IMPROVEMENT USING CIRCULAR HARMONICS BEAMFORMING ON REVERBERATION PROBLEM OF WAVE FIELD RECONSTRUCTION FILTERING

Shoichi Koyama[1], *Timothy Lee*[2], *Ken'ichi Furuya*[3], *Yusuke Hiwasaki*[1], *and Yoichi Haneda*[4]

[1] NTT Media Intelligence Laboratories, NTT Corporation, Tokyo, Japan
[2] University of British Columbia, Vancouver, British Columbia, Canada
[3] Oita University, Oita, Japan.
[4] The University of Electro-Communications, Tokyo, Japan.

ABSTRACT

In real-time sound field transmission systems, the driving signals of a loudspeaker array should be obtained using only the received signals of a microphone array. For efficient transformation of signals of planar or linear arrays, we previously proposed a method of applying a transform filter in the spatio-temporal frequency domain; this filter is defined as the wave field reconstruction (WFR) filter. In linear array configurations, the major artifact of this method is an increase in reverberation time and a decrease in the direct-to-reverberant energy ratio (DRR) because reflections from above and below the microphone array can not be distinguished. We propose a method combining circular harmonics beamforming with the WFR filter as a preprocess in order to match the reproduced DRR to the original one at the time the direct sound wave is properly reproduced. Simulation results indicated that the DRR reproduced using the proposed method was much closer to that reproduced by the method without beamforming.

Index Terms— Wave field reconstruction filter, circular harmonics beamforming, sound field reproduction, wave field synthesis

1. INTRODUCTION

Sound field reproduction methods allow the calculation of driving signals of loudspeakers for reproducing a desired sound field. For real-time recording and reproducing systems, such as telecommunication systems, it is preferable that the driving signals of the loudspeakers are directly transformed from the received signals of microphones arranged in the recording room without any parameters such as source positions, directions, and original signals. We call this type of transformation sound-pressure-to-driving-signal (SP-DS) conversion. We previously proposed an SP-DS conversion method for transforming received signals of a planar or linear microphone array into driving signals of a planar or linear loudspeaker array in the spatio-temporal frequency domain by applying a transform filter, which is defined as a *wave field reconstruction (WFR) filter* [1, 2].

In practical applications, linear array configuration of microphones and loudspeakers is simple to implement; however, some artifacts appear. Evidently, sound sources at different heights cannot be properly reproduced because each microphone in the linear array is assumed to be omni-directional. When sound sources are located at approximately the same height in the recording room, listeners can localize the directions of the sources based on the precedence effect [3] even though the reflections synthesized by the loudspeaker array are different from the original ones. However, in a reverberant environment, this artifact leads to an increase in the reverberation time and a decrease in the direct-to-reverberant energy ratio (DRR), which will be discussed in detail in Sec. 2. Because a different DRR from the original produces a different perceived distance [4], it is important to match the DRR to the original DRR at the time the direct sound wave is properly reproduced. Although several methods have been proposed to improve reproduction accuracy in a reverberant reproduction room [5–9], this type of artifact originating from reverberation in the recording room has never been discussed in detail. Even when other SP-DS conversion methods, such as the least squares (LS) method [10], is applied, the same artifact will appear as long as a linear configuration of the omni-directional microphone array is assumed.

We propose to combine circular harmonics (CH) beamforming [11, 12] with the WFR filter as a preprocessing step. A cylindrical microphone array is used instead of a linear microphone array. Each circular array produces a beamformed signal in which the reflections from the ceiling and floor are reduced. The signal is then used as the input signal of the WFR filter for linear arrays; therefore, this preprocess enables the extraction of a two-dimensional sound field in the recording room. We conducted experiments to simulate the reproduced DRR in a reverberant environment using the proposed method and the method without beamforming for comparison.

2. WFR FILTER AND ITS LIMITATIONS

2.1. Derivation of WFR filter for linear arrays

We will now briefly revisit the derivation of the WFR filter for linear microphone and loudspeaker arrays [1]. As given in Fig. 1, a sound field created by primary sources is captured by

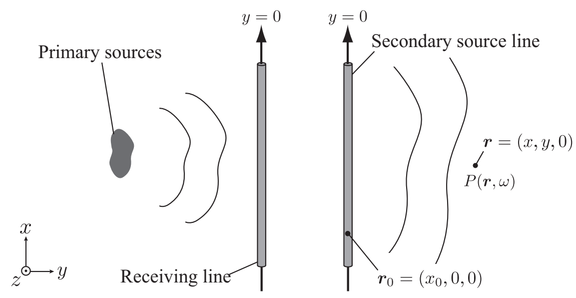

Fig. 1. Sound field reproduction using linear distributions of receivers and secondary sources

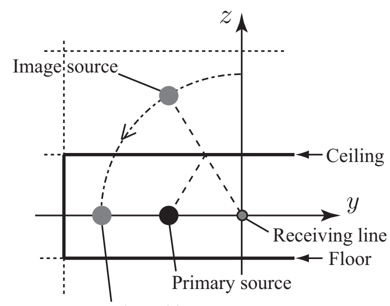

Fig. 2. All primary sources at axisymmetric positions with central axis on receiving line are equivalent to being projected onto x-y-plane at $z=0$ in source area

a linear distribution of receivers (microphones) in the source area (recording room) and is reproduced by a linear distribution of secondary sources (loudspeakers) in the target area (reproducing room). The receivers and secondary sources are assumed to be continuously distributed along the x-axis in the source and target areas, respectively. The received and driving signals in the frequency domain using the Cartesian coordinates are denoted as $P(x_0,0,0,\omega)$ and $D(x_0,0,0,\omega)$, respectively. Here, ω is the frequency. The WFR filter is designed to transform the received signals $P(x_0,0,0,\omega)$ into the driving signals $D(x_0,0,0,\omega)$ used for reproducing the sound field.

The derivation of the WFR filter for the linear receiver and secondary source distributions is based on equating and solving the synthesized sound field, $P_{\mathrm{syn}}(x,y,0,\omega)$, described as the spatial convolution of $D(x_0,0,0,\omega)$ and the transfer function between the secondary source and the position in the target area, $G(x-x_0,y,0,\omega)$ [13], and the desired sound field, $P_{\mathrm{des}}(x,y,0,\omega)$, defined by the Rayleigh I integral in two dimensions. Note that the reproduced sound field is limited on the x-y-plane at $z=0$. These two equations are described in the spatio-temporal frequency domain as follows [1]:

$$\tilde{P}_{\mathrm{des}}(k_x,y,0,\omega) = 2jk_\rho \tilde{P}(k_x,0,0,\omega)\cdot\tilde{G}_{\mathrm{2D}}(k_x,y,\omega) \quad (1)$$

$$\tilde{P}_{\mathrm{syn}}(k_x,y,0,\omega) = \tilde{D}(k_x,\omega)\cdot\tilde{G}(k_x,y,0,\omega), \quad (2)$$

where k_x denotes the spatial frequency in the direction of x, $\tilde{G}_{\mathrm{2D}}(\cdot)$ is the two-dimensional free-field Green function, $k_\rho=\sqrt{k^2-k_x^2}$, $k=\omega/c$ is the wavenumber, and c is the speed of sound. The variable in the spatio-temporal frequency domain is indicated by a tilde. When $G(\cdot)$ is assumed to be monopole for simplicity, the WFR equation is derived as

$$\tilde{D}(k_x,\omega) = \tilde{F}(k_x,\omega)\tilde{P}(k_x,0,0,\omega), \quad (3)$$

where

$$\tilde{F}(k_x,\omega) = 4j\frac{\exp\left(jk_\rho y_{\mathrm{ref}}\right)}{H_0^{(1)}\left(k_\rho y_{\mathrm{ref}}\right)}. \quad (4)$$

Here, $H_0^{(1)}(\cdot)$ is the 0-th order Hankel function of the first kind, and y_{ref} is the line parallel to the secondary sources chosen for proper reproduction of amplitude. Equations (3) and (4) are discretized, and the signal transform is achieved by applying the WFR filter (4) in the spatio-temporal frequency domain as an FIR filter.

2.2. Limitation of WFR filter for linear arrays in reverberant environment

Although the linear array configurations are simple to implement, some artifacts appear. The major artifact is derived from the two-dimensional assumption of the desired sound field, where the primary sources cannot be properly reproduced unless they are on the x-y-plane at $z=0$. Even though sound sources, such as talkers, can be assumed to be approximately at the same height in practical situations in telecommunication systems, the image sources are produced by reverberation [14] at different heights in the reverberant source area. As shown in Fig. 2, all the image sources at axisymmetric positions relative to the central axis on the receiving line are equivalent to being projected onto the x-y-plane at $z=0$ because the receivers are assumed to have an omni-directional characteristic. This artifact increases reverberation time and reduces the DRR in the target area compared with the sound field generated directly by the primary sources, i.e., the original sound field. The WFR filter can properly reproduce the direct sound wave from the primary sources; therefore, listeners can localize the direction of the primary sources based on the precedence effect [3]. However, the difference in DRR may lead to a different perceived distance to a reproduced sound source because the DRR contains the information necessary for sound source distance judgement in human auditory perception [4]. Even if the transform filter is designed using another method, for example, the LS method [10], the same artifact may appear as long as the linear microphone array in the source area is assumed to be omni-directional.

3. DRR MATCHING BY COMBINING WITH CH BEAMFORMING

Initial investigations and simulations suggest that the DRR of the reproduced sound field using the WFR filter is comparable to that of the original sound field if the reflections from the floor, ceiling, and wall in the $y>0$ direction are reduced in the source area. Therefore, it is desirable to extract the sound wave from $y<0$ on the x-y-plane at $z=0$ in reverberant environments. To achieve this, we propose combining a beamforming technique with WFR filtering. The beamform-

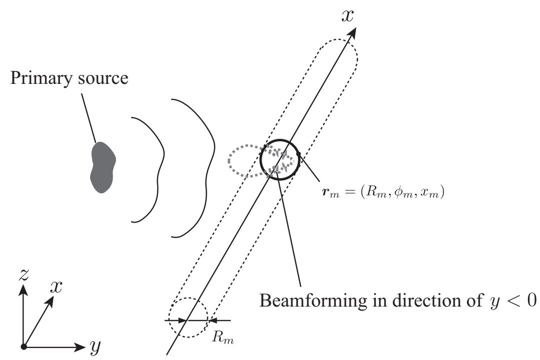

Fig. 3. Beamforming as preprocess of WFR filtering by using cylindrical microphone array

ing technique we investigated is known as CH beamforming [11, 12]. We chose this technique because it exhibits higher performance at low frequencies compared to the conventional delay-and-sum beamforming [15]. As shown in Fig. 3, CH beamforming is applied to the received signals of circular arrays at each x of the cylindrical microphone array as a preprocess of WFR filtering.

3.1. CH beamforming

We reformulate the CH beamforming for the discrete case in this context. Assume that microphones are spaced equally and mounted on a sufficiently long and rigid cylindrical baffle with a radius of R_m (Fig. 3). The position vector of each microphone is denoted as $\boldsymbol{r}_{ij} = (R_m, \phi_j, x_i)$ in cylindrical coordinates, which corresponds to $\boldsymbol{r}_{ij} = (x_i, R_m \cos\phi_j, R_m \sin\phi_j)$ in Cartesian coordinates. Here, $1 \leq i \leq M_x$ and $1 \leq j \leq M_\phi$ are the indexes of each microphone in the directions of x and ϕ, respectively.

CH beamforming is derived based on plane wave decomposition of the captured sound field in the CH domain [11]. When the beamformed signals are denoted as $P_{\mathrm{CHB}}(x_i, \omega)$,

$$P_{\mathrm{CHB}}(x_i, \omega) = \sum_{j=1}^{M_\phi} W(R_m, \phi_j, \omega) P(R_m, \phi_j, x_i, \omega) \quad (5)$$

with

$$W(R_m, \phi_j, \omega) = \sum_{n=-N}^{N} (-j)^{-n} \left(J_n(kR_m) - \frac{J'_n(kR_m)}{H_n^{(1)\prime}(kR_m)} H_n^{(1)}(kR_m) \right)^{-1} \exp\left(-jn(\phi_j - \phi_0)\right), \quad (6)$$

where ϕ_0 is the desired direction of beamforming and is set as π, N denotes the maximum order of circular harmonics, $J_n(\cdot)$ is the Bessel function of order n, $H_n^{(1)}(\cdot)$ is the Hankel function of the first kind of order n, and $J'_n(\cdot)$ and $H_n^{(1)\prime}(\cdot)$ represent the derivatives of $J_n(\cdot)$ and $H_n^{(1)}(\cdot)$. To minimize sampling and truncation errors, N for each ω should follow $N \leq \lceil kR_m \rceil$ and $N < \lceil M_\phi/2 \rceil$, where $\lceil \cdot \rceil$ represents the ceiling function. The issue of noise amplification due to use of high-order coefficients must be properly dealt with [16].

The beamformed signal $P_{\mathrm{CHB}}(x_i, \omega)$ is used as the input of the WFR filter for linear arrays (3). While each output of the CH beamforming enhances the sound wave from $\phi_0 = \pi$, it has an omni-directional characteristic in the direction of x; therefore, direct sound waves from primary sources are properly reproduced.

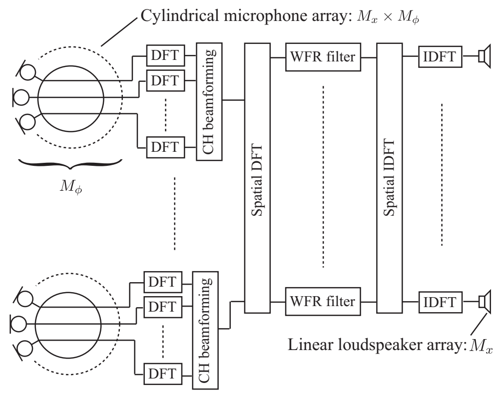

Fig. 4. Block diagram of system using proposed method

3.2. Practical implementation of the proposed method

Fig. 4 shows a block diagram of a system using the proposed method. The omni-directional microphones are aligned on the rigid cylindrical baffle. The sound wave from the direction of $y < 0$ on the x-y-plane at $z = 0$ is enhanced by using CH beamforming. The beamformed signals are transformed into the spatio-temporal frequency domain, where the WFR filter is applied in order to obtain driving signals of the linear loudspeaker array.

4. EXPERIMENT

4.1. Experimental setup

Numerical simulations were conducted to evaluate the performance of the proposed method. For comparison, delay-and-sum beamforming was also explored and tested as an alternative to CH beamforming. Both the source and target areas were assumed to be $(3.84, 7.0, 3.0)$ m with a reflection coefficient of 0.83. The number of microphones in the cylindrical array, $M_x \times M_\phi$, was 64×3. The microphones were spaced 6.0 cm apart along the length of the cylinder, and the radius of the cylinder was 1.8 cm. The center of the cylindrical array was located at $(0.0, 0.0, -0.5)$ m, and its center axis was located at $(y, z) = (0.0, -0.5)$ m. A linear microphone array without beamforming, $M_x = 64$, was also evaluated, and its center axis was located at $(y, z) = (0.0, -0.5)$ m. The linear loudspeaker array consisted of 64 loudspeakers spaced 6.0 cm apart in the target area. Similarly, its center axis was located at $(y, z) = (0.0, -0.5)$ m. The directivity of both the microphones and loudspeakers was assumed to be omni-directional. The primary source was assumed to be a point source located at $(0.0, -1.0, -0.5)$ m. The simulated results were sampled within a 3.0×3.0 m area centered at $(0.0, 2.0, -0.5)$ m with intervals of 0.3 m. The sampling frequency was 48 kHz, and the frequency band of interest was chosen to be between 20 Hz to 3 kHz to eliminate the effects of spatial aliasing. The image source method [14] was used to simulate a reverberant environment.

For simplicity, the beamformed signals were generated using the beam pattern of the beamformer. This is because generating signals scattered on the baffled cylinder is very time-consuming due to the large number of elements and the extensive equations that need to be computed. We verified that this alternative method provided a low-error approximation for all the investigated array sizes. An infinite-length baffled cylindrical microphone array was used to derive the beam pattern for the CH beamforming simulation, and an unbaffled cylindrical microphone array was used for delay-and-sum beamforming.

4.2. Results and evaluation

The reproduced sound field was evaluated using methods and procedures intended to quantify the performance of the proposed method. The results were compared with the original sound field and the sound field reproduced without beamforming using a linear microphone array. The DRR is calculated as [17]:

$$\mathrm{DRR} = 10\log_{10}\left(\frac{\sum_{n=0}^{n_d} h^2(t_n, x_i, y_j)}{\sum_{n=n_d+1}^{\infty} h^2(t_n, x_i, y_j)}\right), \tag{7}$$

where $h(t_n, x_i, y_j)$ is the reproduced impulse response from the primary source to the position in the simulated area, (x_i, y_j), and samples from t_0 to t_{n_d} are assumed to represent only the direct-path propagation. We set t_{n_d} as 16 ms. The value of the DRR changes according to the distance between the measured point and the source; the DRR decreases as the distance increases. Therefore, the DRR plots are generated by taking the average of the DRR values for each y in the simulated area, which is shown in Fig. 5. The plot shows that the methods using beamforming techniques improved the perceived distance to the reproduced sound source compared with the method without beamforming. Additionally, the proposed method using CH beamforming was much closer to the original sound field compared to the method by using delay-and-sum beamforming.

Table 1 lists the average reverberation times, T_{60} [18], in the simulated area. Although the reverberation time was improved for both beamforming techniques, the proposed method showed better performance. However, the difference between the original sound field and the sound field reproduced using the proposed method is still apparent. This is because the late reverberations are not sufficiently reduced, whereas the early reflections are reduced by the CH beamforming. Therefore, the spatial impression of the sound field reproduced using the proposed method may be different from that of the original one.

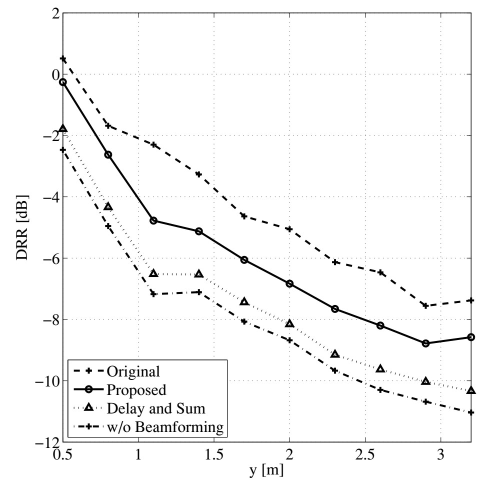

Fig. 5. Comparison of DRR averaged at each y

Table 1. Comparison of averaged reverberation time

	Original	Proposed	Delay and Sum	Without Beamforming
T_{60}	372 ms	473 ms	492 ms	508 ms

5. CONCLUSION

A method for improving the performance of the WFR filter in a reverberant environment was proposed. We proposed combining WFR filtering and CH beamforming in order to match the DRR when direct sound wave is properly reproduced. Each circular array in the cylindrical microphone array enhances the sound wave from the desired direction as a preprocess of WFR filtering. The simulation results indicated that the DRR reproduced using the proposed method was much closer to that of the original sound field compared to the case with the WFR filter without beamforming.

6. RELATION TO PRIOR WORK

The work presented here focused on the SP-DS conversion method for reducing the artifacts originating from reverberation in the recording room. This method is an extension of the our previously proposed method [1, 2]. Several methods have been proposed in order to improve reproduction accuracy in a reverberant reproduction room [5–9]. Even when these methods are applied, however, the artifacts originating from reverberation in the recording room are still unavoidable as long as a linear configuration of omni-directional microphones is assumed. The proposed method is based on a combination of WFR filtering [1, 2] and CH beamforming [11, 12] in order to match the reproduced DRRs with the original ones.

7. 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, "Analytical approach to wave field reconstruction filtering in spatio-temporal frequency domain," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 4, pp. 685–696, 2013 (in press).

[3] Jens Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, The MIT Press, 1996.

[4] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, "Auditory distance perception in humans: A summary of past and present research," *Acta Acustica United with Acustica*, vol. 91, pp. 409–420, 2005.

[5] P. A. Gauthier and A. Berry, "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, 2005.

[6] 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.

[7] T. Betlehem and T. D. Abhayapala, "Theory and design of sound field reproduction in reberverant rooms," *J. Acoust. Soc. Am.*, vol. 117, no. 4, pp. 2100–2111, 2005.

[8] E. Corteel, "Equalization in an extended area using multichannel inversion and wave field synthesis," *J. Audio Eng. Soc.*, vol. 54, pp. 1140–1161, 2006.

[9] S. Spors, H. Buchner, and R. Rabenstein, "Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering," *J. Acoust. Soc. Am.*, vol. 54, no. 12, pp. 354–369, 2007.

[10] 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.

[11] E. Tiana-Roig, F. Jacobsen, and E. F. Grande, "Beamforming with a circular microphone array for localization of environmental noise sources," *J. Acoust. Soc. Am.*, vol. 128, no. 6, pp. 3535–3542, 2010.

[12] H. Teutsch, *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition*, Springer, 2007.

[13] J. Ahrens and S. Spors, "Sound field reproduction using planar and linear arrays of loudspeakers," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 8, pp. 2038–2050, 2010.

[14] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.

[15] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*, P T R Prentice-Hall, 1993.

[16] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 135–143, 2005.

[17] P. A. Naylor and N. D. Gaubitch, *Speech Dereverberation*, Springer, 2010.

[18] E. A. Lehmann and A. M. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model," *J. Acoust. Soc. Am.*, vol. 124, no. 1, pp. 269–277, 2008.