IEICE TRANS. FUNDAMENTALS, VOL.E97–A, NO.9 SEPTEMBER 2014

PAPER Special Section on Spatial Acoustic Signal Processing and Applications

Real-Time Sound Field Transmission System by Using Wave Field Reconstruction Filter and Its Evaluation*

Shoichi KOYAMA[†a)], ***Member***, **Ken'ichi FURUYA**[††], ***Senior Member***, **Hisashi UEMATSU**[†], ***Nonmember***, **Yusuke HIWASAKI**[†], ***Member***, ***and*** **Yoichi HANEDA**[†††], ***Senior Member***

SUMMARY A new real-time sound field transmission system is presented. To construct this system, a large listening area needs to be reproduced at not less than a constant height. Additionally, the driving signals of the loudspeakers should be obtained only from received signals of microphones. Wave field reconstruction (WFR) filtering for linear arrays of microphones and loudspeakers is considered to be suitable for this kind of system. An experimental system was developed to show the feasibility of real-time sound field transmission using the WFR filter. Experiments to measure the reproduced sound field and a subjective listening test of sound localization were conducted to evaluate the proposed system. Although the reproduced sound field included several artifacts such as spatial aliasing and faster amplitude decay, the experimental results indicated that the proposed system was able to provide sound localization accuracy for virtual sound sources comparable to that for real sound sources in a large listening area.

key words: *sound field reproduction, sound field transmission, wave field reconstruction filter, wave field synthesis, spatio-temporal frequency*

1. Introduction

Large scale audio systems are becoming more feasible because of the recent development of acoustic sensors and transducers. Physical reproduction of a sound field is a representative example of such systems and is intended to construct more realistic audio playback. An important application of sound field reproduction is in telecommunication systems as follows: multiple listeners and talkers are respectively located in near-end and far-end rooms, and the sound field in the far-end room is reconstructed in the near-end room so that these two rooms are acoustically concatenated for the listeners. It is assumed that these listeners can perceive the directions and distances of the talkers by using this system. This system may also be applied to live broadcasting. To construct this type of real-time recording and reproduction system (sound field transmission system), the following requirements of sound field reproduction should be fulfilled: 1) a large listening area is achieved, 2) the required number of microphones and loudspeakers is as small as possible, and 3) driving signals of loudspeakers are obtained only from received signals of microphones. The first requirement is necessary because the reproduced region should cover the potential locations of multiple listeners as widely as possible. The second requirement is important for implementation simplicity. Three-dimensional sound field recording and reproduction requires a huge number of microphones and loudspeakers; however, when sound sources are located approximately in a horizontal plane, the sound field in the horizontal ear-plane is the most important from a perception viewpoint [2]. Therefore, only a sound field reproduced at a constant height by using a two-dimensional array of microphones and loudspeakers may be acceptable when the listeners are almost at the same height. The third requirement is preferable because parameters used to obtain the driving signals of loudspeakers, for example, source positions, directions, and original signals, are generally unknown and difficult to obtain. Therefore, direct transformation from received signals into driving signals is necessary. This type of signal transformation is defined as sound-pressure-to-driving-signal (SP-DS) conversion.

Wave field synthesis (WFS) [3]–[6] is a method based on the Kirchhoff-Helmholtz or Rayleigh integrals. WFS and related analytical methods for linear loudspeaker arrays [7], [8] can provide a relatively large listening area at a constant height, and they satisfy the first and second requirements. However, these methods require parameters other than received signals of the microphone array, as previously discussed [9]. Even though several systems using WFS have been developed [6], [10], [11], real-time sound field transmission has not been achieved because the third requirement cannot be directly satisfied by using WFS.

The Ambisonics [12] and higher order Ambisonics (HOA) [13]–[16] methods use spherical or circular loudspeaker array based on spherical harmonic expansion of the sound field. They can be applied as SP-DS conversion through encoding and decoding processes. However, the high reproduction accuracy region is limited to the region neighboring the center of the array. To widen this region, the radius of and the number of elements in the microphone array need to be very large [17], [18]. Therefore, these methods are not suitable because they do not fulfill the first requirement.

Methods based on numerical algorithms for sound pressure control can be applied to SP-DS conversion [19]–

Manuscript received November 22, 2013.
Manuscript revised March 12, 2014.

[†]The authors are with NTT Media Intelligence Laboratories, NTT Corporation, Musashino-shi, 180-8585 Japan.
[††]The author is with the Department of Computer Science and Intelligent Systems, Faculty of Engineering, Oita University, Oita-shi, 870-1192 Japan.
[†††]The author is with the Graduate School of Informatics and Engineering, The University of Electro-Communications, Chofu-shi, 182-8585 Japan.

*Part of this work was presented at AES 52nd Conference [1].

a) E-mail: koyama.shoichi@lab.ntt.co.jp

DOI: 10.1587/transfun.E97.A.1840

[21]. Because these methods do not depend on array configurations, it may be possible to achieve a large listening area at a constant height as in WFS by using linear arrays of microphones and loudspeakers. In these methods, sound pressures at discrete points are controlled to correspond to the desired sound pressures by using the inverse of a known transfer function matrix. However, these methods are extensions of the multi-point control technique, and it is therefore difficult to design and apply the transform filters for real-time sound field transmission systems [9].

An SP-DS conversion method that applies a transform filter called a *wave field reconstruction (WFR) filter* [9] was previously proposed by the authors. This filter is used 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. The WFR filtering method is represented as a form of spatio-temporal convolution, so it has many advantages in terms of filter design, filter size, computational cost, and filter stability. Therefore, the WFR filtering method for linear arrays is considered to be suitable for real-time sound field transmission systems. An experimental system was developed to show the feasibility of real-time sound field transmission. The sound field received by the linear microphone array at the far end is transmitted over an IP (Internet Protocol) network and reconstructed using the linear loudspeaker array and the WFR filter in real time at the near end. Experiments to measure the reproduced sound field and a subjective listening test of sound localization to evaluate the quality of the system were conducted.

This paper is organized as follows. In Sect. 2, the WFR filtering method for linear arrays is revisited. The implementation of the system is presented in Sect. 3. Section 4 reports on measurement experiments and subjective listening tests. Finally, Sect. 5 concludes this paper.

2. WFR Filter for Linear Arrays of Microphones and Loudspeakers

Here, we briefly revisit the derivation of a WFR filter for linear microphone and loudspeaker arrays [9]. As shown in Fig. 1, a sound field created by primary sources is captured by a linear distribution of receivers (microphones) in the source area and is reconstructed by a linear distribution of secondary sources (loudspeakers) in the target area. The receivers and secondary sources are assumed to be continuously distributed along the x-axis in the source and target areas, respectively. Position vectors on the receiving line and secondary source are respectively denoted as $\boldsymbol{r}_{\mathrm{m}} = (x_{\mathrm{m}}, 0, 0)$ and $\boldsymbol{r}_{\mathrm{s}} = (x_{\mathrm{s}}, 0, 0)$, and a sound pressure distribution on the receiving line and driving signals of the secondary sources in the temporal frequency ω are respectively denoted as $P_{\mathrm{rcv}}(\boldsymbol{r}_{\mathrm{m}}, \omega)$ and $D(\boldsymbol{r}_{\mathrm{s}}, \omega)$. SP-DS conversion requires that $D(\cdot)$ be obtained only from $P_{\mathrm{rcv}}(\cdot)$. The WFR filter is derived by equating and solving the synthesized sound field, $P_{\mathrm{syn}}(\boldsymbol{r}, \omega)$, and the desired sound field, $P_{\mathrm{des}}(\boldsymbol{r}, \omega)$, in the spatio-temporal frequency domain. Here, $\boldsymbol{r} = (x, y, z)$ is the position vector in the target area.

The synthesized sound field can be described as:

$$P_{\mathrm{syn}}(\boldsymbol{r}, \omega) = \int_{-\infty}^{\infty} D(\boldsymbol{r}_{\mathrm{s}}, \omega) G(\boldsymbol{r} - \boldsymbol{r}_{\mathrm{s}}, \omega) dx_{\mathrm{s}}, \tag{1}$$

where $G(\boldsymbol{r} - \boldsymbol{r}_{\mathrm{s}}, \omega)$ denotes the transfer function between $\boldsymbol{r}$ and $\boldsymbol{r}_{\mathrm{s}}$. Based on the convolution theorem, the spatial Fourier transform of Eq. (1) with respect to x is represented as [8]:

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

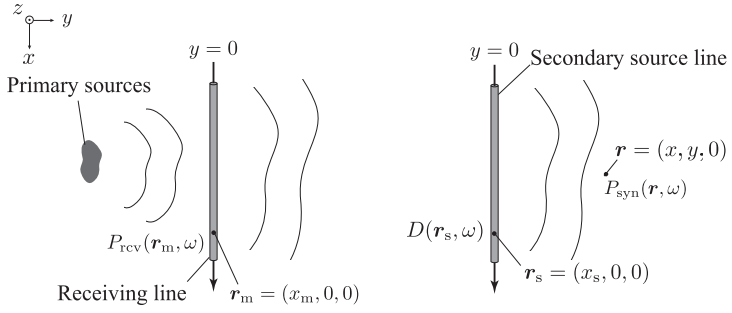
where k_x denotes the spatial frequency. The variables in the spatial frequency domain are hereafter indicated by tildes. Note that the reproduced sound field is limited on the x-y-plane on $z = 0$. In this context, the spatial Fourier transform is defined as:

$$\tilde{P}_{\mathrm{syn}}(k_x, y, 0, \omega) = \int_{-\infty}^{\infty} P_{\mathrm{syn}}(x, y, 0, \omega) e^{jk_x x} dx. \tag{3}$$

On the contrary, it is assumed that the desired sound field is described as the Rayleigh integral of the first kind in two dimensions (2D) as [22]:

$$P_{\mathrm{des}}(\boldsymbol{r}, \omega) = -2 \int_{-\infty}^{\infty} \frac{\partial P_{\mathrm{des}}(\boldsymbol{r}_{\mathrm{s}}, \omega)}{\partial y_{\mathrm{s}}} G_{\mathrm{2D}}(\boldsymbol{r} - \boldsymbol{r}_{\mathrm{s}}, \omega) dx_{\mathrm{s}}. \tag{4}$$

Here, $G_{\mathrm{2D}}(\cdot)$ is the free field Green's function in 2D defined



(a) Source area ($y < 0$) (b) Target area ($y > 0$)

Fig. 1 Sound field in source area is reproduced in target area.

as:

$$G_{2D}(\boldsymbol{r}-\boldsymbol{r}_s,\omega) = -\frac{j}{4}H_0^{(2)}\left(k|\boldsymbol{r}-\boldsymbol{r}_s|\right), \tag{5}$$

where $H_0^{(2)}(\cdot)$ is the 0-th order Hankel function of the second kind. The Fourier transform of Eq. (4) with respect to x can be derived as in [9]:

$$\tilde{P}_{des}(k_x,y,0,\omega) = 2jk_\rho \tilde{P}_{des}(k_x,0,0,\omega)\tilde{G}_{2D}(k_x,y,\omega), \tag{6}$$

where $k_\rho = \sqrt{k^2-k_x^2}$. It is assumed that the desired sound field is invariant with regard to changes in the z-axis. The spatial frequency spectrum on the x-axis of the desired sound field, $\tilde{P}_{des}(k_x,0,0,\omega)$, can be given as that on the receiving line, $\tilde{P}_{rcv}(k_x,0,0,\omega)$.

When the equation of the synthesized and desired sound fields, Eqs. (2) and (6), are simultaneously solved, i.e., $\tilde{P}_{syn}(\cdot) = \tilde{P}_{des}(\cdot)$, the equation that relates $\tilde{P}_{rcv}(\cdot)$ with $\tilde{D}(\cdot)$ is derived as:

$$\begin{aligned}\tilde{D}(k_x,\omega) &= 2jk_\rho \frac{\tilde{G}_{2D}(k_x,y,\omega)}{\tilde{G}(k_x,y,0,\omega)}\tilde{P}_{rcv}(k_x,0,0,\omega)\\ &= 4j\frac{e^{-jk_\rho y}}{H_0^{(2)}(k_\rho y)}\tilde{P}_{rcv}(k_x,0,0,\omega).\end{aligned} \tag{7}$$

For simplicity, it is assumed that each secondary source can be approximated as monopole; therefore, the analytical forms of $\tilde{G}(\cdot)$ and $\tilde{G}_{2D}(\cdot)$ can be used to obtain Eq. (7). Because Eq. (7) depends on y, Eqs. (1) and (4) can be equivalent only on a line parallel to the x-axis. Therefore, the reference line $y = y_{ref}$ must be set, which leads to a faster amplitude decay than desired.

By introducing spatial phase-shift to Eq. (7), it is possible to reproduce an arbitrarily shifted sound field [23]. As shown in Fig. 2, where d_x and d_y respectively denote the shift distances in the x and y directions, $\tilde{D}(\cdot)$ can be obtained by a spatial phase-shift of Eq. (7) as:

$$\tilde{D}(k_x,\omega) = 4j\frac{e^{-jk_\rho y_{ref}}}{H_0^{(2)}(k_\rho y_{ref})}e^{j(k_x d_x + k_\rho d_y)}\tilde{P}_{rcv}(k_x,0,0,\omega). \tag{8}$$

It should be noted that the extrapolated region is assumed to be free-field. When d_y is larger than the distance between the receiving line and the primary sources, the primary sources are virtually reconstructed in front of the secondary source line [24]. However, the sound field in the region between the virtual primary sources and the secondary source line is inverted.

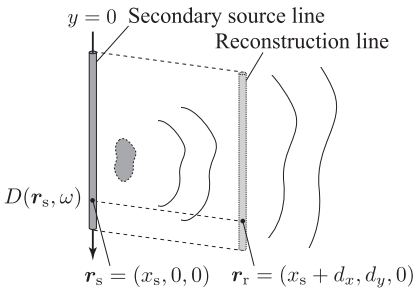


Fig. 2 Reproduction of arbitrary shifted sound field.

3. Implementation of Real-Time Sound Field Transmission System

The WFR filtering method was used to develop an experimental system for real-time sound field transmission. A block diagram of the system is depicted in Fig. 3.

Linear microphone and loudspeaker arrays were set in recording and reproduction rooms, respectively. The microphone array was made up of 64 omni-directional microphones, and each microphone was equally spaced 6 cm apart. Primo EM172 microphones were used. The loudspeaker array was composed of two kinds of loudspeakers. One was for a lower frequency band and there were 32 of these loudspeakers equally spaced 120 mm apart. The other was for a higher frequency band, and there were 64 loudspeakers equally spaced 60 mm apart. Therefore, the spatial Nyquist frequency was about 2.8 kHz [9]. Foster 411222 loudspeakers were used for the higher frequency band, and Yamaha NS-B210 loudspeakers were used for the lower frequency band. The intensity of loudspeakers for the higher frequency band in the target area was approximately omni-directional, with a difference of 1.0 dB below 1.3 kHz. The microphone and loudspeaker arrays were 3.84 m long. A MOTU 24 I/O interface was used for A/D and D/A converters.

The signals obtained by the microphone array in the recording room were transmitted to the reproduction room using UDP (User Datagram Protocol) over an exclusive IP network called GEMnet2 [25]. The signals were encoded as 16-bit PCM (pulse code modulation) data, and the sampling frequency was 48 kHz. Therefore, the transmission rate was about 50 Mbps.

The signals of the microphone array were transformed into driving signals of the loudspeakers by using the WFR filtering method. The WFR filter was designed as a 2D FIR filter by discretizing Eq. (8). Therefore, conversion was achieved as a 2D convolution of the WFR filter and the received signals. The length of the WFR filter was 1024 taps in the time domain and 64 taps in the space domain. Therefore, a 2D FFT of 2048×128 was applied for 2D linear convolution. The Tukey window function was applied as a tapering window whose sides tapered by 10% to reduce artifacts of finite array length approximation. The converted signals were divided into two frequency bands, and the cut-off frequency was 0.6 kHz. These signals were used for the driving signals of the two loudspeaker arrays. The signals of the 32 channels were created by thinning the signals of the 64 channels.

The delay times of the WFR filtering and data buffering were about 21 and 131 ms, respectively. Therefore, the total delay time was about 152 ms. This one-way delay time may be acceptable for telecommunication as well as live broad-

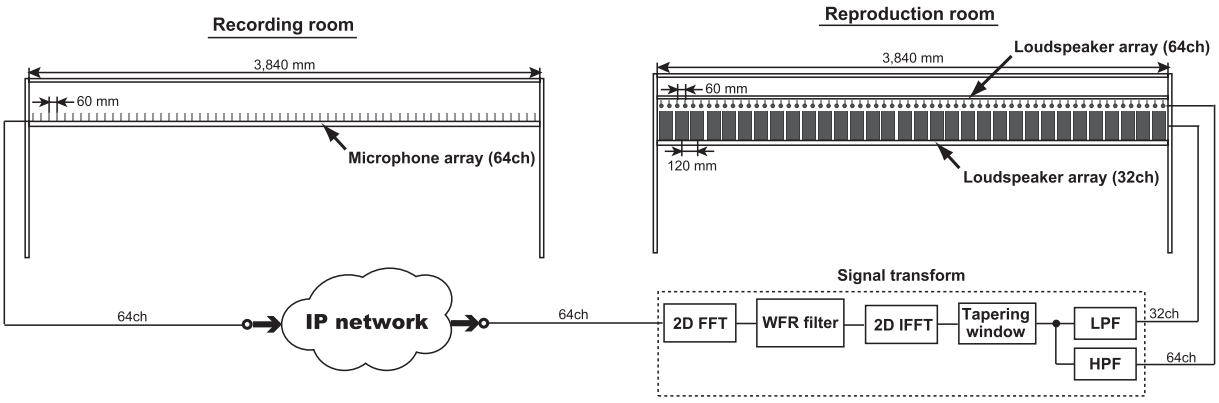


Fig. 3 Block diagram of real-time sound field transmission system.

casting [26]; however, it may be too large for applications such as network music performance [27]. The rate of packet loss was 0.0% when the received data were observed for 600 sec.

Using this proposed system makes it possible to transmit a sound field of about 4 m in real time. The system was also tested between two locations as far apart as about 55 km. In this case, the delay time of the IP network was about 1 ms. It was possible to transmit a sound field to any place that the network was connected.

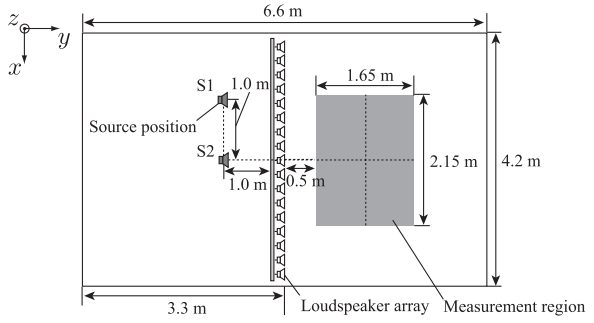


Fig. 4 Setup for measurement experiment.

4. Experiments

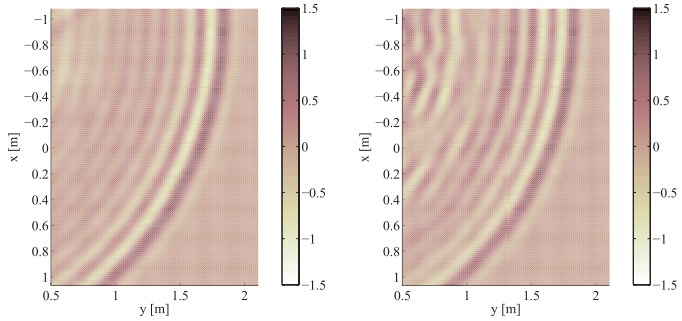
Measurement experiments and a subjective listening test were conducted to evaluate the proposed system described in Sect. 3. A sound field reproduced using this system may have several artifacts. Spatial aliasing artifacts limit the maximum frequency of a properly reproduced sound field. The faster amplitude decay may affect the perceived distance of reproduced sound sources [28]. It has been shown that linear array approximation also leads to an increase in reverberation time and a decrease in the direct-to-reverberant energy ratio (DRR) [29]. Unnecessary reflections may be produced by the framed structure of the array. However, if direct sound waves are properly reproduced in a sufficiently broad frequency band, it can be expected that the listeners in the reproduction room will be able to localize the primary sources in the recording room despite these artifacts. First, reproduction of direct sound waves was validated by visualizing the reproduced sound field. Second, sound localization listening tests were conducted to investigate the perceptual acceptability of the reproduced sound field.

4.1 Measurement of Reproduced Sound Field

Measurement experiments were conducted to compare the original and reproduced sound fields. The setup of the experiments is shown in Fig. 4. In these experiments, only the loudspeaker array with 64 channels was used, and the IP network was not used. The origin of the Cartesian coordinates was set at the center of the room. Ordinary enclosed loudspeakers were used as primary sources and were located at S1 and S2. Reverberation time (T_{60}) of the experimental room was about 167 ms.

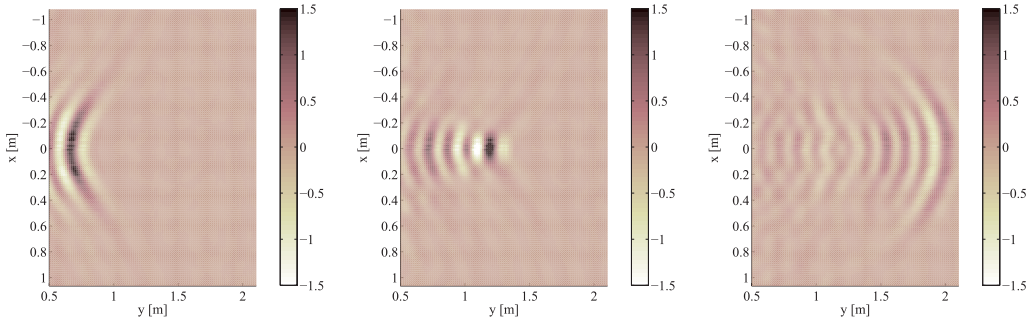
Impulse responses from each loudspeaker to grid points in a planar measurement region were measured. These impulse responses were used to obtain sound pressure distributions of the original and reproduced sound fields by post-processing. The impulse responses were measured by using time-stretched pulse (TSP) signals [30] in 2.2×1.6-m regions (shaded region in Fig. 4) at intervals of 1.5 cm, i.e., 144×108 points. The signals of the linear microphone array were obtained in the same room. The microphone array was set at the same position as the loudspeaker array, and the impulse responses from the loudspeakers at S1 and S2 were measured.

The instantaneous sound pressure distributions of the original and reproduced sound fields in the measurement region are shown in Fig. 5. The amplitudes and phases of the original and reproduced signals were normalized at the center of the measurement region, (0.0, 1.3) m. The primary source was located at S1, and the source signal was a pulse signal that was band-limited below 2.6 kHz to avoid spatial aliasing artifacts. In Fig. 5a, it can be seen that the direct sound wave was propagating from S1 in the measurement region. This direct sound wave was properly reproduced by using the proposed system (Fig. 5b). The disturbed waves that occur after the reproduced direct sound wave are



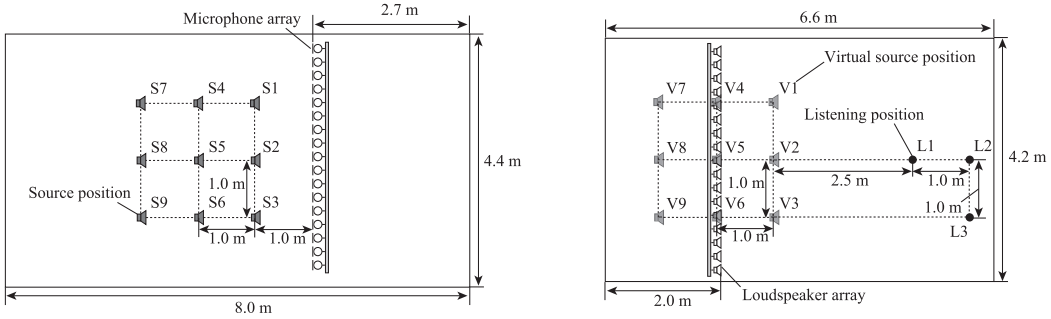
(a) Original (b) Reproduced

Fig. 5 Measured sound pressure distributions of original and reproduced sound fields.



(a) t = 0.0 ms (b) t = 1.5 ms (c) t = 2.3 ms

Fig. 6 Reproduced sound pressure distribution of shifted sound field.



(a) Recording room (b) Reproduction room

Fig. 7 Subjective listening test setup.

thought to be reflections from the framed structure of the array. When the source signal includes frequency spectrum components above 2.8 kHz, the reproduction accuracy of the direct sound wave cannot be maintained. Because the array configurations are linear, sound waves from vertical directions could not be reproduced.

Figure 6 shows the reproduced sound pressure distribution when the source position was S2 and the shift parameter of the WFR filter, (d_x, d_y) in Eq. (8), was set as $(0.0, 2.0)$ m. Three instantaneous sound pressure distributions are shown, and t indicates the time when Fig. 5a is set as $t = 0.0$ ms. The reproduced sound wave was focused at $(0.0, 1.0)$ m; therefore, a virtual sound source was created at this position as expected. When the listeners are in front of the virtual sound source, i.e., $y > 1.0$ m, it is expected that the listeners can localize the source at this position. It should be noted that finite array length approximation also makes the reproduced region smaller [24].

4.2 Subjective Listening Test

A sound localization listening test was conducted in order to perceptually evaluate the proposed system. The subjective listening test setup is shown in Fig. 7. The system described in Sect. 3 including an IP network was used for the experiments. The loudspeaker array was set about 1.2 m above the floor. The ordinary loudspeakers used as primary sources were set in nine positions, S1 – S9, in the recording room.

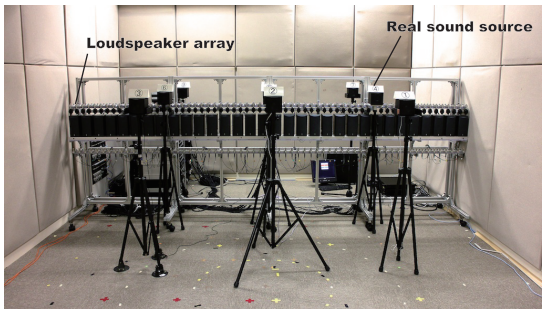


Fig. 8 Photo of reproduction room for subjective listening test.

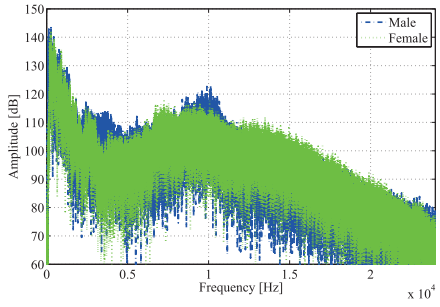


Fig. 9 Examples of frequency characteristics of speech signals.

The shift parameter of the WFR filter, (d_x, d_y) in Eq. (8), was set as (0.0, 2.0) m. Therefore, reproduced sound sources were located at V1 – V9 in the reproduction room. As references, real loudspeakers of the same type used in the recording room were also set at V1 – V9. A listener was positioned at L1 – L3. Because the real loudspeakers were higher than the linear loudspeaker array, and the diameter of the stands they were on was 3 cm (Fig. 8), it can be assumed that the real loudspeakers do not affect the sound field reproduced for the listener, especially at low frequencies. There were six listeners between the ages of 25 and 40. The listener was able to move his or her head while sitting on the chair and to see the position signs at V1 – V9 (Fig. 8). The height of the listeners' ear-plane was 1.1 – 1.3 m. The recording and reproduction rooms were 4.4×8.0×3.0 m and 4.2×6.6×3.1 m, respectively. The reverberation times (T_{60}) of the respective rooms were about 150 and 167 ms. The background noise level of the reproduction room was about 27.9 dBA.

The speech signals of female and male utterances were used as source signals. The source signals were recorded and played at a 48-kHz sampling rate. The duration of each source signal was 4 sec, and the interval between them was 2 sec. Two examples of frequency characteristics of the speech signals are shown in Fig. 9. The signals of real and virtual sound sources were played randomly 144 times at each listening position; therefore, each source signal was played 8 times. The order of the listening positions was L1, L2, and L3, and the listeners had a break when the listening position was changed. The listeners were asked to report the perceived positions of the sound from V1 – V9. The listeners did not receive any feedback about their performance.

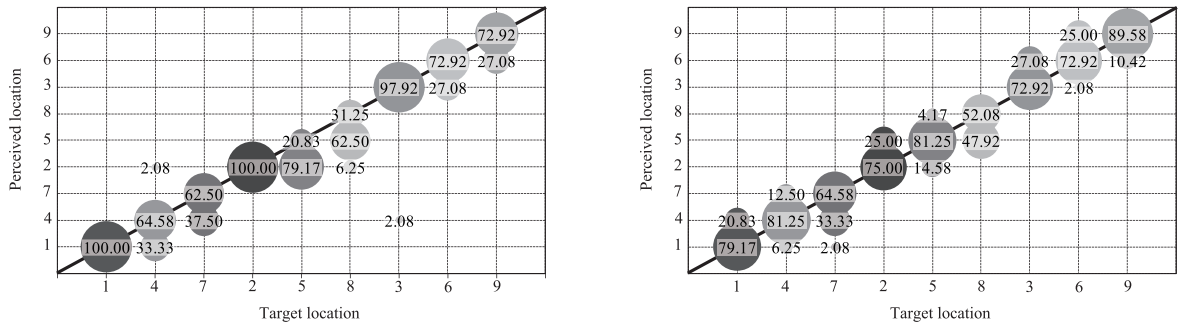
Figures 10, 11, and 12 show the average rate of listeners noting each sound source location. The horizontal and vertical axes are target and perceived locations, respectively. Therefore, the accuracy rate is shown on the diagonal line. The sound source locations on the axes are sorted vertically. The results of L1 (Fig. 10) show that the accuracies of real and virtual source localization were generally the same. The accuracy rate of the virtual sound sources at V1 – V3 was relatively low compared with that of the real sound sources. It seems that real sound sources at short distances were more easily localized than virtual ones. Because it is difficult to distinguish the source locations in line with the front-facing position of the listener, i.e., V2, V5, and V8 [2], [28], the accuracy rate of V5 and V8 for the real sound sources was distinctly low, and that for the virtual sound sources was higher. It can be considered that a clue to distinguish them was provided by the faster amplitude decay due to the linear array approximation that leads to larger amplitude differences in distances. The results of L2 (Fig. 11) showed a similar tendency to those of L1.

The results of L3 (Fig. 12) also showed a similar tendency to those of L1. The accuracy rate of real sound sources at V6 and V9 was distinctly low because these sources were in line with the front-facing position of the listener. Because L3 was almost at the boundary of the reproduction area of V1 for the virtual sound source [24], its accuracy rate was lower than that of the real one.

These results indicate that the proposed system provides sound localization accuracy for virtual sound sources that is comparable to that for real sound sources in a large listening area despite the presence of several artifacts. Although the reproduced frequency band was limited below 2.8 kHz, it seems that the spatial aliasing artifacts do not have a significant effect on the sound localization accuracy; however, this effect may depend on the source signals. The frequency characteristics of the source signal were also affected by the spatial aliasing artifacts, but this effect was not evaluated in these results. Interestingly, the perceived distances of the reproduced sound sources were more accurate than those of the real sound sources. Although the faster amplitude decay interferes with the perfect reconstruction of the sound field, the difference in the source distances was enhanced. Consequently, further investigation of perceptual effects of the reproduced sound field using more advanced subjective evaluation methods is required.

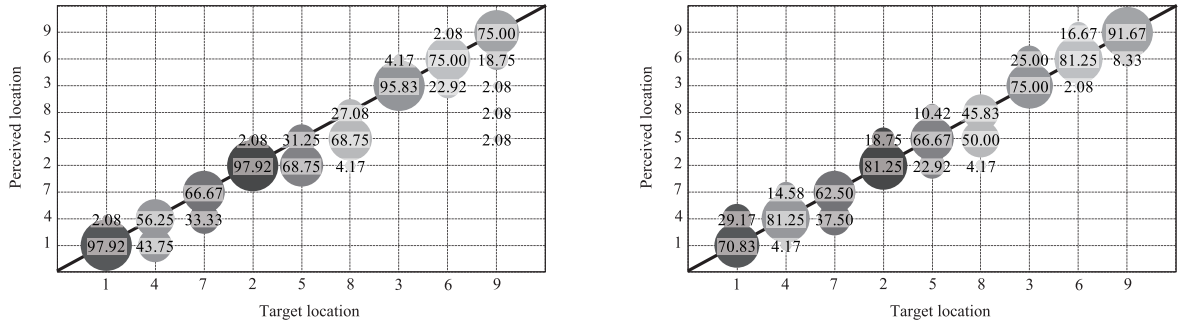
5. Conclusion

A real-time sound field transmission system using linear arrays of microphones and loudspeakers was proposed. The linear array configurations can provide a large listening area at a constant height. The received signals of the microphone array can be directly transformed into the driving signals of the loudspeaker array by using the WFR filter. Therefore, the WFR filtering method for linear arrays is suitable for real-time sound field transmission systems. Experiments to measure the reproduced sound field and a subjective lis-



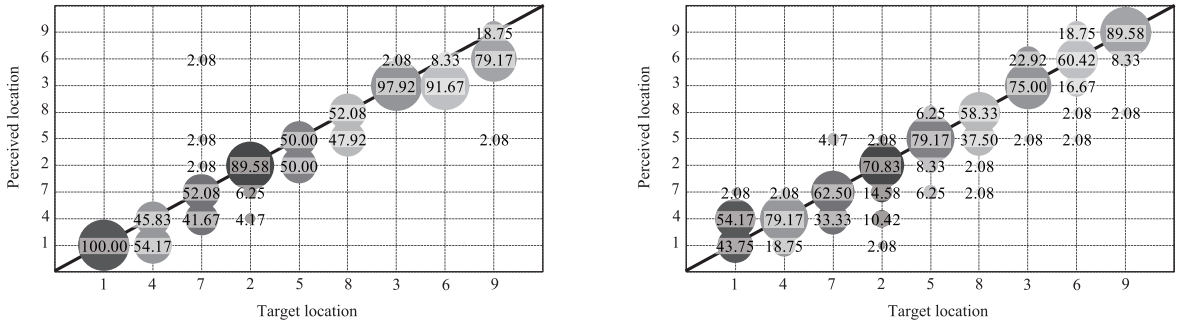
(a) Real (b) Virtual

Fig. 10 Results of subjective listening test at L1 (%).



(a) Real (b) Virtual

Fig. 11 Results of subjective listening test at L2 (%).



(a) Real (b) Virtual

Fig. 12 Results of subjective listening test at L3 (%).

tening test of sound localization were conducted to evaluate the proposed system. Even though the reproduced sound field included several artifacts, the proposed system was able to provide sound localization accuracy for virtual sound sources comparable to that for real sound sources in a large listening area.

References

[1] S. Koyama, K. Furuya, H. Uematsu, Y. Hiwasaki, and Y. Haneda, "Real-time sound field transmission system by using wave field reconstruction filter and its subjective listening test," AES 52nd Conference, Guildford, Sept. 2013.

[2] J. Blauert, Spatial Hearing: The Psychophysics of Human Sound Localization, The MIT Press, 1996.

[3] A.J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," J. Acoust. Soc. Am., vol.93, no.5, pp.2764–2778, 1993.

[4] E. Verheijen, Sound Field Reproduction by Wave Field Synthesis, Ph.D. thesis, Delft University of Technology, 1997.

[5] S. Spors, H. Teutsch, A. Kuntz, and R. Rabenstein, "Sound field synthesis," in Audio Signal Processing For Next-Generation Multimedia Communication Systems, ed. Y. Huang and J. Benesty, ch. 12, Kluwer Academic Publishers, 2004.

[6] D. de Vries, Wave Field Synthesis, AES Monograph, Audio Eng. Soc., 2009.

[7] S. Spors, R. Rabenstein, and J. Ahrens, "The theory of wave field synthesis revisited," Proc. 124th Conv. AES, Amsterdam, The Netherlands, Oct. 2008.

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

[9] 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 Language Process.,

vol.21, no.4, pp.685–696, 2013.
[10] U. Horbach and A. Karamustafaoglu, "Practical implementation of a data-based wave field reproduction system," Proc. 108th Conv. AES, Paris, Feb. 2000.
[11] S. Brix, T. Sporer, and J. Plogsties, "Carrouso — An European approach to 3D-audio," Proc. 110th Conv. AES, Amsterdam, May 2001.
[12] M.A. Gerzon, "Periphony: With-height sound field reproduction," J. Audio Eng. Soc., vol.21, pp.2–10, Jan. 1973.
[13] J. Daniel, "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonics format," Proc. 23rd Conf. AES, Copenhagen, Denmark, May 2003.
[14] M. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," J. Audio Eng. Soc., vol.53, no.11, pp.1004–1025, 2005.
[15] J. Ahrens and S. Spors, "An analytical approach to sound field reproduction using circular and spherical loudspeaker distributions," Acta Acustica United with Acustica, vol.94, pp.988–999, 2008.
[16] Y.J. Wu and T.D. Abhayapala, "Theory and design of soundfield reproduction using continuous loudspeaker concept," IEEE Trans. Audio Speech Language Process., vol.17, no.1, pp.107–116, 2009.
[17] B. Rafaely, "Analysis and design of spherical microphone arrays," IEEE Trans. Audio Speech Language Process., vol.13, no.1, pp.135–143, 2005.
[18] B. Rafaely, B. Weiss, and E. Bachmat, "Spatial aliasing in spherical microphone arrays," IEEE Trans. Signal Process., vol.55, no.3, pp.1003–1010, 2007.
[19] P.A. Nelson, "Active control of acoustic fields and the reproduction of sound," J. Sound Vibr., vol.177, no.4, pp.447–477, 1993.
[20] 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.
[21] M. Kolundžija, C. Faller, and M. Vitterli, "Reproducing sound fields using MIMO acoustic channel inversion," J. Audio Eng. Soc., vol.59, no.10, pp.721–734, 2011.
[22] E.G. Williams, Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography, Academic Press, New York, 1999.
[23] S. Koyama, K. Furuya, Y. Hiwasaki, and Y. Haneda, "Design of transform filter for reproducing arbitrarily shifted sound field using phase-shift of spatio-temporal frequency," Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), pp.381–384, Kyoto, March 2012.
[24] S. Koyama, K. Furuya, Y. Hiwasaki, and Y. Haneda, "Reproducing virtual sound sources in front of a loudspeaker array using inverse wave propagator," IEEE Trans. Audio Speech Language Process., vol.20, no.6, pp.1746–1758, 2012.
[25] M. Shimomura, A. Masuda, H. Uose, N. Inoue, and N. Nakayama, "GEMnet2 R&D testbed network," NTT Technical Review, vol.11, no.1, 2013.
[26] N. Kitawaki and K. Itoh, "Pure delay effects on speech quality in telecommunications," IEEE J. Sel. Areas Commun., vol.9, no.4, pp.586–593, 1991.
[27] A. Carôt, U. Krämer, and G. Schuller, "Network music performance (NMP) in narrow band networks," Proc. 120th Conv. AES, Paris, 2006.
[28] 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.
[29] S. Koyama, T. Lee, K. Furuya, Y. Hiwasaki, and Y. Haneda, "Improvement using circular harmonics beamforming on reverberation problem of wave field reconstruction filtering," Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), pp.276–280, 2013.
[30] Y. Suzuki, F. Asano, H.Y. Kim, and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," J. Acoust. Soc. Am., vol.97, no.2, pp.1119–1123, 1995.

Shoichi Koyama received the B.E., M.S., and Ph.D. degrees in mathematical engineering and information physics from the University of Tokyo, Japan, in 2007, 2009, and 2014, respectively. He joined Nippon Telegraph and Telephone Corporation (NTT) in 2009 and started the career as a researcher in acoustic signal processing at NTT Cyberspace Laboratories (currently Media Intelligence Laboratories). In 2014, he joined the Graduate School of Information Science and Technology, the University of Tokyo, as an Assistant Professor (Research Associate). Dr. Koyama was awarded the Young Researcher Award on Measurement Division by the Society of Instrument and Control Engineers (SICE) in 2009, the Best Young Researcher Paper Award on Sensors and Micromachines Society by the Institute of Electrical Engineers of Japan (IEEJ) in 2010, and the Awaya Prize Young Researcher Award by Acoustical Society of Japan (ASJ) in 2011. He is a member of the IEEE, Audio Engineering Society (AES), ASJ, and IEICE.

Ken'ichi Furuya received his B.E. and M.E. degrees in acoustic design from Kyushu Institute of Design, Fukuoka, Japan, in 1985 and 1987, and his Ph.D. degree from Kyushu University, Japan, in 2005. From 1987 to 2012, he was with the laboratories of Nippon Telegraph and Telephone Corporation (NTT), Tokyo, Japan. In 2012, he joined the Department of Computer Science and Intelligent Systems of Oita University, Oita, Japan, where he is currently a Professor. His current research interests include signal processing in acoustic engineering. Dr. Furuya was awarded the Sato Prize by the Acoustical Society of Japan (ASJ) in 1991. He is also a member of the Acoustical Society of Japan, the Acoustical Society of America, the Audio Engineering Society, and IEEE.

Hisashi Uematsu His biography and photo are not available.

Yusuke Hiwasaki received the B.E. degree in instrumentation engineering and the M.E. and Ph.D. degrees in computer science from Keio University, Yokohama, Japan, in 1993, 1995, and 2006, respectively. Joining NTT in 1995, he started the carrier as a research engineer in low bit-rate speech coding and voice-over-IP telephony. From 2001 to 2002, he was a guest researcher at KTH (Royal Institute of Technology) in Sweden. From 2009, he contributed as an Associate Rapporteur of ITU-T SG16 Q.10, a question on speech coding matters, then became Rapporteur in 2011, and held the position until 2013. He joined NTT FACILITIES Inc. in 2013, and has worked as Senior Manager in Global Business Division. Dr. Hiwasaki received the Technology Development Award from the Acoustical Society of Japan, the Best Paper Award from the IEICE Communications Society, the IEICE Achievement Award, and the Teishin Association Maejima Award in 2006, 2006, 2009, and 2010, respectively. He is also a member of the IEEE, and Acoustical Society of Japan.

Yoichi Haneda received the B.S., M.S., and Ph.D. degrees from Tohoku University, Sendai, in 1987, 1989, and 1999. From 1989 to 2012, he was with the Nippon Telegraph and Telephone Corporation (NTT), Japan. In 2012, he joined the University of Electro-Communications, where he is a Professor. His research interests include modeling of acoustic transfer functions, microphone arrays, loudspeaker arrays, and acoustic echo cancellers. Dr. Hanedareceived paper awards from the ASJ and from the IEICE of Japan in 2002. He is a senior member of the IEEE and IEICE, and a member of ASJ.