Display system for distribution of virtual image sources by using mixed reality technology

Wataru Teraoka[a)]
Yuta Kataoka[b)]
Yasuhiro Oikawa[c)]
Waseda University
3-4-1 Okubo, Shinjuku, Tokyo 169-8555, Japan

Yusuke Ikeda[d)]
Tokyo Denki University
5 Senju-Asahi-Cho, Adachi, Tokyo 120-8551, Japan

ABSTRACT

The closely located four-point microphone method was proposed by Yamasaki in 1989. The method can grasp the spatial and temporal structures of sound reflections by estimating a distribution of image sources from four measured impulse responses. In recent years, mixed reality (MR) technology has rapidly developed and is now more familiar. Many sensors, display devices, and ICT technologies have been implemented in MR equipment, which enable interaction between real and virtual worlds. In this paper, we propose an MR display system for distribution of image sources and directivity patterns of sound reflection, which are obtained by the closely located four-point microphone method. The user can view a real room and the data representation simultaneously by using the MR display. Thus, the user can observe the spatial and temporal structures of sound reflections from any direction and any distance while maintaining a relationship between the positional information of the real room and the data. In addition, the user can view the data with arbitrary scale with the spatial map of the measured room. Thus, the size of the data appearance can be changed by the user. Furthermore, the system can show multiple data sets side-by-side for comparison.

1 INTRODUCTION

There are various ways to measure a sound source position or grasp the spatial and temporal structures of sound reflections. Sound field visualization is an effective technique to

[a)] email: tera-wata1129@fuji.waseda.jp
[b)] email: kataoka8894@akane.waseda.jp
[c)] email: yoikawa@waseda.jp
[d)] email: yusuke.ikeda@mail.dendai.ac.jp

understand spatial sound information. For example, in previous research, acoustical holography[1,2], beam-forming[3,4], and optical methods[5-14] have been proposed and studied.

One of the visualization methods, the closely located four-point microphone method, was proposed by Yamasaki in 1989[15]. The method constructs a distribution of image sources from four impulse responses measured by using four microphones that are not coplanar and by using a short-time correlation technique. The distribution of image sources represents the spatial and temporal structures of sound reflections. However, because the positional information of data is three-dimensional and the distribution of image sources is dense, it is difficult to understand the structure of sound reflections at a glance.

On the other hand, the HoloLens, a display developed by Microsoft, is realized by mixed reality (MR) technology in which real and virtual spaces interact with each other as if the real and virtual worlds were merged. Its application is receiving increased worldwide attention. By taking advantage of MR technology, visualization of three-dimensional sound intensity maps with an optical see-thorough head mounted display (ST-HMD) was recently proposed[16,17].

In this paper, we propose an MR display system for the distribution of image sources and directivity patterns obtained by the closely located four-point microphone method.

2 METHOD

2.1 Closely located four-point microphone method

To grasp the spatial structures of sound fields, the closely located four-point microphone method constructs a distribution of image sources from four impulse responses measured by using four closely-located microphones that are not coplanar and by using a short-time correlation technique[15]. The method is briefly described in the following discussion.

The distances $r_{i,n}$ between the i-th microphone and the n-th image source are represented by

$$r_{i,n} = ct_{i,n} \, , \tag{1}$$

where c is the velocity of sound and $t_{i,n}$ is the arrival time of the n-th image source at the i-th microphone. When four microphones are located on the origin and three points at the same distances from the origin on the rectangular coordinate axis, the coordinates of the n-th image source are estimated by

$$\begin{aligned} X_n &= \frac{d^2 + r_{1,n}^2 - r_{2,n}^2}{2d}, \\ Y_n &= \frac{d^2 + r_{1,n}^2 - r_{3,n}^2}{2d}, \\ Z_n &= \frac{d^2 + r_{1,n}^2 - r_{4,n}^2}{2d}, \end{aligned} \tag{2}$$

where d is the distance between microphones located on the origin and a rectangular coordinate axis. Therefore, if microphone distance d, arrival time $t_{i,n}$, and the speed of sound c are known, positional information of the image sources can be estimated by using Eq. (2). The arrival time $t_{i,n}$ is estimated by a short-time correlation technique.

2.2 Mixed reality technology

In MR technology, the real and virtual worlds in visual information interact with each other as if these two worlds were merged. In this study, we introduced MR technology by using Microsoft HoloLens as shown in Fig. 1. HoloLens is a self-contained holographic computer with an optical ST-HMD and simultaneous localization and mapping (SLAM) techniques[18]. An optical ST-HMD is a stereo transparent display that can overlay 3DCG on the real view. SLAM technology can obtain the viewing position, the direction of the user and the shape of the room with a depth camera and image processing[19]. By using the spatial mesh of a real object's shape, 3DCG objects can be occluded by the spatial mesh of the real objects as if the virtual object were in real space. The spatial mesh continues to be updated and follows the observer's translation and rotation. By continuing to change the 3DCG view on the basis of the spatial mesh and the user's translation and rotation, 3DCG is appropriately overlaid on the view of the freely-moving user.

In this study, the proposed system was developed with Unity 2017, which contains the Unity3D engine of Unity Technologies, and with Microsoft Visual Studio 2017[20]. The signal processing of the proposed system was implemented by C# scripts.

Fig. 1 – Worn Microsoft HoloLens.

2.3 The display system

Figure 2 shows the display procedure of the proposed system. The system displays a three-dimensional distribution of image sources and directivity patterns by using the HoloLens in the following steps:

1) The system prepares the image source distributions by the closely located four-point microphone method.
2) The user chooses one of the distributions with the HoloLens user interface (UI), and then the system loads data in memory.
3) In displaying the distribution of image sources, 3DCG spheres are placed at the positions of the image sources. In displaying directivity patterns, a 3DCG cylinder is directed from the origin to each image source. The lengths of the cylinders represent the energies of the image sources.
4) The arrival times of the image sources are divided into five categories and represented by colors for each category.

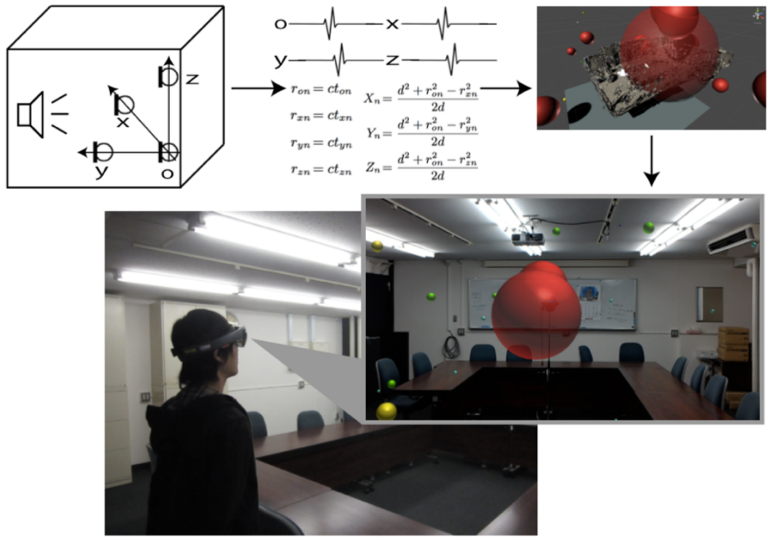

Fig. 2 – Display procedure. First, the impulse responses in the room are measured using four microphones. Second, image sources are estimated using the four-point microphone method to obtain the analysis results. Next, each image source is set as a sphere. Finally, the spheres are displayed on the HoloLens.

Figure 3 shows a scaled distribution of image sources and its color reference. In displaying the image source distribution, the origin of the coordinates shows the sound-receiving point, the center of sphere shows the coordinates of the image source, and the diameter of the sphere shows the energy of the image source. The arrival time from each image source to the receiving point is expressed by colors.

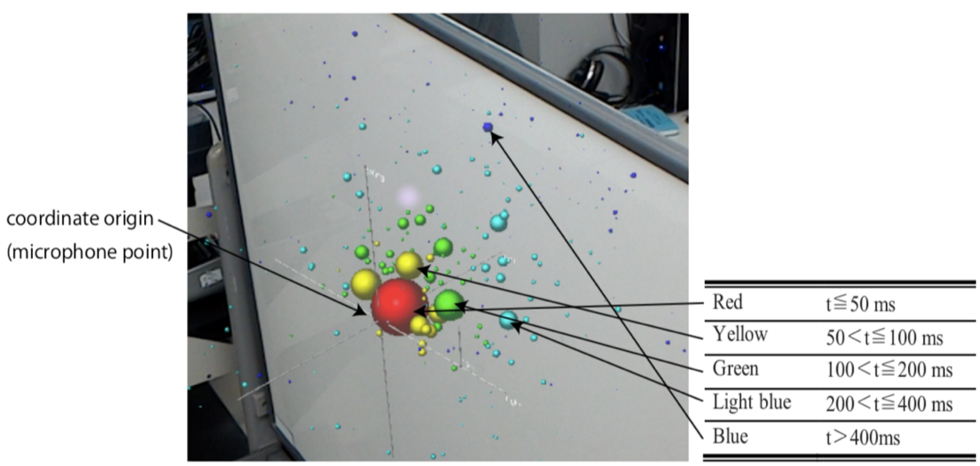

Fig. 3 – Distribution of image sources and its color reference.

Figure 4 shows a scaled directivity pattern. In displaying directivity patterns, the length of the cylinder is proportional to the logarithm of the energy of the image source. The direction of the cylinder shows the direction of the image source. The color shows arrival time in the same manner as displaying the distribution of image sources.

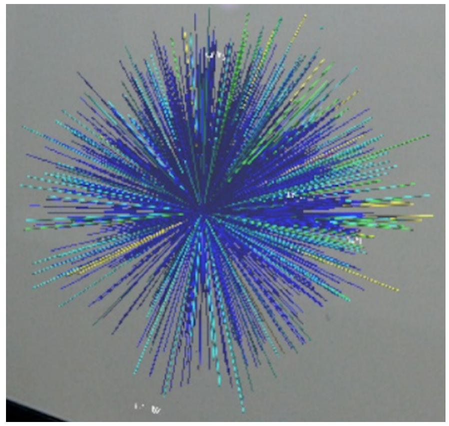

Fig. 4 – Directivity pattern.

2.4 Animation of image sources

To help the user understand the spatial and temporal structure of sound reflection, we implemented an additional display form of image source distribution as shown in Fig. 5. Each image source moves in animation from its estimated position to the microphone position. In the computer display, when its arrow key is down, the image sources start to move. When the key is up, the animation stops. The image sources also move in reverse to return to the original position.

As shown in Fig. 5, we can understand that spheres move forward to the microphone in the order of red, yellow and green. We can also understand the size, which direction and when each image source arrives.

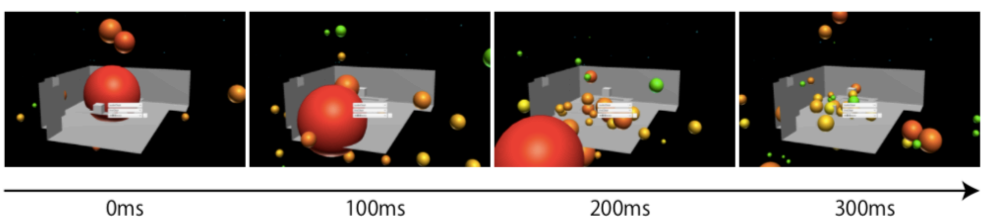

Fig. 5 – Animation of image sources moving from the original positions to the origin.

3 EXPERIMENTS

To evaluate the proposed system, we performed some experiments and overlaid the analysis results using the closely located four-point microphone method on the measured room.

3.1 Measurement of impulse responses and spatial map

We measured the impulse responses by the closely located four-point microphone method in an ordinary meeting room at Waseda University. Table 1 shows the measurement conditions. A loudspeaker was located at a distance of 1.5 m from the floor and at a distance of 3.0 m from the nearest wall. The four-point microphone was located at a distance of 1.2 m from the floor and at a distance of 3.0 m from the loudspeaker. A time-stretched pulse (TSP) signal was used as the measurement signal.

SLAM sensors attached to an optical ST-HMD continuously acquire the space shape of the room. The measured shape of the space is represented as a wireframe. We loaded and saved the spatial mesh as an object file and overlaid it on the data.

We measured as per the following steps:

1) A TSP signal was played by the loudspeaker via an audio interface and a power amplifier.
2) The sound signals were recorded by four microphones with a PC via an audio interface.
3) After calculation of the impulse responses, the distribution of image sources was obtained by using the closely located four-point microphone method.
4) The spatial map of the measured room was obtained by the optical ST-HMD.
5) Distribution of image sources was overlaid on the room mesh.

Table 1 – Measurement conditions.

Measured room	59-402 Meeting room, Nishi-Waseda campus, Waseda University
Equipment	Opposed speaker (Fostex FE204) Power amplifier (YAMAHA P4050) Audio interface (MOTU 8M) Microsoft HoloLens MacBook Pro (2.9 GHz Intel Core i5, 8 GB 1600 MHz DDR3) Four microphones (AUDIX TM1)
Distance between microphones [cm]	5.0
Measurement signal	Time stretched pulse
Temperature [°C]	29.4
Sampling frequency [Hz]	48000
Interpolation for calculation	16

3.2 Results of experiments with the MR display

We describe the experimental display of image source distribution and directivity patterns in several ways. In Sect. 3.2.1, we show the experimental results of the image source distribution overlaid on the measured room to understand the sound reflections of the measured room. In Sect. 3.2.2, we display the image source distribution and directivity patterns with an arbitrary scale to understand entire sound reflections. In Sect. 3.2.3, we also display the distribution of image sources and directivity patterns by using the other devices in order to observe the data by using familiar devices.

3.2.1 MR display of image source distribution with real scale

Figure 6 shows experimental result of image source distribution with the MR display. The left-side figure in Fig. 6 shows the measured real room. The right-side figure shows the MR display through HoloLens in the measured room. When a user wears the HoloLens, many spheres can be observed in the display. Because the HoloLens provides depth information of the 3DCG to us by using binocular disparity, the sphere object appears in the room with a natural view. Thus, we can understand where sound reflections occurred and how large the sound reflections were by using the positional and size information of image sources.

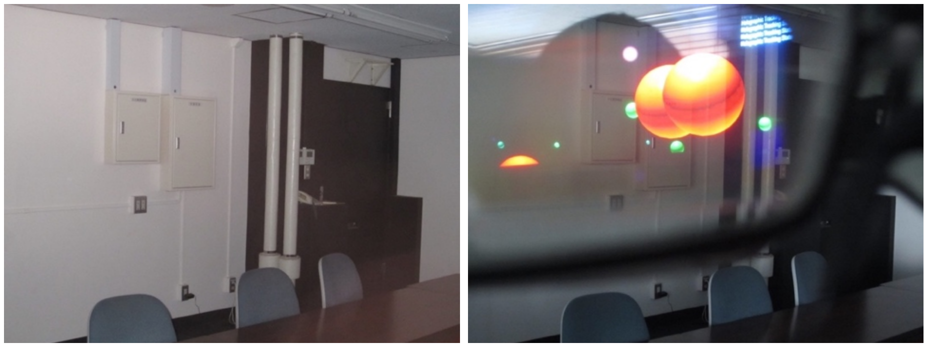

Fig. 6 – Measured room and MR display overlaid on the measured room with Microsoft HoloLens.

Figure 7 shows two types of display method for distribution of image sources. The left-side figure shows image source spheres whose diameters are proportional to the energies of the image sources. As shown in the left-side figure, two large sound reflections were observed.

On the other hand, the right-side figure shows image source spheres whose diameters are proportional to common logarithms of the energies of the image sources. As shown in the right-side figure, sound reflections with energy can be also observed. Thus, distant and small sound reflections should be observed by using the logarithm scale, because it is expected that the logarithm scale approximately represents relationship between stimulus and human perception from the Weber-Fechner law.

Fig. 7 – Linear scale and logarithm scale display.

3.2.2 MR display of image sources and directivity patterns with arbitrary scale

Figure 8 shows the MR display of distribution of image sources and directivity patterns with an arbitrary scale. In this small-scaled distribution of image sources, we can grasp the whole distribution of sound reflections. Although image source distribution is too complicated to understand its structure at a glance, the depth information of data provided by the MR display helps us to understand the three-dimensional structure. In the display of directivity patterns, we can grasp the magnitude of each direction's sound reflections. Furthermore, we can compare multiple data sets at the same time as shown in Fig. 8.

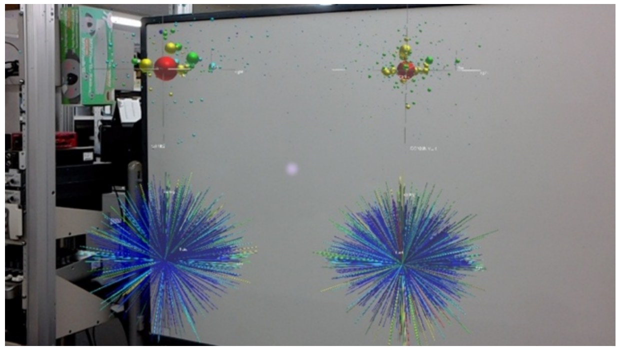

Fig. 8 – MR display of image source distribution and directivity patterns with arbitrary scale. The upper and lower objects represent image source distribution and directivity patterns, respectively. Three-dimensional sound-reflection data can be observed in any direction.

Figure 9 shows an image source distribution overlaid on the small-scale model of a hall. The small-scale model is a 1/20 scale of the Tome festival hall in Miyagi prefecture, Japan. Image source distribution was constructed from the measurement data of the Tome festival hall in 1994 by using the closely located four-point microphone method[21]. The upper-left and upper-right figures show the entire display of the image source distribution overlaid on the small-scale model. The lower-left and lower-right figures show the appearance at the closer points of view. In the upper area of the lower-right figure, yellow-and-green spheres are located outside of the small-scale model. The green-line meshes around these spheres show the form of the outside. In the lower part of the lower-right figure, a red sphere is located inside the small-scale model. The green-line meshes around the red sphere also show the shape of the inside. Thus, when we observe data and the model at a close distance, we can observe the data inside of the model.

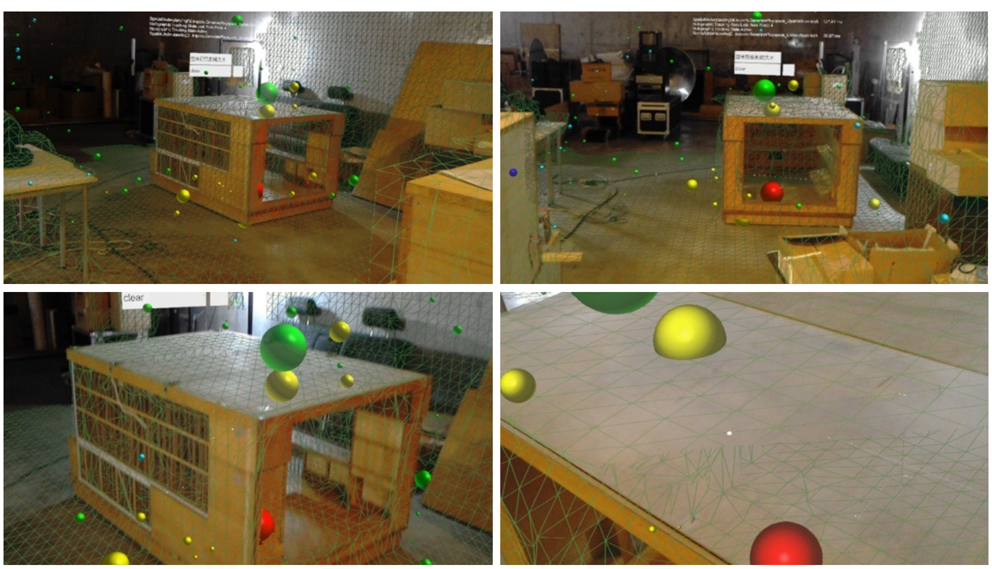

Fig. 9 – Scale display of distribution of image sources overlaid on the model. White objects and text over the model in the upper figures are part of the UI that determines which data are shown.

3.2.3 Display by using other devices

In addition, the image source distribution and the directivity patterns of sound reflections can be represented by using other familiar devices, such as a smartphone, a tablet, or a PC. Figure 10 shows an arbitrary-scale display of image source distribution and directivity patterns by using a smartphone, tablet, or computer. The user can translate, scale, and rotate the represented data. Figure 11 shows observation results in multiple directions.

In particular, iPhone devices with iOS11 and later can overlay a real-scaled distribution of image sources on the view of real space by using ARKit, which is Apple's augmented reality (AR) framework and can easily provide AR with a monocular camera on the iPhone. Figure 12 shows a sample of image source spheres in an iPhone display.

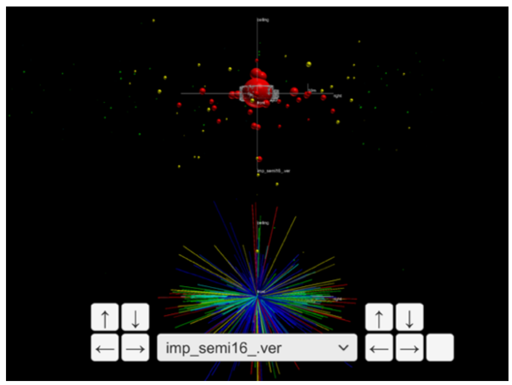

Fig. 10 – Display on a smartphone, tablet, or computer.

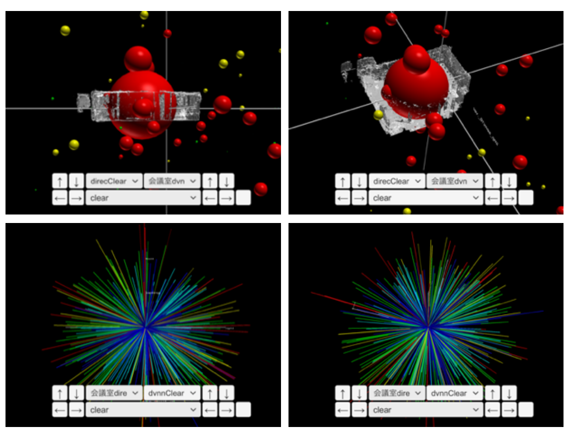

Fig. 11 – Display with multiple directions on a smartphone, tablet, or computer.

Fig. 12 – Smartphone Display by using ARKit.

4 CONCLUSIONS

In this paper, we propose a display system to overlay a distribution of image sources on a real room and a small-scaled model. By using the proposed system, the temporal and spatial structure of sound field reflection of the room can be understood immediately. Because of the MR technology, users can understand intuitively the features of the sound reflections and easily compare different data. The proposed system helps users observe the data in any direction. By selecting the logarithm scale of represented data, sound reflections with small energies can be observed. Thus, users can understand and grasp the spatial information of the sound field in greater detail. Furthermore, for an image source at a position far from the microphone in the arbitrary-scale display mode, users can understand and grasp a whole feature of the sound reflections. In the arbitrary-scale display mode, multiple data sets can be compared by placing the data side-by-side.

In future work, to represent the measured data immediately in the measured room, we will integrate the image source estimation and display systems.

5 REFERENCES

1. J. D. Maynard, E. G. Williams, and Y. Lee, "Nearfield acoustic holography: I. theory of generalized holography and the development of nah," *J. Acoust. Soc. Am.*, **78**(4), 1395–1413, (1985).
2. S.-H. Park and Y.-H. Kim, "Visualization of pass-by noise by means of moving frame acoustic holography," *J. Acoust. Soc. Am.*, **110**(5), 2326–2339, (2001).
3. H. Kook, G. Moebs, P. Davis, and J. Bolton, "An efficient procedure for visualizing the sound field radiated by vehicles during standardized passby tests," *J. Sound and Vib.*, **233**(1), 137–156, (2000).
4. J. C. Chen, K. Yao, and R. E. Hudson, "Source localization and beamforming," *IEEE Signal Proc. Mag.*, **19**(2), 30–39, (2002).
5. Y. Oikawa, M. Goto, Y. Ikeda, T. Takizawa, and Y. Yamasaki, "Sound field measurements based on reconstruction from laser projections," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, **4**, iv/661–iv/664, (2005).

6. K. Yatabe and Y. Oikawa, "PDE-based interpolation method for optically visualized sound field," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 4771–4775, (2014).

7. N. Chitanont, K. Yaginuma, K. Yatabe, and Y. Oikawa, "Visualization of sound field by means of Schlieren method with spatio-temporal filtering," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 509–513, (2015).

8. K. Yatabe and Y. Oikawa, "Optically visualized sound field reconstruction using Kirchhoff–Helmholtz equation," *Acoust. Sci. & Tech.*, **36**(4), 351–354, (2015).

9. K. Yatabe and Y. Oikawa, "Optically visualized sound field reconstruction based on sparse selection of point sound sources," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 504–508, (2015).

10. Y. Ikeda, N. Okamoto, T. Konishi, Y. Oikawa, Y. Tokita, and Y. Yamasaki, "Observation of traveling wave with laser tomography," *Acoust. Sci. & Tech.*, **37**(5), 231–238, (2016).

11. K. Ishikawa, K. Yatabe, Y. Ikeda, Y. Oikawa, T. Onuma, H. Niwa, and M. Yoshii, "Interferometric imaging of acoustical phenomena using high-speed polarization camera and 4-step parallel phase-shifting technique," *Proc. Int. Congr. High-Speed Imaging Photonics (ICHSIP)*, 331–336, (2016).

12. Y. Oikawa, K. Yatabe, K. Ishikawa, and Y. Ikeda, "Optical sound field measurement and imaging using laser and high-speed camera," *Proc. Int. Congr. Noise Control Eng. (INTER-NOISE 2016)*, 258–266, (2016).

13. K. Ishikawa, K. Yatabe, N. Chitanont, Y. Ikeda, Y. Oikawa, T. Onuma, H. Niwa, and M. Yoshii, "High-speed imaging of sound using parallel phase-shifting interferometry," *Opt. Express*, **24**(12), 12922–12932, (2016).

14. N. Chitanont, K. Yatabe, K. Ishikawa, and Y. Oikawa, "Spatio-temporal filter bank for visualizing audible sound field by Schlieren method," *Appl. Acoust.*, **115**, 109–120, (2017).

15. Y. Yamasaki and T. Itow, "Measurement of spatial information in sound fields by closely located four point microphone method," *J. Acoust. Soc. Jpn. (E)*, **10**(2), 101–110, (1989).

16. A. Inoue, Y. Ikeda, K. Yatabe, and Y. Oikawa, "Three-dimensional sound-field visualization system using head mounted display and stereo camera," *Proc. Mtgs. Acoust.*, **29**(1), 025001, (2016).

17. A. Inoue, K. Yatabe, Y. Oikawa, and Y. Ikeda, "Visualization of 3D Sound Field using See-Through Head Mounted Display," *SIGGRAPH '17*, Posters, (2017).

18. https://www.microsoft.com/hololens

19. H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part I," *IEEE Robot. & Autom. Mag.*, **13**, 99–110, (2006).

20. https://unity3d.com

21. Y. Yamasaki, "Acoustics of TOME Festival Hall, MIYAGI," *Architectural Acoustics and Noise Control*, **24**(3), 63–66, (1995) (in Japanese).