Audio Engineering Society

Conference Paper 46

Presented at the International Conference on Spatial & Immersive Audio,

2023 August 23-25, Huddersfield, UK

Virtual 3D microphone arrays
by virtual sound sources detected from an FOA response.

Masataka Nakahara[1,2], Yasuhiko Nagatomo[3], and Akira Omoto[4,1]

[1] ONFUTURE Ltd, Tokyo 141-0032 Japan
[2] SONA Corporation, Tokyo 164-0013 Japan
[3] Evixar Inc, Tokyo 104-0033 Japan
[4] Kyushu University, Fukuoka 815-8540 Japan

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

ABSTRACT

A microphone array, also known as a miking, is one of the most important techniques for sound production, and various types of arrays are used for practical recording work. Since the miking strategies are closely related to the spatial impression of the sound content, the sizes of the microphone arrays have become larger especially for recent 3D immersive sound productions. We propose the method to reduce the labor to obtain the spatial responses of the large-sized microphone arrays. Our method requires only one A-format microphone instead of many numbers of microphones that are usually placed at different positions and angles, and generates responses of various types of microphone arrays from the recorded single FOA response, W, X, Y, and Z. The A-format microphone is used as a sound intensity probe to obtain three orthogonal sound intensities from measured W, X, Y, and Z impulse responses. According to the geometrical acoustics, the spatial and temporal characteristics of room reverberation can be defined by the virtual sound sources that are located outside the room. We detect them from the obtained sound intensities calculated from the FOA responses, and generate a reverberation, that we call VSVerb. Since the VSVerb is a scene-based 4-pi reverberation generated from the detected virtual sound sources of the target room, it can be decoded into any type of microphone response. This paper shows the generated examples of 48 virtual microphone responses that are used for five types of 3D microphone arrays, OCT-3D, PCMA-3D, 2L-Cube, Decca Cuboid and Hamasaki Cube H=0[m]/1[m]. The responses are calculated from a single FOA response. In the last part, the comparison of impulse responses between virtual and real responses is also shown.

1 Introduction

The demand for immersive streaming audio content of live performances has grown in recent years. However, there are fewer opportunities to set up large-sized microphone arrays at the venue. To solve this problem, we propose a method to generate virtual responses of various types of 3D microphone arrays from a single FOA response measured by an A-format microphone. We implement this using the VSVerb technique. VSVerb (Virtual Sources' reVerb) is the 4-pi scene-based sampling reverb technique that was developed based on our previous research VSV (Virtual Source Visualizer), that detects virtual sound sources from measured sound intensities [1-8].

2 Virtual Sound Sources

We tend to understand a reflection sound as a sound particle coming along the red path shown in (a) of Figure 1. However, this is a way of understanding auditory phenomena through a visual information. If we have no visual information, for example, if we close our eyes, we will understand that a reflection sound as a delayed sound coming from a sound source located outside of a room, red circles in (b) of Figure 1. Since we are surrounded by many pieces of walls in a room, our ears see many sound sources that are located outside the room instead of seeing walls with our eyes. Figure 2 shows an example of detected sound sources located outside a room.

The sound sources represent dominant reflections of a room. In geometrical acoustics, they are called Virtual Sound Sources. If we can detect all virtual sound sources of a room, we can obtain complete acoustic information about the reverberation of the room.

3 VSVerb method

The VSVerb is a method to restore a 4-pi scene-based reverberation of a target room from virtual sound sources detected in sound intensities that are calculated from impulse responses measured by an A-format microphone in a target room [1-8].

Although an A-format microphone is used for the measurement, no spherical harmonic technique is applied. Since the spatial resolution of FOA is too low to restore a fine 4-pi reverberation, we use a sound intensity technique instead of Ambisonics processing. We extract acoustic information of virtual sound sources from the obtained sound intensities. However, they do not have frequency-oriented information, because a virtual sound source is a theoretical idea that comes from geometrical acoustics. To add frequency characteristics to the reverberation, we analyze the sound intensities in low, mid, and high frequency bands.

The processing flow of the VSVerb is as follows;

1. Impulse responses are measured in a target room by using an A-format microphone. The microphone is used as a sound intensity probe, not as an Ambisonics tool.
2. Impulse responses are filtered by BPFs and divided into Low, Mid and High frequency bands.
3. Filtered impulse responses are converted to the B-Format signals, W, X, Y and Z, and Hilbert transforms are operated.
4. The instantaneous active sound intensities, Ix≈WX, Iy≈WY and Iz≈WZ, are calculated in Low, Mid and High bands. Then a time averaging operation is performed using an appropriate width of a time window.

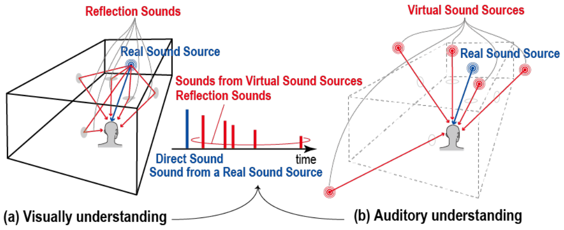

Reflection Sounds
Real Sound Source
Sounds from Virtual Sound Sources
Reflection Sounds
Direct Sound
time
Sound from a Real Sound Source
Virtual Sound Sources
Real Sound Source
(a) Visually understanding
(b) Auditory understanding

Figure 1. Understanding reflection sounds through (a) visual information or (b) auditory information.

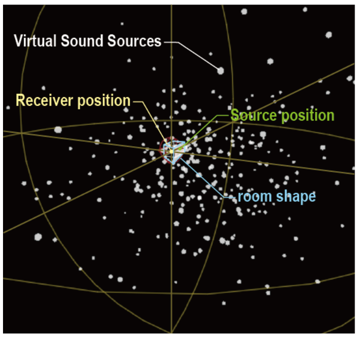


Figure 2. A measurement example of virtual sound sources generated by the walls of a room.

5. Virtual sound sources, i.e. dominant reflections, are detected from averaged sound intensities using the "Speed Detection" method [4,8].
6. Phase characteristics of virtual sound sources, either + or -, are estimated for detected virtual sound sources [7,8].
7. 4-pi reverberations in Low, Mid and High bands are obtained as spatial information of virtual sound sources.
8. Spatial properties of sound sources are translated into time domain, then the 4-pi scene-based reverberation, VSVerb, is generated in Low, Mid and High bands.
9. The VSVerb is decoded into a specific playback channel format. Since each reflection of the VSVerb knows its arrival direction, all reflections can be devided into their proper channel areas.

Since the VSVerb is a 4-pi scene-based reverb, various kinds of post-processing can be applied [6,9].

The positions of virtual sound sources remain the same unless the real source position is moved. Moving around in a room while listening to reverberant sound is the same action as moving around in the world surrounded by fixed virtual sound sources with listening to the difference of the reverberation timbre that caused by the changes in the

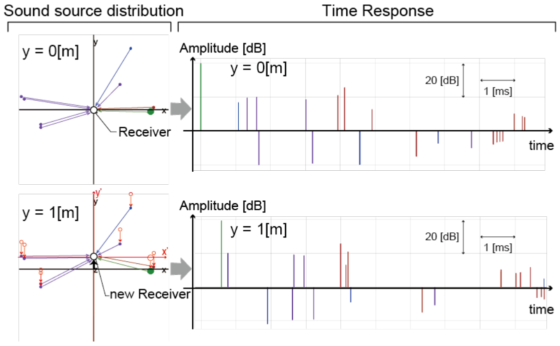


Figure 3. VSVerb's post-processing; Moving a receiver position.

relationships between positions of virtual sound sources and a receiver. Using this logic, we can generate a new reverberation at the other receiver position by using the same reverb resources. The VSVerb generates new reverberation at the different receiver position by moving virtual sound sources in the opposite direction to the movement of the receiver (Figure 3).

The VSVerb's receiver is designed to be omni-directional from a room acoustics point of view. However, if you desired, its directivity can also be changed by multiplying the strengths of the virtual sound sources by the directivity weights, the absolute values of (1)-(4). Using this technique, VSVerb can generate different types of microphone responses placed at the receiver position (Figure 4).

Figure 8	$\cos(\theta+\theta_0)\cos(\varphi+\varphi_0)$	(1)
Cardioid	$0.500+0.500\cos(\theta+\theta_0)\cos(\varphi+\varphi_0)$	(2)
Super Cardioid	$0.375+0.625\cos(\theta+\theta_0)\cos(\varphi+\varphi_0)$	(3)
Hyper Cardioid	$0.250+0.750\cos(\theta+\theta_0)\cos(\varphi+\varphi_0)$	(4)

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

Using the techniques described above, we can generate the immersive responses of various types of microphone arrays from a single FOA response.

4 Venue, Mic Arrays and Playback Format

A total of 48 microphone responses from five types of 3D microphone arrays were tried to be generated virtually from a single FOA response. The physical parameters of the 3D microphone arrays, and the playback formats were used from the previous work of Hyunkook Lee of APL, University of Huddersfield [9,10]. Figures 5 to 7 are shown the details of the venue (St. Paul's Concert Hall), the 3D microphone arrays (OCT-3D, PCMA-3D, 2L-Cube, Decca Cuboid and Hamasaki Cube H=0[m]/1[m]), and the playback format (5.0.4ch) of their work. Their parameters are also summarized in Table 1.

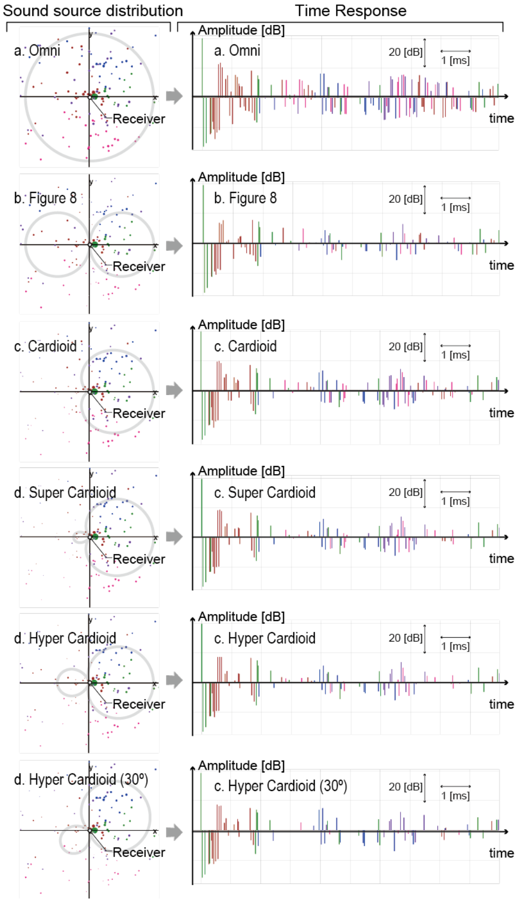


Figure 4. VSVerb's post-processing; Changing the directivity of a receiver.

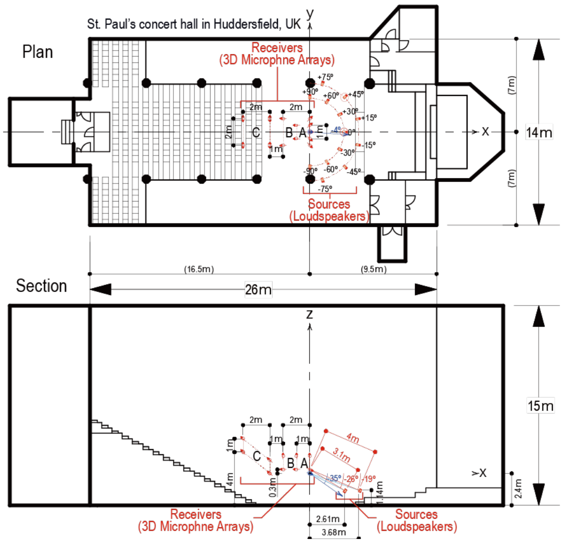


Figure 5. The venue (St. Paul's Concert Hall), the sources positions (loudspeakers) , and the receiver positions (microphone arrays).

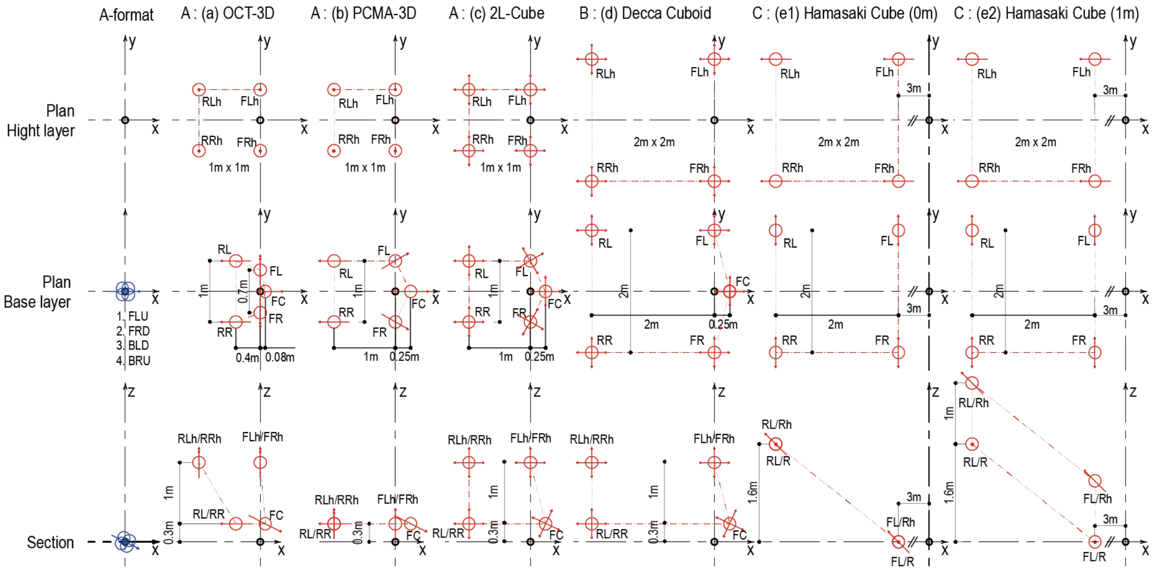
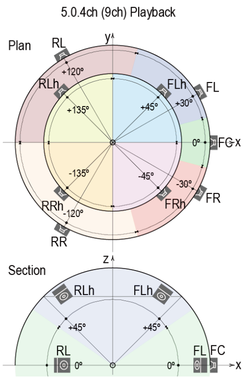


Figure 6. Located positions of the A-format microphone and the microphones of five 3D microphone arrays.

		Recording						Playback	
A-format (AMBEO) (x,y,z) =(0m,0m,0m) (azm,elv) = (-4°,-35°)		A : 1m x 1m			B : 2m x 2m	C : 2m x 2m		5.0.4ch	
		(a) OCT-3D	(b) PCMA-3D	(c) 2L-Cube	(d) Decca Cuboid	(e1) Hamasaki Cube hight 0m	(e2) Hamasaki Cube hight 1m		
1	FL (x,y,z) (azm,elv)	supercardioid (4018) (0m, +0.35m, +0.3m) (+90°, 0°)	cardioid (4011) (0m, +0.5m, +0.3m) (+30°, -25°)	omni (4006) (0m, +0.5m, +0.3m) (+30°, -25°)	omni (4006) (0m, +1m, +0.3m) (+30°, -25°)	figure-8 (CCM8) (-3m, +1m, 0m) (+90°, 0°)	← ← ←	elv, azm coverage : azm coverage : elv	0° , +30° (+15° , +75°] [-90° , +35°]
2	FR (x,y,z) (azm,elv)	supercardioid (4018) (0m, -0.35m, +0.3m) (-90°, 0°)	cardioid (4011) (0m, -0.5m, +0.3m) (-30°, -25°)	omni (4006) (0m, -0.5m, +0.3m) (-30°, -25°)	omni (4006) (0m, -1m, +0.3m) (-30°, -25°)	figure-8 (CCM8) (-3m, -1m, 0m) (-90°, 0°)	← ← ←	elv, azm coverage : azm coverage : elv	0° , -30° [-75° . -15°) [-90° , +35°]
3	FC (x,y,z) (azm,elv)	cardioid (4011) (+0.08m, 0m, +0.3m) (0°, -25°)	cardioid (4011) (+0.25m, 0m, +0.3m) (0°, -25°)	omni (4006) (+0.25m, 0m, +0.3m) (0°, -25°)	omni (4006) (+0.25m, 0m, +0.3m) (0°, -25°)	- - -	- - -	elv, azm coverage : azm coverage : elv	0° , 0° [-15°, +15°] [-90° , +35°]
4	(LFE)	-	-	-	-	-	-	-	-
5	RL (x,y,z) (azm,elv)	cardioid (4011) (-0.4m, +0.5m, +0.3m) (180°, 0°)	cardioid (4011) (-1m, +0.5m, +0.3m) (180°, 0°)	omni (4006) (-1m, +0.5m, +0.3m) (180°, 0°)	omni (4006) (-2m, +1m, +0.3m) (180°, 0°)	figure-8 (CCM8) (-5m, +1m, +1.6m) (+90°, 0°)	← ← ←	elv, azm coverage : azm coverage : elv	0° , +120° (+75° , +180°] [-90° , +35°]
6	RR (x,y,z) (azm,elv)	cardioid (4011) (-0.4m, -0.5m, +0.3m) (180°, 0°)	cardioid (4011) (-1m, -0.5m, +0.3m) (180°, 0°)	omni (4006) (-1m, -0.5m, +0.3m) (180°, 0°)	omni (4006) (-2m, -1m, +0.3m) (180°, 0°)	figure-8 (CCM8) (-5m, -1m, +1.6m) (-90°, 0°)	← ← ←	elv, azm coverage : azm coverage : elv	0° , -120° (-180° , -75°) [-90° , +35°]
7	FLh (x,y,z) (azm,elv)	supercardioid (4018) (0m, +0.5m, +1.3m) (0°, +90°)	supercardioid (4018) (0m, +0.5m, +0.3m) (0°, +90°)	omni (4006) (0m, +0.5m, +1.3m) (0°, +90°)	omni (4006) (0m, +1m, +1.3m) (0°, +90°)	cardioid (4011) (-3m, +1m, 0m) (180°, +45°)	← (0m, +1m, +1m) ←	elv, azm coverage : azm coverage : elv	+45° , +45° [0° , +90°] (+35°, +90°]
8	FRh (x,y,z) (azm,elv)	supercardioid (4018) (0m, -0.5m, +1.3m) (0°, +90°)	supercardioid (4018) (0m, -0.5m, +0.3m) (0°, +90°)	omni (4006) (0m, -0.5m, +1.3m) (0°, +90°)	omni (4006) (0m, -1m, +1.3m) (0°, +90°)	cardioid (4011) (-3m, -1m, 0m) (180°, +45°)	← (0m, -1m, +1m) ←	elv, azm coverage : azm coverage : elv	+45° , -45° [-90° , 0°) (+35°, +90°]
9	RLh (x,y,z) (azm,elv)	supercardioid (4018) (-1m, +0.5m, +1.3m) (0°, +90°)	supercardioid (4018) (-1m, +0.5m, +0.3m) (0°, +90°)	omni (4006) (-1m, +0.5m, +1.3m) (0°, +90°)	omni (4006) (-2m, +1m, +1.3m) (0°, +90°)	cardioid (4011) (-5m, +1m, +1.6m) (180°, +45°)	← (-2m, +1m, +2.6m) ←	elv, azm coverage : azm coverage : elv	+45° , +135° (+90° , +180] (+35°, +90°]
10	RRh (x,y,z) (azm,elv)	supercardioid (4018) (-1m, -0.5m, +1.3m) (0°, +90°)	supercardioid (4018) (-1m, -0.5m, +0.3m) (0°, +90°)	omni (4006) (-1m, -0.5m, +1.3m) (0°, +90°)	omni (4006) (-2m, -1m, +1.3m) (0°, +90°)	cardioid (4011) (-5m, -1m, +1.6m) (180°, +45°)	← (-2m, -1m, +2.6m) ←	elv, azm coverage : azm coverage : elv	+45° , -135° (-180° , -90°) (+35°, +90°]

Table 1. Located positions and angels of recording microphones and playback loudspeakers (5.0.4ch).



ch #	Azimuth [cover area]	Elevation [cover area]
01 FL	30° [15°, 75°]	0° [-90°, 35°]
02 FR	-30° [-75°, -15°]	↑
03 FC	0° [-15°, 15°]	↑
(04 LFE)	-	-
05 RL	120° [75°, 180°]	↑
06 RR	-120° [-180°, -75°]	↑
07 FLh	45° [0°, 90°]	45° [35°, 90°]
08 FRh	-45° [-90°, 0°]	↑
09 RLh	135° [90°, 180°]	↑
10 RRh	-135° [-180°, -90°]	↑

Figure 7 The assumed playback environment; 5.0.4.

5 Generating VSVerb

In the first step, virtual sound sources were detected and their time responses in three frequency bands were generated from the four impulse responses (FLU, FRD, BLD, and BRU) measured by an A-format microphone. We used the FOA responses available from the APL open resources on the University of Huddersfield website [11]. The APL provides 13 types of FOA responses from the 13 loudspeaker positions (Figure 5), and we chose the response where the loudspeaker is located at the 0° position.

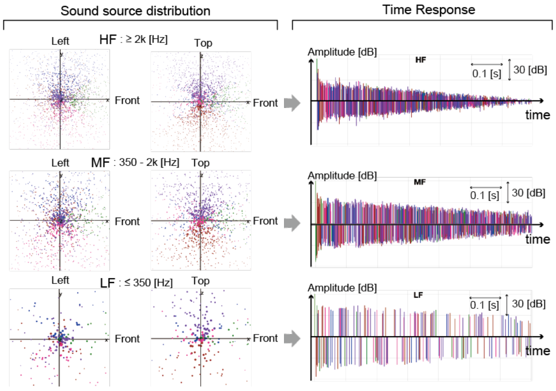


Figure 8. Detected virtual sources and generated VSVerbs in low, mid and high frequency bands.

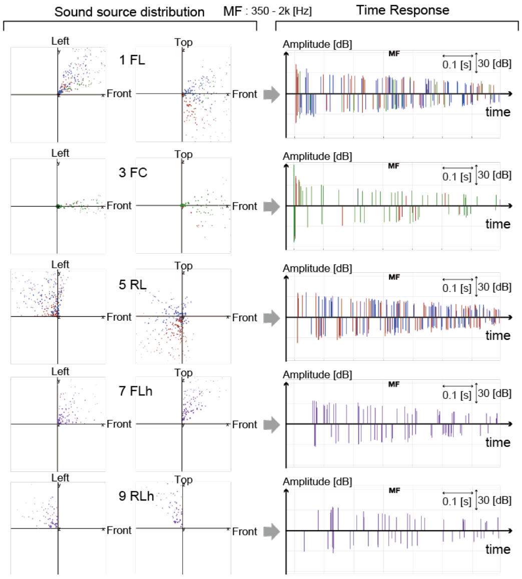


Figure 9. 5.0.4 decoded VSVerb; FL, FC, RL, FLh, RLh

Figure 8 is shown the results of the obtained virtual sound sources and their time responses when the source is located at the center position (0º) of the St. Paul’s Concert Hall. The colors in the figure indicate the six categories of arrival directions of the reflection sounds as follows;

Green	Front	$-30^\circ \le \theta \le 30^\circ$	$-30^\circ \le \varphi \le 30^\circ$
Orange	Back	$-30\,^\circ\le \theta \le 30^\circ$	$150^\circ\le \varphi \le 180^\circ$ $-180^\circ\le \varphi \le -150^\circ$
Blue	Left	$-30^\circ \le \theta \le 30^\circ$	$30^\circ \le \varphi \le 150^\circ$
Pink	Right	$-30^\circ \le \theta \le 30^\circ$	$-150^\circ \le \varphi \le -30^\circ$
Purple	Up	$30^\circ \le \theta \le 90^\circ$	$-180^\circ \le \varphi \le 180^\circ$
Brown	Down	$-90^\circ \le \theta \le -30\,^\circ$	$-180^\circ \le \varphi \le 1\ 80^\circ$

(θ : elevation angle , φ : azimuth angle)

Figure 9 is shown the FL, FC, RL, FLh, RLh channels of 5.0.4ch decoded VSVerbs. For the playback, the obtained reverberation must be divided into proper coverage areas provided by the playback channels without overlap, from a sound field restoration point of view.

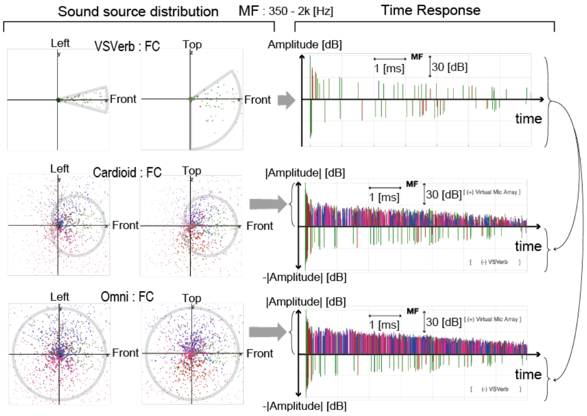


Figure 10. Comparison of reflection sounds (FC)

6 Virtual Array Responses

Figure 10 is shown three types of reflection sounds provided by the FC channel. It can be seen that the cardioid and the omni-directional receivers provide a larger number of reflection sounds coming from non-front directions, while the VSVerb provides reflection sounds only from the front direction. The microphone arrays, also known as miking techniques, has been using for practical recording work, and they are considered to provide such a rich and dense reverberation than an original sound field. The VSVerb method may be theoretically correct, but from an audio engineering point of the view, some acoustic decoration by the overlapped reverberation could be considered preferable for the sound content production.

Taking into account the positions and directivities of the microphones, we generated virtual microphone responses of five 3D microphone arrays from the single VSVerb data.

The responses of FL, FC, RL, FLh, and RLh channels in mid frequency band (350-2k[Hz]), and their virtual source distributions are shown in Figures 11 to 15; OCT-3D, PCMA-3D, 2L-Cube, Decca Cuboid, and Hamasaki Cube.

We can see that all the responses contain rich reflection sounds coming from various directions other than the target channel areas, and we can also find their differences as follows.

Figures 11 and 12 show that the OCT-3D and PCMA-3D provide sharp reverberation shapes from cardioid microphones.

On the other hand, Figures 13 and 14 show that the 2L-Cube and Decca Cuboid provide rich reverberation from omni-directional microphones.

Finally, Figure 15 shows that the Hamasaki Cube, for the use of ambience recording, provide small-head reverberation from the side-facing figure 8 microphones.

Although all the virtual responses were generated from the same VSVerb, they can express well the acoustical differences of different microphone arrays.

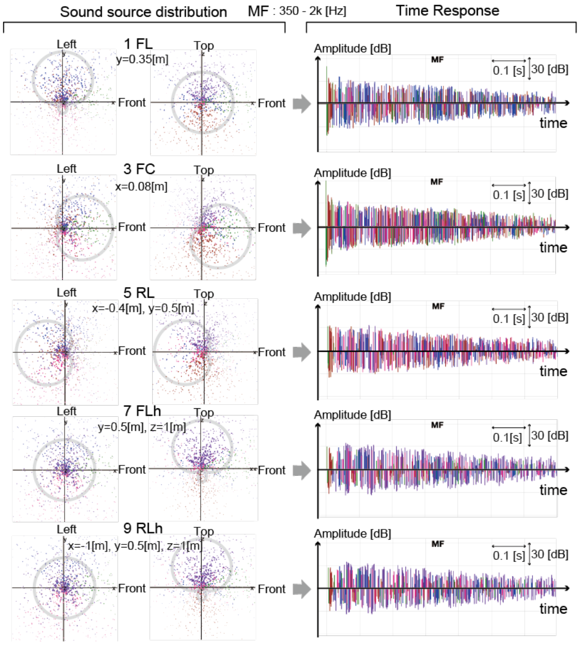


Figure 11. Virtual Mic Array responses.
(a) OCT-3D : FL, FC, RL, FLh, FRh

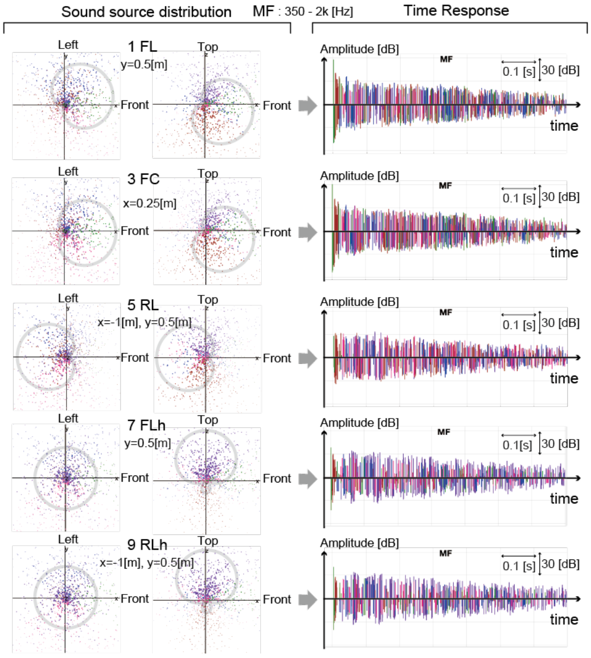


Figure 12. Virtual Mic Array responses.
(b) PCMA-3D : FL, FC, RL, FLh, FRh

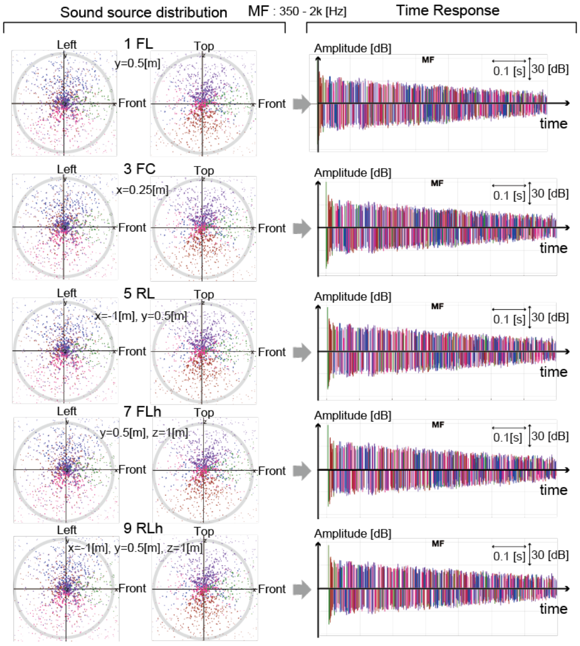


Figure 13. Virtual Mic Array responses.
(c) 2L-Cube : FL, FC, RL, FLh, FRh

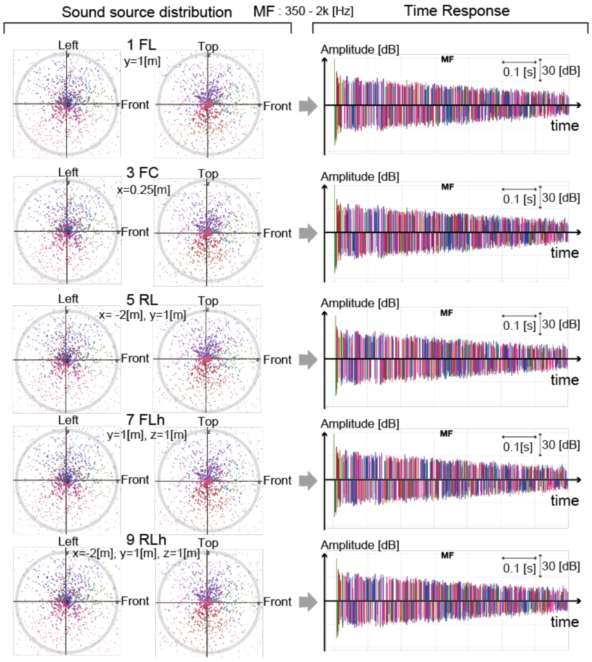


Figure 14. Virtual Mic Array responses.
(d) Decca Cuboid : FL, FC, RL, FLh, FRh

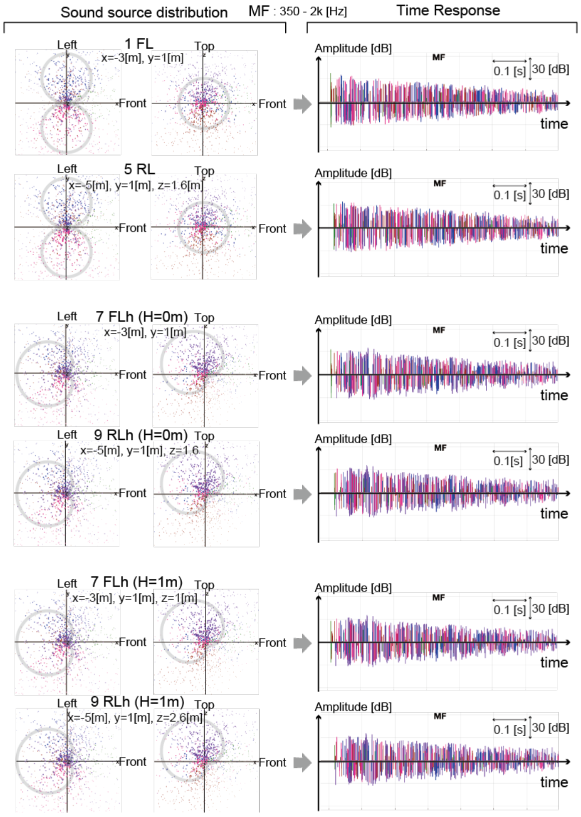


Figure 15. Virtual Mic Array responses.
(e) Hamasaki Cube : FL, RL,
FLh(0[m]), FRh(0[m]) / FLh(1[m]), FRh(1[m])

7 Virtual Responses vs Measured IRs

To compare the virtual responses with the measured impulse responses, the generated responses in the low, mid, and high frequency bands were summed up into single impulse responses by filtering through the LPF, BPF, and HPF of linear phase FIRs.

Figure 16 is shown the comparison between virtual and real impulse responses of the FL, FC, RL, FLh, and RLh channels of three types of 3D microphone arrays. In terms of the real impulse responses, we used the wav archives provided by APL, University of Huddersfield [11].

In the Figure 16, the upper part of the blue responses shows the absolute values of the virtual impulse responses, and the lower part of the red responses shows the absolute values of the measured impulse responses. Figure 17 is also shown the zoomed values from 0 to 0.1[s] to compare the details of the waveforms. From the results in the Figures 16 and 17, we can see that the virtual impulse responses can copy the waveforms of the real impulse responses well.

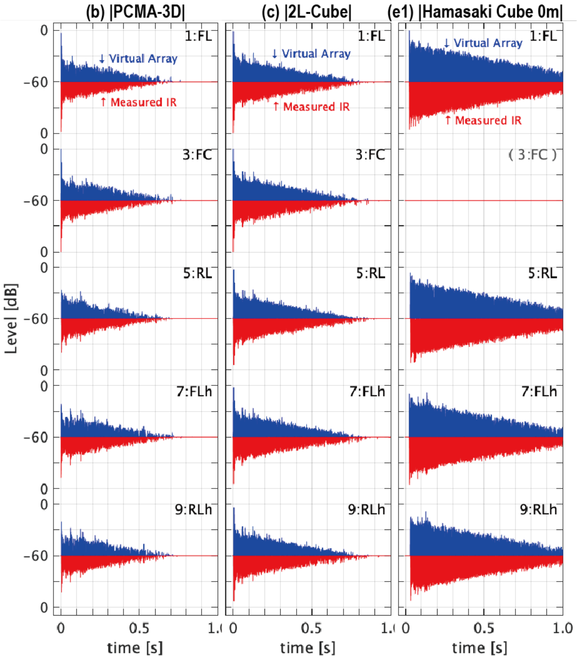


Figure 16. Virtual vs Measured responses
t = 0 ~ 1[s] : FL, FC, RL, FLh, RLh

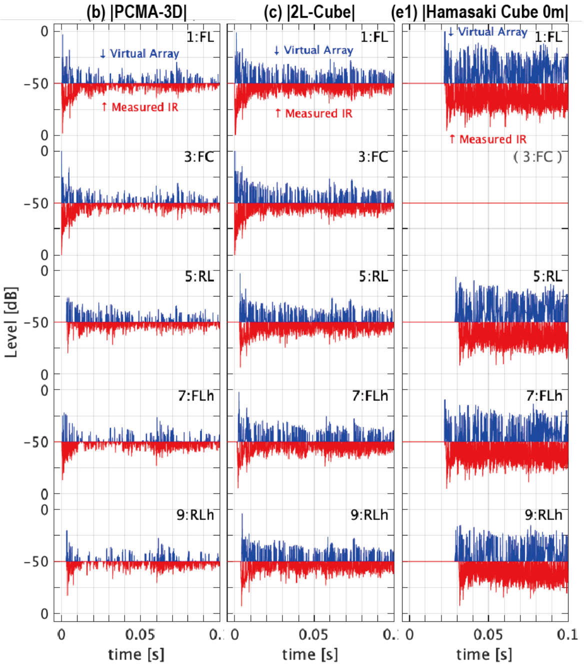


Figure 17. Virtual vs Measured responses
t = 0 ~ 0.1[s] : FL, FC, RL, FLh, RLh

8 Listening Impression

The 5.0.4ch virtual and measured impulse responses were convolved with a dry sound and encoded into binaural signals using VIRTUOSO [12] for the simple listening check. The author's subjective impression is that the virtual responses well represent the acoustic difference between the 3D microphone arrays as well as the measured impulse responses. On the other hand, in terms of similarity between virtual and measured responses, some of them show slight differences.

Figure 18 compares the frequency responses of the virtual and the measured responses of the FL channel. We can see that the measured impulse responses (red) include the frequency characteristics of the measurement loudspeaker and microphones, while the virtual responses (blue) show flat characteristics. The VSVerb removes the acoustic characteristics of the measurement tools and noise components, leaving only the reflection sounds. On the other hand, measured impulse responses always include the acoustic characteristics of the measurement tools and the on-site noise. If we try to examine the similarity between virtual and measured reverberation by subjective evaluation, the timbre differences caused by the measurement tools may lead to an incorrect judgment. We may not focus on the difference in reverberation, but we may be attracted to the difference in the tonal characteristics of the measurement tools. From this point of the view, it is considered to be preferable to compare them not by the subjective evaluation but by the objective measurements.

9 Concluding Remarks

Using the VSVerb, acoustic responses of any types of microphone arrays can be generated from a single FOA response. The author will continue to examine the similarity between virtual and measured responses through objective measurements.

Acknowledgments

The author would like to thank Prof. Hyunkook Lee and APL of the University of Huddersfield for providing valuable resources and kindly answering to my questions about 3D-MARCo.

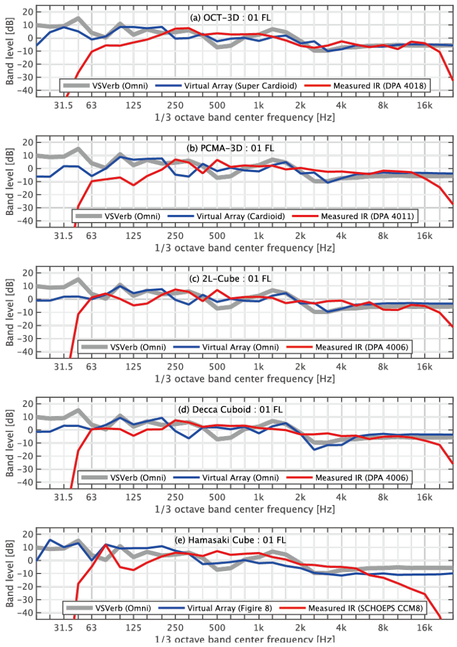


Figure 18. Comparison of frequency responses (FL) between Virtual Mic Arrays and Measured IRs.

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, "A simple evaluating method of a reproduced sound field by a measurement of sound intensities using Virtual Source Visualizer," eB, AES 143rd Convention, NY (2017)

[3] M. Nakahara, A. Omoto and Y. Nagatomo, "A simple evaluating method of a reproduced sound field by a measurement of sound intensities using Virtual Source Visualizer," eB, AES 143rd Convention, NY (2017)

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

[5] 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)

[6] 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)

[7] M. Nakahara, A. Omoto and Y. Nagatomo, " Development of a 4-pi sampling reverberator, VSVerb. - Source Reduction," eB, AES 145th Convention, NY (2018)

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

[9] 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)

[10] H. Lee and D. Johnson, "An open-access database of 3D microphone array recordings," eB, AES 147th Convention, New York (2019)

[11] H. Lee and D. Johnson, " 3D Microphone Array Comparison: Objective Measurements," J. Audio Eng. Soc., vol. 69, no. 11, pp. 871–887, (2021).

[12] H. Lee and D. Johnson, " 3D MICROPHONE ARRAY RECORDING COMPARISON (3D-MARCo)," Resources of APL, University of Huddersfield (2019). https://research.hud.ac.uk/institutes-centres/apl/resources/ https://zenodo.org/record/3477602

[13] VIRTUOSO https://apl-hud.com/product/virtuoso/