A study on the auralization system using flexible rendering for virtual reality-based acoustic simulation

Yeongchan LEE* and Jaeho RYU**

*Department of Design and Engineering, Graduate School of Nano IT Design Fusion,
SEOUL NATIONAL UNIVERSITY OF SCIENCE&TECHNOLOGY, 232, Gongneung-ro, Nowon-gu, Seoul, Korea
**School of Architecture,
SEOUL NATIONAL UNIVERSITY OF SCIENCE&TECHNOLOGY, 232, Gongneung-ro, Nowon-gu, Seoul, Korea
E-mail: jhryu@seoultech.ac.kr

Received: 18 March 2019; Revised: 18 June 2019; Accepted: 2 September 2019

Abstract
It is not easy for architects to consider the effect of the acoustic environment of a plan in the early design phase because it requires great expertise to properly use acoustic simulation software. The importance of a room’s acoustics has increased, so a solution that makes the accessibility of acoustic design work for architects is demanded. We developed a kind of acoustic simulation system that can provide a service for architects using virtual reality technology. In this paper, we described how to implement real-time sound representation with the system, which is connected to the visual data of a virtual reality headset and the acoustic simulation. Although there have been many studies using virtual reality technologies in acoustic simulation that have been conducted for a long time, some limitations have been unresolved, such as cost and portability problems. This study tried to realize flexible rendering using vector base amplitude panning (VBAP) and multi-channel speakers with an auralization method. In this paper, we propose a system that implements spatial sound around one person by using 11 speakers and VBAP, and we made a prototype to verify system function. We applied an acoustic ray tracing algorithm by using the Unity game engine and MAX digital signal processing software to control multiple speakers at the same time for VBAP. The hardware part of this system consists of an audio interface, two digital to analog converters (DAC), 11 mono amplifiers, and 11 full-range loudspeakers. The three-dimensional layout of the loudspeakers was based on an icosahedron to maximize the performance of VBAP. We tested the prototype of the system to verify its effectiveness and confirmed the change of sound pressure level according to the distance from the sound source and the sound absorption coefficient of the wall. In a reliability test of the performance of sound image localization by human cognitive ability, the result was a match rate of 89.4%, which proved the usefulness of this system for sound simulation for architects with low cost and easy portability.

Keywords : Virtual reality, Acoustic simulation, Auralization, VR headset, Vector base amplitude panning, Flexible rendering, Multi-channel speaker

1. Introduction

Although sound is invisible, it has a big effect on the health of the human body. So, people try to make a good acoustic environment to get rid of noise in their house and make a space to enjoy better sound. That is usually shown in a society that has a high average income, because the symptoms of diseases caused by noise do not appear immediately, and are not serious enough to endanger life. However, with the recent development of transportation and communication technology, the average global income level has risen, and so the number of people who know the importance of the acoustic environment is continuously increasing. The development of multimedia technology also makes people realize the importance of the acoustic environment of a room. One example is ultra-high definition (UHD). UHD technology, which was developed in 2005 has led to bigger TVs, and as a result, many speakers are
needed for the large “sweet spot” (Jang et al., 2015). In this way, as advanced technology is popularized, spaces must be able to accommodate it, but the performance of current buildings is not sufficient. In the case of apartment buildings, which occupy most of the urban housing types, it needs to be improved because vibrations are transmitted very well, and most importantly, architects must consider the acoustic design of a building from the beginning of architectural design. However, most architects still regard acoustic design as the engineer’s task, which is carried out after the design process is finished. Actually, it requires considerable expertise for a designer to analyze the results of acoustic simulation. This makes it difficult for architects to consider the acoustic environment in their usual work, except for particular cases such as theater design. Therefore, a solution is needed for architects that increases the accessibility of acoustic design work, and we tried to improve the existing acoustic design work to make it more intuitive by using both a head-mounted display (HMD) and virtual reality (VR) technology. The HMD for VR is based on the principles of stereoscopic depth perception, and it is called a VR headset. Compared to the CAVE system (Cruz-Neira, 1992), which uses a big screen wall, a VR headset is small and inexpensive, so it is very effective. In 2016, the Oculus Rift CV1 and HTC VIVE VR headsets, which have a wide viewing angle and many sensors, were released. They provide a better immersive experience than before, so they have been used in many types of research. We also recognized the ability of the VR headset and used it in this study.

This study is about a method of implementing real-time sound that is connected to the visual data of a VR headset by using a flexible rendering method and multi-channel speakers as an auralization method of acoustic simulation. This paper consists of five parts. In the first part, we derive the work process of the virtual reality-based acoustic simulation with a theoretical analysis of acoustic simulation and auralization. In the second part, we compare and analyze immersive sound technologies to choose a way to output the auralization. The third part describes the proposal that resulted based on the contents of the first and second parts. The fourth part is about the process of making the prototype. In the fifth part, we describe the test and its results.

2. Theoretical analysis

<table>
<thead>
<tr>
<th>Year</th>
<th>Author</th>
<th>Title</th>
<th>HMD</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1992</td>
<td>Gardner</td>
<td>A real-time multichannel room simulator</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4ch</td>
</tr>
<tr>
<td>1995</td>
<td>Huopaniemi</td>
<td>Real-Time Binaural Room Simulation</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>1996</td>
<td>Huopaniemi et al.</td>
<td>DIVA Virtual Audio Reality System</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>2004</td>
<td>Funkhouser et al.</td>
<td>A beam tracing method for interactive architectural acoustics</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>2007</td>
<td>Lentz et al.</td>
<td>Virtual Reality System with Integrated Sound Field Simulation and Reproduction</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4ch</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5ch</td>
</tr>
<tr>
<td>2008</td>
<td>Yang et al.</td>
<td>Development of three-dimensional sound effects system for virtual reality</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4ch</td>
</tr>
<tr>
<td>2014</td>
<td>Pelzer et al.</td>
<td>Integrating real-time room acoustics simulation into a CAD modeling software to enhance the architectural design process</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.1ch</td>
</tr>
<tr>
<td>2017</td>
<td>Poirier-Quinot et al.</td>
<td>EVERTims: Open source framework for real-time auralization in architectural acoustics and virtual reality</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>2018</td>
<td>Imran et al.</td>
<td>Auralization of Airborne Sound Transmission for Coupled Rooms in Virtual Reality</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>2018</td>
<td>Katz et al.</td>
<td>Objective and perceptive evaluations of high-resolution room acoustic simulations and auralizations</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>

Y: available(HMD), support(Reproduction) / N: not available(HMD), not support(channel)
In order to establish clear goals, we conducted literature surveys and analyzed previous studies related to our study. Table 1 is the result of that, and there were 11 cases except for overlapping ones among acoustic simulation-related studies that satisfied the keywords “real-time auralization” and “virtual reality” in Google Scholar. Recent studies have a tendency to be conducted by specific research groups using RAVEN (Pelzer et al., 2014) by RWTH Aachen University and EVERTims (Poirier-Quinot et al., 2017) from the IRCAM Institute. They have two common points. One is that working processes could be commonly divided into two parts, a ray tracing part and an auralization part. The other is that they used binaural rendering, which is one of the immersive sound technologies in sound representation.

### 2.1 Auralization process for virtual reality-based acoustic simulation

The purpose of acoustic simulation in architectural design is to check whether the expected performance of a sound field in space meets the first design goal and to prevent further losses afterward by fixing mistakes before the construction phase. The room impulse response (RIR) should be measured to visualize and analyze the sound field (Kang, 2012). The RIR is a chronological record of the energies of a sound wave that starts from a sound source reflected in space and reach a measurement point. With this, it is possible to know the straight distance from the sound source to the measurement point with direct sound measured first, and the distance to the nearest wall can be known through the early reflection sound. In the past, it measured the RIR directly by using a scale model, speakers, and microphones, but now the expected RIR is calculated by using a computing model. The calculation method of the RIR can be categorized into three parts, a ray-based method from geometric acoustics, a wave-based method, and statistical energy analysis for a high frequency model (Alpkocak and Sis, 2010). The first three-dimensional algorithm for a computer is the ray tracing method (RTM), one of the ray-based methods (Krokstad et al., 1968). Then, full implementation of the image source method (ISM) was proposed in (Allen and Berkley, 1979). Although they are the oldest among the methods, most of the software for architectural acoustic simulation still uses them by mixing them (hybrid method) because it has less of a computational burden. Fig. 1 shows the RTM and the ISM.

![Fig. 1](image)

Auralization is a technology that provides an experiencing of a sound field in an arbitrary position in a room model by computing the convolution of anechoic sound and the RIR (Amengual et al., 2016). Since it is necessary to know the RIR at a specific position for auralization, it is appropriate to use ISM, and the hybrid method that uses RTM in calculating late reverberation is used most. It is used in acoustic simulations essentially because it has the advantage that a non-specialist such as a client can listen to the expected sound of a designed space. However, real-time auralization techniques that fully interact with users are still under development. The above-mentioned RAVEN and EVERTims, and CATT-acoustic are software that support both real-time auralization and visual data in a space (Noisternig et al., 2008). They use binaural rendering for sound representation, and discussion about it is in the next part.
2.2 Immersive sound technologies and flexible rendering

Immersive sound technology refers to all the techniques that utilize the characteristics of hearing among methods that use the five senses to increase the immersive experience in virtual reality. Basically, using the factors that affect the perception of a space such as direction of sound, distance from sound source, reverberation time, etc., it controls the sense of reality of the listener in virtual reality. Immersive sound technology can be categorized into two parts by the number of channels used because it is based on existing 3D sound technologies. One is binaural rendering, which uses two channels, and the other is multi-channel using two or more channels. The multi-channel method described in this paper is a recent method that has improved the drawback of existing channel-based methods. It includes flexible rendering, speaker array rendering and ambisonics. Flexible rendering and speaker array rendering are theoretical classification names, and there are many ways to use them. Therefore, typical representative examples are used in this comparison, and those are vector base amplitude panning (VBAP) (Pulkki, 2001) and the beamforming system, which was developed by ETRI (Jang et al., 2015). Also, we exclude ambisonics from this comparison because it is specialized to record 3D sound for scene-base audio. Table 2 shows the results of comparing their pros and cons from a qualitative point of view. Binaural rendering has the disadvantage that the ears need to be covered ears, and the multi-channel method has the disadvantages that the sound is exposed to air and devices are difficult to carry. However, the most important thing in immersive sound technology is the ability to implement sound at the location where the listener feels it in virtual reality, and this is called sound image localization. In sound image localization, binaural rendering has the problem that there is an individual difference by using a head-related transfer function (HRTF) (Masterson, 2011), and VBAP also has the fundamental problem that a sound image is localized around the speaker (Gerzon, 1992). Although the overall performance of beamforming is excellent, much cost should be invested to use it well. In addition, since binaural rendering also causes problems in use by wrapping around the ears, we decided to use VBAP as the auralization method for this study.

![Fig. 2 Process of real-time auralization (Lee and Ryu, 2018).](image)

![Fig. 3 Binaural rendering (left) implements sound around the eardrums by using the HRTF. VBAP (center) implements the sound image by controlling the sound pressure level (SPL) of the three speakers. The beamforming system (right) from ETRI can implement a sound image around a listener by using techniques such as wave field synthesis based on Huygens' principle, focused sound, and beamforming using wall reflection.](image)
3. Proposal

Based on the contents of chapter 2, we propose a method of implementing real-time sound that is connected to the visual data of a VR headset by using flexible rendering, which uses VBAP and multi-channel speakers as an auralization method of acoustic simulation. The method can be divided into three parts. The first one is a process to calculate ray tracing and changes in the arrival time of a sound ray by changing the position of a sound source and a listener in virtual reality. The second one is a process that generates a digital audio signal for VBAP by multiplying the speaker power and each speaker’s distance ratio from the intersection of a triangle and a sound ray to each speaker. The third process is for converting the signal generated in the previous process into an analog signal through the hardware system and playing the sound in a real speaker system.

3.1 Software prototype

This prototype was devised as a popular platform that can be easily accessed by non-specialists for continuous development and convenience of modification. We used Unity 2018.2 as an authoring tool for virtual reality contents to implement ray tracing in virtual reality and connect with the VR headset. MAX 7 of Cycling 74 was used for the generation and transmission of digital audio signals for auralization. MAX is visual programming software that has a feature for digital signal processing (DSP), and it is optimized for music and media art.
3.1.1 Ray tracing and visualization using Unity

For real-time auralization and visualization from a first-person viewpoint by using the VR headset, the RIR must be calculated at the same rate as the refresh rate of the VR headset. So, it is difficult to use ways that require much calculation, and we used the ISM in this study. The working process for ray tracing simulation is as follows. (1) Defining the position of a sound source and a listener in a random room model. (2) Calculating the path of the direct sound. (3) Generating a mirror image of the sound source. (4) Calculating the reflection point and the mirror image-listener path. (5) Visualization of the sound source-reflection point-listener path. In Unity, scripts are needed to perform a specific command. Therefore, we wrote a script for the ray tracing algorithm using C# language. First of all, we implemented the path of the direct sound by using the vector that is made by the position variable of the sound source and the listener. The visualization of the path was implemented by using the line renderer provided as a component of Unity. Then, we generated the mirror image of the sound source. Implementing a mirror image of the sound source is where the sound source is copied at a symmetrical position with respect to each wall. If the space has a regular shape, each wall of the room is parallel to the X, Y, and Z axes. So, we defined the copy of the sound source as a child object of a wall object, which is the criterion of mirroring, and inputted -1, the value that reverses the local scale for the parent object. If the wall object isn’t parallel to any axis, such as an irregular shape, it can be solved by moving the object to the direction of the normal vector of the wall by a distance perpendicular from the object to the wall. Lastly, we needed to find the reflection point on the wall. The point is the intersection of the wall and the line that connects the position of the mirror image to the position of the listener. Therefore, we generated an infinite line, the starting point of which was the position of the mirror image and the direction from the mirror image to the listener, and we defined the crash point of the line and the wall as a reflection point. The sound source-reflection point-listener path could be generated by using it, and we could calculate the arrival time of a reflected sound by using the length of the path and the speed of sound.

![Fig. 5](image_url) The upper image show the results of implementation of the first-order reflection. The bottom left image shows the way of mirroring in the case of a room that has an irregular shape. The bottom right image is implementation it.
3.1.2 Controlling design variables for architects

The formula for calculating the arrival time of a reflection sound is Eq. (1). \(S\) is the arrival time of an acoustic ray, \(D\) is the distance of the ray’s path, and \(V\) is the sound speed. Since the change of temperature in the room is not dramatic, the sound speed can be considered a nearly fixed value. However, since the distance is proportional to the size of the space, the difference in arrival time between the sound rays occurs due to the size of the space and reverberation occurs. Because the size of the space can continuously change during the architectural design process, the sound field of the space will constantly change, and we implemented the function to confirm this in our prototype. The real-time distance of the path is the sum of the real-time distance from the sound source to the reflection point and the real-time distance from the reflection point to the listener. The real-time arrival time of an acoustic ray can be calculated by specifying an arbitrary room temperature value and connecting the distance value. In Unity, the size of the space can easily be changed, so the user can easily understand the relationship between space and arrival time.

\[
S = \frac{D}{V}
\]  
\(S\): Arrival time, \(D\): Distance, \(V\): Sound speed = 331.5+0.6t (m/s), \(t\) = Celsius temperature

The surface material of a space also changes the sound field. Every material has its own sound absorption rate, and the reflection of sound can be controlled by changing the material of a part such as a wall. The sound absorption rate is the result of quantifying the degree of sound energy that is absorbed by a specific material. In the ISM, the attenuation of the spherical wave is applied because sound moves from a mirror image that is generated by a wall to the listener’s position. Therefore, the inverse square law should be applied to the entire path (Kim, 2008). Equation (2) is the formula for calculating the application of the sound absorption rate, and Eq. (3) is for calculating the sound pressure level (SPL) at the specific position. With it, this prototype can calculate attenuation in the simple condition.

\[
SPL(\alpha) = SPL \times (1 - \alpha) \tag{2}
\]
\(SPL\): Sound pressure level (dB), \(SPL(\alpha)\): Attenuated SPL (dB)
\(\alpha\): Sound absorption rate

\[
SPL = PWL + 10 \log \frac{Q}{4\pi r^2} \tag{3}
\]
\(SPL\): Sound pressure level (dB), \(PWL\): Sound power level (dB)
\(Q\): Directivity factor (\(Q=1\) free-field, \(Q=2\) single reflecting surface, \(Q=4\) two reflecting surfaces, corner)
\(r\): Distance from source (m)
For accurate sound image localization, we had to know the location of the actual speaker in virtual reality. Because the VR headset is the only one that connects reality to virtual reality in this prototype, a method to recognize the location of the actual speaker by using the position of the VR headset was needed. So, we made the virtual structure the same size as the real one, and it was subordinate to an object of the VR headset. However, we designed it so that the initial data was continuously used because the actual structure doesn’t move when the VR headset rotates. Therefore, when this prototype is applied, the direction and position of the VR headset object in virtual reality should match the real one before operation. The sound image is implemented with the VBAP technique on each face made by the linear members of the prototype. Because we simplified the sound waves into rays, we had to find the intersection point of each face of the prototype and the ray and convert it to a value with respect to the power of the speakers in each corner of the triangle. We applied the intersect triangle function, which is frequently used to implement the action for object detection, to the prototype. By using it, we could convert the intersect point to the values $u$, $v$, and $r$, which are expressed as a percentage of the distance from each corner of the triangle (Fig. 7).

\[
T \quad [r=1-(u+v)]
\]

\[
M_2 \quad [v=1]
\]

\[
M_1 \quad [u=1]
\]

Fig. 7 Conversion of intersection points to $u$, $v$ values. The colored balls are the intersection points.

### 3.1.3 Digital signal processing using MAX

To send the data in Unity to MAX in real time, communication between the two programs should be smooth. Communication between Unity and MAX was implemented by using open sound control (OSC). OSC is a protocol for real-time communication between a computer and a sound synthesizer. It uses the user datagram protocol (UDP), and it was developed by the UC Berkeley Center for New Music and Audio Technology (CNMAT). Since an open source library for the MAX platform already existed, we used it.

Fig. 8 A sample image of the data transmission between Unity and MAX. It shows that $u$ and $v$ are transmitted from Unity to MAX through UDP communication, and the data in Unity can control the SPL of a speaker.
3.2 Hardware prototype

This prototype uses many channels at the same time. Therefore, we used an Alesis digital audio tape (ADAT), which can send eight channels at once, and an RME Digiface USB was used as an audio interface. For the digital to analog converter (DAC), a Beringer ADA8200 preamp was used. Eleven 4Ω, 6w (root mean square or RMS) full-range speaker modules were used. To match the number of speakers, 11 mono amplifiers were used, and they output a maximum of 18w by using 6~12V.

Fig. 9 A diagram for controlling the SPL of each speaker. The block at the top of the image receives data from Unity. It defines the port for receiving the data. The address of each triangle is written in the block below, and it sends $u$, $v$, and $r$ to each block, which controls the power of a signal. This continuously changing signal is synthesized with source sound, and it is sent to each speaker.

Fig. 10 Components of the prototype. Top left: audio interface; Top center: loudspeaker; Bottom left: DAC; Bottom center: mono amplifier; Right: the circuit to connect the amplifier to the RCA cable from the DAC.
To optimize the performance of VBAP, the speaker system must have a triangular configuration that is structurally stable. Therefore, an icosahedron, that has the shape closest to a sphere, was chosen from among the convex deltahedra of which all the faces are equilateral triangles. A sphere is structurally stable because it has the smallest surface area. Also, the range of motion of the human body is close to a sphere. Therefore, it is reasonable to use the shape closest to a sphere. Table 3 shows the results of analyzing the convex deltahedra mathematically. Inscribed sphere refers to the range of motion of the human body. In the case of the deltahedron, which doesn’t have an inscribed sphere, we assumed that the length of a side of a triangle is 10 and made the largest sphere that does not arbitrarily invade the shape. We calculated the volume ratio of each deltahedron to their inscribed sphere, and the degree that is close to the sphere was represented numerically.

<table>
<thead>
<tr>
<th>Image</th>
<th>Volume of inscribed sphere</th>
<th>Volume of deltahedron</th>
<th>Ratio (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tetrahedron</td>
<td>$\frac{4}{3} \pi \left(\frac{1}{2\sqrt{6}}a\right)^3$</td>
<td>$\frac{\sqrt{2}}{12}a^3$</td>
<td>30.22</td>
</tr>
<tr>
<td>Octahedron</td>
<td>$\frac{4}{3} \pi \left(\frac{1}{\sqrt{6}}a\right)^3$</td>
<td>$\frac{\sqrt{2}}{3}a^3$</td>
<td>60.45</td>
</tr>
<tr>
<td>Icosahedron</td>
<td>$\frac{4}{3} \pi \left(\frac{1}{8\sqrt{3}}a\right)^3$</td>
<td>$\frac{3 + \sqrt{5}}{12}a^3$</td>
<td>82.87</td>
</tr>
<tr>
<td>Triangular bipyramid</td>
<td>82.59</td>
<td>235.70</td>
<td>35.04</td>
</tr>
<tr>
<td>Pentagonal bipyramid</td>
<td>301.54</td>
<td>603.00</td>
<td>50.00</td>
</tr>
<tr>
<td>Snub disphenoid</td>
<td>370.96</td>
<td>859.49</td>
<td>43.16</td>
</tr>
<tr>
<td>Triaugmented triangular prism</td>
<td>479.45</td>
<td>1140.12</td>
<td>42.05</td>
</tr>
<tr>
<td>Gyroelongated square dipyramid</td>
<td>810.87</td>
<td>1428.79</td>
<td>56.75</td>
</tr>
</tbody>
</table>

The prototype was designed considering the range of arm motion of a person 170 to 175 cm tall, which is the average height of a Korean male aged 25 to 34. This methodology isn't just for people in a specific group, so we tried to follow the average. This prototype is for testing the change of sound image by movement in virtual reality, so the subjects have to recognize changes in the center of the prototype where all sounds reach uniformly. Only hand controllers are used for movement in virtual reality. Therefore, we designed a shape that satisfies both the center height considering the average male height and the longest length when he spread his arms (Fig. 11). Actually, the size of the prototype isn't important. Even if the size is changed, the location of a sound image doesn't change because the path of the sound doesn't change. The size is only to control the number of people who can use it at the same time. The shape of the prototype is a result of considering not only the user, but also the ease of changing the system size. The prototype basically follows the icosahedron, so all the parts are connected organically. Therefore, the size of parts can be calculated mathematically, and that makes it easy for enlargement and reduction in design.
3.3 Comparison with other conventional method

To gain an objective understanding of the prototype's performance, we compared it with other conventional methods such as aixCAVE and Virtual Acoustics (VA) applications of RWTH Aachen University because those have been used in many research studies related to architectural acoustics. Because it was difficult to compare the user experience of each device directly, we referred to related papers and the website of the RWTH Aachen University IT Center. As the research achievements of RWTH Aachen University, both the aixCAVE and VA applications have the same work process in the sound reproduction of virtual acoustics, and they are as follows; (1) The binaural room impulse responses are calculated by real-time room acoustics simulation software. (2) The binaural room impulse responses and virtual sound source are convolved, and it is converted to a binaural stimulus. (3) With the playback system, the sound reproduction system is completed. (4) If it is played in the speaker system, it goes through the process of applying crosstalk cancellation (Vorländer, 2014).

aixCAVE is one of the CAVE systems that is a representative immersive virtual reality device, and it is the biggest CAVE system in existence. The shape of aixCAVE is a cube of which five sides except the top side are a projection screen, and the initial sound system was a four-channel speaker system located at each top corner of the cube (Schröder, 2010). The new model in 2012 was bigger, and eight IR-optical tracking cameras and automatic sliding doors were installed. The sound system was improved in 2015. Twelve loudspeakers, nine subwoofer speakers, 22 microphones and a PU block for absorbing reflection sound were installed at the top of that (Wefers et al., 2015). Compared with aixCAVE, the prototype has the advantage of portability and lower production cost. Although we did not install a floor speaker in this prototype for convenience of the test, it is possible to implement spatial sound that surrounds the user with 12 speakers including a floor speaker. The speaker system of aixCAVE was only installed on the ceiling. Therefore, it is hard to make the effect of sound images that hover around the user, and it will reduce the immersion. If it uses the headphones as the sound system, the disadvantage is resolved. However, immersion reduction with the absence of the top side screen and the individual difference by the use of binaural rendering still remain. Table 4 is a comparison of the hardware specifications.

VA is software for a virtual acoustics environment created at the Institute of Technical Acoustics (ITA) at RWTH Aachen University. VA is a real-time auralization framework for room acoustics, and it supports various programming languages and connection with Unity for ease of use. Aspöck and his team introduced two cases of applications using the HMD and several loudspeakers in their paper (Aspöck et al., 2018). No details were given, but the pictures and their description helped us to infer its performance. One example is a system using the HMD's headphones and one subwoofer speaker. The other is a system using one projection screen and loudspeaker arrays that surround the user.
The former has an individual difference problem. However, the latter is similar to our prototype, and the method that uses one projection screen can decrease visual immersion and it cannot confirm whether sound image localization is possible in the gaps among the loudspeakers.

4. Test

To verify the effectiveness of the proposed method, we conducted two tests by using the prototype. One was to verify the reliability of the prototype, and the other was an evaluation of the sound image localization performance by human perception. The tests were conducted in a general room without a sound absorbing facility. For the experimental equipment, the prototype, an Oculus Rift VR headset, and a desktop computer that has Intel i7-8700 CPU, GTX 1060 GPU, DDR4 16Gb RAM, and Microsoft Windows 10 OS were used.

4.1 Prototype performance test

We used a Uni-Trend UT353 sound level meter to measure the SPL. The sound source that was used in this test was a beep sound that did not change in loudness. We measured the SPL at a distance of 5 cm from the front of the speaker because the probability of interference by reflected sound increases as the distance from the speaker increases.
We arbitrarily chose three speakers and measured the change of the SPL of an actual speaker when the speaker power value is changed in MAX. The data, minus the average SPL of the laboratory from the measured SPL, was used in the analysis. Figure 13 shows that the change in MAX is applied almost equally to all speakers.

![Figure 13](image)

**Fig. 13** SPL change of an actual speaker by changing the speaker power value in MAX.

### 4.2 Evaluation of sound image localization performance by human

An evaluation test was conducted with 10 people aged 25 to 35 who were familiar with VR technology. Because the test was not for evaluating the resolution, that is, the ability to find the accurate location of the sound image, we chose anyone who did not have serious visual or hearing impairments that interfered with everyday life. For accurate measurement of the sound image localization performance, we didn’t simulate other reflections except for the direct sound because sound reproduction by the side and rear reflection sound may interfere with this test. The evaluation method is as follows. (1) The system makes a sound image at the center of one triangle among all 15 triangles in the
prototype. (2) The subject points to the triangle where they think the sound image is located. (3) It shows the degree that is close to the triangle where the sound image was located as a score. The score is from 0 to 5 points (Fig. 14). The difference between 4 points and 3 points is the difference in the number of shared speakers. The triangle of the 4 points shares two speakers with that of the 5 points, while that of the 3 points shares one speaker. For the scoring to yield meaningful results in the test, we prevented the subjects from facing each other and randomized the order of the triangles where the sound image was located. Table 5 shows the results of the evaluation, and there was an 89.4% match rate.

<table>
<thead>
<tr>
<th>Age</th>
<th>Gender</th>
<th>Height (cm)</th>
<th>Triangle number (score)</th>
<th>Total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Male</td>
<td>25</td>
<td>175</td>
<td>5(3) 3(5) 5(5) 8(5) 4(5) 2(5) 11(5) 5(4) 3(5) 1(5) 47</td>
</tr>
<tr>
<td>B</td>
<td>Female</td>
<td>28</td>
<td>156</td>
<td>6(5) 3(5) 5(5) 8(5) 4(5) 2(5) 11(5) 6(4) 12(3) 1(5) 47</td>
</tr>
<tr>
<td>C</td>
<td>Female</td>
<td>29</td>
<td>158</td>
<td>6(5) 11(4) 15(4) 8(5) 14(3) 8(3) 11(5) 15(5) 3(5) 1(5) 44</td>
</tr>
<tr>
<td>D</td>
<td>Female</td>
<td>35</td>
<td>150</td>
<td>6(5) 10(3) 14(3) 8(5) 13(4) 8(3) 11(5) 15(5) 10(3) 8(3) 39</td>
</tr>
<tr>
<td>E</td>
<td>Male</td>
<td>31</td>
<td>180</td>
<td>6(5) 12(3) 5(5) 8(5) 4(5) 1(4) 11(5) 15(5) 3(5) 8(3) 45</td>
</tr>
<tr>
<td>F</td>
<td>Male</td>
<td>32</td>
<td>174</td>
<td>6(5) 13(3) 5(5) 8(5) 4(5) 9(4) 10(4) 15(5) 10(3) 2(4) 43</td>
</tr>
<tr>
<td>G</td>
<td>Male</td>
<td>32</td>
<td>173</td>
<td>6(5) 3(5) 5(5) 8(5) 4(5) 1(4) 11(5) 15(5) 10(3) 1(5) 47</td>
</tr>
<tr>
<td>H</td>
<td>Male</td>
<td>29</td>
<td>183</td>
<td>6(5) 3(5) 5(5) 8(5) 13(4) 9(4) 11(5) 15(5) 11(4) 2(4) 46</td>
</tr>
<tr>
<td>I</td>
<td>Male</td>
<td>32</td>
<td>177</td>
<td>5(3) 3(5) 5(5) 8(5) 4(5) 9(4) 11(5) 15(5) 3(5) 1(5) 47</td>
</tr>
<tr>
<td>J</td>
<td>Female</td>
<td>27</td>
<td>168</td>
<td>5(3) 10(3) 15(4) 8(5) 13(4) 1(4) 11(5) 15(5) 11(4) 1(5) 42</td>
</tr>
</tbody>
</table>

5. Conclusion

In this study, we have proposed a method that uses an authoring tool for virtual reality contents, an auralization engine, and VBAP to check the possibility of flexible rendering in a virtual reality-based acoustic simulation. A prototype was made by using Unity, MAX, and an 11ch speaker system. Then we compared it with other conventional methods, aixCAVE and VA applications at RWTH Aachen University. Lastly, we conducted an evaluation of the sound image localization performance by human perception, and the following conclusions were obtained by using the prototype.

1. It can calculate the ray tracing of the direct sound and the first order reflection sound by using ISM.
2. It can calculate a sound effect such as an arrival time difference, absorption, and attenuation.
3. It can generate a digital audio signal for auralization using VBAP.
4. The sound image localization performance of the direct sound is reliable.
5. The shape of the prototype is optimized for the simulation method that uses the VBAP technique and makes it easy to get enlargement and reduction in design.
6. Compared with aixCAVE, the prototype has the advantage of portability and requires less production and can implement spatial sound that surrounds the user.

However, there are still unsolved problems in this study. Problems such as the process that matches the direction and position of the VR headset object in virtual reality with the real one and the situation where more than one ray is located in one triangle need to be researched further. The conclusions show that the method we proposed is appropriate in a free sound field simulation, and we have verified the effectiveness of this study.

Table 5  Results of evaluation of sound image localization performance by human.
Acknowledgement

This work has been conducted with the support of the “Project for Nurturing Advanced Design Professionals,” a R&D project initiated by the Ministry of Trade, Industry and Energy of the Republic of Korea.

References

RWTH Aachen University IT Center, aixCAVE at RWTH Aachen University, RWTH (online), available from <http://www.itc.rwth-aachen.de/cms/IT-Center/Forschung-Projekte/Virtuelle-Realitaet/Infrastruktur/~fgqa/aixCAVE?tidx=1>, (accessed on 15 June, 2019).