Visualization system for sound field using see-through head-mounted display

Atsuto Inoue1,*, Yusuke Ikeda2, Kohei Yatabe1 and Yasuhiro Oikawa1,†

1Department of Intermedia Art and Science, Waseda University, 3–4–1 Ohkubo, Shinjuku-ku, Tokyo, 169–8555 Japan
2Department of Information Systems and Multimedia Design, Tokyo Denki University, 5 Senju-asahi-cho, Adachi-ku, Tokyo, 120–8551 Japan

(Received 31 January 2018, Accepted for publication 4 July 2018)

Abstract: For the visualization of a sound field, a widely used method is the superimposition of the sound information onto a camera view. Although it effectively enables the understanding the relationship between space and sound, a planar display cannot resolve depth information in a straightforward manner. In contrast, a see-through head-mounted display (STHMD) is capable of representing three-dimensional (3D) vision and natural augmented reality (AR) or mixed reality (MR). In this paper, we propose a system for the measurement and visualization of a sound field with an STHMD. We created two visualization systems using different types of STHMDs and technologies for realizing AR/MR and a measurement system for a 3D sound intensity map, which can be used together with the visualization system. Through three visualization experiments, we empirically found that the stereoscopic viewing and the convenient viewpoint movement associated with the STHMD enables understanding of the sound field in a short time.

Keywords: Sound intensity, Marker detection, Simultaneous localization and mapping (SLAM), Augmented reality (AR), Mixed reality (MR)

PACS number: 43.35.Sx, 43.58.+z [doi:10.1250/ast.40.1]

1. INTRODUCTION

Visualization is an effective technique for understanding invisible physical phenomena such as sound. Sound field visualization technology is utilized for noise source detection, acoustical design, and other applications. In previous studies, sound field visualization has been realized with various measurement methods, such as acoustical holography [1,2], beam-forming [3,4], and optical methods [5–9].

One of the most effective methods to visualize the measured sound field information is the superimposition of the sound information and a camera image onto a video [10]. Compared with a simple graph of the sound field information, the relationship between the sound propagation and surrounding environment can be more easily assessed. However, considering three-dimensional (3D) sound field information, it is difficult to visualize the depth information because a planar display cannot represent the depth information under normal conditions. While using a planar display, observations from multiple viewpoints are required to comprehend the 3D sound field information. This takes considerable effort, as cameras and measuring instruments need to be moved for each observation viewpoint.

In recent years, stereoscopic head-mounted displays (HMDs), which are wearable 3D displays, have evolved rapidly [11,12]; in particular, see-through HMDs (STHMDs), which can superimpose 3D computer graphics (3DCGs) on the eye view of the wearer [13], have received much interest. By combining STHMD and technologies to detect the head movements of the wearer, it is possible to visualize 3DCGs in real space with depth perception as if the 3DCGs existed in the real world. These technologies are known as augmented reality (AR) [14] or mixed reality (MR) [15]. Thus, by using AR or MR with 3D measurements, information regarding invisible physical phenomena can be extended to the real world via STHMD.

In this paper, we propose a 3D measurement and visualization system for sound field information using a STHMD. The stereoscopic view and free viewpoint move-
Can help realize the 3D assessment of sound propagation. We developed two types of sound field visualization systems using different STHMDs and technologies for AR/MR and then created a 3D sound intensity map as a visualization target for sound information [16,17]. We also verified the effectiveness of the proposed visualization systems.

2. PROPOSED VISUALIZATION SYSTEMS

In the following section, we describe the details of the two developed visualization systems for the 3D sound intensity map. The technologies used in each system are summarized in Table 1.

2.1. Visualization System with Video STHMD and Marker Detection

2.1.1. Video STHMD and marker detection

The video STHMD realizes a real-time stereo view, including binocular disparity, with a stereo camera installed on the front of the HMD [13,18]. Thus, the users can view the scene in the front via the stereo camera as if they do not wear a HMD. To represent certain additional information, the 3DCGs are superimposed on the stereo view of the real world.

The view of the 3DCGs needs to be transformed with respect to the relative position between the user and the target position. Thus, the positional information of the target needs to be measured to achieve a natural stereoscopic view. This information is determined based on marker detection technology often used in AR systems [19,20]. The positional relationship between the camera and the pre-registered marker is acquired with respect to the size and the shape of the marker in the camera image.

2.1.2. System configuration

Figure 1 shows the observation of a loudspeaker emitting sound and its sound field using the developed visualization system. During the preparations, a number of markers are arranged around the visualization target. The number of markers can be increased according to the observation viewpoint range up to the detection distance limits for each marker. After the size and the image pattern of the marker are registered in the system, these positional relationships are also registered by a scanning camera. When the camera of the video STHMD detects one or more markers, the 3DCGs representing the sound field are superimposed on the video image based on the position of the detected markers.

The stereo video camera and the HMD of the video STHMD are connected to a computer. The process of superimposing the 3DCGs on the stereo video is performed in virtual space in the computer. In virtual space, a virtual stereo camera and a virtual stereo screen are arranged to face each other along the Z-axis in world coordinates, as shown in Fig. 1. The virtual stereo screen projects the stereo camera view, and the view of the virtual stereo camera is projected to the HMD. When the 3DCGs of the sound field are arranged between the virtual screen and the camera, the 3DCGs overlap with the virtual screen from the viewpoint of the virtual stereo camera.

The position and rotation of the 3DCGs in the virtual space always change in accordance with the position and shape of the detected marker in the stereo video. For example, when the marker is enlarged in the video, the output view of the 3DCGs becomes larger by moving toward the virtual stereo camera. When the detected marker moves to the right in the video, the 3DCGs move in the positive direction of the X-axis. When the shape of the marker changes, the angle of the 3DCGs from the virtual video changes by rotating the 3DCGs in the virtual space.

In our system, the video STHMD consists of an Oculus Rift CV1 immersive HMD and an Ovrvision Pro2 stereo camera. We used the Unity3 development platform to build the 3D virtual space for AR applications with the Oculus Rift CV1. We also employed the ArUco library to detect the marker position using the camera vision. From the evaluation results of the recognition accuracy of this marker library by Garrido-Jurado et al. [21], it is clear that the translation error of the marker position is small enough to be ignored when the distance is less than 0.7 m. Because the microphone array is held by hand, the distance is sufficiently close to the camera. The camera must be close to at least one spatial marker to ensure the accuracy of the camera position. Thus, the number of spatial markers will increase depending on the extent of the sound field to be visualized.

2.2. Visualization System with Optical STHMD and SLAM

2.2.1. Optical STHMD and SLAM

An optical STHMD is a stereo transparent display that is able to superimpose 3DCGs on the real world [13,22]. To naturally superimpose the 3DCGs, the optical STHMD requires tracking of the head.

Table 1 Configurations of two visualization systems.

<table>
<thead>
<tr>
<th>Section</th>
<th>Type of STHMD</th>
<th>AR/MR technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Video STHMD</td>
<td>Marker detection</td>
</tr>
<tr>
<td>2.2</td>
<td>Optical STHMD</td>
<td>Simultaneous localization and mapping (SLAM)</td>
</tr>
</tbody>
</table>

1 https://www.oculus.com/rift
2 http://ovrvision.com
3 https://unity3d.com
The simultaneous localization and mapping (SLAM) technique has been studied to determine the positional state of a robot and to construct a model of the environment with sensors such as a depth camera [23,24]. Recently, these techniques have attracted attention for realizing MR and AR without a marker. The SLAM technology enables changing the view of a 3DCG model according to the position orientation and the environment of the user, using only the sensors installed on a wearable or portable display. Thus, this technology acquires not only the shape, such as the surrounding walls and floor of the space, but also the position of the observer with respect to the recognized space.

2.2.2. System configuration

Figure 2 shows the visualization of the sound field information around a loudspeaker emitting noise using the proposed system with the optical STHMD and SLAM. After starting up our system, the SLAM sensors attached to the optical STHMD constantly acquire the shape of the space, as well as the position and view direction of the user. The measured shape of the space is represented as gray wire frames.

First, the user configures the reference position of the sound field 3DCGs by detecting a marker or by indicating the position with a gesture. Then, the system loads the 3D sound field information data and displays it as 3DCGs as if it exists in the real world. In the previous system using the video STHMD, the system always needed to detect the marker for head tracking. However, the system with the optical STHMD and SLAM first configures the reference position of the sound information simultaneously, after which it does not require any marker because the SLAM
achieves head tracking by measuring the surrounding space.

The 3DCG video displayed on the optical STHMD is generated in 3D virtual space, as shown in Fig. 2. When the SLAM sensors are activated, a spatial shaped wire frame is generated in this space. Furthermore, when the reference position of the sound field information is configured, the 3DCGs appear at the same position in virtual space with respect to the measured spatial shape. The virtual stereo camera moves in this space according to the movement of the user. The image of the virtual stereo camera is projected to the optical STHMD.

In our system, we used the Microsoft HoloLens, which is an optical STHMD with SLAM sensors. The HoloLens has a co-processor (Microsoft Holographic Processing Unit) to create a spatial map. The spatial map can be updated in approximately three seconds. The position and orientation of the user can be detected in real time with the spatial map and an inertial measurement unit, which includes a 3D gyro sensor and a 3D acceleration sensor. The positional information of the user can be updated in display frame rates of 30 fps or 60 fps. The spatial map is not updated frequently with respect to the display frame rate. However, once the spatial map is obtained, it is not necessary to update the spatial map frequently, unless the objects are moved in the room. The software for the system was made with Unity, which supported the development of this device.

3. 3D SOUND INTENSITY MAP MEASUREMENT SYSTEM WITH MARKER DETECTION

In the following section, we present our proposed interactive measurement system for the 3D sound intensity map using a handy four-point microphone and marker detection. The sound field can be measured and visualized simultaneously using both the proposed measurement system and the video STHMD visualization system presented in Sect. 2.1.

The sound intensity at each measurement point is represented by a conical 3DCG. The tip of the conical 3DCG shows the direction of sound intensity, and its color and height represent the sound intensity level.

3.1. System Configuration

Figure 3 shows the method for measuring and creating the sound intensity map around a loudspeaker with our proposed systems. The left pane of Fig. 3 shows the real-time visualization process in 3D virtual space. To obtain the position of the measurement point in the visualization target sound field, our measurement system requires two types of markers.

The first is a marker attached to a handy four-point microphone, which is known as the microphone marker. The microphone marker with respect to the camera from the detection of the pattern of this marker via the camera [19,20]. Thus, using the pre-measured positional relationship between the microphone marker and the measurement point, which is at the center of the four microphones, the position of the

---

4 [https://www.microsoft.com/hololens]
measured point with respect to the camera can be acquired as well.

The other type of markers is known as a “spatial markers,” which are arranged in the sound field to be visualized. During preparation, the positional relationship between the spatial markers is registered by the scanning camera. Then, the user configures the positional relationship between the spatial markers and the coordinate axis as the reference of the sound intensity map \((x, y, z)\). With this preparation, regardless of which spatial marker is recognized, the position of the intensity map reference coordinates \((x, y, z)\) can be obtained.

When these two types of markers are simultaneously detected in the video, the system can store the positional and rotational relationships between the measurement point and the intensity map coordinates \((x, y, z)\). At the same time, the system records four sound signals with the four-point microphone. The 3D sound intensities in arbitrary frequency bands can be calculated from the recorded signals, as shown in the Appendix. From the calculation results and the stored positional information, a colonial 3DCG object of the sound intensity can be created based on the sound intensity coordinates \((x, y, z)\). The direction component of the sound intensity rotates according to the saved rotation information [25].

The measurement is automatically repeated with time intervals of 0.3 s. Table 2 shows the processing time required for the measurement under various measuring conditions and frame rates for the video STHMD. As the real-time factor (RTF) is less than one in a time interval of 0.3 s, a real-time measurement can be performed.

Table 2 Information of the required time [ms] to measure at a point, real-time factor (RTF) when the measurement update time is set to 0.3 s, and video frame rates [fps] under various settings.

<table>
<thead>
<tr>
<th>Analysis Length</th>
<th>Sample Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>44,100</td>
</tr>
<tr>
<td>Time [ms]</td>
<td>RTF</td>
</tr>
<tr>
<td>2,048</td>
<td>102</td>
</tr>
<tr>
<td>4,096</td>
<td>147</td>
</tr>
<tr>
<td>8,192</td>
<td>238</td>
</tr>
</tbody>
</table>

Table 3 Comparison of sound field visualization systems.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sound field information</th>
<th>Measurement method</th>
<th>Display</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAE software &amp; systems, Bionic XS-56</td>
<td>2D sound pressure map</td>
<td>Handy microphone array</td>
<td>Single</td>
</tr>
<tr>
<td>VisiSonics, 5/64 Audio Visual Camera</td>
<td>2D sound pressure map</td>
<td>Deferred microphone array</td>
<td>STHMD</td>
</tr>
<tr>
<td>Scantec, SoftdB I-Track</td>
<td>2D sound intensity level or pressure map</td>
<td>Handy sensor scanning</td>
<td>Single</td>
</tr>
<tr>
<td>Microllown Tech., SCAN&amp;PAINT 3D</td>
<td>3D sound intensity or pressure map</td>
<td>Handy sensor scanning</td>
<td>Single cam</td>
</tr>
<tr>
<td>Our measurement system</td>
<td>3D sound intensity map</td>
<td>Handy sensor scanning</td>
<td>STHMD</td>
</tr>
</tbody>
</table>

When the four-point microphone is moved to a position far from the other measured points with respect to the preconfigured distance, the system automatically measures the intensity. Thus, by moving the microphone in the sound field only, a 3D sound intensity map is created immediately, provided that the sound field can be assumed to be stationary, or the emitting and recording sounds are synchronized. In addition, the movement of the microphone is assumed to be slow enough to ignore the Doppler effect.

Table 3 shows features of visualization systems proposed previously and our proposed system. Because the sound field is scanned by a handy microphone array, the sound field needs to be assumed to be stationary in our system. Compared with other systems, our system is the first to measure 3D sound field information and display it using a 3D display device.

In our system, the four-point microphone system is composed of an Audix TM1 measurement microphone and a MOTU 8M audio interface. All microphone signals are calibrated using a B&K Type 4231 sound calibrator. The four sound signals and the positional information of the measurement points can be saved in a file, and both proposed visualization systems in Sect. 2 can read this file.

4. EXPERIMENTS

Three experiments were conducted to evaluate our systems and to show their possible applications. The first experiment was a simple visualization of a well-known sound field emitted from two loudspeakers to evaluate the suitability of our system. In the second experiment, a leaked sound field, emitted from loudspeakers installed in a car, was visualized to show a possible application. Finally, in the third experiment, the noise from a car engine was visualized to show an example of a sound field not emitted from a loudspeaker. In each experiment, the set value was changed not to exceed RTF. Table 4 shows typical setting values.

Table 4 Typical conditions in all experiments.

| Distance between microphones [m] | 0.05 |
| Marker image size [m] | 0.05 x 0.05 |
| Microphone | Audix TM-1 |
4.1. Two Loudspeakers Emitting Band Noises

In the first experiment, we examined the consistency of the measurement system by visualizing the steady sound field of well-known sound sources. We visualized the sound intensity map around two full-range loudspeakers (YAMAHA MS101III) emitting different band noises. Two loudspeakers, A and B, were arranged in parallel 0.4 m apart, as shown in the upper row of Fig. 4, and generated 500 Hz and 1,000 Hz oct. band noises, respectively. The sound pressure level of the emitted noise was modulated to

![Fig. 4](image_url)

Visualization of the sound intensity map with proposed system around two loudspeakers emitting different noises in several analysis frequency ranges. All figures show right-eye view images. The color bar represents the sound intensity level range to be visualized and the colors corresponding to those values.
75 dB at 1 m in front of the loudspeaker. We measured 362 points by moving the microphone in 5-min intervals. The results were visualized in three analysis frequency ranges: 500-, 750-, and 1,000-Hz oct. bands. Further measurement conditions are shown in Table 5.

Figure 4 shows the four right eye view images using the two developed visualization systems in the respective analysis frequency bands. In both visualization systems, it was confirmed from the front and bird’s-eye views that the sound in the 500-Hz oct. band was emitted from loudspeaker A. Similarly, the high sound intensity from the front of loudspeaker B was observed in the 1,000-Hz oct. band. In the visualization of the 750-Hz oct. band, the source of the sound was both loudspeakers. Considering the output sound source, these results are correct.

In the visualization using a video STHMD and marker detection, all sound intensity 3DCGs were straightforwardly superimposed on the camera view. However, the 3DCGs on the opposite side of the loudspeaker can be confused with those in front of the loudspeaker, which is a problem, especially from the side and backside view in Fig. 4. Nevertheless, in the optical see-through system, it was possible to hide the 3DCGs on the opposite side of the loudspeakers from the viewpoint and display the 3DCGs in the front side only by using the shape of the loudspeaker, which is detected by the spatial detection sensors. Therefore, as the 3DCGs were obstructed by actual objects and walls, a significantly large amount of data can be observed without confusion.

One advantage of the measurement with video STHMD is the mobility of the microphone and the display. Compared with the measurement while displaying 2D information, it is more convenient to simultaneously measure and observe the sound field because the measurement results are confirmed via HMD in real time.

### 4.2. Sound Leakages with Loudspeakers in a Vehicle

An important application of sound visualization technology is the assessment of sound leakages from products or buildings. In this experiment, we visualized sound leakages from the side of a standing car (Subaru Impreza G4) with white noise emitted from loudspeakers in the vehicle. We measured the sound intensities at 1,087 measurement points for 20 min, using the video STHMD visualization system. The measured data were visualized with the optical STHMD system in four analysis frequency bands: 315-, 400-, 800-, and 1,000-Hz 1/3 oct. bands. Further measurement configurations and conditions are shown in Table 6.

Figure 5 shows three right-eye images via the optical STHMD visualization system in each analysis frequency band. The positions with strong sound leaks differ depending on the analysis frequency band. One of the places where the sound leaks strongly was around the door gap. In the 315-Hz 1/3 oct. band visualization, the sound leaked strongly from the door gap in the vicinity of the two woofers attached to the lower part of the car. In the 800-Hz 1/3 oct. band visualization, the sound leaked only from the vicinity of the front woofer. In the 1,000-Hz 1/3 oct. band, the sound leaks could be observed where the tweeter was attached to the dashboard of the car. Nevertheless, in the 400-Hz 1/3 oct. band visualization, the sound leaked from the gap between the doors, not from around the loudspeaker.

Comparing observation via the STHMD and the planar display, we empirically found that the STHMD allows us to understand the small-sized 3DCG objects of the sound intensities. In this experiment, the observational area was wider than that of the previous experiment in Sect. 4.1. Thus, as it is required to observe the sound intensity 3DCGs from a far viewpoint, several 3DCG objects may become small. However, it is difficult to see small objects with a general size flat display. The existence of STHMD can be conveniently perceived by binocular vision and can easily be observed with changing viewpoints if there is an concerning place.

### 4.3. Car Engine Sound

In the last experiment, we visualized the steady sound emitted from the engine of a car (Subaru Impreza G4). We measured the sound intensity map around the engine unit at 1,374 points with an opened car hood and at 1,462 points with a closed hood. The time for each measurement was 30 min. The measured data were visualized in three analysis frequency ranges: 250, 630, and 1,000-Hz 1/3 oct. bands. Further measurement configurations and conditions are shown in Table 7.
Fig. 5  Sound intensity map of sound leakage of a standing vehicle. The loudspeakers in the vehicle emitted white noise. All figures show right-eye view images of the proposed system of optical STHMD.
Fig. 6  Sound intensity map around the car engine rotating at 2,000 rpm. The upper and lower three rows are showing the respective cases when the car hood is opened and closed. All figures show right-eye view images of the proposed system of an optical STHMD.
Figure 6 shows right eye images in our optical STHMD visualization system from four different viewing angles and in four analysis frequency bands. When the car hood was opened, low-frequency sound was strongly emitted from the bottom of the car and high-frequency sound was emitted from the engine unit. As can be seen from the side view, the sound emitted from the engine was diffracted by the car hood.

When the car hood was closed, the sound intensity level was attenuated compared with the case when the car hood was opened. A lower-frequency sound was similarly leaked from the bottom of car and a higher-frequency sound was strongly emitted from the air intake section.

It should be noted that the stereoscopic view and the convenient viewpoint movement associated with the STHMD allow us to understand the highly dense sound intensity 3DCGs. An example of the sound field around the engine when the car hood is opened can be seen in Fig. 6. It is difficult to recognize sound intensity 3DCGs near the engine on a planar display because 3DCGs on the side closer to the observer hide the 3DCGs on the far side. When we used the STHMD, it was possible to differentiate between high-frequency sounds emitted from around the timing belt and low-frequency sounds emitted from the entire engine.

5. CONCLUSION

We proposed a 3D visualization system for a sound field using an STHMD. We developed two visualization systems and a 3D sound intensity map measurement system that can be used with the visualization system. It was confirmed that the binocular vision and movement of the viewpoint allow us to understand the 3D sound field information from three visualization experiments. Nevertheless, the measuring range of the measurement system using marker detection is limited to the range where markers can be recognized. Therefore, we created a sound field information mapping system with an optical STHMD visualization system. Because of the benefits of SLAM, the measurable range extended within the recognized spatial range. In addition, visualization of 3D sound field information other than sound intensity was taken into consideration.

ACKNOWLEDGEMENT

This research has been supported by the Kayamori Foundation of Informational Science Advancement.

REFERENCES

APPENDIX: Sound Intensity Calculation Using a Four-point Microphone

In general, the 3D sound intensity was measured using four or more non-coplanar closely located microphones [26], while another choice is an array of PU probes [27]. In this paper, we used four omni-directional microphones [26], while another choice is an array of PU probes [27]. In four or more non-coplanar closely located microphones was calculated by the cross-spectral method [28]:

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

(A-1)

where \( \rho \) is the atmospheric density, \( \Delta r \) is the distance between the two microphones, \( f_1 \) and \( f_2 \) are the lower and upper frequency of the analyzed frequency range, respectively, and \( S_{12} \) is the cross spectrum of two sound signals.

The four microphones can measure six sound intensities in the direction of all six pairs of the four microphones. These are shown by \( I_{01}, I_{02}, I_{03}, I_{12}, I_{13}, I_{23} \) in Fig. A-1. By synthesizing these six sound intensities, we calculated a 3D sound intensity at the center \( O \) of the four microphones. When defining the coordinates as shown in Fig. A-1, the 3D sound intensity \( \vec{I} = (I_x, I_y, I_z) \) can be calculated by

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

(A-2)

The norm of \( \vec{I} \) refers to the magnitude of the sound intensity. This can be converted to the sound intensity level as,

\[ L_I = 10 \cdot \log_{10} \frac{|I|}{I_0} \]  

(A-3)

where \( I_0 = 10^{-12} \text{ W/m}^2 \) is a reference intensity.

Fig. A-1 Image of the four-point microphone, which is one of the sound intensity measurement sensor. \( I_{01}, I_{02}, I_{03}, I_{12}, I_{13}, I_{23} \) represents the sound intensity between the two microphones. By combining these, it is possible to calculate the 3D sound intensity at center \( O \).