7. International Immersive Projection Technologies Workshop
9. Eurographics Workshop on Virtual Environments
J. Deisinger, A. Kunz (Editors)

Immersive Sound Field Simulation in Multi-screen Projection Displays

T. Ogi[1,2], T. Kayahara[1,2], M. Kato[2], H. Asayama[3] and M. Hirose[2]

1 MVL Research Center, Telecommunications Advancement Organization of Japan, Tokyo, Japan
2 Intelligent Modeling Laboratory, The University of Tokyo, Tokyo, Japan
3 Real Wave Research Japan, Tokyo, Japan

Abstract

This paper describes the immersive sound field simulation technology that represents an interactive sound field in the multi-screen projection display. In this method, convolution filters, that were calculated based on the wave equation, are replaced in real-time using the multi-channel digital signal processor, and the simulated sounds are displayed using the 16-channel speakers. In addition, compensation filters are used to reduce the influence of the screen attenuation. This system was applied to the video avatar communication, and the effectiveness of this method was evaluated.

Keywords:

Sound Field Simulation, Immersive Projection Display, Multi-channel Digital Signal Processor, Multi-channel Speaker System

Categories and Subject Descriptors (according to ACM CSS): H.5.5 [Sound and Music Computing]: Systems; I.3.7 [Three-Dimensional Graphics and Realism]: Virtual Reality

1. Introduction

Multi-screen immersive projection displays such as the CAVE have become very popular as virtual reality visual display systems [1]. This kind of display system generates a high presence virtual world by projecting stereo images onto surrounding wide screens. Although several acoustic display systems, such as the Avango or blue-c, have been developed to be used in conjunction with immersive projection displays [2][3][4], a standardized technology has not yet been established.

Typical acoustic display technologies that are often used in existing virtual reality systems are the Head Related Transfer Function (HRTF) [5][6] or the amplitude panning method [7]. These methods can localize the virtual sound interactively, by convoluting the HRTF with the sound source signal or by applying different amplitude signals to several speakers respectively. However, these methods have limitations in displaying a high presence sensation, because they cannot represent the influence of sound reflected against a wall or the floor in the virtual world.

On the other hand, numerical simulation methods based on geometric or wave acoustics have been used to calculate a high presence sound field that includes the influence of sound reflection [7][8]. Since these methods calculate an accurate sound field by taking the properties of sound treated as particles or waves into account, they are often applied to acoustics design in architecture such as in designing concert halls or theaters. However, these methods cannot be used to simulate an interactive sound field in real-time, because they require a large amount of calculation time.

Most of the existing acoustic display systems are constructed considering the trade-off between the real-time calculation and the presence of the simulated sound. The purpose of this study is to develop a sound field simulation technology that includes both real-time interaction and sound reflection in the virtual world. In addition, a requirement of this method is that it can be used in a multi-screen immersive projection display in order to generate high presence virtual worlds with visual and acoustic sensations.

In the CAVE-like multi-screen display, several users can experience the virtual world simultaneously. Therefore,

we consider that a multi-channel speaker system that can represent the sound field for multiple users is the most suitable for the acoustic display system used in immersive projection displays. However, in multi-screen displays, sound outputted from speakers placed behind the screens is attenuated due to having to pass through the screens. Therefore, we must develop an acoustic display technology that can present an accurate sound field by compensating for the influence of screen attenuation.

In this study, immersive sound field simulation that can be used in an immersive projection display was developed. This method can generate an interactive sound field by replacing the convolution filters in real-time. In addition, the screen attenuation of the outputted sound is compensated using a compensation filter. This paper describes the principle of the proposed method, its implementation in the CABIN, an experimental evaluation and its use in virtual reality applications.

2. Immersive Sound Field Simulation

2.1. Sound Field Simulation

In the proposed method, the impulse response between the sound source and the user is calculated by solving Kirchhoff's integral equation, which is formulated from the wave equation. Although several numerical calculation methods such as finite element methods or the boundary element method have been proposed to solve Kirchhoff's integral equation, a finite sound ray integration method was used [9]. Since this method does not generate meshes and the sound wave is approximated by a large number of sound ray vectors radiating in all directions from the sound source, the impulse response can be calculated in a relatively short time. In this study, about 20,000,000 sound ray vectors were radiated from the sound source, and the impulse response for 150ms was calculated. Figure 1 shows the principle of this sound field simulation method.

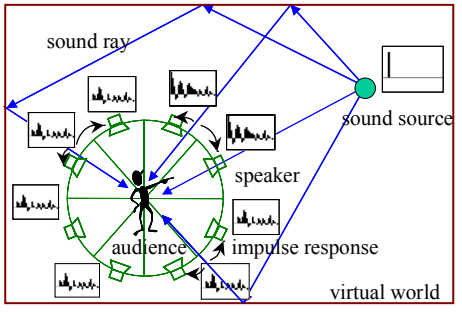

Figure 1 *Principle of sound field simulation.*

When a multi-channel speaker system is used to output the sound field, the impulse response calculated for each sound ray should be applied to the speakers that are placed around the arriving sound ray. In this method, 16 speakers are placed three-dimensionally, and the impulse responses are divided based on the vector base amplitude panning (VBAP) method among three speakers in a triangular formation [10]. That is, the division among the three speakers, α_1, α_2 and α_3, is determined so that the following equation is satisfied.

$$\boldsymbol{x} = \alpha_1 \boldsymbol{s}_1 + \alpha_2 \boldsymbol{s}_2 + \alpha_3 \boldsymbol{s}_3 \qquad (1)$$

where $\boldsymbol{s}_1$, $\boldsymbol{s}_2$, and $\boldsymbol{s}_3$ are the position vectors from the audience to the three loudspeakers, and $\boldsymbol{x}$ is a vector to the intersection point between the arriving sound ray and the plane containing the three speakers (Figure 2).

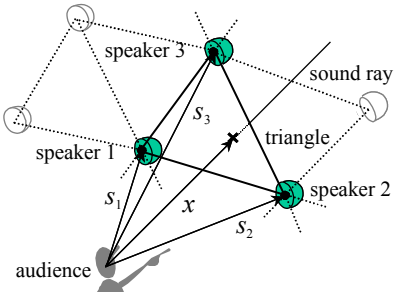

Figure 2 *Division of impulse response among speakers*

By integrating the impulse responses divided among the speakers for all sound rays, the transfer function is finally constructed. These transfer functions are convoluted with the sound source signal, and then a high presence sound field that includes sound reflection in the virtual world is generated.

2.2. Interaction with the Sound Field

Since the impulse response is calculated for the specific positional relationship between the sound source and the audience, it can represent only a static sound field. When the sound source or the audience moves in the virtual world, an impulse response for each positional relationship is necessary. However, this calculation requires a large amount of computing time, so that it cannot be performed in real-time. Therefore, in this method, the impulse responses are calculated beforehand for every possible position of the sound source and the audience, and the selected data is changed dynamically to realize the interactive sound field.

In practical use, the virtual world is divided into cubic meshes, and both the sound source and the audience are placed on grid points. Then, the impulse responses between

the sound source and the audience are calculated for every possible grid point. When the sound source or the audience moves in the virtual world, the nearest impulse response data to their present positions is selected and used. In this study, an SGI Origin2000 parallel supercomputer, which has 16 CPUs, was used to calculate the impulse response data. Although this calculation needs tens of minutes to several hours for each case, it can be efficiently parallelized for multiple cases and the data set of impulse responses for various positions can be calculated automatically.

2.3. 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 movement of the sound source and the audience. In this study, a multi-channel digital signal processor named the Wave Engine was developed to meet these requirements.

Figure 3 shows a functional diagram of the Wave Engine. In this system, a digital sound is inputted and it is convoluted with the 16-channel filters in real-time. Each channel consists of a 32-tap multi-tap filter and a 512-tap FIR filter. The multi-tap filter is used to convolute the impulse response, and the FIR filter is used to convolute the compensation filter discussed in section 3.2. These calculations are conducted in real-time using an ADSP-21065L digital signal processor (Analog Devices, Inc.). Each multi-tap filter can store 3,072 convolution filters in its memory, and the available filter can be replaced dynamically. In this system, the impulse response data calculated for various conditions are stored in the memory and the available data is changed in each channel.

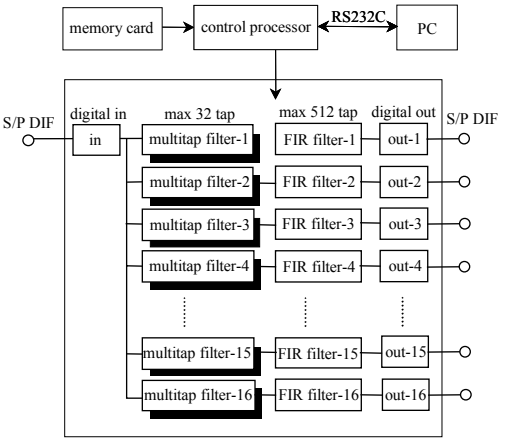

Figure3 *Functional diagram of the Wave Engine*

In this process, the filter data is replaced using cross fading in order to change the sound field smoothly without discontinuous noises. Figure 4 shows the process of the cross fading method. In this method, arbitrary shapes of the fade-in and fade-out curves can be used, by defining the coefficient values in the fade-in and fade-out tables. When the convolution filter is changed, the sound signal is convoluted with the current and next filter data, and the outputs are multiplied by the coefficient values defined in the fade-in and fade-out tables respectively. Then, the sum of these signals is outputted. By changing the coefficient values selected from the fade-in and fade-out tables, the output sound can be changed smoothly.

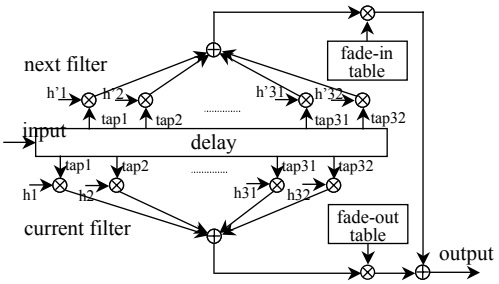

Figure 4 *Cross fading method using multi-tap filter*

Figure 5 shows the construction of the system of the immersive sound field simulation display developed in this study. The sound source is inputted to the Wave Engine through an A/D converter as digital data, and the 16-channel output sounds are transmitted to the speakers through D/A converters. The Wave Engine is connected to the PC through an RS-232C interface, and the convolution filter data is controlled by a command transmitted from the

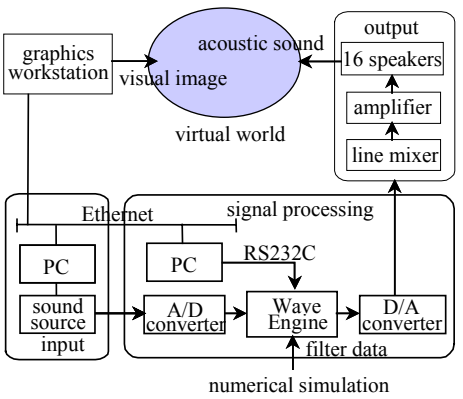

Figure 5 *Construction of acoustic display*

PC. This PC also communicates with a graphics workstation through the Ethernet, so that the control of the virtual sound can be synchronized with the visual image.

3. Implementation in the CABIN

3.1. Speaker Arrangement

In this study, the immersive sound field simulation technology was implemented in the multi-screen immersive projection display known as CABIN [11]. The CABIN is a CAVE-like cubic display that has five screens, one at the front and one each on the left, right, ceiling and floor. Though this technology uses a multi-channel speaker system, these speakers must be placed at positions where they do not disturb the projected images. In this system, 16 speakers were placed around the screens outside the projection area. Figure 6 shows the multi-channel speaker system used in the CABIN, and Figure 7 shows the speaker arrangement of this system. Eight speakers are placed at the height of the user's ear (1.5m from the floor), and a speaker is placed at the center of each side of the ceiling screen and each side of the floor screen. These speakers are each assigned to a channel of the Wave Engine, and they are used to output the simulation sound.

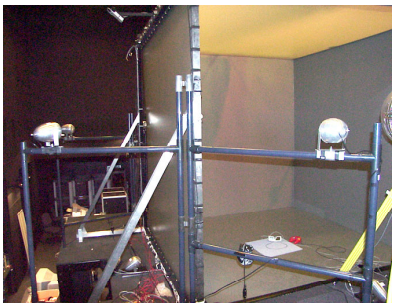

Figure 6 *Multi-channel speaker system in the CABIN*

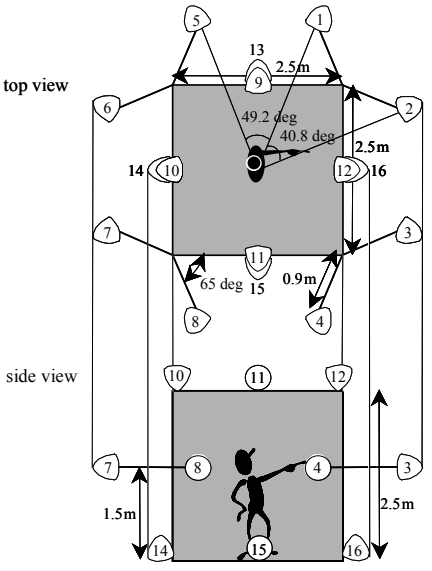

Figure 7 *Speaker arrangement*

In order to represent an accurate sound arriving from an arbitrary direction, the speaker should be placed in the same direction. However, it is impossible to place a speaker in every direction. In this system, the sound wave is approximated by using a finite number of speakers and by applying divided impulse responses to neighboring speakers. Therefore, we must discuss how an accurate sound field is generated in the display space of the CABIN by using 16 speakers. In this section, numerical simulation is used to compare the approximate sound wave produced by using neighboring speakers with the actual sound wave arriving from the source position. The sound pressure $p(x,t)$ at position x from the sound source and time t is given by the following equation:

$$p(x,t) = e^{j(2\pi f(t-x/v)+\varphi)}/x \qquad (2)$$

where v is the sound velocity, f is the frequency, and φ is the initial phase. The sound source is assumed to be placed at the height of the user's ear, and only the horizontal components are calculated.

Figures 8 and 9 show the sound pressure contours, when sound waves with frequencies of 200Hz or 800Hz arrive from a point midway between speakers 1 and 2, and from a point between the speakers divided in the ratio 4:1. The results of the actual sound are plotted using "+", and the results of the approximate sounds are plotted using "x". The dotted circle with diameter 1.0m shows the center area of the CABIN. From these graphs, the approximate sound waves in the center area of the display space are almost consistent with the actual sound waves arriving from the source positions. By considering these results for each

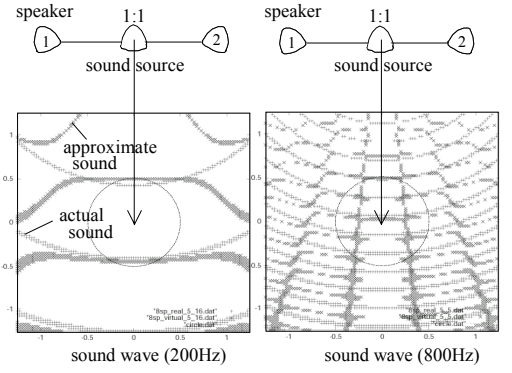

Figure 8 *Sound pressure contour (source 1:1)*

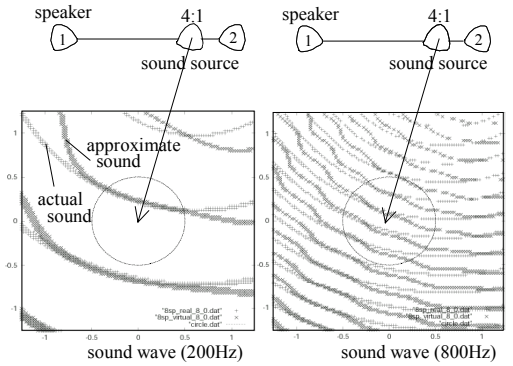

Figure 9 *Sound pressure contour (source 4:1)*

sound ray, we can understand that a fairly accurate sound field can be constructed in the center area of the CABIN using the 16-channel speaker system.

3.2. Compensation of Screen Attenuation

With this speaker arrangement, the influence of sound attenuation due to the screens is an unavoidable problem, because the speakers are placed behind the screens. Therefore, in order to generate an accurate sound field around the user, compensation for the influence of screen attenuation should be done. In this study, the impulse response from each speaker to the audience was measured, and the inverse functions of the impulse responses were used for the compensation filter.

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 upon the measurement position, the impulse response data measured at the user's position would also change when the user moves in the display space of the CABIN. Therefore, in this study, the impulse responses were measured immediately after passing through each screen (30cm from the screen) so that only the influence of the screen attenuation was considered.

Figure 10 shows an example of the measured impulse response data. From this data, only the direct sound component was extracted by reducing the data after the first reflection using the window function, and the compensation filter was constructed by calculating the inverse function. Figure 11 shows the compensation filter generated from the impulse response data. When this compensation filter was used, an impulse response with the frequency characteristic shown in Figure 12 was measured. In this graph, a flat gain characteristic is shown, although the frequency characteristic was distorted without the compensation. Therefore, we can see that the sound attenuation is effectively compensated, by using the inverse function filter of the impulse response. In this system, compensation filters for each speaker were prepared and stored in the FIR filters of the Wave Engine.

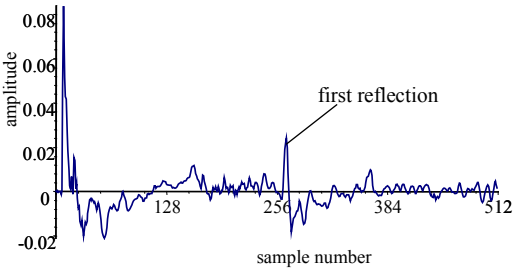

Figure 10 *Measured impulse response data*

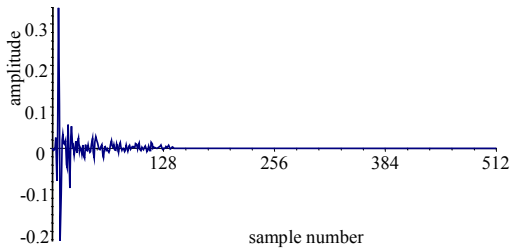

Figure 11 *Compensation filter for screen attenuation*

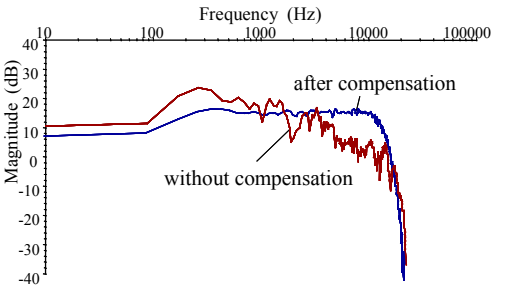

Figure 12 *Impulse response using compensation filter*

4. Experimental Evaluation

4.1. Preparatory Experiment

In this system, a high presence virtual sound field can be generated in a multi-screen immersive projection display. However, in order to construct a practical virtual world, we must discuss how many convolution filters are necessary to move the sound source smoothly or how naturally the simulated sound field can be changed. In this study, we conducted an experiment to evaluate the smoothness of the sound source movement, when the number of 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 Figure 13, and the subjects were asked to stand near the left-side wall of the room, in order to investigate the influence of sound reflection. First, under these conditions, a preparatory experiment on sound localization was conducted. In this experiment, a white noise sound source was located at several positions along lines at distances of 5.0m and 1.0m in front of the subjects, and the subjects were asked to indicate the perceived directions of the virtual sound source by moving a marker in the virtual world.

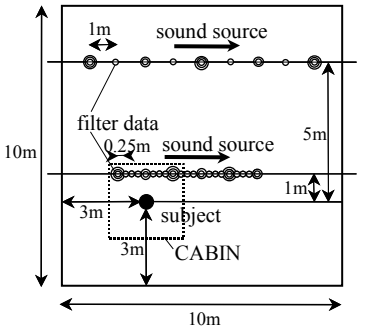

Figure 13 *Virtual room used in the experiment*

Figure 14 shows the result of this experiment for five subjects. In this graph, average values and standard deviations of the localized directions are shown. Zero degrees denotes the direction in front of the subject and positive values are to the right. Although the perceived directions include some errors, we can consider that they are caused by the influence of sound reflection, because the perceived directions tend to shift toward the nearest wall. From these results, we can conclude that this system can represent a high presence virtual sound field that includes sound localization and the accuracy of the localized sound is sufficient to conduct an experiment on the sound source movement.

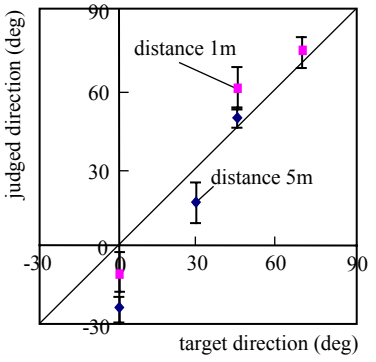

Figure 14 *Result of the sound localization*

4.2. Experiments on Sound Source Movement

Next, in these experiments, the sound source was moved from left to right on the lines at distances of 5.0m and 1.0m from the subject. In this case, an image of the loudspeaker was also visualized at the sound source position synchronized with the sound movement. When the sound source was moved at a distance of 5.0m, convolution filters were replaced at 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 a distance of 1.0m, convolution filters were changed at intervals of 0.25m, 0.5m and 1.0m, and the sound source was moved at speeds of 0.5m/s and 1.0m/s. Under these conditions, the filter data were changed at the same frequency for sounds moving at 5.0m or at 1.0m. These filter data were replaced using a linear cross fading method taking a time of 4.5ms. The subjects stood at the center of the CABIN, and they were asked to evaluate the smoothness of the change in the sound field using a five-grade system for each experimental condition and give their response verbally. The number of the subjects was five, and the moving sound source was displayed twice for each experimental condition in random order.

Figure 15 shows the results of this experiment. In these graphs, average values and standard deviations of the evaluated grades for each experimental condition are shown. When the sound source was moved at a distance of 5.0m, the analysis of the variance showed a significant difference at the 5% level both for the results when the filter data interval was changed and when the velocity of the sound source was changed. But, in the case of moving sound source at a distance of 1m, there was no significant difference for the results when the sound source velocity was changed, though a difference was found at the 5% level when the filter data interval was changed. We consider that this result was due to saturation of the grade given in the evaluation for the experimental conditions in which the filter data interval was smaller than 0.5m.

In this experiment, the subjects perceived a smoother movement of the sound source, when the interval of the convolution filters was shorter. In addition, when the virtual sound source was located farther from the subjects and it moved faster, they could perceive a smoother movement for the same interval of the convolution filters. Therefore, we understand that shorter intervals for the convolution filters are necessary to simulate a smooth change in the sound field, when the sound source is located nearer to the user and when it moves slower. 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 and the positional relationship between the sound source and the audience. However, from this experiment, we can conclude that by replacing the convolution filters at intervals of about 1.0m, a smooth movement of the sound source in the ordinary virtual room can be realized.

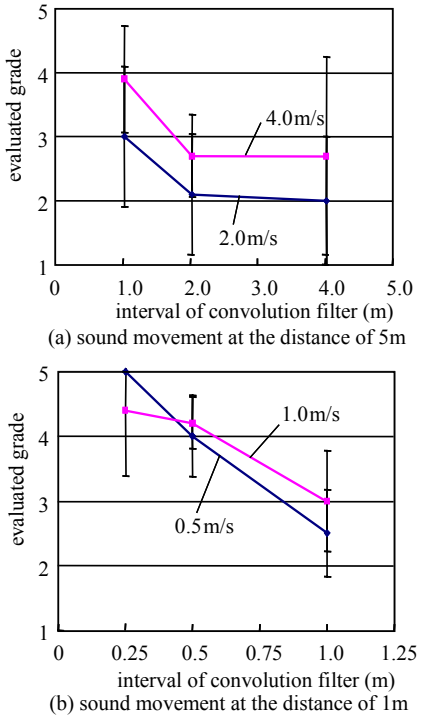

Figure 15 *Results of sound source movement*

5. Virtual Reality Application

In order to verify the effectiveness of the immersive sound field simulation technology, it was integrated into several virtual reality applications. For example, it was applied to the video avatar communication system [12]. The video avatar is a high presence communication technology used in the networked immersive projection display environment. In this method, the user's figure is captured by the video camera and it is transmitted to another site through the network. Then, this image is superimposed on the shared virtual world at the three-dimensional position and is used for communication. In this system, immersive sound field simulation is used so that remote users are able to communicate with each other sharing both visual and acoustic information. In particular, it is expected that the users can communicate naturally by recognizing each other's location and the acoustic characteristics of the shared virtual world from the sound cue.

Figure 16 shows a user communicating with a video avatar in the immersive sound field simulation environment generated in the CABIN. In the demonstration system, a virtual room was modeled and the impulse responses between each grid point of the meshes divided at intervals of 1.0m 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. Then, the users were able to communicate with each other using the localized voice in the shared virtual world. In particular, even when the remote user moved behind a wall and his figure was not seen, the user could recognize where he was from the localized voice. Thus, remote users were able to share the virtual world with a high presence sound field.

Figure 16 *Video avatar communication*

6. Conclusions

In this study, immersive sound field simulation technology that can control the virtual sound field interactively by replacing convolution filters at intervals was developed. The filters were calculated on the basis of the wave equation. In order to use this technology in a multi-screen projection display, a multi-channel digital signal processor named the Wave Engine was developed, and a compensation filter to reduce the influence of screen attenuation was introduced. By using this system, we conducted an experiment to evaluate the smoothness of a moving sound source, and we confirmed that the sound field in an ordinary virtual room could be changed smoothly by using convolution filters calculated to be at intervals of about 1.0m.

However, in order to implement this method in a practical application, it is necessary to reduce the calculation load of making a data set of the convolution filters. Although filter data

calculated at larger intervals could be used for a sound source placed at a far distance from the audience, the problem of the calculation load would become more severe in constructing large-scale virtual worlds. In future work, we are planning to develop an interpolation method to calculate approximate convolution filter data as well as to develop a higher performance computation method in order to construct a data set more easily.

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. Bargar, R., Choi, I.: Model-based Interactive Sound for an Immersive Virtual Environment, Proc. of International Computer Music Conference, pp.471-474 (1994).
3. Eckel, G.: Applications of the Cyberstage Spatial Sound Server, Proc. of the Audio Engineering Society 16th International Conference on Spatial Sound Reproduction (1999).
4. Naef M., Staadt, O., Gross M.,: Spatialized Audio Rendering for Immersive Virtual Environments, Proc. of VRST2002 (2002).
5. 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).
6. Wenzel, E.M.: Localization in Virtual Acoustic Displays, PRESENCE, Vol.1, No.1, pp.88-107 (1992).
7. Krockstadt, U.: Calculating the acoustical room response by the use of a ray tracing technique, Journal of Sound and Vibrations, 8,18, (1968).
8. 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).
9. 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).
10. Pulkki, V.: Virtual Sound Source Positioning Using Vector Base Amplitude Panning, Journal of the Audio Engineering Society, Vol.45, No.6, pp.456-466 (1997).
11. 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).
12. Ogi, T., Yamada, T., Kano, M., Hirose, M.: Immersive Telecommunication Using Stereo Video Avatar, Proc. IEEE VR2001, pp.45-51 (2001).