ISRA 2019 AMSTERDAM

PROCEEDINGS of the
International Symposium on Room Acoustics

15 to 17 September 2019 in Amsterdam, Netherlands

NAG NEDERLANDS AKOESTISCH GENOOTSCHAP

Sound field Reproduction using Convolving Directional Dry Signals and Directional Impulse Responses

Akira OMOTO[1]; Tomohiro SHIMIZU[2]; Hiroshi KASHIWAZAKI[3]

[1] Faculty of Design, Kyushu University, Japan

[2] Graduate School of Design, Kyushu University, Japan

[3] Graduate School of Design, Kyushu University, Japan

ABSTRACT

The overall performance of a loudspeaker-based sound-field reproducing system is examined. We assume it essential to satisfy four conditions. (A) The reproduction is based on physical principles to ensure the fundamental performance. (B) The system is robust against unavoidable disturbances, such as the presence of listeners. (C) There is room to accept additional direction, such as an intentional change in reverberation or frequency characteristics. (D) The system has high affinity with other stimulation, such as visual information. In a practical examination, a 24-channel hedgehog-shaped narrow-directional microphone array and a 24-channel loudspeaker array in which loudspeakers are arranged at intervals of 45° in the azimuth angle and in three layers are used as a platform. Using this system and considering condition (A), the reproduction of the sound field in a concert hall is attempted. The directional impulse responses measured in 24 directions are convoluted with conventional stereo recording signals or signals recorded in different directions to cope with the directivity of musical instruments. The performance is examined in terms of the reproducibility of physical parameters and subjective evaluation.

Keywords: Sound field reproduction, Directional microphone

1. INTRODUCTION

Sound-field reproduction systems can be roughly categorized into two types. The first type of system is based on physical principles and aims at the accurate reproduction of physical quantities. The second type of system is sometimes referred to as having a psychological basis and arbitrarily modifies and creates phantom images of sources and the surrounding environment to reproduce an imaginary sound field as desired. The latter type is a creative and artistic system.

The first type of system uses wave-based methods such as wave field synthesis (1,2), ambisonics (3,4), boundary surface control (BoSC) (5), and derived versions as standard methods. Especially in the reproduction of sound in a concert hall, the analysis and synthesis of impulse responses or recorded sound based on human perception, such as the use of SIRR (6,7), DirAC (8,9), and SDM (10), are also powerful approaches.

We previously examined the validity of a system based on the boundary surface control principle. Our system captures the sound field using a 80-channel fullerene-shaped microphone array and reproduces the sound field using 96-channel loudspeakers arranged in a nonagonal enclosure called the Sound Cask (11) or using 48-channel loudspeakers arranged in what is called the Sound Block.

More recently, we introduced a more straightforward system that uses 24-channel narrow-directional microphones and 24-channel loudspeakers (12). This system is designed to realize a robust and versatile reproduction system in which the targets of reproduction are the sound fields of not only concert halls but also various acoustical environments. A microphone array having similar structure was adopted by NHK for the capture of ambient sound with a 22.2-channel system (13).

A principal idea of our current 24-channel system is to well reproduce sound with a high-quality microphone, such as a microphone used for recording music, with as little signal processing as possible. We assume it vital to satisfy four conditions in improving the overall performance of the

[1] omoto@design.kyushu-u.ac.jp
[2] Currently with Foster Co., Ltd.
[3] hrs@penr.in

system. (A) The reproduction is based on physical principles to ensure the fundamental performance of the reproduction. (B) The system is robust against unavoidable disturbances, such as the presence of listeners. (C) There is room to accept additional direction, such as intentional changes in the locations of sources, reverberation, and frequency characteristics. (D) The system has high affinity with other stimulations, such as visual information.

The present paper attempts the hybrid use of wave-based reproduction methods, such as BoSC and Ambisonics, beamforming, and the natural directional characteristics of the microphone to realize condition (A). In this trial, additional processing is only applied at low frequency while high-frequency components are directly reproduced by a loudspeaker aligned in a direction almost corresponding to the direction of the main lobe of a corresponding microphone. This direct and straightforward reproduction is a natural expansion of a six-channel reproduction system (14).

As another method of realizing condition (A) and partly realizing condition (C), the present paper attempts reproduction after convolving dry signals with directional impulse responses. The directional impulse responses used in this method are previously measured in the primary sound field with 24-channel narrow-directional microphones. Additionally, to cope with the directivity of the sound source, such as a musical instrument, several dry signals are recorded at surrounding microphone positions. The final reproduction sounds are obtained by convolving the directional impulse responses and dry directional signals. The performance of the method is verified by comparing the physical quantities derived from the original and reproduced impulse responses. A subjective evaluation examines the effectiveness of using combinations of dry and directional signals and directional impulse responses.

2. SYSTEM STRUCTURE

2.1 Microphone and loudspeaker arrays

Figure 1(a) shows a microphone array for capturing sound-field information comprising 24 narrow-directional microphones arranged at intervals of the azimuth angle of 45º and at three elevation angles of 0º and ±45º. Figure 1(b) shows a 24-channel speaker array with speakers having alignments almost corresponding to those of the microphones. The directional characteristics (polar pattern) are almost cardioid at frequencies lower than 1 kHz. The main lobes become narrower as the frequency increases.

Two systems of speakers having different geometry were used in the present experiments and demonstrations. The radius of the central horizontal ring of speakers and the height of the upper speakers are respectively 2.0 and 2.4 m for the smaller system and 2.5 and 3.0 m for the larger system.

2.2 Reproduction Methodology

Figure 1: Platform for the examination: (a) hedgehog-shaped narrow-directivity microphone array and (b) loudspeaker array having three heights.

2.2.1 Minimum Signal Processing

The most straightforward method of reproduction is direct reproduction (referred to as **direct**

hereafter), in which the sound captured by the 24-channel microphones is directly reproduced by the 24-channel loudspeakers positioned with almost corresponding alignments as the microphones without any processing of the signals. Additionally, several processing schemes have been attempted to eliminate unexpected overlap in the low-frequency range; i.e., inverse filtering based on the boundary surface control principle (referred to as **BoSC** hereafter) and the combination of beamforming in the low-frequency range and the direct emission of higher-frequency components (referred to as **bf+direct** hereafter). We presented typical results showing the reproducibility of the wavefronts in the case of BoSC in earlier work (12).

The beamformer was designed adopting spherical harmonics decomposition to obtain a uniform beam pattern regardless of the frequency and direction (15). The spherical harmonics of our microphone can be approximately modeled as a cardioid-like pattern, and the coefficients up to second order are used in calculations. The detailed design procedure was presented in a previous study (16). The ideal beam pattern is assumed to have a cone shape with an open angle of 45º. In this case, the synthesis coefficients c_n are obtained as $c_0 = 0.4256, c_1 = 0.7091, c_2 = 0.8457$. The direction of the beam was oriented in the 24 directions of the microphones. The upper limiting frequency was set at 1.6 kHz and natural directional characteristics were used for higher frequencies.

2.2.2 **Reproduction by Convolution**

The system can not only simply capture and reproduces the sound field but also reproduce the sound field by convolving the directional impulse responses. The present paper attempts the reproduction of the directional information of the sound field using impulse responses measured by the microphone array shown in Fig. 1(a). This method makes two approximations. First, the impulse responses measured at a specific microphone comprise a series of reflections mainly coming from the direction of the axis of the microphone. Second, a series of reflections can be effectively reproduced by the loudspeaker having almost the corresponding alignment.

In the measurement of impulse responses, the direction of the source (loudspeaker) was varied and the convolution with dry signals recorded at different positions was examined. Figure 2(a) is a schematic diagram of the reproduction method. The figure shows that the dry signal measured in the direction of (θ_i, ϕ_i) is denoted $d(\theta_i, \phi_i)$. If the impulse response measured for the loudspeaker facing the direction of (θ_i, ϕ_i) and k-th microphone in the array is expressed as $h_k(\theta_i, \phi_i)$, the k-th output signal y_k can be expressed as

$$y_k = \sum_i h_k(\theta_i, \phi_i) * d(\theta_i, \phi_i), \tag{1}$$

where i is arbitrarily selected and the asterisk indicates convolution. The loudspeakers emit the signals y_k $(k = 1,2,\cdots,24)$ after the above processing; i.e., **direct**, **BoSC**, or **bf+direct**.

If the source signal is provided in monaural or stereo format, the directional impulse responses can be used as an up-mix processor with a spatial reverberator that can reproduce the surrounding sound environment. Figure 2(b) is a conceptual diagram. In this case, the ambient components are generated by convolving the monaural signal and the impulse responses measured using the loudspeaker having a specific direction and the 24-channel microphones. The direct signal is separately supplied to the appropriate loudspeakers with panning if necessary.

In the case of reproduction using 24-channel original signals, the processing described above, such as **BoSC** or **bf**, requires matrix-shaped filtering. Meanwhile, in the case of the up-mix procedure with the monaural signal, the processing can be summarized as the 24-channel convolution of the calculated finite impulse response filter, which can be realized by sampling reverberator plugins in commonly used Digital Audio Workstations.

We assume that this up-mix methodology with the spatial reverberator minimally satisfies condition (A) mentioned above and might be useful for satisfying condition (C), there being room to accept an additional presentation, because we can modify the degree of reverberation for arbitrary directions. Similarly, the generation of an effective spatial reverberator using the sound intensity responses, derived from the impulse responses measured at closely located microphones in three orthogonal directions, is attempted. The comparison of such methods is left as future work (17).

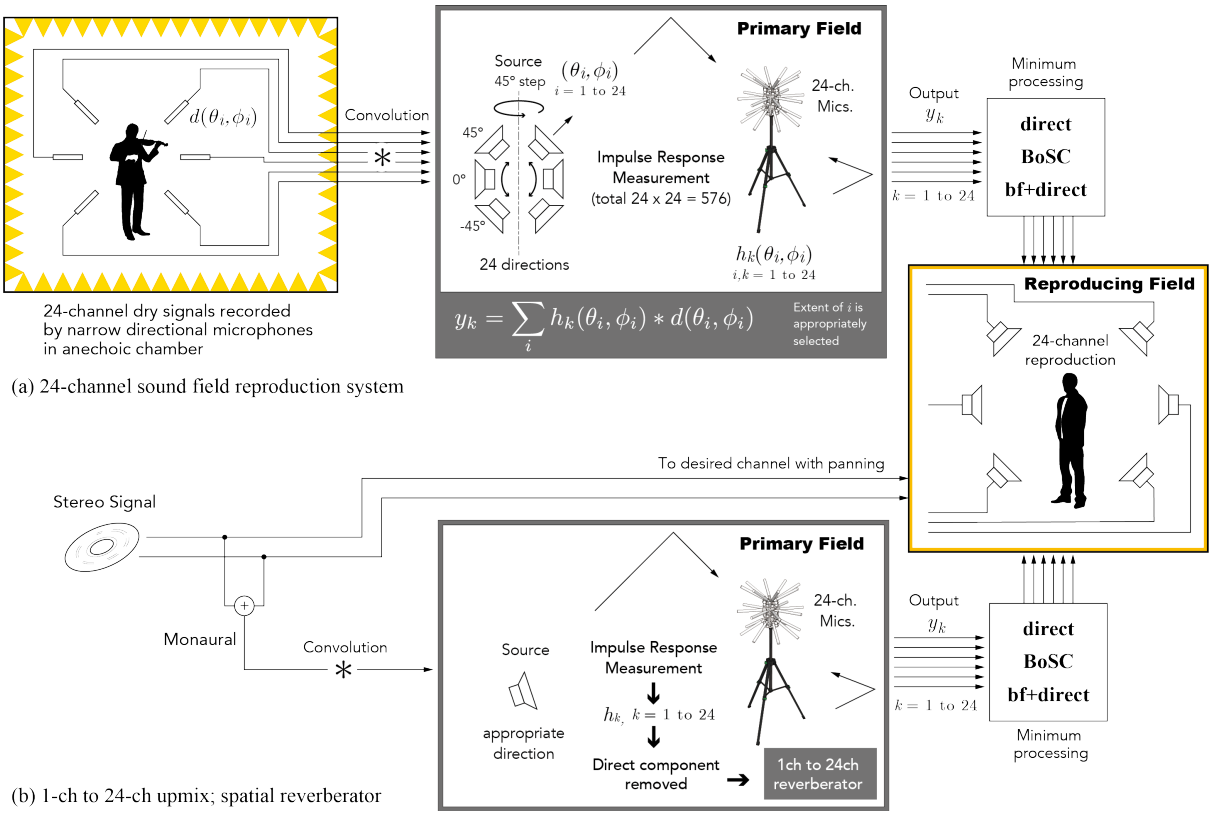

Figure 2: Schematic diagram of reproduction methods.

2.3 Impulse responses used in the examination

The impulse responses used in the examination were measured in a large concert hall having more than 1800 seats. The loudspeaker was located at the center of the stage while the receiving position was set around the central seat in the 10th row. Setting the source position at the center is usually avoided in the measurement of physical parameters. However, in this case, a central position was selected to obtain generic responses useful for the spatial reverberator.

The microphone had sensitivity even in the backward direction and the measured impulse responses thus included undesired direct sound. The components of direct sound should be removed when using the up-mix function. The duration of the removed signal was set to 30 ms, considering the geometry of the hall.

The front face of the loudspeaker was aligned in 24 different directions during the measurements. The quantity i in Fig. 2 (a) therefore takes values from 1 to 24. The front face was rotated in intervals of 45º (eight directions) in terms of the azimuth angle and at elevations of 0° and ±45° (three directions). As described later, the dry signals of several musical instruments were recorded in the same 24 directions. We assume that the dry signal that was recorded in a specific direction; e.g., the signal recorded by the microphone aligned upward and 45º to the right, is convolved with the impulse response measured with the loudspeaker aligned to the same direction, upward and 45º to the right. The combination of dry directional signals and directional impulse responses was examined later in preliminary subjective evaluations.

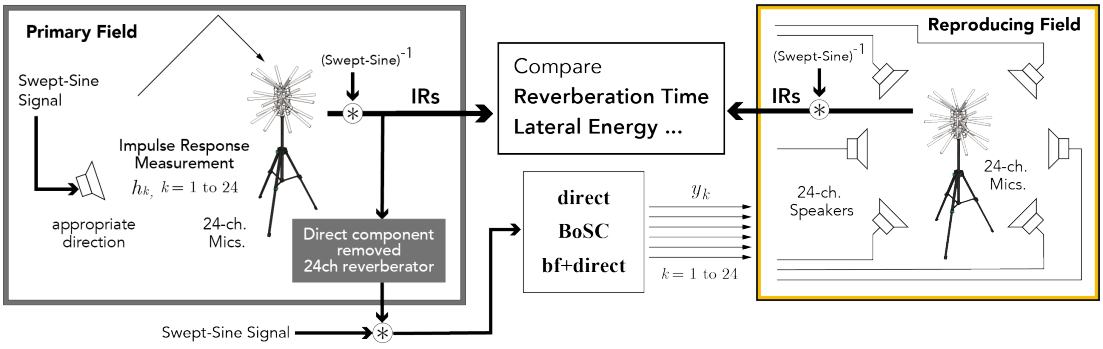

Figure 3: Block diagram of the experiment.

2.4 Basic performance:

reproduction of physical quantities

The reproducibility of several physical parameters was first examined. The block diagram is shown in Fig. 3. The 24 impulse responses measured by the microphone array in the primary field (concert hall) were taken as references. These references were obtained by measuring with the front face of the loudspeaker pointed toward the front of the stage.

The recorded swept sine signals in the primary field were emitted by the 24-channel loudspeakers to reproduce the field after the signals were passed through the above-mentioned spatial reverberator using the measured impulse responses with 30 ms of direct sound removed.

Furthermore, additional filters of **BoSC** and **bf+direct** were examined. The 24-channel microphone array was again set at the center of the loudspeaker array and the reproduced version of impulse responses was obtained.

Figure 4 shows the results of reverberation times calculated by each response and average values for the 24 microphones. There is a characteristic pattern in the results, especially at 1 and 2 kHz, in that the reverberation time was longer for microphones 3, 7, 11, 15, 19, and 23; i.e., the microphones pointing toward the sides. All reproduction methods followed the same trends, and there were no appreciable differences in the results except at the lowest frequency, 125 Hz. There were small differences in the average reverberation times for all strategies.

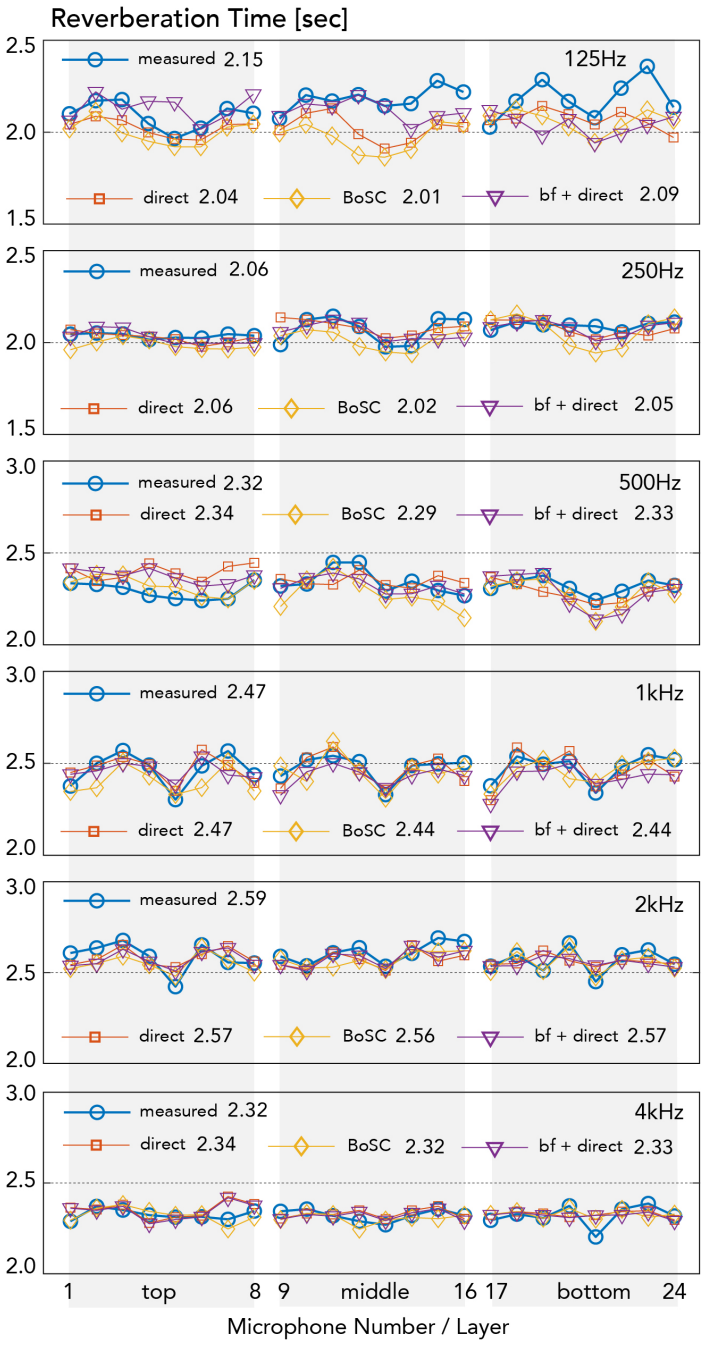

Figure 4: Reverberation times calculated from the impulse responses measured in the hall, and the reproduced responses. Average values are shown for each condition.

Figure 5 shows the pseudo lateral energy fraction (PLF), which is the ratio of the sum of the squared values of impulse responses observed by two microphones pointing toward the sides to that of the squared values of impulse responses observed by all microphones, as expressed in the following equation and at the top of the figure. The calculation was carried out for each layer of microphones and the three figures give the results for the top, middle, and bottom layers. The definition of the PLF is

$$\mathrm{PLF} = \sum_{L \text{ and } R} \int_0^\infty p_{L \text{ or } R}^2(t)dt \Big/ \sum_{i=1}^{8} \int_0^\infty p_i^2(t)dt. \qquad (2)$$

The average values for frequencies between 125 Hz and 4 kHz are shown in the figure. Provided that the difference limen of this fraction is 0.05 (i.e., the same as the limen for the conventional lateral energy fraction), **BoSC** provided better results for the top and middle layers. There were no appreciable differences in results for the middle layer among strategies.

The results show that fundamental characteristics are reasonably reproduced using each method,

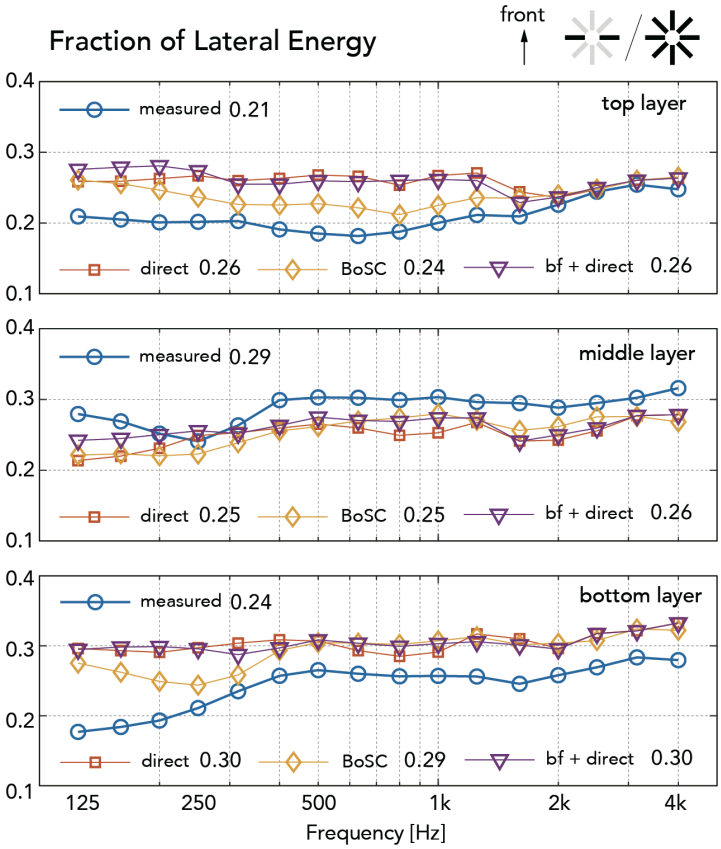

Figure 5: Reverberation times calculated using the impulse responses measured in the hall, and the reproduced responses.

even the simple **direct** method. This suggests that it is difficult to judge the superiority of any of the methods.

If the reverberation time of an environment is short, equalization by inverse filtering might be a reasonable solution while beam forming might work better for other environments. Alternatively, the straightforward, direct reproduction might be better for 'rough' reproduction when many people listen simultaneously, for example. Additionally, there must be an affinity with the reproduced content. An examination adopting a subjectivity evaluation is currently being carried out.

3. CONVOLUTION WITH DIRECTIONAL DRY SIGNALS

Twenty-four surrounding microphones recorded dry signals of a violin and saxophone. The microphone array was the same shown in Fig. 1(a). The signals were convoluted with impulse responses measured using the loudspeaker whose front face was pointed in directions almost corresponding with the 24 directions of the recording microphones. As a preliminary experiment, we examined several patterns of convolution, in which the combination of a single source and the responses, the multiple source signals, and the multiple responses was attempted. An example of the slanted direction of the loudspeaker and examples of recordings in an anechoic chamber are shown in Fig. 6. The impression of reproduced sound was examined by subjective evaluation for various combinations of the source signal and impulse responses.

3.1 Conditions of Reproduction Signals

The working conditions used in the evaluation are summarized in Table 1. Condition 1 uses a monaural dry signal as an anchor, condition 2 uses only one signal and the response, and condition 3 uses all dry signals and all impulse responses; i.e., a total of 24 × 24 = 576 convolutions are carried out. Conditions 4, 5, and 6 use nine dry signals and responses (9 × 24 = 216) for the front, left, and right directions. Condition 7 uses the four directions in which the highest equivalent sound pressure levels are observed.

Figure 6: Measurement of impulse responses with a slanted loudspeaker (left) and recordings with 24-channel narrow-directional microphones for the saxophone (center) and violin (right).

Table 1: Conditions of the subjective evaluation

#: index	Azimuth angle θ	Horizontal angle ϕ	No. of signals
1: anchor	Monaural dry source at $(\theta, \phi) = (0°, 0°)$		1
2: front mono	$\theta = 0°$	$\phi = 0°$	1
3: all	$-180° \leq \theta < 180°$	$-45° \leq \phi < 45°$	24
4: front nine	$-45° \leq \theta < 45°$	$-45° \leq \phi < 45°$	9
5: left nine	$45° \leq \phi < 135°$	$-45° \leq \phi < 45°$	9
6: right nine	$-135° \leq \phi < 45°$	$-45° \leq \phi < 45°$	9
7: selected four	Selected four directions for highest levels		4

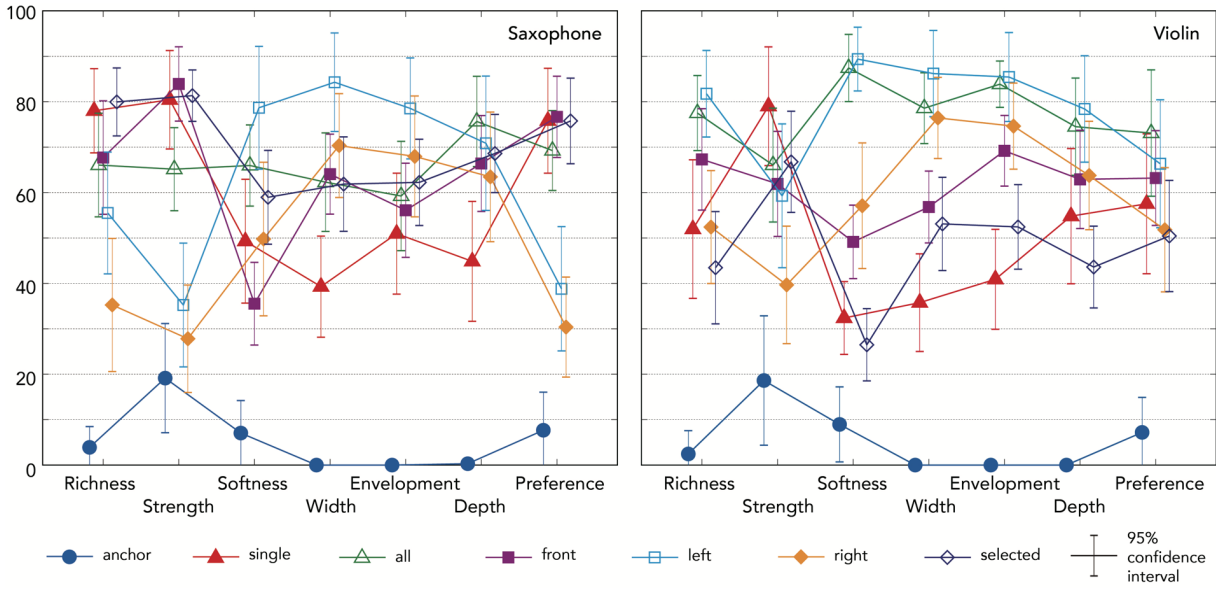

Figure 7: Results of the subjective evaluation.

3.2 Results of Evaluation

The evaluation adopted the MUSHRA (MUltiple Stimuli with Hidden Reference and Anchor) based method. No reference signal was assumed in the experiment. Twenty-one participants (18 male and three female) were asked to score the reproduced sound from 0 to 100 for seven evaluation terms; i.e., richness, strength, softness, width, envelopment, depth, and overall preference.

Figure 7 presents the results of the evaluation. A *t*-test was carried out to verify significant differences between evaluation terms. Relatively low scores were observed for timbre, richness, strength, and softness in cases 5 (using the nine left signals) and 6 (using the nine right signals) (see Table 1) for the saxophone. In these cases, there were no significant differences with the anchor. In the case of the violin, cases 3 (using all signals) and 5 (using the nine left signals) had high scores for softness. Higher scores were observed in the case of the violin for terms related to spatial impressions, such as width, envelopment, and depth.

Roughly speaking, case 2 (using a single front signal or a single convolution) had lower scores for spatial aspects while case 5 (using the nine left signals) had higher scores. Additionally, case 3 (using all signals) had a high score for the violin but not necessarily the highest. The results for case 7 (selecting four dominant directions) were low in most cases.

4. CONCLUDING REMARKS

A 24-channel narrow-directional microphone array and 24-channel loudspeaker array with a stacked-ring layout were used to construct a versatile and high-performance system for reproducing a sound field. This system allows for the simple and straightforward reproduction of a sound field; e.g., emitting a recorded signal from a loudspeaker having a direction almost corresponding to that of the microphone that recorded the signal.

Twenty-four-directional impulse responses can be obtained using the 24-channel narrow-directional

microphone array. A reproduction method that convolves these directional impulse responses with dry signals recorded in various directions or uses up-mix procedures with few channel signals can be assumed. For each method, minimal signal processing, such as inverse filtering based on the boundary surface control principle or beamforming based on spherical harmonics decomposition, was introduced to compensate for insufficient directional characteristics in the low-frequency range.

The present paper examined the reproduction performance using several physical parameters with a reproduction scheme similar to the up-mix procedure. Additionally, subjective evaluations were made to examine the validity of the convolution of dry and directional signals with directional impulse responses. Results suggest that 1) the system reproduces fundamental quantities, such as the directional reverberation time and the fraction of energy coming from lateral directions, and 2) the convolution of the directional impulse response and the dry and directional signals contribute positively to the spatial impression of reproduced sound.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI under grant numbers JP 25282003 and JP 17H00811. We thank Glenn Pennycook, MSc, from Edanz Group for editing a English draft of this manuscript.

REFERENCES

1. Berkhout AJ, de Vries D, Vogel P. Acoustic control by wave field synthesis. J Acoust Soc Am.1993; 93: 2764–2778.
2. Gauthier P, Berry A. Adaptive wave field synthesis for sound field reproduction: Theory, experiments, and future perspectives. J Audio Eng Soc. 2007; 55: 1107–1124.
3. Poletti MA. Three-dimensional surround sound systems based on spherical harmonics. J Audio Eng Soc. 2005; 53: 1004–1025.
4. Trevino J, Koyama S, Sakamoto S, Suzuki Y. Mixed-order ambisonics encoding of cylindrical microphone array signals, Acoust Sci & Tech. 2014; 35: 174–177.
5. Ise S. A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems. Acta Acust united Ac. 1999; 85: 78–87.
6. Merimaa J, Pulkki V. Spatial impulse response rendering I: Analysis and synthesis. J Audio Eng Soc. 2005; 53: 1115–1127.
7. Pulkki V, Merimaa J. Spatial impulse response rendering II: Reproduction of diffuse sound and listening tests. J Audio Eng Soc. 2006; 54: 3–20.
8. Pulkki V. Spatial sound reproduction with directional audio coding. J Audio Eng Soc. 2007; 55: 503–516.
9. Vilkamo J, Vilkamo J, Lokki T, Pulkki V. Directional audio coding: Virtual microphone-based synthesis and subjective evaluation. J Audio Eng Soc. 2009; 57: 709–724.
10. Tervo S, Patynen J, Kuusinen A, Lokki T. Spatial decomposition method for room impulse responses. J Audio Eng Soc. 2013; 61: 17–28.
11. Omoto A, Ise S, Ikeda Y, Ueno K, Enomoto S, Kobayashi M. Sound field reproduction and sharing system based on the boundary surface control principle. Acoust Sci & Tech. 2015; 36: 1–11.
12. Kashiwazaki H, Omoto A. Sound field reproduction system using narrow directivity microphones and boundary surface control principle. Acoust Sci & Tech. 2018; 39: 295–304.
13. Sasaki Y, Nishiguchi T, Ono K. Development of multichannel single-unit microphone using shotgun microphone array. Proc ICA 2016;5-9 September 2016; Buenos Aires, Argentine 2016. paper ICA-2016-155.
14. Yokoyama S, Ueno K, Sakamoto S, Tachibana H. 6-channel recording/reproduction system for 3-dimensional auralization of sound fields. Acoust Sci & Tech. 2002; 23: 97– 103.
15. Meyer J, Eiko G. A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield. Proc. ICASSP 2002; 13-17 May; Orlando, Florida, USA 2002; p. 1781–4.
16. Kashiwazaki H, Omoto A. Attempt to improve the total performance of sound field reproduction system: Integration of wave-based methods and simple reproduction method. Proc ICA 2019; 9-13 September 2019; Aachen, Germany 2019.
17. Nakahara M, Omoto A, Nagatomo Y. Development of a 4pi sampling reverberator, VSVerb -Application to in-game sounds-. Proc AES Dublin 2019; 20-23 March; Dublin, Ireland 2019. EB04-7.