeakers and the control points, which has a very high computational cost.

We derive a transform filter for reproducing arbitrary shifted sound fields by combining the sound field extrapolation method and the method presented in [1]. The extrapolation of the sound field is achieved by phase-shift of the spatio-temporal frequency; therefore, the location control of the sound field to be reproduced corresponds to the phaseshift of the transform filter in the spatio-temporal frequency domain. When the location of the sound field to be reproduced is changed, the computational cost for recalculating the transform filter is substantially lower than with the LS method. A method for controlling the location of the sound field to be reproduced back and forth is also presented in [9, 10], but the method formulated in the spatio-temporal frequency domain is novel. We conducted numerical simulations to compare the proposed method with the LS method.

2. TRANSFORM METHOD FOR REPRODUCING ARBITRARY-SHIFTED SOUND FIELD

We focus on the use of linear distributions of receivers and secondary sources, as shown in Fig. 1. A similar formulation is applicable to planar distributions.



Fig. 1. Sound field reproduction using linear distributions of receivers and secondary sources.

2.1. Transform sound pressure distribution observed using receivers into driving signals of secondary sources

First, we revisit the method we previously proposed [1]. When the sound pressure in the half-space of y > 0 (the target area) is reproduced to correspond to the sound field in the half-space of y < 0 (the source area) as shown in Fig. 1, this method makes it possible to transform sound pressure distribution obtained on the receiving line into the driving signals of secondary sources on the secondary source line. The secondary sources and receivers are assumed to be continuously distributed on the secondary source and receiving lines at the *x*-axis, respectively. The transform equation is derived by setting up a simultaneous equation of the sound field created by secondary sources and the Rayleigh I integral as the desired sound field in the spatio-temporal frequency domain as follows.

The sound field created by secondary sources is given by

$$P(\boldsymbol{r},\omega) = \int_{-\infty}^{\infty} D(\boldsymbol{r}_0,\omega) G(\boldsymbol{r}-\boldsymbol{r}_0,\omega) dx_0, \qquad (1)$$

where $\mathbf{r} = (x, y, z)$ is the position vector in the target area, $\mathbf{r}_0 = (x_0, 0, 0)$ is the position vector on the x-axis, $P(\mathbf{r}, \omega)$ is the sound pressure of frequency ω at \mathbf{r} , $D(\mathbf{r}_0, \omega)$ is the driving signal of the secondary source at \mathbf{r}_0 , and $G(\mathbf{r} - \mathbf{r}_0, \omega)$ is the transfer function between \mathbf{r} and \mathbf{r}_0 . Equation (1) represents a convolution of $D(\cdot)$ and $G(\cdot)$ in the spatial domain, x. Therefore, the spatial Fourier transform of (1) with respect to x is represented as [6]

$$\tilde{P}(k_x, y, z, \omega) = \tilde{D}(k_x, \omega) \cdot \tilde{G}(k_x, y, z, \omega), \qquad (2)$$

where k_x denotes the spatial frequency. From here on, the variable in the spatial frequency domain is indicated with a tilde. Note that the phase of the kernel of the spatial Fourier transform is positive in this context [6, 1].

However, we assume the sound field in the target area is determined by the Rayleigh I integral in two dimensions on the x-y-plane at z = 0. The Rayleigh I integral in two dimensions is described as

$$P(\boldsymbol{r},\omega) = -2 \int_{-\infty}^{\infty} \frac{\partial P(\boldsymbol{r}_{0},\omega)}{\partial y} G_{2\mathrm{D}}(\boldsymbol{r}-\boldsymbol{r}_{0},\omega) dx_{0}, \quad (3)$$

where $G_{\rm 2D}(\boldsymbol{r}-\boldsymbol{r}_0,\omega)$ is the two-dimensional free-field Green



Fig. 2. Reproduction of arbitrary-shifted sound field. Reconstruction line is shifted from secondary source line in target area.

function. The abbreviated notation $\partial/\partial y$ indicates the directional gradient in the direction of y at r_0 . The spatial Fourier transform of (3) with respect to x is represented as

 $\tilde{P}(k_x, y, z, \omega) = 2jk_{\rho}\tilde{P}(k_x, 0, 0, \omega) \cdot \tilde{G}_{2D}(k_x, y, \omega),$ (4) where $k_{\rho} = \sqrt{k^2 - k_x^2}, k = \omega/c$ is the wave number, and c is the sound velocity. This equation is only valid when primary sources are on the *x*-*y*-plane at z = 0 [1].

From (2) and (4), the driving signals of secondary sources can be derived [1]:

$$\tilde{D}(k_{x},\omega) = 2jk_{\rho}\frac{G_{2D}(k_{x},y,\omega)}{\tilde{G}(k_{x},y,0,\omega)}\tilde{P}(k_{x},0,0,\omega)
= 4j\frac{\exp(-jk_{\rho}y)}{H_{0}^{(2)}(k_{\rho}y)}\tilde{P}(k_{x},0,0,\omega),$$
(5)

where $H_0^{(2)}(\cdot)$ is the 0-th order Hankel function of the second kind. For simplicity, $G(\mathbf{r} - \mathbf{r}_0, \omega)$ is assumed to have monopole characteristics. Equation (5) is a computable expression because $\exp(\cdot)$ and $H_0^{(2)}(\cdot)$ are proportional in large arguments. This equation requires only the sound pressure distribution on the x-axis, $P(\mathbf{r}_0, \omega)$, to calculate driving signals. However, (5) depends on y in the target area. Therefore, the reference line $y = y_{ref}$, which is parallel to the secondary sources on the x-y-plane at z = 0, must be set. This leads to a faster than ideal amplitude decay.

Although it is assumed that primary sources are on the xy-plane at z = 0, the driving signals of secondary sources can be transformed from the received signals on the receiving line in the spatio-temporal frequency domain by (5).

2.2. Shifting reconstruction line by using phase-shift of spatial frequency

By using the driving signals calculated by (5), the sound field obtained on the receiving line is reconstructed without the amplitude errors on the secondary source line. To shift the reproduced sound field in the target area, the sound field obtained on the receiving line is reconstructed on an arbitrary shifted line, as shown in Fig. 2. We call this line a reconstruction line. The position of the reconstruction line is denoted as $r_r = (x_0 + d_x, d_y, 0)$, where d_x and d_y are the shift distances in the directions of x and y, respectively. Therefore, a trans-



Fig. 3. Block diagram of proposed method. Transform filter is applied as convolution in spatio-temporal frequency domain.

form filter for shifting the reconstruction line to \boldsymbol{r}_r needs to be obtained.

The sound pressure on the secondary source line can be estimated from that on the reconstruction line by using the phase-shift of the spatio-temporal frequency

 $\tilde{P}(k_x, 0, 0, \omega) = \tilde{P}(k_x, d_y, 0, \omega) \exp(j(k_x d_x + k_\rho d_y))$, (6) where $\tilde{P}(k_x, 0, 0, \omega)$ and $\tilde{P}(k_x, d_y, 0, \omega)$ are the respective spatial frequencies of the sound pressure distributions at r_0 and r_r . Estimation in the direction of x is a well known property of the Fourier transform. Estimation in the direction of y can be explained by the principle of near-field acoustical holography in two dimensions [11].

Substituting (6) into (5), we derive the transform equation for shifting the reconstruction line as

$$\tilde{D}(k_x, \omega) = 4j \frac{\exp(-jk_{\rho}y_{\text{ref}})}{H_0^{(2)}(k_{\rho}y_{\text{ref}})} \exp(j(k_x d_x + k_{\rho} d_y)) \tilde{P}(k_x, d_y, 0, \omega).$$
(7)

By substituting the spatial frequency of the sound pressure on the receiving line into that on the reconstruction line, we can reproduce an arbitrary-shifted sound field in the target area.

As a general acoustic-holographic objective, such as imaging of sound pressure distribution, (6) is only valid when the distance between the receiving line and primary sources is smaller than d_y . However, in the proposed method, when d_y is larger than the distance between the receiving line and primary sources, the primary sources are virtually reconstructed in front of the secondary source line [9, 10]. The region between the virtual primary sources and the secondary source line is not ensured because the secondary sources emit a sound wave that travels towards the reconstruction line.

3. IMPLEMENTATION USING LINEAR MICROPHONE AND LOUDSPEAKER ARRAYS

In practice, linear arrays of microphones and loudspeakers are used as linear distributions of receivers and secondary sources. Therefore, spatial discretization and truncation of receiver and secondary source lines must be introduced. These approximations result in artifacts, but further investigations into such artifacts are not presented in this paper. In this section, a filter representation of the proposed method for practical implementation is discussed. A block diagram of the proposed method is depicted in Fig. 3. The sound pressure distribution is obtained using an equally distributed omni-directional microphone array. Received signals of the microphone array are denoted as $P_n(\omega)$ in the temporal frequency domain, and n denotes the index of microphone positions. An omni-directional loudspeaker array is assumed. Driving signals of the loudspeaker array are denoted as $D_n(\omega)$ in the temporal frequency domain, and ndenotes the index of loudspeaker positions. The discrete spatial Fourier transforms of $P_n(\omega)$ and $D_n(\omega)$ are denoted as $\tilde{P}_m(\omega)$ and $\tilde{D}_m(\omega)$, respectively. Here, m denotes the index of spatial frequency, $k_{x,m}$. According to (7), the transformation of the microphone array's received signals into the loudspeaker array's driving signals is derived as

 $\tilde{D}_m(\omega) = \tilde{F}_m(\omega) \cdot \tilde{P}_m(\omega), \qquad (8)$

where

$$\tilde{F}_{m}(\omega) = 4j \frac{\exp\left(-j\sqrt{k^{2}-k_{x,m}^{2}}y_{\text{ref}}\right)}{H_{0}^{(2)}\left(\sqrt{k^{2}-k_{x,m}^{2}}y_{\text{ref}}\right)} \exp\left(j\left(k_{x,m}d_{x}+\sqrt{k^{2}-k_{x,m}^{2}}d_{y}\right)\right).$$
(9)

The shift distance of the sound field to be reproduced is controllable by setting d_x and d_y in the transform filter (9).

A method based on a least squares algorithm (LS method) is also applicable to reproduce arbitrary shifted sound fields when only the linear distribution of sound pressure is known [8]. In this method, sound pressure at control points on the reconstruction line is controlled using an inverse transfer matrix. However, as we previously presented [1], the size of the filter designed based on the LS method and the amount of calculation of its filtering are much larger than those of the proposed method. Additionally, recalculation of the transform filter in the LS method is complicated when the position of the reconstruction line is changed because the inverse transfer matrix also needs to be recalculated. When the number of channels in each array is N_{ch} , the computational cost of recalculation is $O(N_{ch}^3)$ in each frequency ω . On the other hand, that of the proposed method is $O(N_{ch})$ because the recalculation is simply achieved by the phase-shift of (9).

4. EXPERIMENTS

Numerical simulations were conducted under the free-field assumption. We compared them with the LS method [8]. The coordinates of the numerical simulations follow those in Fig. 1. Identical linear microphone and loudspeaker arrays, each with 96 channels, were located at y = 0. The directivity of the array elements was assumed to be omni-directional. The elements of the arrays were equally spaced at 8 cm, so the total length of the arrays was 7.68 m. Sound pressure generated by a point source as a primary source located at (0.0 m, -1.0 m) was observed with a microphone array at y = 0 in the source area.

The simulation results in a 3×3 m region at intervals of 1.5 cm are shown in Fig. 4. The source signal was a 1-



Fig. 4. Numerical simulation results of sound pressure distribution in target area. SDRs of proposed and LS methods were 6.3 and 6.9 dB, respectively.

kHz sinusoidal wave. The shift distances $(d_x \text{ and } d_y)$ were set as 0.2 m and 1.8 m, respectively. Therefore, the virtual point source was reproduced at (0.2 m, 0.8 m) by using both methods. The amplitude was normalized at the center of the simulated area. We define the signal to distortion ratio (SDR) as that of the original sound pressure distribution to the error of the reproduced sound pressure distribution. SDR can be written as

SDR =

$$-10 \log \frac{1}{N_t} \sum_{k} \frac{\sum_{i} \sum_{j} |p(t_k, x_i, y_j) - p_{\text{org}}(t_k, x_i, y_j)|^2}{\sum_{i} \sum_{j} |p_{\text{org}}(t_k, x_i, y_j)|^2}, (10)$$

where $p(t_k, x_i, y_j)$ and $p_{org}(t_k, x_i, y_j)$ are the reproduced and original sound pressure distributions in the time domain, respectively, (x_i, y_j) denotes discrete positions in the simulated area, t_k denotes discrete times, and N_t is the averaged number of times. The sampling frequency was 48 kHz and Nt was set as $N_t = 480$. The SDRs in the region between the virtual point source and secondary source line, i.e., y < 0.8 m, and that within 0.1 m of the virtual point source were not calculated. In this case, the SDRs of the proposed and LS methods were 6.3 and 6.9 dB, respectively.

The relation between the SDR and frequency of the source signal is shown in Fig. 5. The SDRs of the proposed method were slightly lower than that of the LS method below the spatial Nyquist frequency, 2.1 kHz, but they generally remained almost the same. Above the spatial Nyquist frequency, the SDRs of the LS method were higher than those of the proposed method within a small range. While the filter size and calculation cost of the proposed method were much smaller than those of the LS method, the reproduction accuracies were almost the same.

5. CONCLUSION

A method for reproducing an arbitrary shifted sound field was proposed. We derived the transform equation from the received signals of the microphone array to the driving signals of the loudspeaker array in order to shift the reconstruction line. The transform filter is determined only by setting shift-



Fig. 5. Relation between SDR and frequency of source signal. SDRs were almost the same in both method below spatial Nyquist frequency.

ing parameters. The position of the reconstruction line is controlled by phase-shift of the transform filter in the spatiotemporal frequency domain. Numerical simulations were conducted to compare the proposed method with the LS method. The reproduction accuracies were almost the same in both methods below the spatial Nyquist frequency, but the filter size and amount of calculation of the proposed method are much smaller than those of the LS method.

6. REFERENCES

- [1] S. Koyama, K. Furuya, Y. Hiwasaki, and Y. Haneda, "Design of transform filter for sound field reproduction using microphone array and loudspeaker array," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.* (WASPAA), New Paltz, Oct. 2011, pp. 5–8.
- [2] S. Koyama, K. Furuya, Y. Hiwasaki, and Y. Haneda, "Sound field recording and reproduction using transform filter designed in spatio-temporal frequency domain," in *Proc. 131st Conv. AES*, New York, Oct. 2011.
- [3] A. J. Berkhout, D. D. Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, vol. 93, no. 5, pp. 2764–2778, 1993.
- [4] D. D. Vries, *Wave Field Synthesis*, AES Monograph. Audio Eng. Soc., 2009.
- [5] S. Spors, R. Rabenstein, and J. Ahrens, "The theory of wave field synthesis revisited," in *Proc. 124th Conv. AES*, Amsterdam, The Netherlands, Oct. 2008.
- [6] J. Ahrens and S. Spors, "Sound field reproduction using planar and linear arrays of loudspeakers," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 18, no. 8, pp. 2038–2050, 2010.
- [7] J. J. López, A. González, and L. Fuster, "Room compensation in wave field synthesis by means of multichannel inversion," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, New Paltz, Oct. 2005.
- [8] N. Kamado, H. Hokari, S. Shimada, H. Saruwatari, and K. Shikano, "Sound field reproduction by wavefront synthesis using directly aligned multi point control," in *Proc. 40th Conf. AES*, Tokyo, Oct. 2010.
- [9] S. Koyama, Y. Hiwasaki, K. Furuya, and Y. Haneda, "Inverse wave propagation in wave field synthesis," in *Proc. 40th Conf. AES*, Tokyo, 2010.
- [10] S. Koyama, Y. Hiwasaki, K. Furuya, and Y. Haneda, "Inverse wave propagation for reproducing virtual sources in front of loudspeaker array," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Barcelona, 2011, pp. 1322–1326.
- [11] E. G. Williams, Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography, Academic Press, 1999.