Audio Engineering Society
Convention Express Paper 204

Presented at the 156th Convention
2024 June 15-17, Madrid, Spain

This Express Paper was selected on the basis of a submitted synopsis that has been peer-reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This Express Paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib) all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Immersive recording using a virtual microphone array generated from an FOA response.

Masataka Nakahra[1,2], Toru Kamekawa[3], and Akira Omoto[4,1]

[1] *ONFUTURE Ltd, Tokyo 141-0032 Japan*
[2] *SONA Corporation, Tokyo 164-0013 Japan*
[3] *Tokyo University of the Arts, Tokyo 120-0034 Japan*
[4] *Kyushu University, Fukuoka 815-8540 Japan*

Correspondence should be addressed to Masataka Nakahara (nakahara@onftr.com)

ABSTRACT

We propose a cost-effective method to generate acoustic responses of arbitrary types of microphone arrays from a single set of FOA responses. The method generates a 4pi scene-based reverb of a target room by sampling acoustic information of virtual sound sources of the room. The virtual sound sources are detected from sound intensities calculated from impulse responses that are measured by an A-format microphone. We use an A-format microphone to obtain FOA responses, but our method does not use spherical-harmonics operation. Our method is based on geometrical acoustics, not based on Ambisonics. We call the 4pi scene-based reverb generated from the FOA responses "VSVerb." Once we generated the VSVerb, we can also recalculate spatial room impulse responses (SRIRs) at designated locations in the room with the specific directivity characteristics and a facing angle of the receiver. By generating multiple SRIRs at the virtual microphone positions, we can virtually obtain the acoustic response of any type of immersive microphone array. We call this virtual microphone array "V2MA" (VSVerb Virtual Microphone Array). In this paper, we present an overview of the method and some examples of the V2MAs generated from a single set of FOA responses measured at one position of the audience chair in a concert hall. In the last part of the paper, an immersive recoding technique for a string quartet using multiple V2MAs is also introduced.

1 Introduction

We have developed a measurement and analysis method to visualize the spatial property of a room acoustic response, including the invention of the method to detect virtual sound sources from a measured sound intensity. We call this method "VSV" (Virtual Source Visualizer) [1, 2].

Then, we started to develop the method to generate a 4pi scene-based sampling reverb by using the VSV method. We call the method and/or the generated 4pi scene-based reverb "VSVerb" [3-6].

Recently, by using the VSVerb method, we further develop a virtual recording technique by generating SRIRs captured by virtually created microphone arrays in a room [7]. We call this virtual microphone array "V2MA" (VSVerb Virtual Microphone Array). The V2MA enables us to create immersive audio content using any type of microphone array from one set of FOA responses measured by an A-format microphone in the target room.

2 Virtual Sound Sources

According to geometrical acoustics, reflection sounds in a room, (A) in Figure 1, are replaced with sounds from virtual sound sources appearing outside the room, (B) in Figure 1. By translating the spatial characteristics of the virtual sound sources into the time domain, we can obtain a SRIR, Spatial Room Impulse Response. The distances from the receiver

and the strengths of the virtual sound sources are converted to arrival times and amplitudes of reflections of a SRIR. Thus, the spatial characteristics of the virtual sound sources are a kind of fingerprint of the room reverberation.

3 Outline of the V2MA

Figure 2 is for showing the overview of the V2MA, the method for generating SRIRs of designated microphone arrays from a set of FOA responses measured in a room [7]. The flow is as follows.

1. Impulse responses are measured in a target room using an A-format microphone, and they are filtered into Low, Mid, and High frequency bands. Then, filtered responses are converted to FOA signals, W, X, Y, and Z.
2. The instantaneous active sound intensities, Ix≈W·X, Iy≈W·Y, and Iz≈W·Z, are calculated from the Hilbert-transformed FOA signals.
3. Virtual sound sources are detected from time-averaged instantaneous sound intensities [2], and their phase characteristics, either + or -, are also estimated [5,6]. Then, the spatial characteristics of the virtual sound sources at the original microphone position are obtained
4. Positions and strengths of the virtual sound sources are updated with reference to the new receiver positions, i.e., the microphone positions of the virtual microphone array, V2MA. The positions of the new virtual sources are obtained by moving the original sources in the opposite direction of the receiver's movement. The strengths are of the new virtual sources are obtained by changing original strengths in inverse proportion to the differences between the distances of the new and original sources to the receiver.
5. The strengths of the virtual sound sources are weighted according to the microphone's directivity such as follows.

Omni	1 (no changes)	(1)
Figure 8	$\cos(\theta+\theta_0)\cos(\varphi+\varphi_0)$	(2)
Cardioid	$0.500+0.500\cos(\theta+\theta_0)\cos(\varphi+\varphi_0)$	(3)
Super Card.	$0.375+0.625\cos(\theta+\theta_0)\cos(\varphi+\varphi_0)$	(4)
Hyper Card.	$0.250+0.750\cos(\theta+\theta_0)\cos(\varphi+\varphi_0)$	(5)

(θ and φ : elevation and azimuth angles of virtual sources; θ_0 and φ_0 : elevation and azimuth angles of the receiver)

6. Spatial properties of virtual sound sources are translated into time responses, and they are convolved with a source signal in each channel of the V2MA. An immersive audio content is created by balancing them.

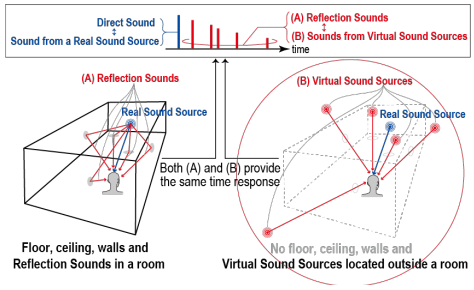


Figure 1. Reflection sounds and Virtual Sound Sources.

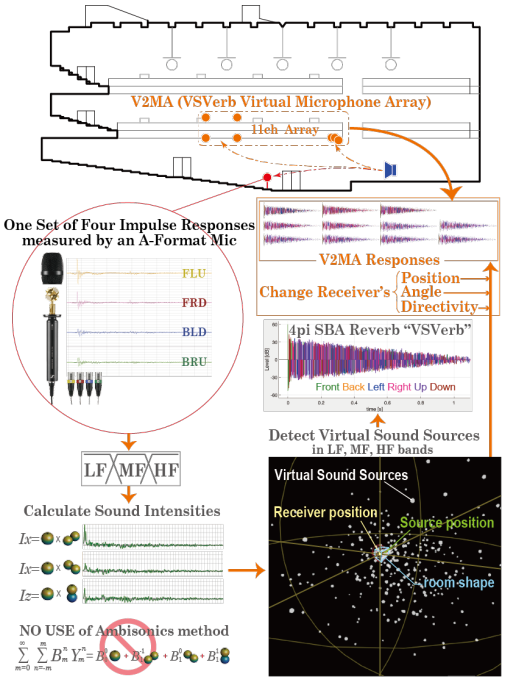


Figure 2. Overview of generating V2MA responses.

4 V2MA-1 : Detection of Virtual Sound Sources of a concert hall

Impulse responses were measured with an A-format microphone, Sennheiser's AMBEO VR Mic, at the Fukuoka Symphony Hall. The microphone was placed at the height of an audience chair and received a swept-sine signal provided by a loudspeaker, Genelec's 8351, on the stage (Figure 3).

Figure 4 is for showing the detected virtual sound sources of the hall in LF (≤350Hz), MF (250-2kHz) and HF (≥2kHz) bands. In the figure, the origin point of the x, y, and z axes indicates the receiver position, and the colored dots indicate the virtual sound sources. The distance between the dots and the origin point is proportional to the arrival time of reflection sounds, and the size of the dots indicates the strength of the reflection sounds. Since the figure shows the

virtual sound sources within about 500m from the origin point, the reflection sounds within about 1.5s are shown in the figure. Figure 4 shows that the virtual sound sources in LF are sparser than the those in MF. The figure also shows that the area where the virtual sound sources exist in HF is narrower than the area in MF. These results indicate that the number of reflection sounds becomes smaller in LF and the reverberation time becomes shorter in HF.

Figure 5 is shown the 4pi scene-based reverbs, VSVerbs generated from the detected virtual sound sources shown in Figure 4.

Note that, in the figures of this paper, six colors indicate six areas of arrival directions of reflection sounds as follows.

Green	Front	$-45° \le \theta \le 45°$	$-45° \le \varphi \le 45°$
Orange	Back	$-45° \le \theta \le 45°$	$135° \le \varphi \le 180°$ $-180° \le \varphi \le -135°$
Blue	Left	$-45° \le \theta \le 45°$	$45° \le \varphi \le 135°$
Pink	Right	$-45° \le \theta \le 45°$	$-135° \le \varphi \le -45°$
Purple	Up	$45° \le \theta \le 90°$	$-180° \le \varphi \le 180°$
Brown	Down	$-90° \le \theta \le -45°$	$-180° \le \varphi \le 180°$

(θ : elevation angle , φ : azimuth angle)

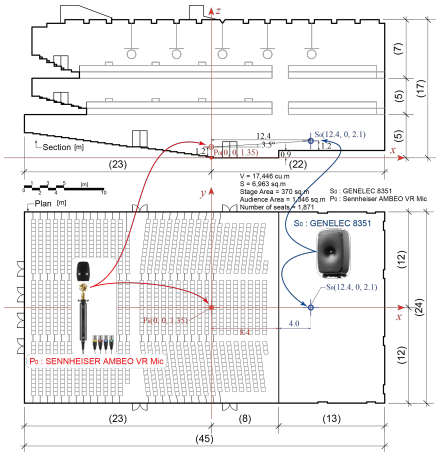


Figure 3. Measurement of FOA responses at Fukuoka Symphony Hall (P_0 : Receiver, S_0 : Source).

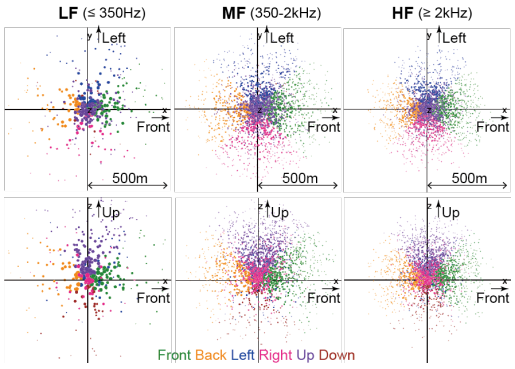


Figure 4. Virtual Sound Sources of the Fukuoka Symphony Hall in LF, MF, and HF (Receiver is at P_0).

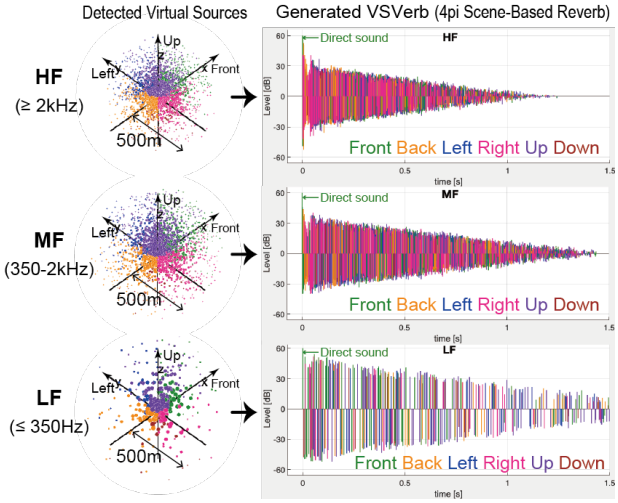


Figure 5. 4pi Scene-Based Reverb "VSVerb" at P_0 of the Fukuoka Symphony Hall in LF, MF, and HF.

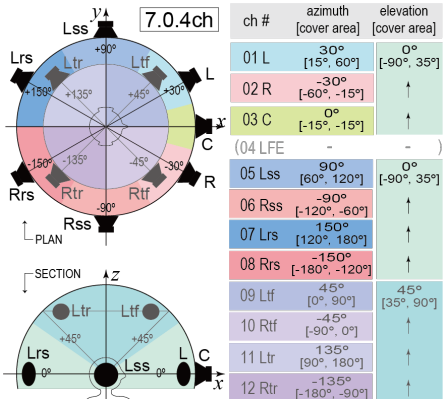


Figure 6. Target playback format, 7.0.4ch, and the loudspeaker placement.

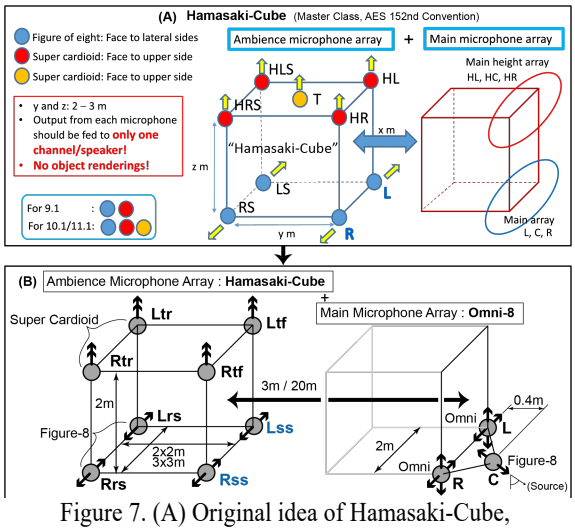


Figure 7. (A) Original idea of Hamasaki-Cube, and (B) 7.0.4ch V2MA consisting of the Omin-8 and Hamasaki-Cube microphone arrays.

5 V2MA-2 : Playback Configuration and Microphone Arrays

Figure 6 is shown the target playback configuration i.e., the loudspeaker placement, and Figure 7 is shown the microphone array of the V2MA we tested in this paper. We chose the 7.0.4ch playback and the 11ch microphone array consisting of the virtual Omni-8 and the Hamasaki-Cube for the use of main and ambience microphone arrays, respectively [8, 9].

6 V2MA-3 : Main Microphone Distance and recalculated Virtual Sources

To determine the main microphone position of the V2MA, i.e., the C position of the virtual Omni-8, a 2ch listening test was performed as a preliminary experiment.

Figure 8 is for showing the positions of the 2ch microphones for the listening test. We generated V2MA responses of the 2ch AB microphone arrays at six positions at the 4m heights, M_{10}, M_{12}, M_{14}, M_{16}, M_{18}, M_{20}, whose distances are 1.0, 1.2, 1.4, 1.6, 1.8, 2.0 times the critical distance of the hall.

Based on the authors' listening impressions of the balance between direct and reverberant sound, the position of M_{14}, at 1.4 times the critical distance of the hall, was chosen for the main microphone position.

"ALL" in Figure 9 is for showing the virtual sound sources and the VSVerb in MF (350-2kHz) at the main microphone position, that are recalculated from the original ones shown in Figure 5.

If we would like to restore the hall tone at the main microphone position using the 7.0.4ch playback environment, we should simply divide the virtual sound sources into 11 areas. Figure 9 is showing for the divided results in L, C, Lss, and Ltf areas. This is the original manner of the VSVerb, and is the correct method for the purpose of restoring the original sound field based on the theory of room acoustics. But we know that it does not always provide sufficient sound quality for the purpose of producing sound content. Science is important, but it does not always provide the best solution for engineering.

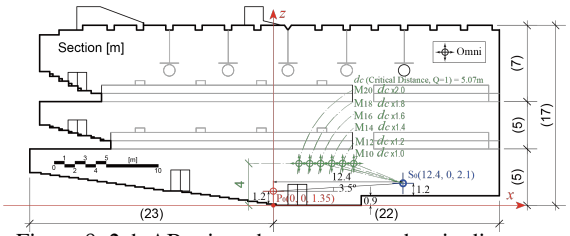


Figure 8. 2ch AB microphone arrays at the six distances from the source S_0; M_{10}, M_{12}, M_{14}, M_{16}, M_{18}, M_{20}.

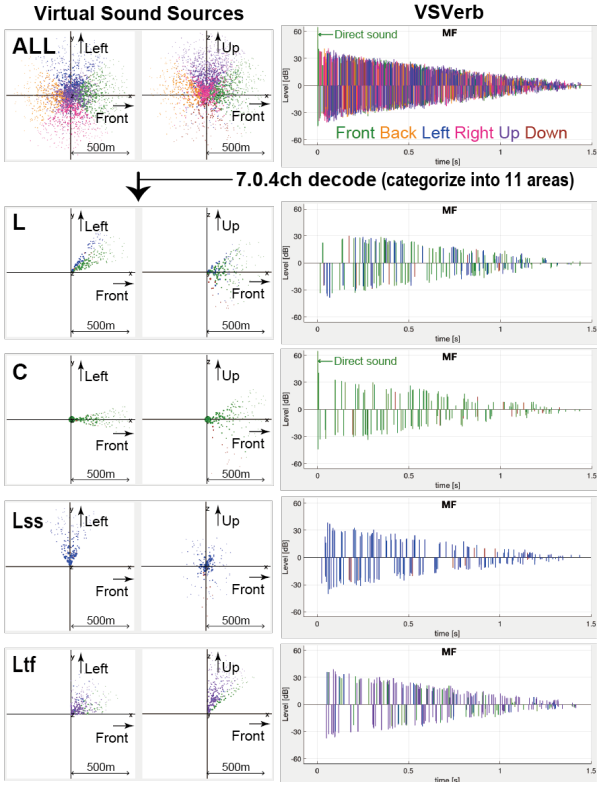


Figure 9. ALL : 4pi VSVerb at the main mic position.
L, C, Lss, Ltf : VSVerbs rendered into 7.0.4ch. (MF)

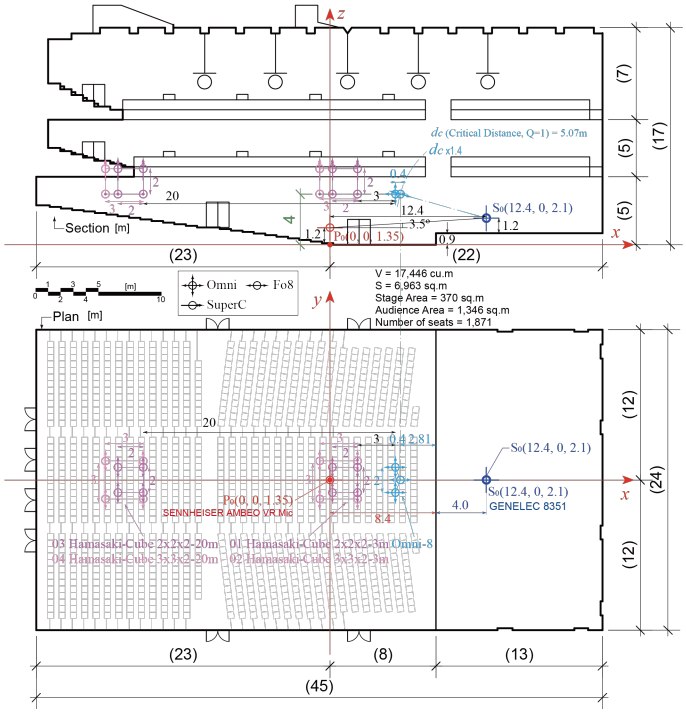


Figure 10. Microphone positions of the 11(7.0.4)ch V2MA consisting of the virtual Omni-8 and Hamasaki-Cube.

7 V2MA-4 : Generating 11ch outputs from Omni-8 + Hamasaki-Cube

Figure 10 is for showing 11 microphone positions of the 7.0.4ch V2MA consisting of the virtual Omni-8 and Hamasaki-Cube. In the figure, four types of Hamasaki-Cubes are shown to evaluate four responses of the V2MAs. They have two cube sizes and two distances from the main microphone array, Omini-8, as follows.

01 Cube size (WxDxH) = 2x2x2m, Distance from Omni-8 = 3m.

02 Cube size (WxDxH) = 3x3x2m, Distance from Omni-8 = 3m.

03 Cube size (WxDxH) = 2x2x2m, Distance from Omni-8 = 20m.

04 Cube size (WxDxH) = 3x3x2m, Distance from Omni-8 = 20m.

To obtain the 7.0.4ch V2MA response, 11 responses were calculated from the updated virtual sound sources at each microphone position.

We must note that the response of the V2MA can be calculated virtually at any position, even outside the room, but the further away from the original microphone position, the more acoustical errors (lies) the response contains. For example, the responses of the virtual Hamasaki-Cubes at the 20m positions can be calculated here without any problems, but the 03 and 04 of the sectional view in Figure 10 suggest that their height microphones cannot be placed in the real hall. It should be noted again that the generating responses far from the original microphone position is possible in terms of audio effects, but acoustically it is more likely to be wrong.

Figure 11 is for showing the calculated results of the responses in MF (350-2kHz) obtained by the main microphone array, the virtual Omni-8, that consists of two omni-directional and one figure of 8 microphones for the L, R, and C channels.

The responses show that the L and R of the omnidirectional microphones capture almost the same sound fields (virtual sound source distributions) including a strong direct sound, but they have slight differences in the time responses. We can know that the omnidirectional microphone array provides "decorrelated" same reverberation with strong direct sound.

On the other hand, the C of the figure of 8 microphone provides a different sound field (virtual sound source distribution), and the time response shows less reverberation, but not only from the front, but also from the back. We can know that the C of the Omni-8 microphone array provides a dry and clear direct sound.

Figure 12 is for showing the calculated results of the Lss responses in MF (350-2kHz) obtained by the ambience microphone array, the virtual Hamasaki-Cube that consists of four figure of 8 microphones for a low layer (Lss, Rss, Lrs, Rrs) and four super cardioid microphones for a high layer (Ltf, Rtf, Ltr, Rtr).

Since the figure of 8 microphones of the Hamasaki-Cube are placed facing to the lateral side (left and right wall direction), we know that the microphones of the low layer avoid capturing strong direct sound. Especially at the 20m position, Lss seems to provide only the reverberation sound (Figure 12).

On the other hand, the high channels of the Hamasaki-Cube provide much more direct sound than the low layer even at the 20m position. Figure 13 is for showing the calculated results of the Ltf responses in MF (350-2kHz). Since the super cardioid microphones are used for the high layer, the direct sound cannot be removed even though the microphones are placed facing to the upward.

Since the Hamasaki-Cube requires no omnidirectional microphones, each channel provides different reverberant field.

According to the calculated results of the V2MA responses of the virtual Hamasaki-Cube, we can know that the Hamasaki-Cube provides high "decorrelated" sounds between channels, especially when they are placed far position from the source.

It is ascertained that theV2MA is useful not only for content production but also for learning acoustic features of microphone arrays.

8 Hamasaki-Cube and Direct Sound

Figure 14 is for showing the off-axis azimuth angles between the source direction and the null directions of the low layer's figure of 8 microphones of the virtual Hamasaki-Cube examined in the previous chapter. We can see that the closer the microphones move to the source position, the larger their off-axis angles become. This is probably one of the main reasons why the low layer of the virtual Hamasaki-Cube provides much more direct sound at 3m than at the 20m (Figure 12).

In order to further reduce the direct sound from the low layer of the virtual Hamasaki-Cube, we tried to place the figure of 8 microphones of the low layer of the Hamasaki-Cube so that make their null axes point to the source direction. Figure 15 is shown that the angled placements of the low layer microphones of the virtual 2x2x2m and 3x3x2m Hamasaki-Cubes at the 3m position. The calculated results are also shown in Figure 16. They show that the angled placement provides a little reduction of the direct sound. In terms of direct sound reduction, we can know that the far placement is much more effective than the angled placement.

It is ascertained that the V2MA is useful not only for content production but also for trial-and-error testing of a microphone array prior to field assembly.

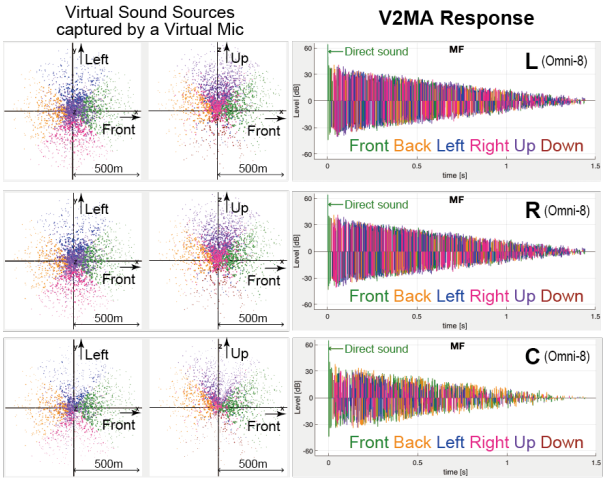


Figure 11. 7.0.4ch V2MA responses.
L, C, and R channels of the virtual Omni-8 in MF.

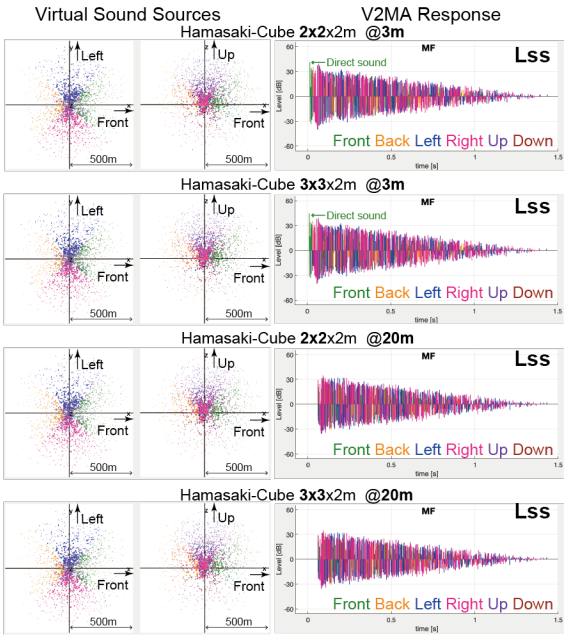


Figure 12. 7.0.4ch V2MA responses.
Lss channel of the virtual Hamasaki-Cube in MF.
(a) 2x2x2m and (b) 3x3x2m Cubes at 3m distance.
(c) 2x2x2m and (d) 3x3x2m Cubes at 20m distance.

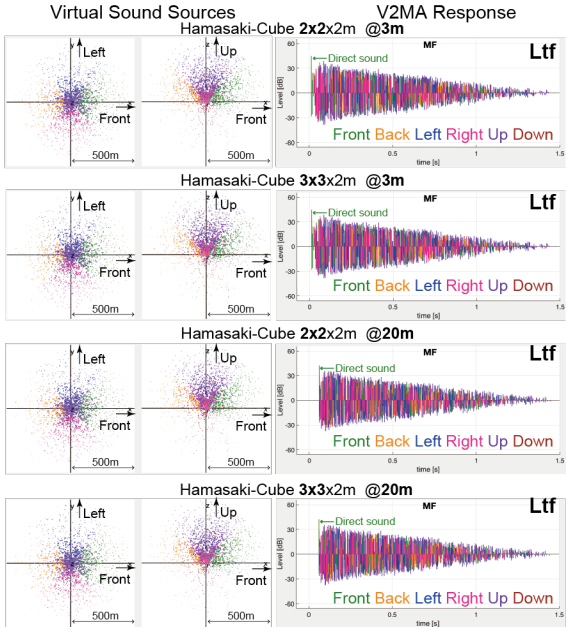
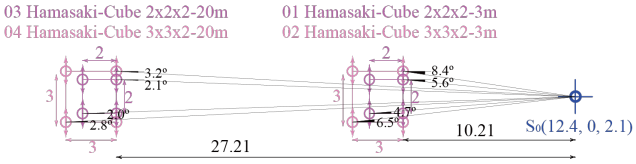


Figure 13. 7.0.4ch V2MA responses.
Ltf channel of the virtual Hamasaki-Cube in MF.
(a) 2x2x2m and (b) 3x3x2m Cubes at 3m distances.
(c) 2x2x2m and (d) 3x3x2m Cubes at 20m distances.



off-axis azimuth angles from the null directions of Figure-8

	Lss / Rss	Lrs / Rrs
01 Hamasaki-Cube 2x2x2-3m	±5.6°	±4.7°
02 Hamasaki-Cube 3x3x2-3m	±8.4°	±6.5°
03 Hamasaki-Cube 2x2x2-20m	±2.1°	±2.0°
04 Hamasaki-Cube 3x3x2-20m	±3.2°	±2.8°

Figure 14. Off-axis azimuth angles between the source direction and the null directions of the figure of eight microphones of the low layer of virtual Hamasaki-Cubes

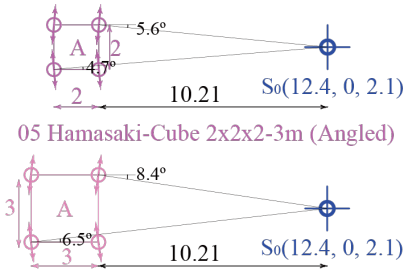


Figure 15. “Angled” virtual Hamasaki-Cubes whose sizes and distances are 2x2x2m-@3m and 3x3x2m-@3m. Null directions of the low layers are faced to the source direction to reduce the direct sound.

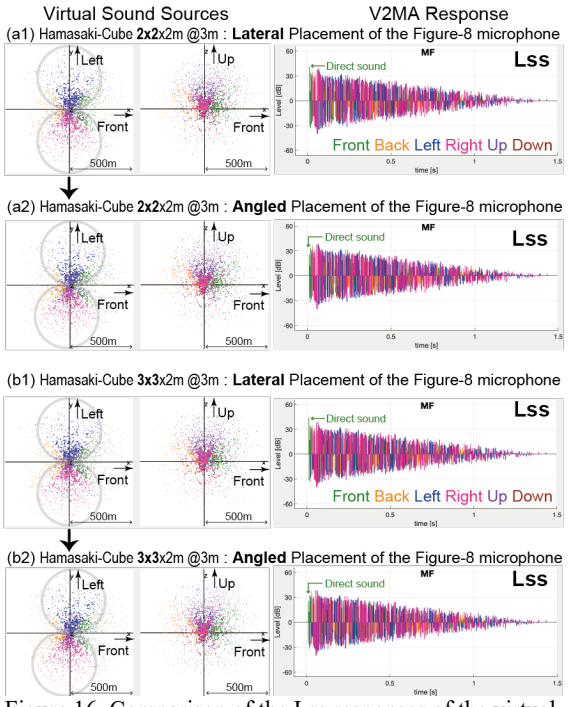


Figure 16. Comparison of the Lss responses of the virtual Hamasaki-Cubes between "lateral" and "angled" placements of low layer's figure of eight microphones.

9 Listening impression

We evaluated to listen to the immersive sound of the following six 7.0.4ch V2MAs by convolving a dry source of a solo violin performance. As a result, 01 and 04 received high scores. 02 provided a somewhat bright sound for a violin solo, and 03 showed a slight imbalance of spaciousness and reverberation. 05 and 06 did not provide the null effect as much as we expected.

01 Omni-8 + 2x2x2m Hamasaki-Cube at 3m (Figure 10)

02 Omni-8 + 3x3x2m Hamasaki-Cube at 3m (Figure 10)

03 Omni-8 + 2x2x2m Hamasaki-Cube at 20m (Figure 10)

04 Omni-8 + 3x3x2m Hamasaki-Cube at 20m (Figure 10)

03 Omni-8 + 2x2x2m angled Hamasaki-Cube at 3m (Figure 15)

04 Omni-8 + 3x3x2m angled Hamasaki-Cube at 3m (Figure 15)

10 V2MA for a String Quartet

We also tried to listen to the V2MA shown in the previous chapter by convolving a string quartet performance, but the direct sound showed an unnatural sound image. Since the V2MA response was generated by the measured IRs using a point source, a single loudspeaker placed at the center position of the concert hall stage, the string quartet was reproduced as a small point source.

To improve the unnaturalness of a string quartet reproduction, we should use four kinds of V2MA responses that are generated from four FOA responses using four loudspeaker positions, Vn1, Vn2, Va, and Vc instead of the S0, as shown in Figure 17. However, unfortunately, this time we only have one set of the FOA responses using the S0 source position, as shown in Figure 3. So, we tried another idea to generate the four V2MA responses for the four instruments of a string quartet.

Figure 18 is for showing the four different positions of the same V2MA consisting of the virtual Omni-8 and the Hamasaki-Cube. In the figure, each V2MA moves in the opposite direction from S0 to V1n1, Vn2, Va, and Vc shown in Figure 17. Although the sound fields captured by the four V2MAs in Figure 18 are not exactly the same as those shown in Figure 17, the L, R, and C channels are expected to provide the same phantom sound images of the Vn1, Vn2, Va, and Vc as those captured in Figure 17.

Figure 19 is for showing the calculated results of the early part of the main microphone responses, i.e., L, R, and C responses of the virtual Omni-8. Since the distances between the main microphones and the source S0 are from 7.4m to 8.7m, the level differences of the direct sound among L, R, and C are small values, less than 0.6dB. On the other hand, the differences of the arrival times of the direct sound among L, R, and C are from 0.8ms to 1.5ms, and they are expected to affect the left-right localization of the direct sound.

According to the listening test by the authors, stereophonic localization of the string quartet was found to be improved.

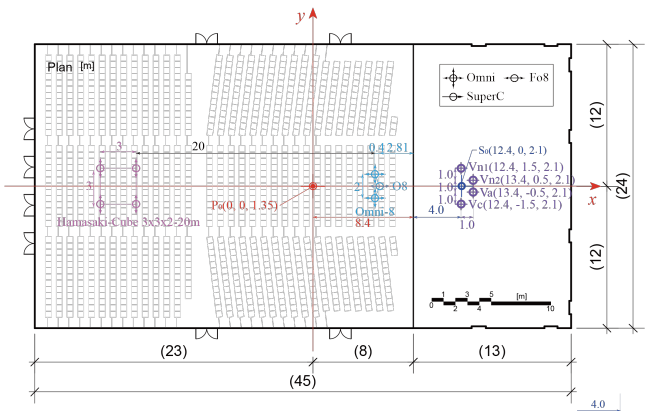


Figure 17. Target recording field for a string quartet. One set of the V2MA (Omni-8+Hamsaki-Cube) and the four sources, Vn1, Vn2, Va, and Vc, are assumed.

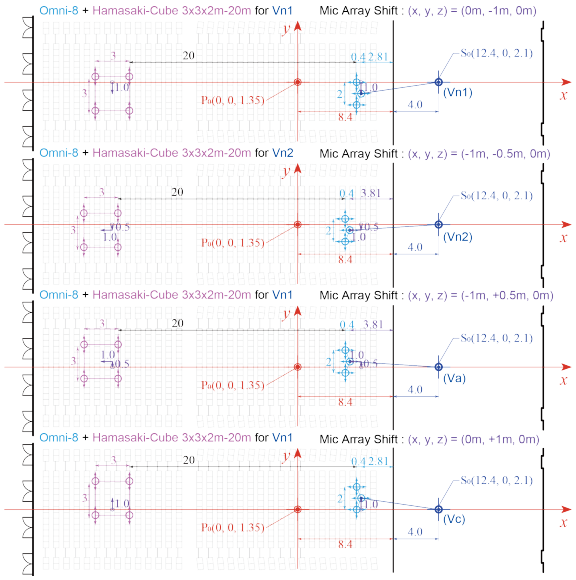


Figure 18. Strategy for adapting the "single source" V2MA response to the "four sources" response.

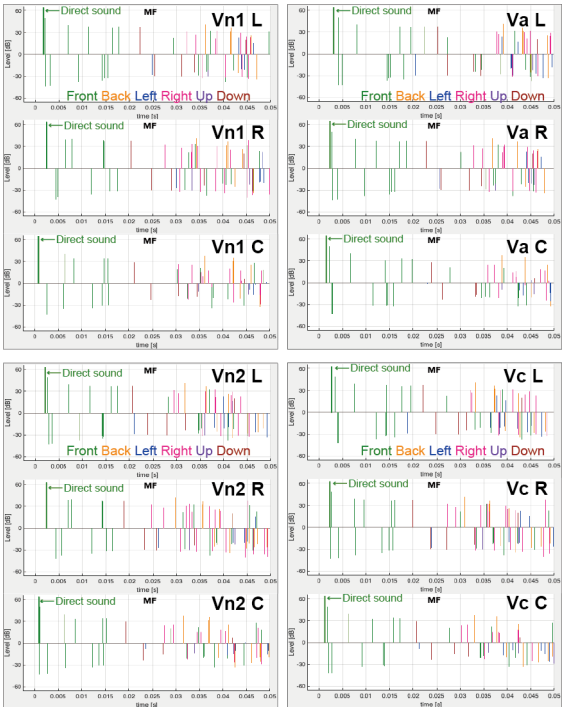


Figure 19. Four sets of the V2MA responses of L/C/R channels in MF for four sources, Vn1, Vn2, Va, and Vc.

11 Conclusions

This paper has proposed the method, V2MA, that generates SRIRs captured by a virtual microphone array from a single set of FOA responses, measured in the target space. From the listening impressions by the authors, the V2MA provides sufficient sound quality and spaciousness of the designated virtual microphone array. The V2MA is useful not only for sound production, but also for education of the miking methods and for the trial-and-error testing of a microphone array prior to field assembly.

The authors plan to continue studying the physical differences in the response between the V2MA and the real microphone arrays.

Acknowledgement

The authors would like to thank Kimio Hamasaki, who provided detailed information about the Hamasaki-Cube and helped with the evaluation of the generated V2MA responses.

References

[1] M. Nakahara, A. Omoto, and Y. Nagatomo, "VSV (Virtual Source Visualizer), a practical tool for 3D-visualizing acoustical properties of spatial sounds," eB, AES 140th Convention, Paris (2016)

[2] M. Nakahara, A. Omoto, and Y. Nagatomo, "Accurate extraction of dominant reflections from measured sound intensity responses in a room," eB, AES 142th Convention, Berlin (2017)

[3] M. Nakahara, A. Omoto, and Y. Nagatomo, "Development of a 4-pi sampling reverberator, VSVerb. -preliminary experiments," eB, AES 144th Convention, Milan (2018)

[4] M. Nakahara, A. Omoto, and Y. Nagatomo,, "Development of a 4-pi sampling reverberator, VSVerb. -Application to in-game sounds," eB, AES 146th Convention, Dublin (2018)

[5] M. Nakahara, Y. Nagatomo, and A. Omoto, "Development of a 4-pi sampling reverberator, VSVerb. -Phase Detection," eB, AES 148th Convention, Virtual Vienna (2020)

[6] M. Nakahara, Y. Nagatomo and A. Omoto, "VSVerb, a practical method of reproducing spatial reverberation from virtual sound sources captured in a room," 24th ICA, Gyeongju (2022)

[7] M. Nakahara, A. Omoto, and Y. Nagatomo, "Virtual 3D microphone arrays by virtual sound sources detected from an FOA response," Proc, AES International Conference on Spatial & Immersive Audio, Huddersfield (2023)

[8] T. Kamekawa, "Impression Differences by Placement of Front and Rear Microphones for Multichannel Stereo Recording," Proc, AES 118th Convention, Barcelona (2005)

[9] K. Hamasaki, "Microphone Techniques for Natural Music Recording," Master Class, AES 152nd Convention, Hague / Online (2022)