
1. INTRODUCTION

The visualization of sound field helps us understand the sound source position and sound propagation. It is useful for detecting noise sources, acoustical design, acoustical education, etc.. There are various visualization technologies, such as acoustical holography,¹ beam-forming method,² and optical methods.^{3–12} Most of technologies superimpose the measured sound information onto the image or video taken by a camera.¹³ Compared with the presentation method by diagram, it is easy to understand the relation of the sound propagation and surrounding environment. The visualization with image helps us to understand sound field on the spot in real time. However, considering a three-dimensional (3D) sound field, it is difficult to visualize the depth information of sound field because a two-dimensional display can not show the depth information under normal condition. Another problem is that the visualized area must be limited within camera's view. Thus, it is necessary to visualize from various viewpoints to understand an entire 3D sound field. It takes time and effort to move equipment for repeated measuring.

Meanwhile, 3D display devices such as head mounted display (HMD)¹⁴ have been developing rapidly. HMD provides us with natural stereoscopic vision with a small display for each eye and uses several tracking sensors to change the vision consistently with the user's head motion. In addition, when a stereo camera is attached to the front of HMD and the output of stereo video is sent to two small displays in real time, the 3D vision is reproduced at our own viewpoint. It is called video see-through head mounted display (VST-HMD).¹⁵ By superimposing the 3D computer graphics model onto the stereo video, it realizes natural augmented reality (AR).¹⁶

In this paper, we describe the system of 3D sound field visualization with AR by using VST-HMD. The 3D sound information is represented by 3D model and superimposed onto stereo video. By means of stereoscopic vision and free moving viewpoint, we can understand the entire sound field more intuitively. In order to prove the effectiveness of our proposal, we proposed a real-time system of measuring and visualizing one of the aspects of sound information: 3D sound intensity. We conducted experiments to visualize the sound field generated by loudspeakers and motorcycles.



Figure 1: Video See-Through Head Mounted Display and external head tracking infrared camera.

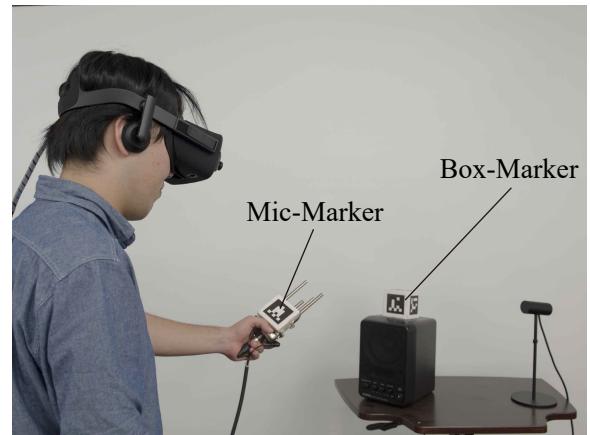


Figure 2: User measuring 3D sound intensity map around a loudspeaker.

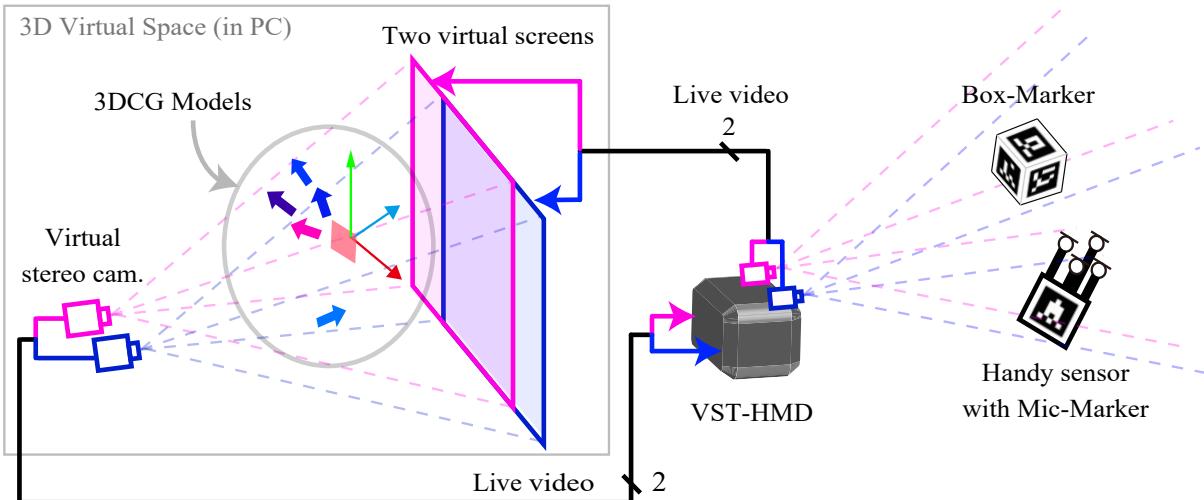
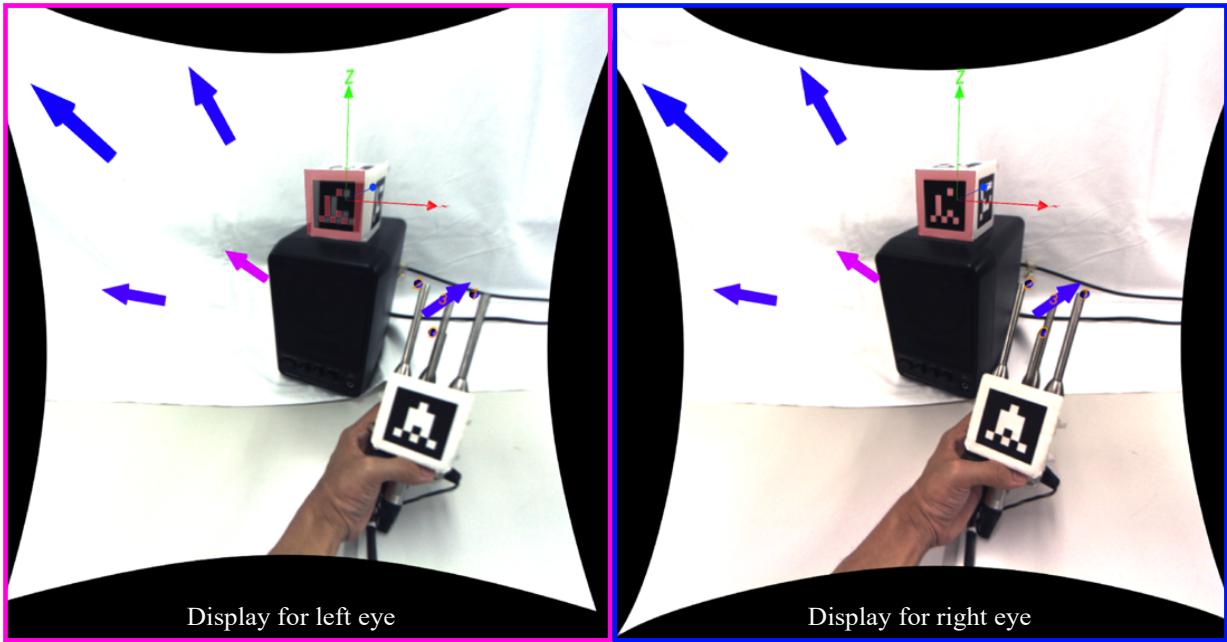


Figure 3: The stereoscopic view for user and processing flow of live videos from stereo camera to head mounted display. Resolution of each video is 960×950 pixel.

2. SYSTEM

Figure 1 shows an VST-HMD and an external head tracking infrared camera. Figure 2 shows the user wearing VST-HMD holds a handy microphone with an AR marker. This maker, which is called Mic-Marker, informs the position and angle of the microphone to the system by using image processing.¹⁷ When the system detects Mic-Marker by left video, it begins to measure 3D sound intensity and represent it by an arrow 3D model with its color indicating level. The model is superimposed onto a stereo video at the measured position. The system keeps the position of 3D model as if it exists at the measured position in real world, even when the user moves the head. To change stereo video consistently with the user's head motion, the system detects the user's

head position by using head tracking sensors and another AR marker which is called Box-Marker. When the handy microphone is away from the other measured points, the system automatically measures the sound intensity and visualizes it. Thus, the system makes a 3D sound intensity map only through a microphone being moved by the user. We assume that the sound field is steady because the microphone is moved to scan the sound fields.

A. VIDEO SEE-THROUGH HEAD MOUNTED DISPLAY

Figure 3 shows stereo image in HMD and a processing flow of the stereo live video. It is taken by a stereo camera and outputted to the stereo display through a personal computer (PC). We make a virtual space in the PC to superimpose 3D models to stereo video. The space has a virtual stereo camera and two virtual screens. The distance between left and right of virtual stereo camera is the same as that between left and right of stereo camera attached to the HMD. The virtual stereo screen is arranged to face a virtual stereo camera at long distance. The live video taken by the left camera is projected to the left virtual screen. The left virtual camera captures only the left virtual screen without capturing the right screen. Similarly, the right virtual camera only captures the right virtual screen. When a 3D model is placed between the virtual stereo camera and the screens, it is superimposed onto the stereo video. The 3D model is moved to change its position and size in stereo video. For example, as the 3D model is moved closer to the stereo camera, it looks bigger in size. In this study, we use Unity5 game engine in order to create 3D virtual space. The VST-HMD consists of Oculus Rift CV1 and Ovrvision Pro.

B. THREE DIMENSIONAL SOUND INTENSITY MEASUREMENT

It is possible to measure 3D sound intensity by using four or more non-coplanar closed microphones. The array of PU prove¹⁸ is the other choice. In this paper, we use four omnidirectional microphones which are arranged in tetrahedron figure as Fig.4 shows. The sound intensity in the direction of two microphones is calculated by Cross-Spectral Method.¹⁹

$$I_r = -\frac{1}{2\pi\rho\Delta r} \int_{f_1}^{f_2} \frac{\text{Im}\{S_{12}(\omega)\}}{\omega} d\omega, \quad (1)$$

where ρ is atmospheric density and Δr is the distance of two microphones. f_1 and f_2 are lower and upper frequency of an analysed frequency range, respectively. A range of analysed frequency. S_{12} is a cross spectrum of two sound signals. $\text{Im}\{\cdot\}$ means an imaginary part.

The microphones can measure six sound intensities because four microphones can make six pairs. Those are represented by $I_{01}, I_{02}, I_{03}, I_{12}, I_{13}$, and I_{23} as Fig.5 shows. By synthesizing those, we calculate a 3D sound intensity at the center O of four microphones.

$$\begin{cases} I_x &= (I_{01} - I_{02} - 2 \cdot I_{12} - I_{13} + I_{23}) / 4, \\ I_y &= (I_{01} + I_{02} + I_{03}) / \sqrt{6}, \\ I_z &= (-I_{01} - I_{02} + 2 \cdot I_{03} + 3 \cdot I_{13} + 3 \cdot I_{23}) / 4\sqrt{3}. \end{cases} \quad (2)$$

The norm of eq. (2) refers to the power of sound intensity. It can be converted to sound intensity level,

$$L_i = 120 + 10 \cdot \log_{10} |I|. \quad (3)$$

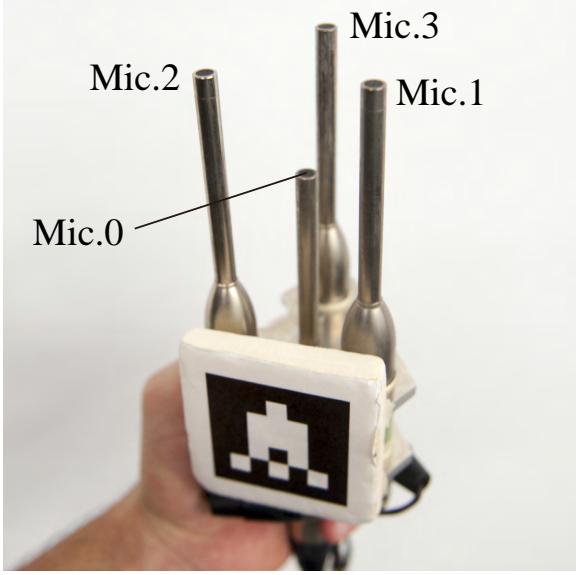


Figure 4: Handy microphone consists of tetrahedron-figured four omnidirectional microphones and Mic-Marker. The distance between two microphones is 5 cm.

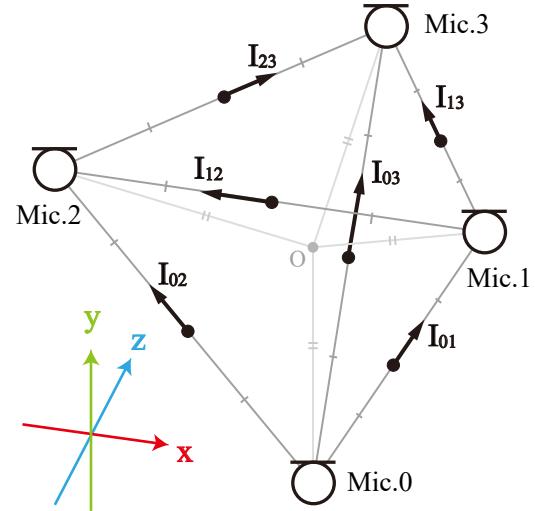


Figure 5: Structure of handy microphone and basic coordinate of 3D sound intensity. O is the center point of four microphones. $I_{01}, I_{02}, I_{03}, I_{12}, I_{13}$, and I_{23} are sound intensities between two microphones. The x -axis and y -axis are parallel with I_{12} and the line between $Mic.0$ and O , respectively.

The sound intensity level is represented by the color of 3D arrow. In this study, we use MOTU 8M audio-interface in order to record four sound signals. The ASIO driver streams these signals from MOTU 8M to Unity5.

C. AR MARKER DETECTION

The positional and rotational relation between VST-HMD and a handy microphone is changed at every video frame, because the user always moves her/his head and the microphone. Therefore, in order to superimpose the new arrow onto the measured point in stereo video, it is necessary to acquire the positional and angular information of the handy microphone. The system acquires the information by using image detection of Mic-Marker in the left camera's video. The image and real size of Mic-Marker are configured in advance. The positional information is acquired from the position and size of the Mic-Marker in the camera's left video. The angular information is also acquired from the figure of the marker. While the Mic-Marker is detected by image detection processing, the coordinate (x, y, z) in Fig.6 moves in accordance with the position and angle of Mic-Marker in the left video. The origin position for the coordinate stays on the line of each virtual camera and the center of Mic-Marker in the paired screen. The position of a new measured arrow is immutable in the coordinate because the positional relation between O and Mic-Marker is fixed.

In addition, we must consider the adjustment of stereo video for the user's head movement. To display 3D arrows fixed at the measured point, it is effective to use a new coordinate with

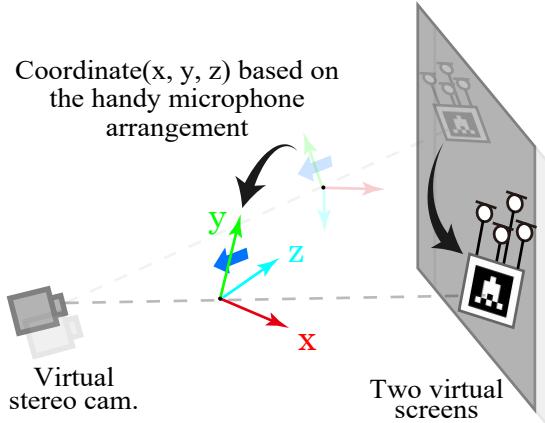


Figure 6: 3D virtual space while detecting Mic-Marker. Coordinate (x, y, z) moves consistently with Mic-Marker's motion in left virtual screen. New measured arrow is represented on the basic of (x, y, z).

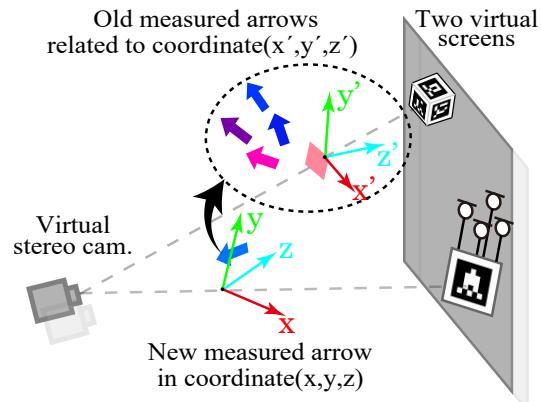


Figure 7: 3D virtual space while detecting Mic-Marker and Box-Marker. Coordinate (x', y', z') moves consistently with Box-Marker's motion in left virtual screen. New measured arrow is represented when handy microphone leaves at the configured measurement distance from other measured points.

consideration of the real positional movement of stereo camera. To make this new coordinate, we use a Box-Marker which contains six AR markers near the measurement area. We assume a local coordinate (x', y', z') with the origin located between each virtual camera and the center of Box-Marker in the paired screen. The coordinate (x', y', z') moves in accordance with Box-Marker. The Mic-Marker and the coordinate (x, y, z) move in the similar way. By relating the measured arrows to the coordinate (x', y', z'), the arrow keeps the position at the measured position, because the Box-Marker does not move in the real world.

In sound intensity measurement, the system saves the measured four signals and the positional and rotational relation between the handy microphone and the Box-Marker. When the user changes the configuration of analysis and visualization, the system recalculates all intensity and changes the view of 3D arrows. In this study, we use ArUco AR marker detection library.

D. HEAD TRACKING SENSORS

When a stereo camera is far from the Box-Marker or something disturbs the image detection of the Box-Marker, the head tracking sensors in the HMD, which consists of gyro, acceleration and magnetic field sensor, and external tracking infrared camera, complement the relational information between stereo camera and Box-Marker.²⁰ Figure 8 shows 3D virtual space while using head tracking sensors. When the Box-Marker cannot be detected, coordinate (X, Y, Z), which relates the virtual stereo camera and screens, starts moving in accordance with the user's head motion without the Box-Marker's information. When the user moves forward, the distance between the virtual stereo camera and the 3D models becomes short. Therefore, the 3D model becomes bigger

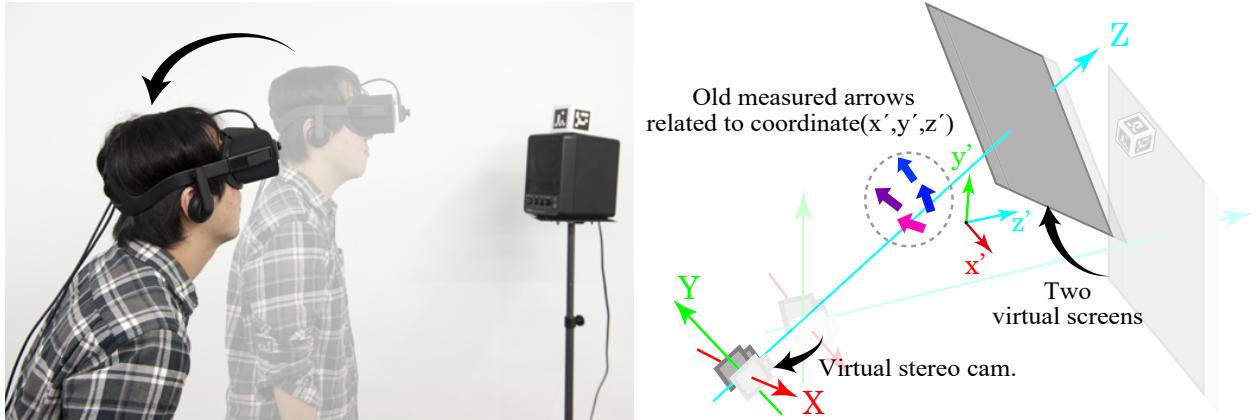


Figure 8: User’s motion and 3D virtual space while not detecting Box-Marker. Coordinate (X, Y, Z) which relates virtual stereo camera and screens moves consistently with the user’s head motion. Whereas, measured arrows which related to coordinate (x' , y' , z') are fixed.

in the stereo video. When the user changes the view angle with the head movement, the system just rotates (X, Y, Z). However, when only using the head tracking sensors to detect a head position, the 3D models are gradually deviate in every frame because the error of these sensors is large. Therefore, the user should regularly correct the head position by using image detection of the Box-Marker.

3. EXPERIMENT

We conducted experiments of visualization by creating 3D sound intensity maps of the several type of sound fields in order to evaluate the effectiveness of the proposal. Table 1 shows common conditions in all experiments. To adjust the condition in each experiment, we measured atmospheric density ρ and configured the analysed frequency range f_1 and f_2 , the distance of measuring points, and the colorbar range. We calibrated four microphones by using 1 kHz sine signal before the experiments.

Table 1: Common conditions of all experiments

Microphone	AUDIX TM-1
Direction of microphones [m]	0.05
Sampling Rate [Hz]	44100
Analysis Length [ms]	92.9
Distance of stereo camera [m]	0.06
Frame rate [fps]	60
Size of AR markers [m]	0.05 × 0.05

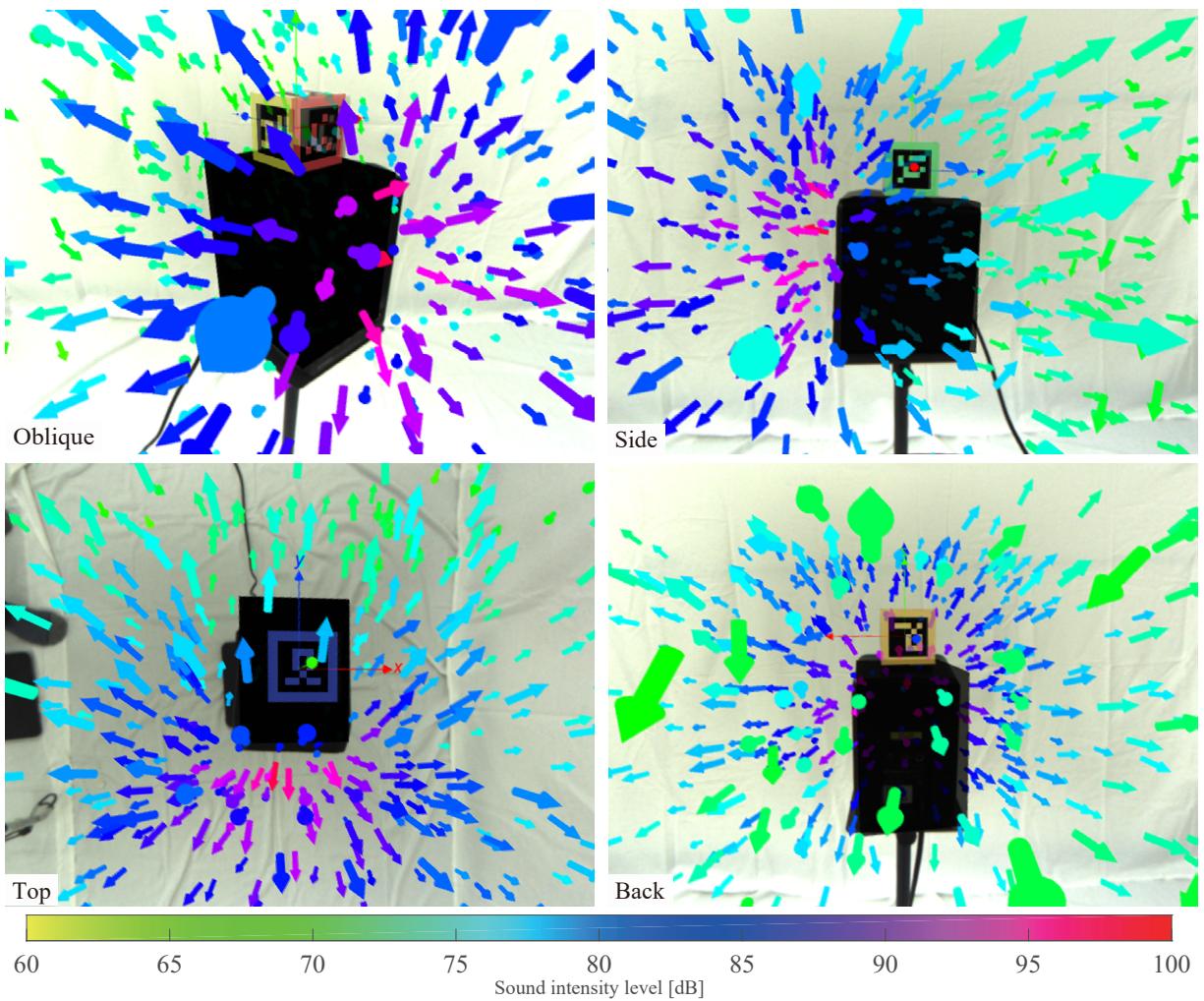


Figure 9: 3D sound intensity map around a full-range loudspeaker analysed in a frequency range: the octave bands of with the center frequency of 1 kHz. Black box 3D model is overlaid onto the loudspeaker to hide arrows behind that. The pictures were snapped from oblique, side, top, and back point of view.

A. FULL-RANGE LOUDSPEAKER WITH WHITE NOISE

In the first experiment, we visualized the 3D sound intensity map around a full range loudspeaker (YAMAHA MS101III) with white noise signal. This experiment was conducted in a general meeting room with background noise level of 57.2 dB. The measured atmospheric density ρ was 1.169 kg/m³. We configured that the sound pressure level of white noise at 1.0 m the distance in front of the loudspeaker to be 75 dB, and the measurement distance to be 0.05 m. It takes about 8 minutes to measure the sound intensities at 457 points. The intensity map is analysed in a frequency range: the octave band of with the center frequency of 1 kHz.

Figure 9 shows the measured 3D sound intensity map in display for the left eye. Although Fig. 7 shows only 2D information, we can understand the sound propagation of the noise by observing several pictures in different angle views. The noise was radiated from the front of the loudspeaker

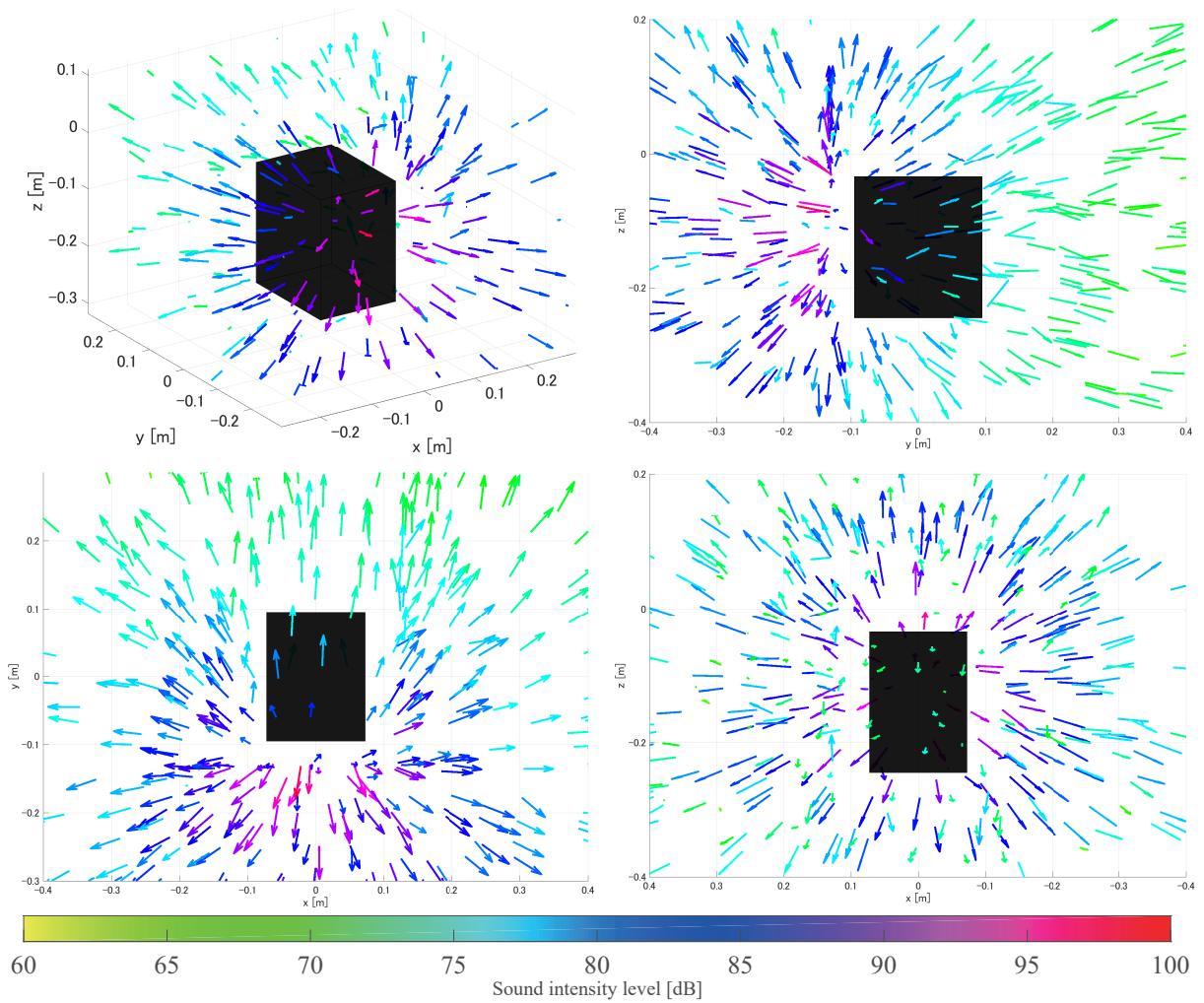


Figure 10: 3D sound intensity map on diagram. The sound intensity data is the same as in Fig. 9. Black box means the loudspeaker.

and diffracted at the side and behind the loudspeaker. In the proposed system, it is possible not only to observe from several view points with our head's movement, but also to perceive the depth information by stereoscopic vision. It helps us clearly understand the positional relation of all arrows and the direction of each arrow.

Figure 10 shows the diagram of the same 3D sound intensity map. Since the background visual information is lost, we must make a 3D visual information by 3D computer graphic with our own hands. In the proposed system, the background visual information is captured by a stereo camera in the real world. In addition, we can perceive the result more intuitively because the display is wearable and movable.

B. 2WAY LOUDSPEAKER WITH WHITE NOISE

We made the 3D sound intensity map around a 2-way loudspeaker (YAMAHA MSP5) with white noise signal. This experiment was conducted in the same conditions as 3.1 except the mea-

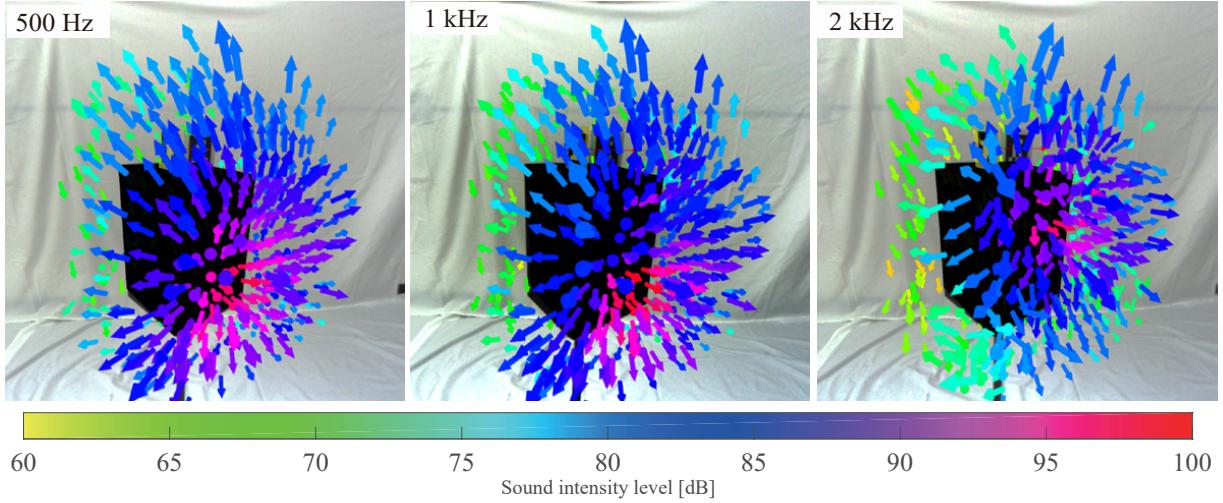


Figure 11: 3D sound intensity maps around a 2-way loudspeaker analysed in different frequency ranges: the octave bands of with the center frequency of 500 Hz, 1 kHz, and 2 kHz.

surement distance is 0.02 m. It takes about 5 minutes to measure intensities at 493 points. We analysed the intensity map in three frequency ranges: the octave bands of with the center frequency of 500 Hz, 1 kHz, and 2 kHz.

Figure 11 shows the 3D sound intensity maps analysed in three octave bands. Those pictures was captured in display for left eye. As the pictures show, the lower frequency sound is propagated from the lower side of loudspeaker. On the other hand, the higher frequency sound is propagated from the upper side of loudspeaker. There are many 3D arrows in high density because the measurement distance is configured to be short. In the proposed system, the depth view information by stereoscopic vision helps us easily understand the positional relations and directions of the congested arrows.

C. MOTORCYCLE

Sound intensity map helps us specify the position of noise sources such as the noise of motor engine. In this experiment, we visualized 3D sound field around two types of motorcycles: KAWASAKI Ninja250(Type A) and YAMAHA MT-25(Type B). Those motorcycles have major differences in the engine unit covers. The Type A has a cover for the engine. However, the Type B has no cover. This experiment was conducted in the field and the background noise level was 57.0 dB. The measured atmospheric density ρ was 1.167 kg/m^3 . We configured the measurement distance to be 0.03 m and measured the sound intensities at about 500 points. The intensity map was analysed in the two frequency ranges: the octave bands of with the center frequency of 500 Hz and 2 kHz.

Figure 12 shows the measured sound intensity maps. As the figure shows, the largest sound intensity is around the engine in both motorcycles. In the sound intensity maps in 2 kHz octave band, the sound propagated from the topside of Type B is suppressed compared with Type A. The quantity of measurement points is almost the same as in the experiment in sec. 3.2. However, it takes about 10 minutes to measure all points in this experiment. This can be explained with the

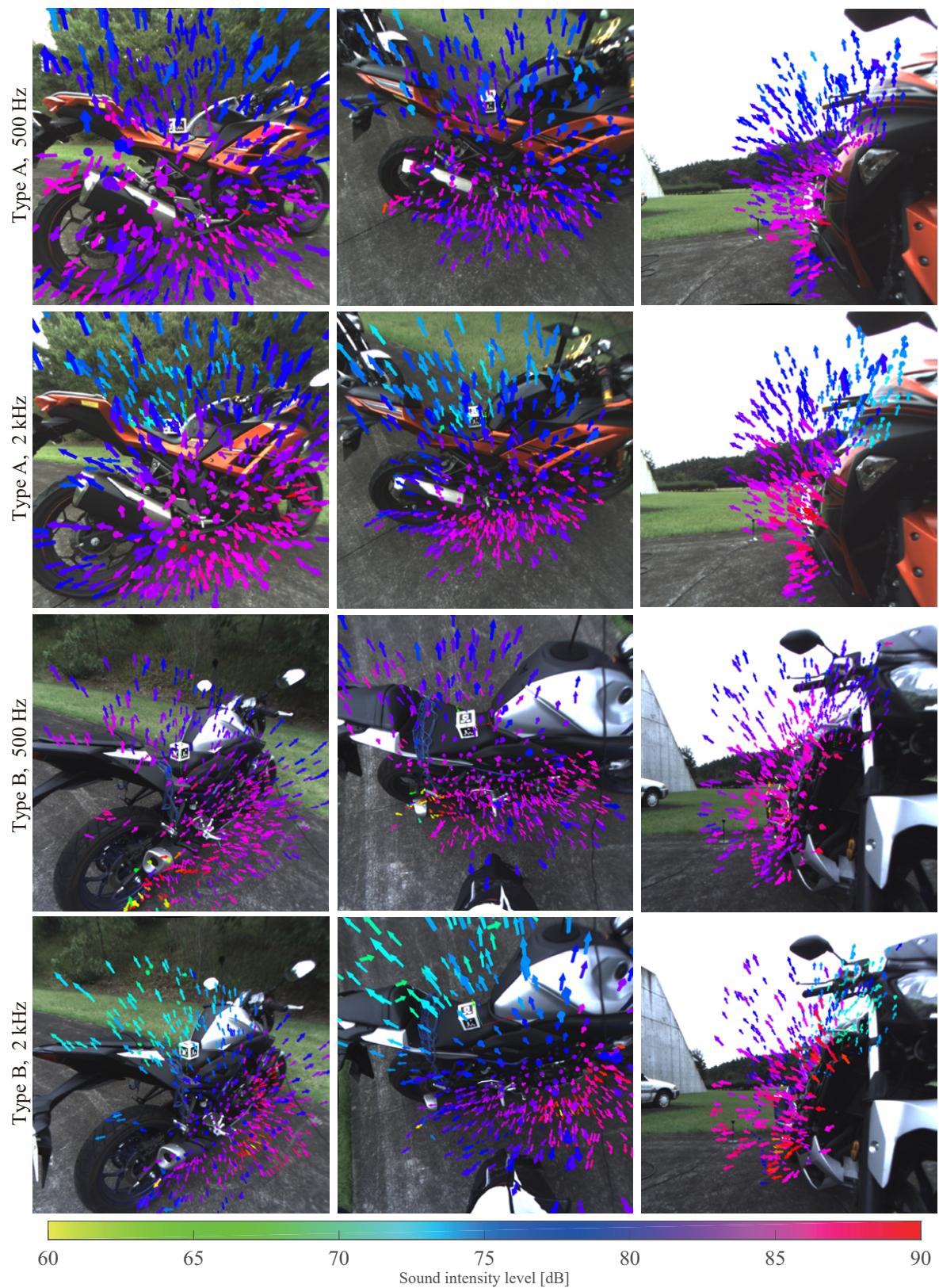


Figure 12: 3D sound intensity maps around two type of motorcycles, analysed in two frequency ranges: the octave bands of which the center frequency are 500 Hz and 2 kHz.

detection of the head position with an external tracking infrared camera which does not operate due to sunlight.

4. CONCLUSION

We proposed a measurement and visualization system of 3D sound intensity map by using the AR marker detection and VST-HMD. The stereoscopic viewpoint vision and free head movement help us understand the sound field more intuitively and reveal interesting aspects. However, there are two problems in actual use. First, the detection range of AR marker is not wide enough to visualize the sound field in large space. Secondly, the head tracking infrared sensors cannot be used in the outdoor conditions because of the sunlight disturbance.

To solve those problems, we will improve the proposed system by applying marker-less AR technology²¹ in future work. In addition, it is necessary to conduct subjective experiments to find out the usefulness of VST-HMD when compared to 2D display. We would like to consider visualizing other sound information and using Optical See-Through HMD.²²

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**, 1395–1413 (1985).
 - ² Y. T. Cho and M. J. Roan, “Adaptive near-field beamforming techniques for sound source imaging,” *J. Acoust. Soc. Am.* **125**, 944–957 (2009).
 - ³ Y. Oikawa, M. Goto, Y. Ikeda, T. Takizawa, and Y. Yamasaki, “Sound field measurements based on reconstruction from laser projections,” in *IEEE Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, (2005), Vol. 4, pp. iv/661–iv/664.
 - ⁴ Y. Ikeda, N. Okamoto, T. Konishi, Y. Oikawa, Y. Tokita, and Y. Yamasaki, “Observation of traveling wave with laser tomography,” *Acoust. Sci. & Tech.* **37**, 231–238 (2016).
 - ⁵ K. Yatabe and Y. Oikawa, “PDE-based interpolation method for optically visualized sound field,” in *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, (2014), pp. 4771–4775.
 - ⁶ N. Chitanont, K. Yaginuma, K. Yatabe, and Y. Oikawa, “Visualization of sound field by means of Schlieren method with spatio-temporal filtering,” in *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, (2015), pp. 509–513.
 - ⁷ K. Yatabe and Y. Oikawa, “Optically visualized sound field reconstruction using Kirchhoff–Helmholtz equation,” *Acoust. Sci. & Tech.* **36**, 351–354 (2015).
 - ⁸ K. Yatabe and Y. Oikawa, “Optically visualized sound field reconstruction based on sparse selection of point sound sources,” in *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, (2015), pp. 504–508.
-

-
- ⁹ 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,” in *31st Int. Congr. High-Speed Imaging Photonics (ICHSIP)*, (2016).
- ¹⁰ Y. Oikawa, K. Yatabe, K. Ishikawa, and Y. Ikeda, “Optical sound field measurement and imaging using laser and high-speed camera,” in *Proc. 45th Int. Congr. Noise Control Eng. (INTER-NOISE 2016)*, (2016), pp. 258–266.
- ¹¹ 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**, 12922–12932 (2016).
- ¹² 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).
- ¹³ D. F. Comesana, “Scan-based sound visualisation methods using sound pressure and particle velocity,” Ph.D. thesis, University of Southampton (2014).
- ¹⁴ I. E. Sutherland, “A head-mounted three dimensional display,” in *Proc. Jt. Comput. Conf., Part I*, (ACM, New York, NY, USA, 1968), AFIPS ’68 (Fall, part I), pp. 757–764.
- ¹⁵ M. Kanbara, T. Okuma, H. Takemura, and N. Yokoya, “A stereoscopic video see-through augmented reality system based on real-time vision-based registration,” in *Virtual Reality, 2000. Proc. IEEE*, (2000), pp. 255–262.
- ¹⁶ R. T. Azuma, “A survey of augmented reality,” *Presence: Teleoperators and virtual environments* **6**, 355–385 (1997).
- ¹⁷ H. Kato and M. Billinghurst, “Marker tracking and hmd calibration for a video-based augmented reality conferencing system,” in *Proc. 2nd IEEE ACM Int. Workshop (IWAR)*, (1999), pp. 85–94.
- ¹⁸ F. Jacobsen and H.-E. de Bree, “A comparison of two different sound intensity measurement principles,” *J. Acoust. Soc. Am.* **118**, 1510–1517 (2005).
- ¹⁹ J. Y. Chung, “Cross-spectral method of measuring acoustic intensity without error caused by instrument phase mismatch,” *J. Acoust. Soc. Am.* **64**, 1613–1616 (1978).
- ²⁰ S. M. LaValle, A. Yershova, M. Katsev, and M. Antonov, “Head tracking for the oculus rift,” in *IEEE Int. Conf. Robotics and Automation (ICRA)*, (2014), pp. 187–194.
- ²¹ Y. Genc, S. Riedel, F. Souvannavong, C. Akinlar, and N. Navab, “Marker-less tracking for ar: A learning-based approach,” in *Mixed and Augmented Reality, 2002. ISMAR 2002. Proceedings. International Symposium on*, (IEEE, 2002), pp. 295–304.
- ²² J. P. Rolland and H. Fuchs, “Optical versus video see-through head-mounted displays in medical visualization,” *Presence: Teleoperators and Virtual Environments* **9**, 287–309 (2000).
-