INTERACTIVE SOUND FIELD SIMULATION IN A MULTI-SCREEN IMMERSIVE PROJECTION DISPLAY

Tetsuro Ogi[1,2], Takuro Kayahara[1,2], Masafumi Kato[2], Hiroshi Asayama[3] and Michitaka Hirose[2]

[1] Telecommunications Advancement Organization of Japan, tetsu@iml.u-tokyo.ac.jp
[2] The University of Tokyo
[3] Real Wave Research Japan

ABSTRACT

This paper describes the interactive sound field simulation method that can control the virtual sound field interactively in the multi-screen immersive projection display. In this method, convolution filters are replaced in real-time using the multi-channel digital signal processor, and the simulated sounds are displayed using the 16-channel speaker system. This system was applied to the communication system using the shared virtual world, and the effectiveness of this method was evaluated.

1. INTRODUCTION

Multi-screen immersive projection displays such as the CAVE or CABIN have become very popular as virtual reality visual display systems [1][2]. These displays can generate high presence virtual worlds by projecting stereo images onto the surrounding wide screens. However, a standardized virtual reality acoustic display technology that generates a high presence sound field has not yet been established. In particular, development of a high presence acoustic display system that can be used in conjunction with a multi-screen immersive projection display is desired.

As typical acoustic display technology that is used in the existing virtual reality systems, HRTF (Head Related Transfer Function) is often used [3][4]. This method can localize the virtual sound interactively by convoluting the sound source signal with the HRTF data in real-time. However, this method has limitations in displaying high presence sensation, because it cannot represent an influence of the sound reflection against the wall or the floor in the virtual world.

On the other hand, a numerical simulation method based on the wave equation is used to generate a high presence sound field that includes the influence of the sound reflection [5][6]. Since this method calculates accurate acoustic influence using the numerical calculation, it is often applied to the architectural acoustics design such as designing concert hall or theater. However, this method cannot be used to simulate a sound field that is changed interactively, because it requires a large amount of calculation time.

The purpose of this study is developing an acoustic display technology that can simulate an interactive sound field that includes an influence of the sound reflection in the virtual world. In addition, this method is required to be used with the multi-screen immersive projection display in order to generate high presence virtual worlds.

In this study, a high presence interactive sound field simulation method that can be used with an immersive projection display was developed. This method can generate an interactive sound filed by replacing the convolution filters in real-time. This paper describes the principle of the proposed method, implementation to the multi-screen immersive projection display, experimental evaluation of the prototype system and the application to the virtual reality applications.

2. INTERACTIVE SOUND FIELD SIMULATION

2.1 Basic Principle

In the proposed method, the sound field of the virtual world is calculated by solving the Kirchhoff's integral equation that is formulated from the wave equation. In order to solve the Kirchhoff's integral equation, several numerical calculation methods such as finite element method or boundary element method have been proposed. In this study, the finite sound ray integration method that is one of the approximate calculation methods was used [7]. In this method, the sound wave is approximated by a large number of sound ray vectors radiating in all directions from the sound source, and the sound field is calculated by integrating the influence of the sound reflection against the walls and the floor for all sound rays.

The simulated sound field is reproduced using a multi-channel speaker system so that the users in the multi-screen immersive projection display can experience the virtual sound simultaneously. In this method, the calculated velocity potential for each sound ray is divided among a set of speakers that are placed around the arriving sound ray. By integrating the velocity potential for all sound rays, the impulse response for each speaker is calculated as shown in Fig.1. This impulse response is convoluted with the sound source signal, and then the sound field around the audience can be simulated.

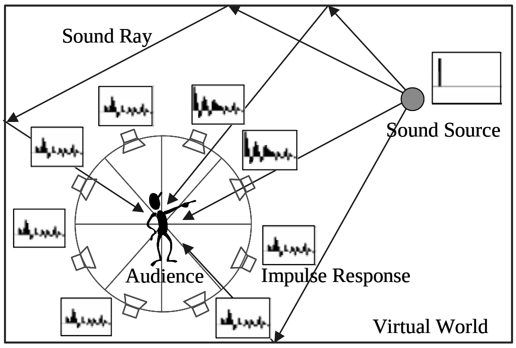


Fig.1 Basic Principle of Sound Field Simulation Method

In this method, since the impulse response is calculated for the specific positional relationship between the sound source and the audience, it must be calculated for each position relationship when the sound source or the audience moves in the virtual world. However, this calculation requires a large amount of calculation time, and the real-time calculation is impossible. Therefore, in this method, the impulse response is calculated beforehand for every possible position of the sound source and the audience, and it is changed dynamically to realize the interactive sound filed simulation.

In the practical use, the virtual world is divided into meshes, and both the sound source and the audience are placed on the grid points. The impulse response is calculated between each two grid-points selected for the sound source and the audience positions. Then, the positions of the sound source and the audience can move freely by using the nearest impulse response data to their present positions.

2.2 Multi-channel Digital Signal Processor

In order to realize the above-mentioned simulation, it is necessary that the multi-channel filters of the impulse responses are convoluted with the sound source in real-time, and the filter data are replaced interactively according to the change of the positional relationship between the sound source and the audience. In this study, a multi-channel digital signal processor named Wave Engine was developed to meet these requirements.

Figure 2 shows the appearance of the Wave Engine and Fig. 3 shows the construction of functions. In this system, one digital sound is inputted and it is convoluted with the sixteen multi-channel filters. Each channel consists of two kinds of filters that are connected in series, namely 32 tap multi-tap filter and 512 tap FIR filter. The multi-tap filter is used for the convolution calculation of the impulse response, and the FIR filter is used for the compensation of the screen attenuation that is discussed in section 3.2. In addition, since two convolution filters are connected to one output channel, the control of the direct sound and the synthesis of the reflection can be calculated individually.

Fig.2 Multi-channel Digital Signal Processor: Wave Engine

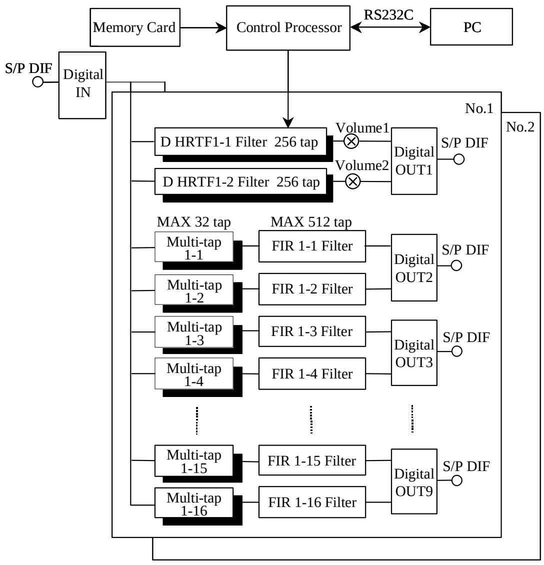


Fig.3 Construction of Functions of the Wave Engine

The multi-tap filter can store about 3,000 convolution filter data in its memory, and the available filter can be replaced dynamically. When it is used to simulate an interactive virtual sound field, several impulse response data for various positions of the sound source and the audience placed on the grid-points are stored in it. This filter data is replaced using the cross fading in order to change the sound field smoothly without discontinuous noise.

Figure 4 shows the system construction of the virtual reality acoustic display system using the Wave Engine. The sound source is inputted to the Wave Engine through the A/D converter as digital data, and the 16 channel output sounds are transmitted to the speakers through the D/A converters. The Wave Engine is connected to the PC through the RS-232C, and the change of the convolution filter data is controlled by the command transmitted from the PC. In the virtual reality system, this PC is communicated with the graphics workstation that generates computer graphics images through the Ethernet, so that the control of the virtual sound can be synchronized with visual images.

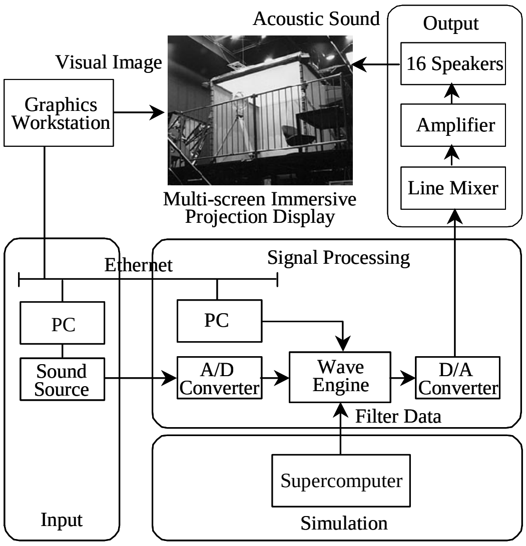


Fig.4 System Construction of the Virtual Reality Acoustic Display System Using Wave Engine

3. IMPLEMENTAION IN THE CABIN

3.1 Speaker Arrangement

In this study, the interactive sound field simulation method was implemented in the multi-screen immersive projection display known as CABIN. The CABIN is a CAVE-like display that has five screens at the front on the left, right, ceiling and floor. Though this sound simulation method uses a multi-channel speaker system, these speakers must be placed at the positions where they would not disturb the projection images.

In this system, sixteen speakers were placed around the screens outside the projection area. Figure 5 shows the multi-channel speaker system used in the CABIN, and figure 6 shows the speaker arrangement of this system. Eight speakers are placed at the height of the user's ear, and four speakers are placed at the center of each side of the ceiling screen and the floor screen respectively.

Each speaker is assigned to each channel of the Wave Engine, and is used to output the simulation sound. In this system, the closed-space around the user is divided into triangular meshes by the positions of the speakers, and the velocity potential calculated for each sound ray is divided among three speakers that are placed around the arriving sound ray.

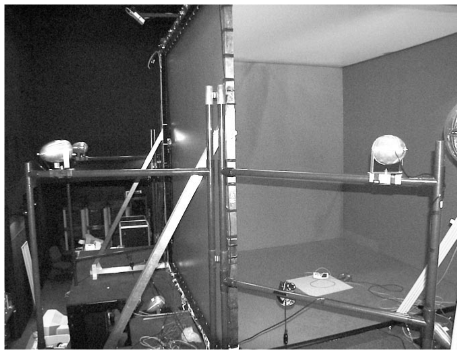


Fig.5 Multi-channel Speaker System Used in the CABIN

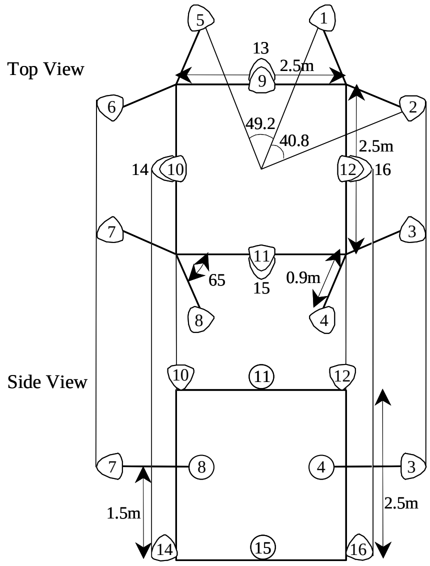


Fig.6 Speaker Arrangement of the Multi-channel Speaker System

The division rations of α_1, α_2 and α_3 among three speakers are determined so that they satisfie the following equation.

$$x = \alpha_1 s_1 + \alpha_2 s_2 + \alpha_3 s_3 \qquad (1)$$

where s_1, s_2, and s_3 are position vectors from the position of the audience to the three speakers, and x is a vector to the intersection point between the arriving sound ray and the plane that contains the three speakers (Fig.7).

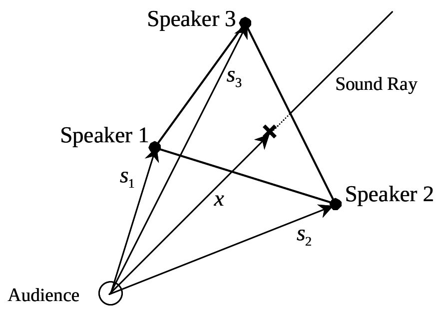


Fig.7 Division of Velocity Potential among Three Speakers

3.2 Compensation of Screen Attenuation

In the above mentioned speaker arrangement, the influence of the sound attenuation due to the screen is an unavoidable problem, because the speakers are placed behind the screen. Therefore, in order to generate an accurate sound field around the user, it is necessary to compensate the influence of the screen attenuation.

In this study, the impulse response in the display space of the CABIN was measured, and the inverse function was used for the compensation filter. However, the impulse response measured in the CABIN contains the influence of the sound reflection between screens as well as the sound attenuation. Since this influence depends on the measurement position, the impulse response measured at the user's position would also change when he moved in the display space.

Therefore, in this study, the impulse response was measured immediately after passing through the screen at the position between the user standing at the center of the CABIN and each speaker. From this data, only the direct sound component was extracted and it was used to make a compensation filter. Figure 8 is an example of the impulse response data measured after passing through the screen, and Fig. 9 is a compensation filter that was made from the direct sound component of the measured data. When this compensation filter was used, the frequency characteristic of the impulse response shown in Fig. 10 was measured.

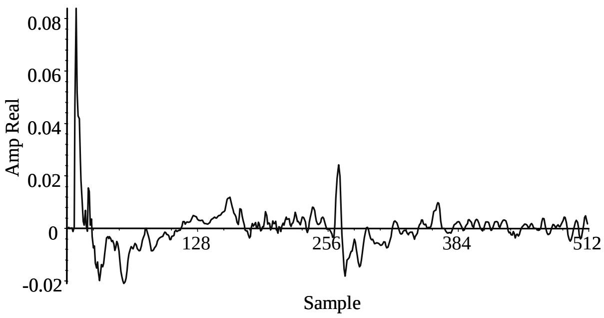


Fig.8 Example of Impulse Response Data Measured after Passing through the Screen

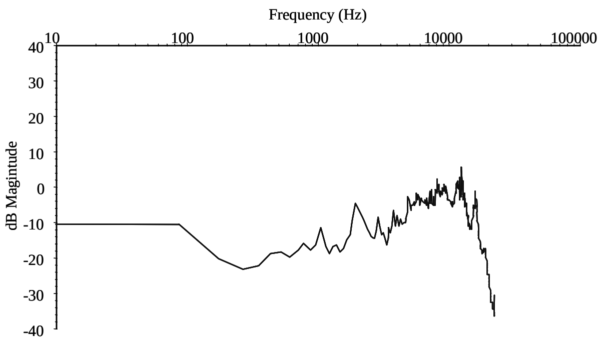


Fig.9 Compensation Filter Made from the Measured Impulse Response Data

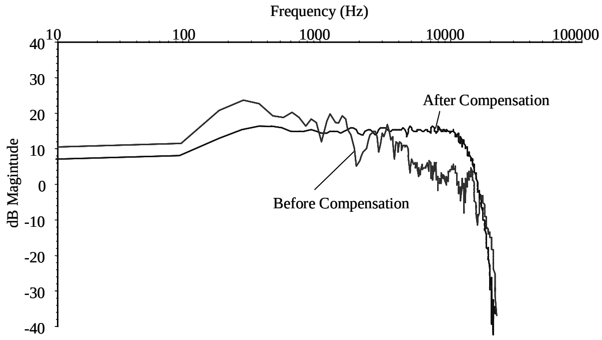


Fig.10 Frequency Characteristic of the Impulse Response Using the Compensation Filter

In this graph, a flat gain characteristic is shown, although the frequency characteristic was disordered when the compensation filter was not used. Therefore, we can understand that the sound attenuation was effectively compensated, by using the inverse function filter of the impulse response.

4. PERCEPTION EXPERIMENT

In this system, high presence virtual sound field can be changed interactively in the multi-screen immersive projection display. However, in order to generate a practical virtual world, we must discuss how many convolution filters are necessary to move the sound source smoothly and how naturally the simulated sound field can be changes. In this study, we conducted an experiment to evaluate the smoothness of the movement of the virtual sound source, when the number of the convolution filters was changed or the speed of the moving sound source was changed.

In this experiment, we modeled a virtual room with dimensions of 10m by 10m by 3m as shown in Fig. 11, and a virtual sound source was moved within it along the horizontal direction at distances of 5m and 1m in front of the subject. The subjects were asked to stand near the wall to investigate the influence of the sound reflection.

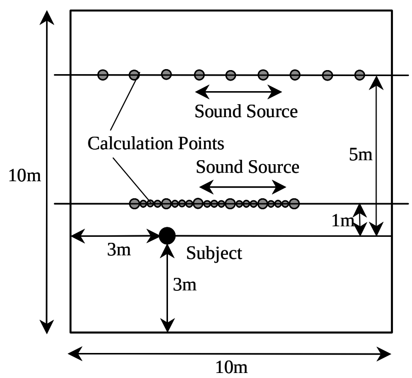
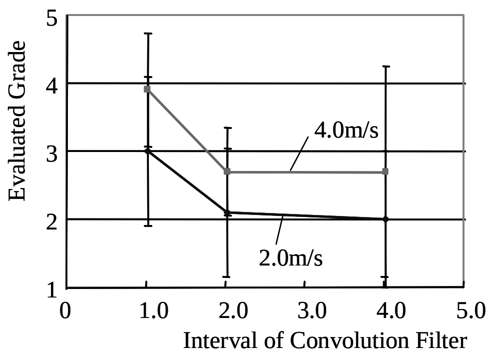


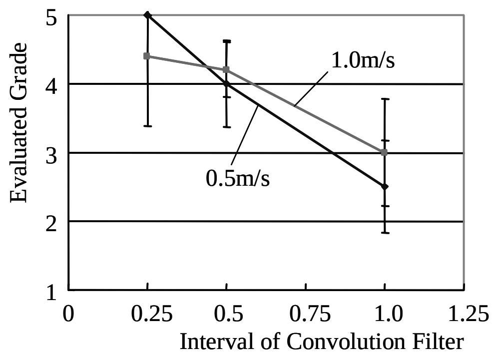
Fig.11 Virtual Room Used in the Perception Experiment

When the sound source was moved at the distance of 5m, convolution filters were prepared at various intervals of 1.0m, 2.0m and 4.0m, and the sound source was moved at speeds of 2.0m/s and 4.0m/s. On the other hand, when it was moved at the distance of 1m, convolution filters were prepared at intervals of 0.25m, 0.5m and 1.0m for the sound source moving at 0.5m/s and 1.0m/s. The subjects evaluated the smoothness of the change in the sound field using a five-grade system for each experimental condition. The number of subjects was five, and the virtual sound source was displayed twice for each experimental condition in random order.

Figure 12 shows the results of this experiment. In these graphs, average values and standard deviations of the evaluated grades for each experimental condition are shown. In this experiment, the subjects could fell smoother movement of the sound source, when the interval of the prepared convolution filters was shorter. In addition, when the virtual sound source was located farther from the subjects and it moved faster, they could feel smoother movement for the same intervals of the convolution filters. Namely, we can understand that the convolution filters at shorter intervals are necessary to simulate a smooth change in the sound field, when the sound source is located nearer the user and moves slower.



(a) Sound Movement at the Distance of 5m



(b) Sound Movement at the Distance of 1m

Fig.12 Results of the Experiment on the Virtual Sound Source Movement

The exact value of the interval at which we should prepare the convolution filters depends on the conditions of the virtual world such as the reflection coefficients of the walls or the positional relationship between the sound source and the audience. However, from this experiment, we can conclude that the convolution filters at intervals of about 1.0m are necessary to realize a smooth change in the sound field of the ordinary virtual room environment.

5. VIRTUAL REALITY APPLICATION

In order to evaluate the effectiveness of the interactive sound field simulation, this method was applied to the virtual reality applications. For example, it was integrated into the video avatar communication system [8].

The video avatar is a high presence communication technology used in the shared virtual world. In this method, the user's figure is captured by a stereo video camera and it is transmitted to the other site through the broadband network. This image is superimposed on the shared virtual world as a video avatar, and is used for the communication.

In this system, the interactive sound field simulation was integrated into the video avatar system so that the remote users can communicate with each other using visual and acoustic information. In particular, it is expected that the users can recognize the other user's position or the acoustic characteristic of the virtual world using the acoustic cue.

In Fig. 13, the user is communicating with the video avatar using the interactive sound field simulation. In the demonstration system, a virtual room was constructed and the convolution filters for several positional relationships between the sound source and the user located at the points shown in Fig. 14 were calculated. When the user moved in the virtual world, the localized sound of the avatar's voice was also moved with the avatar's image by replacing the convolution filters according to the user's movement.

Fig.13 Video Avatar Communication Using the Interactive Sound Field Simulation

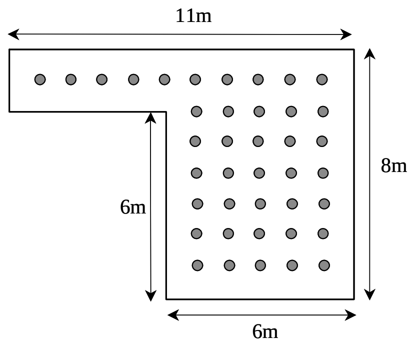


Fig.14 Grid-points for Preparing Convolution Filters

In this demonstration, even when the remote user stood behind the wall and the user could not see the avatar's figure in the virtual world, he could recognize the localized voice of the remote user. Thus, this technology was able to improve the high presence sensation of sharing the virtual world.

6. CONCLUSIONS

In this study, acoustic simulation method that can control the virtual sound field interactively by replacing the convolution filters calculated based on the wave equation was developed. In order to use this method in the multi-screen immersive projection display, the multi-channel digital signal processor named Wave Engine was developed, and it was implemented in the CABIN using a 16-channel speaker system. By using this system, we conducted an experiment on evaluating the smoothness of the moving sound source, and we could confirm that the sound field of the virtual room could be changed smoothly by using the convolution filters calculated at intervals of about 1.0m.

Future work will include making convolution filters more easily by using the interpolation method or dividing the convolution processes into the calculations for direct sound and reflection.

REFERENCES

[1] Cruz-Neira, C., Sandin, D.J., DeFanti, T.A.: Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE, *Proceedings of SIGGRAPH'93*, pp.135-142 (1993).

[2] Hirose, M., Ogi, T., Ishiwata, S., Yamada, T.: Development and Evaluation of the Imersive Multiscreen Display CABIN, *Systems and Computers in Japan*, Vol.30, No.1, pp.13-22 (1999).

[3] Wightman, F.L., Kistler, D.J.: Headphone Simulation of Free-Field Listening I- Stimulus Synthesis, *The Journal of the Acoustical Society of America*, Vol.85, No.2, pp.858-867 (1989).

[4] Wenzel, E.M.: Localization in Virtual Acoustic Displays, *PRESENCE*, Vol.1, No.1, pp.88-107 (1992).

[5] Sakamoto, S., Tachibana, H.: Calculation of Impulse Response in 3-D Sound Field with Absorption Boundary by the Finite Difference Method, Proc. of Inter-noise97, pp.1567-1600 (1997).

[6] Takane, S., Suzuki, Y., Sone, T: A Study on the Estimation of Impulse Responses in an Exclosure by Using Boundary Element Method, Proc. WESTPRAC V, pp.595-600 (1994).

[7] Asayama, H., Kimura, S., Sekiguchi, K.: Revised Finite Sound Ray Integration Method based on Kirchhoff's Integral Equation, The Journal of the Acoustical Society of Japan (E), Vol.10, No.2, pp.93-100 (1989).

[8] Ogi, T., Yamada, T., Kano, M., Hirose, M.: Immersive Telecommunication Using Stereo Video Avatar, Proc. IEEE VR2001, pp.45-51 (2001).