PROCEEDINGS of the
23rd International Congress on Acoustics
9 to 13 September 2019 in Aachen, Germany

Attempt to improve the total performance of sound field reproduction system: Integration of wave-based methods and simple reproduction method

Hiroshi KASHIWAZAKI[(1)], Akira OMOTO[(2)]

[(1)]Graduate School of Design, Kyushu University, Japan, hrs@penr.in
[(2)]Faculty of Design, Kyushu University, Japan, omoto@design.kyushu-u.ac.jp

Abstract

The sound field reproduction system can be applied to various applications. The primary objective of the system is to physically reproduce an arbitrary sound field correctly. However, physical accuracy is not the only requirement in a situation where the system is actually used. Through several applications, we assumed the following four factors as total performance; A) physical accuracy, B) robustness against disturbances, C) flexibility for additional direction, and D) capability of integration with visual media. We used 24-channel narrow directional microphone array and 24-channel loudspeaker as a platform to be considered and worked to improve overall performance. As one of the reproduction methods, directional information can be easily reproduced by amplitude panning which relies on the directivity of the microphone themselves, but the microphone leakage effect occurs in the low frequency. On the other hand, wave-based methods such as boundary surface control or higher-order ambisonics are effective for physical accuracy, but control of high frequency is difficult. Therefore, if wave-based method and simple reproduction method are combined, there is a possibility to reproduce a wide frequency range well. We discuss the total performance of this reproduction method through ITD/ILD measurement and visualization of a wide range of wavefront.

Keywords: Sound field reproduction, Narrow directivity microphone, Boundary surface control, Higher-order Ambisonics

1 INTRODUCTION

1.1 Total performance

The authors had attempted to improve the performance of the immersive sound field reproduction system based on the physical principle[1, 2]. An example is the Sound Cask which used 96-channel loudspeaker array and 80-channel microphone array and based on the Boundary Surface Control principle developed by Ise. In such a system, the physical reproduction accuracy of the sound field was mainly focused.

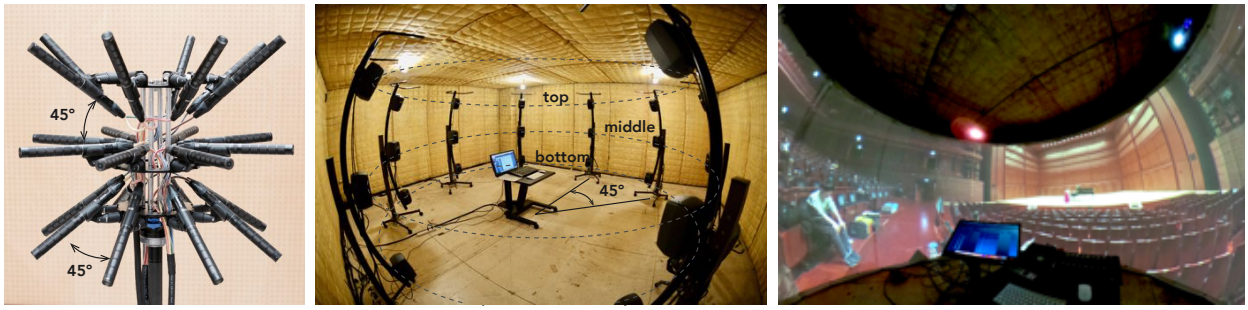
Then, using such a reproduction system, we wanted to use, for example, the following things;

- We want to mix the signal of the main microphone for sound field reproduction and change it to a more core sound.
- We want to use sound field reproduction technology as an auxiliary role of adding reverberation as in the conventional audio mixing.
- We want to use it as a simulator in combination with visual media.

For such usage, the following functions and performance are required;

- Physical reproduction performance.
- Robustness to disturbances and stability in environments with multiple people.
- Flexibility to operation on production such as mixing or adding reverb.
- Capability of integration with visual media.

Currently, the above items are assumed to contribute to “the total performance” of the system.



(a) (b) (c)

Figure 1. Systems overview; (a) the Hedgehog microphone array which consists of 24 narrow directional microphones, (b) the loudspeaker array of stacked-ring layout in three layers, and (c) image when presenting sound and video using a cylindrical screen installed inside the loudspeaker array.

1.2 Platform of examination

There are several sophisticated multi-channel reproduction strategies such as SIRR[3], DirAC[4], SDM[5], and Ambisonics[6] including higher order. Based on such an assumption describe in the previous section, we built an extremely simplified and flexible system that is easy to handle for mixing and reverb, adding screen, etc., as shown in Figure 1.

For sound field recording, a 24-channel narrow directional microphone array is used. It consists of upper, middle and lower layers at 45-degree intervals, and eight microphones are arranged at 45 degrees in the azimuth direction. The distance between the microphone elements is around 140 mm in the horizontal direction and around 160 mm in the longitudinal direction. So, the spatial Nyquist frequency when performing array processing is expected to be about 1000 Hz to 1200 Hz.

24-channel loudspeaker array with a stacked-ring layout for playback. It also consists of upper, middle and lower layers at 20-degree intervals, and eight microphones are arranged at 45 degrees in the azimuth direction. The distance from the listening position to each speaker is 2 m.

1.3 Simple reproduction method

The most basic and easy way to reproduce using this platform is to output the recorded signal directly from the loudspeaker located almost corresponding direction. The amplitude weighting due to narrow directivity is expected to work as natural amplitude panning. This method can be interpreted as an extension of the 6-channel system proposed by Yokoyama[7].

The directivity of the microphone directly affects the reproduction performance. In the case of the element used for the Hedgehog microphone, directivity becomes narrow at 2 kHz or more. So high performance can be expected in high-frequency range. On the other hand, wide directivity in the low-frequency range causes bass-boost and loss of spatial resolution.

1.4 Wave-based integration strategies

Therefore, a wave-based method is introduced to improve this. In contrast to the Direct method, which is suitable for high-frequency control, the wave-based method can be said to easy to control low-frequency range in terms of spatial aliasing.

In this manuscript, examines whether the integration of the wave-based method is effective. Tree wave-based methods are implemented in section 2, and those methods are evaluated for performance by measuring interaural time and level differences and the wavefront in section 3. Section 4 concludes on the possibility of integration strategies with stable and high performance.

2 APPLICABLE WAVE-BASED REPRODUCING METHOD

2.1 Boundary surface control

Boundary surface control (BoSC, hereafter) is a reproduction method based on the Kirchhoff–Helmholtz integral equation[1]. Based on this equation, by matching the sound pressure and the gradient on the closed boundary surface virtually installed in the primary sound field with the secondary field, the sound field inside the boundary is reproduced. In an actual system, only the sound pressure is controlled using microphones arranged on the boundary, because there is no need to measure the pressure gradients due to the uniqueness of the solution at most frequencies[8, 9]. Therefore, the pressures at the control points on the boundary controlled by the multi-channel inverse filter $\boldsymbol{H}(\omega)$ and the loudspeakers surrounding the boundary. Moreover, Kimura[10] theoretically shows that it can be controlled even by using an array with directional microphones.

The number of inverse filters is usually the product of the number of microphones and the number of loudspeakers. In the case of using the above system, the 576 ($= 24 \times 24$) filters are necessary, so we call it inv576 filter.

2.2 Reduced filter version of BoSC

Reducing the scale of the inverse filter[11] is also attempted. Especially when using a narrow directional microphone, few loudspeakers affect to the control point dominantly due to its directivity. Thus, the number of filters can be reduced reasonably with less performance degradation.

There are two possible advantages to reducing the inverse filter. First, since the calculation cost is reduced, it becomes easy to reproduce in real time. Second, since the control is performed using a dominant loudspeaker corresponding to the microphone direction, it seems that a stable filter with high robustness can be obtained. We have tried to reduce from 576 to 104, and it is called inv104.

2.3 Mixed-order Ambisonics

Higher-order Ambisonics is based on sound field representation by spherical harmonics expansion. The sound field is approximately reproduced by matching the expansion coefficients of the finite order of the primary field with the secondary field.

In the Ambisonics encoding process, the expansion coefficients are estimated from the recorded signal of the microphone array. The directivity of the narrow directional microphone used in the Hedgehog array can be considered as a cardioid-like pattern in the low-frequency range. Therefore, the microphone signal $s(\boldsymbol{r},k)$ at $\boldsymbol{r}(r,\theta,\phi)$ is modeled by the following equation[12].

$$s(\boldsymbol{r},k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} B_{n,m}(k) b_n(kr) Y_{n,m}(\theta,\phi) \tag{1}$$

$$b_n(kr) = \alpha j_n(kr) - (1-\alpha) i j_n'(kr) \tag{2}$$

where $j_n(kr)$ and $j_n'(kr)$ are spherical Bessel function and its derivative respectively, and $Y_{n,m}(\theta,\phi)$ is real-valued spherical harmonics. $B_{n,m}(k)$ is expansion coefficients. α and $(1-\alpha)$ are the weighting factor of the sound pressure and its radial derivative component respectively. Here, $\alpha = 0.31$ based on the fitting with the polar pattern under 500 Hz.

When there are 24 microphones, coefficients up to the 3rd order could be calculated. In practice, we used the 3H2V-MOA configuration[13], where the components of $B_{3,0}(k)$ is removed, for the following two reasons. First, the $B_{3,0}(k)$ component is estimated with poor accuracy, which may be attributed to the arrangement of the microphone array. Second, on the decoding side, the loudspeaker density seems to be insufficient for full-sphere decoding, because the loudspeaker array arrangement is a stacked-ring type. Hereafter, it is called MOA.

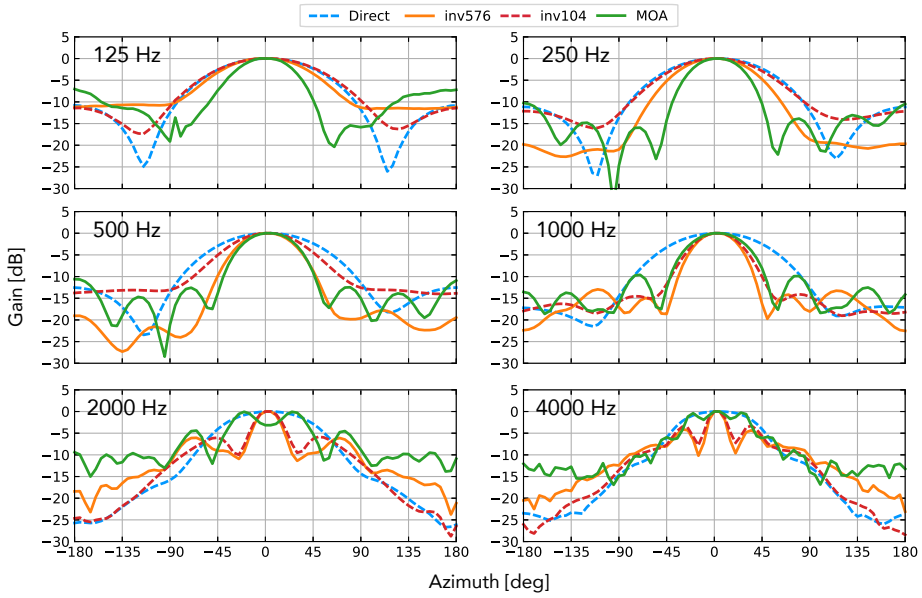


Figure 2. Equivalent directivity obtained by measuring the speaker output levels while rotating the Hedgehog array in the primary field

3 PERFORMANCE COMPARISON OF REPRODUCTION METHODS

In order to examine and compare the features of simple reproduction method and wave-based methods, the reproduction performance of wavefront and interaural time and level difference (ITD, ILD), the robustness of the filter, equivalent directivity of the array were measured.

3.1 Conditions

The primary field is an anechoic room, and the sound is emitted from a loudspeaker and recorded with the Hedgehog microphone. At this time, the Hedgehog microphone is rotated at an interval of 5 degrees on the turntable, in order to equivalently record the situation where the sound signal comes from various angles. Also, the wavefront around the recording position was measured by the laboratory made MEMS microphone array[11], and the reference value of ITD and ILD were measured by the dummy head.

The recording signal is processed by the four methods described above: Direct, inv576, inv104, and MOA. At this time, all of the three wave-based processes are implemented as 4096-tap FIR filters obtained from the IFFT of the transfer characteristic calculated for every 2^{16} frequency samples. As the equivalent directivity of the array, the level of the loudspeaker signal for each sound source direction of the original sound field is obtained. Then each reproduction signal is outputted by the loudspeaker array in the semi-anechoic room, and the wavefront and ITD, ILD were measured.

3.2 Result

3.2.1 Equivalent directivities

Figure 2 shows the equivalent directivities of one of the middle-layer loudspeaker output. The maximum value is normalized to 0 dB.

In the case of Direct, it shows almost the directivity of the microphone itself and has a cardioid-like pattern up to 1000 Hz, and the directivity gradually becomes narrow above 2000 Hz. inv576 has a main lobe that becomes narrower as the frequency is higher, and basically, it is narrower than Direct. The width of the main lobe of

inv104 is equal to that of Direct up to 250 Hz and becomes narrower above 500 Hz. Characteristically, MOA has a main lobe of equal width regardless of the frequency up to 1000 Hz. However, at 2000 Hz or more, the directivity pattern fluctuates greatly. This seems to be because the cardioid model (Eq. (1)) can not be applied at 2000 Hz or higher.

3.2.2 Wavefront

Figure 3 shows the primary and reproduced wavefront of the impulse sound passed through the band-pass filter. A visualization area is a horizontal plane of 0.72 m $\times$ 0.72 m, and the moment when the wavefront almost reaches the center is captured.

In Direct, waves propagate from several speakers and interfere with each other. And the wave in the same direction as the primary field becomes stronger at 2000 Hz or higher due to the directivity of the microphone becomes narrow. In inv576, wavefronts very similar to the primary field in all of the measured bands. However, since the theoretical spatial Nyquist frequency is about 1000 Hz to 1200 Hz, the reproduction accuracy of the wavefront is expected to be lower in the situation where waves higher than spatial Nyquist frequency come from the direction in which the loudspeaker is not actually placed in the secondary field, for example, 22.5 degrees. inv104 is an intermediate similarity between Direct and inv576. In MOA, at 1000 Hz or less, a wavefront in the central area is similar to the primary field. However, reproduction accuracy is low above 2000 Hz.

3.2.3 Interaural time and level differences

The ITD was calculated using the cross-correlation function for the left and right of the dummy head signals passed through the 1600 Hz LPF. In the calculation of ILD, window functions of frequency dependent length were applied to the response to remove the inevitable reflection from the floor.

Figure 4(a) shows that inv576 reproduces ITD best. In Direct and inv104, the difference becomes smaller by about 0.1 ms for the reproduction of 90 degrees and 270 degrees. It seems that the loudspeaker output other than the primary direction, as seen in Figure 3, makes the ITD smaller. MOA's ITD is reproduced about 0.1 ms larger than the primary.

Figure 4(b) shows ILD. At 250 Hz, all methods generally reproduce the characteristics of the primary field, but the MOA is slightly smaller. inv576 and MOA are relatively reproducible at 500 Hz, and inv104 performs as well as inv576 and MOA at 1000 Hz. At 2000 Hz or higher, it is difficult to reproduce stable in all directions regardless of which reproduction method is used. Still, at 8000 Hz, Direct, inv576, and inv104 relatively follow the primary field.

3.2.4 White noise gain

White noise gain (WNG) is a measure for estimating the robustness of microphone array process against microphone self-noise, position error, and amplitude and phase variation. WNG represents the SNR improvement rate for uncorrelated white noise among microphones and is calculated by the following equation[14].

$$\mathrm{WNG}(k) = \frac{|\boldsymbol{W}^H \boldsymbol{d}|^2}{\boldsymbol{W}^H \boldsymbol{W}} \tag{3}$$

where $\boldsymbol{W}$ represents the transfer function of the processing filter associated with anyone loudspeaker output, and $\boldsymbol{d}$ represents the transfer function from the sound source to the microphone array. A positive WNG represents an improvement in SNR.

Figure 5 shows the WNG of the front loudspeaker output in the case where the sound source of the primary field is also in the front direction. In Direct, since the speaker output is the corresponding microphone signal itself, so the output SNR is equal to the microphone SNR, and the WNG becomes 0 dB. In inv576, WNG does not decrease significantly below 1000 Hz, but the value fluctuates above 1000 Hz. Also, the WNG drops at 150 Hz, due to the inverse filter was high gain, since the direct sound and floor reflections of the middle layer speaker output weaken at the listening position at this frequency. On the other hand, inv104's WNG is higher

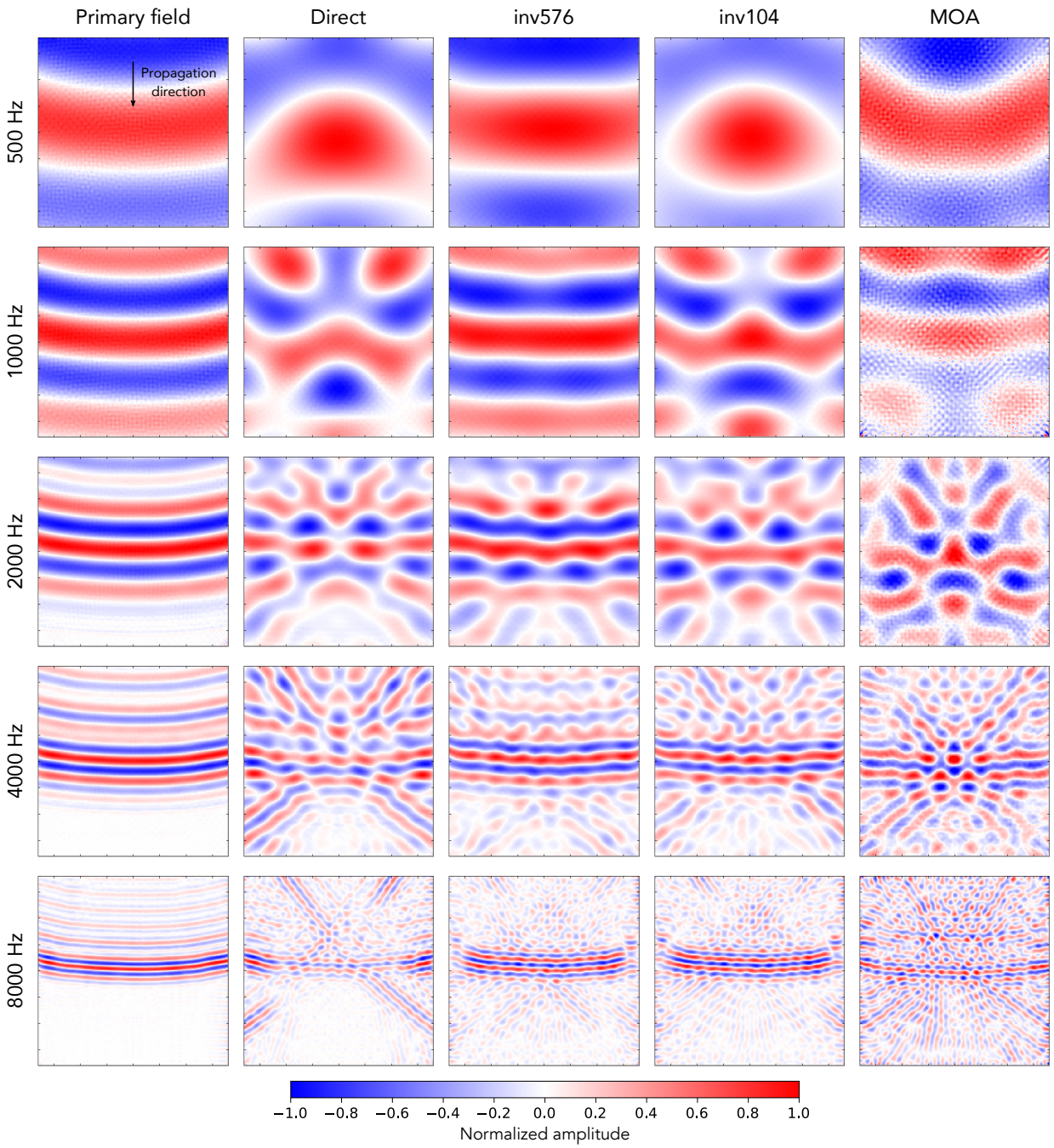


Figure 3. Horizontal wavefront measured in the primary field and the reproduced field with Direct, inv576, inv104, and MOA. The sound source was an impulse passed through one-octave bandpass filter emitted from a loudspeaker placed 2 m away from the center of the visualization area. The frequency described on the left side represents the center frequency of the bandpass filter.

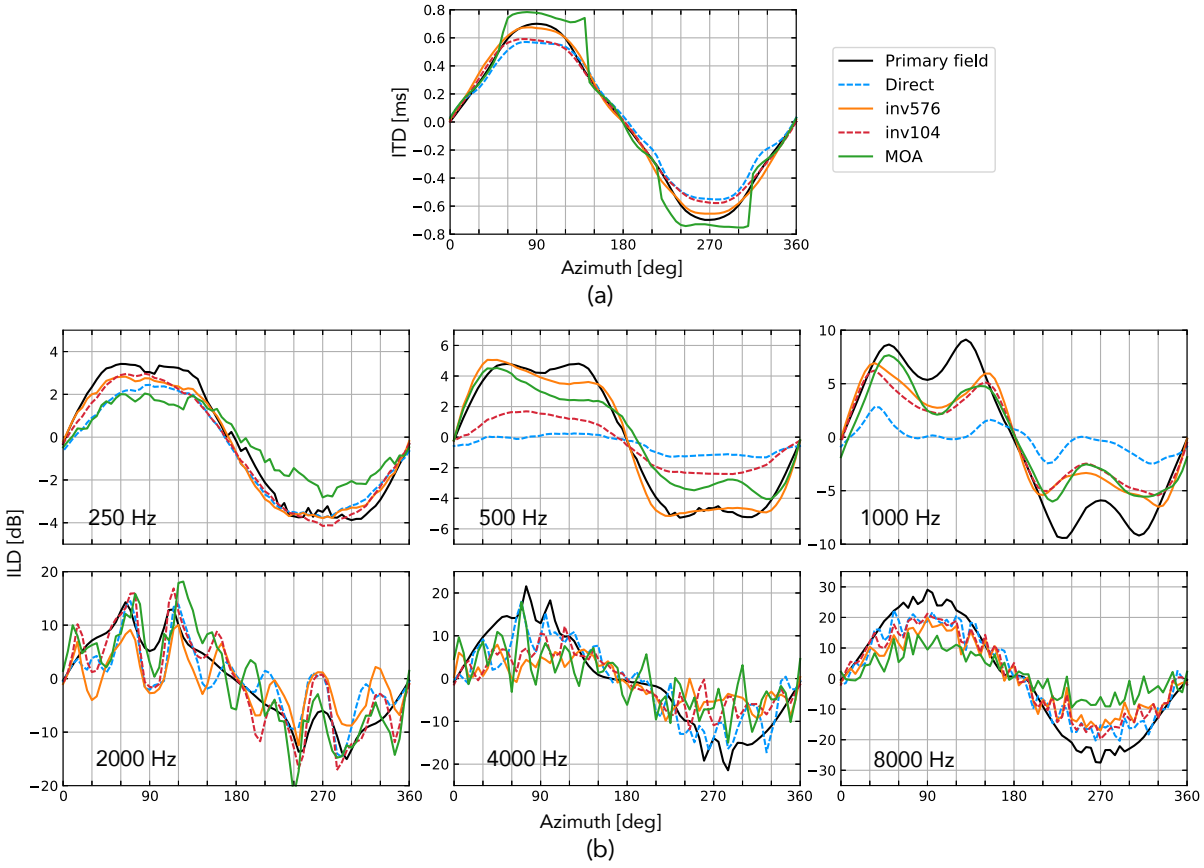


Figure 4. Interaural time and level difference; (a) Interaural time difference in primary and reproduced filed which calculated using the cross-correlation function for the left and right of the dummy head signals passed through the 1600 Hz LPF. (b) Interaural level difference in primary and reproduced field calculated every 250 Hz to 8,000 Hz 1/1 octave bands.

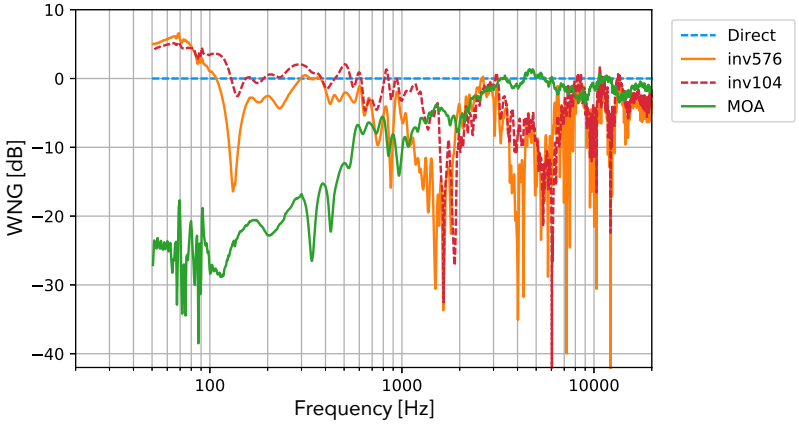


Figure 5. Measured white noise gain for the front loudspeaker signal in the case where the sound source of the primary field is also in the front direction. A positive WNG indicates an improvement in SNR.

than inv576 from 100 Hz to 1500 Hz, and the notch at 150 Hz is suppressed. MOA's WNG is very low at less than 1000 Hz. Above those frequencies, there is less concern for SNR degradation than inv576 or inv104.

4 CONCLUDING REMARKS

Attempts have been made to improve the overall performance of sound field reproduction systems, where we focus not only on A) physical accuracy but also on B) robustness against disturbances, C) flexibility for additional direction, and D) capability of integration with visual media. As a platform of examination, a system using the narrow directional microphone array has been built, and a simple reproduction method called Direct is basically adopted. The performance of the Direct method directly is affected by the directivity of the microphone. Therefore, the wave-based method was applied and the reproduction performance was evaluated.

inv576 has a high reproducibility of wavefront and ITD. In particular, in the band below 1.6 kHz, where ITD is an important cue for sound image perception in the horizontal direction, and stable reduction can be expected without significant WNG drop. inv104 is an intermediate performance between Direct and inv576 and should be used when a low computation is required. MOA was good accuracy on wavefront and ILD below 1 kHz, but WNG decreased. We want to make a stable filter, such as using only low orders ambisonics in this frequency range, and evaluate it again.

The issue is that it is difficult to obtain high ILD accuracy in the 2 kHz to 4 kHz band. Although the performance of the ILD is said to be important above 1.6 kHz, Direct can reproduce ILD with high accuracy only at 8 kHz or higher. As a technical solution, it is considered to use a microphone with narrow directivity from the lower frequency range, or to enhance the high-frequency performance of the wave-based method with a higher density loudspeaker and microphone array. In the actual listening situation, the impact of this problem will differ depending on the playback content, and it is also necessary to examine in detail by subjective evaluation.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Numbers JP 25282003, JP 17H00811.

REFERENCES

[1] Ise, S. A principle of sound filed control based on the Kirchhoff–Helmholtz integral equation and the theory of inverse systems. Acta Acoust. Acustica, Vol 53, 1999, pp 250–257.
[2] Omoto, A.; et al. Sound field reproduction and sharing system based on the boundary surface control principle, Acoust. Sci. & Tech., Vol 36, 2015, pp 1–11.
[3] Merimaa, J.; Pulkki, V. Spatial impulse response rendering I: analysis and synthesis, J. Audio Eng. Soc., Vol 53(12), 2005, pp 1115–1127.
[4] Pulkki, V. Spatial sound reproduction with directional audio coding, J. Audio Eng. Soc., Vol 55(6), 2007, pp 503–516.
[5] Tervo, S.; et al. Spatial decomposition method for room impulse responses, J. Audio Eng. Soc., Vol 61(1/2), 2013, pp 17–28.
[6] Poletti, M. Three-dimensional surround sound systems based on spherical harmonics, J. Audio Eng. Soc., Vol 53(11) pp 1004–1025.
[7] Yokoyama, S.; Ueno, K.; Sakamoto, S.; Tachibana, H. 6-channel recording/reproduction system for 3-dimensional auralization of sound fields, Acoust. Sci. & Tech. Vol 23, 2002, pp 97–103.
[8] Kleinman, R.; Roach, G. Boundary integral equations for the three dimensional Helmholtz equation, SIAM Review, Vol 16, 1974, pp 214–236.
[9] Furuya, K.; Ichinose, Y. Sound field control in a closed space using boundary-pressure-method, Technical Report of IEICE, EA90-15, 1990, pp 25–32.
[10] Kimura, T. Theoretical study and numerical analysis of 3D sound field reproduction system using directional microphones and Boundary Surface Control, Trans. Virtual Reality Soc. Jpn, 5, 2010, pp 231–241.
[11] Kashiwazaki, H.; Omoto, A. Sound field reproduction system using narrow directivity microphones and boundary surface control principle, Acoust. Sci. & Tech., Vol 39 (4), 2018, pp 295–304.
[12] Balmages, I.; Rafaely, B. Open-sphere designs for spherical microphone arrays, IEEE Trans. Audio Speech Lang. Process., Vol 15(2), 2007, pp 727–732.
[13] Travis, C. A new mixed-order scheme for Ambisonic signals, Proc. Ambisonics Symp. Graz, 2009, 6 pp.
[14] Brandstein, M.; Ward, D. Microhpone Arrays - Signal Processing Techniques and Applications, Springer, 2001.