Web- and Mobile-Based Environment for Designing and Presenting Spatial Audiovisual Content

Mami Yamanaka, Makoto Uesaka, Yoshiteru Ito, Shigeyuki Horikawa,
Hikari Shiozaki, and Tomohito Yamamoto[*]

Department of Information and Computer Science,
Kanazawa Institute of Technology,
7-1 Oogigaoka, Nonoichi, Ishikawa, 921-8501 Japan
`tyama@neptune.kanazawa-it.ac.jp`

Abstract. Many types of VR systems have been developed to provide spatial views and surround sound to express high levels of presence. Recently, high-definition TV and 5.1 ch surround sound systems have been made available for watching 3D movies at home. In the near future, more realistic display systems such as "Super Hi-Vision" will be developed and introduced into our homes. However, these types of visual or auditory display systems may require the allocation of a large space for fixed, specialized equipment, and they tend to be expensive. Moreover, compared to the amount of free contents on the Web, highly realistic contents for such systems are still lacking. In this study, we propose a spatial audiovisual display system that comprises multiple mobile devices and a Web-based design system, which allows average users to create and share spatial audio content on the Web.

Keywords: visual and auditory display, mobile device, sound space, sharing, Web.

1 Introduction

Many types of VR systems have been developed that provide spatial views and surround sound for expressing high levels of presence [1–3]. Some of these systems are now available as consumer products, e.g., we can watch 3D movies using high-definition TVs and 5.1 ch surround sound systems at home. The popularity of these systems means that their content is generally available widely. In the near future, more realistic display systems such as "Super Hi-Vision [4]" will be developed and some will be introduced into our homes.

However, these types of visual or auditory display systems sometimes require the allocation of large spaces for fixed, specialized equipment, and they tend to be expensive. Moreover, compared to the amount of free contents on the Web, highly realistic contents for such systems are still lacking. However, mobile devices such as smartphones and tablet PCs are now widespread and service providers such as

---

[*] Corresponding author.

YouTube allow the uploading and sharing of content on the Web. Thus, it may be possible to build an immersive reality system on mobile devices, which users can experience at any time and in any place. It may also be possible to create and share highly realistic content via a Web browser so the interactions between users will facilitate creativity and the volume of available content will increase exponentially.

In this context, we have developed a prototype spatial audiovisual display system using multiple mobile devices [5]. We have also developed a Web-based design system that allows average users to create and share spatial audio content [6]. In this study, we improved their usability, portability, and capacity for expression based on these systems. A combination of these two systems may facilitate the provision of an integrated environment to general users who can enjoy more realistic content.

2   Spatial Audiovisual Display

2.1   Background

In our previous system, we used mobile devices to adjust the scale and composition of the display system to suit the playback environments for users. In this study, we improved this system in four ways. First, we introduced a tablet PC as a visual display to allow greater portability. Second, we implemented a reverberant effect in the auditory display to allow for more effective expression. Third, we implemented a synchronization process between devices to ensure stable operation. Finally, we improved the consistency between virtual and real space.

2.2   System Overview

Figure 1 shows an overview of our proposed system. This system comprises multiple mobile devices for each speaker, a laptop PC with an Head Mount Display (HMD),
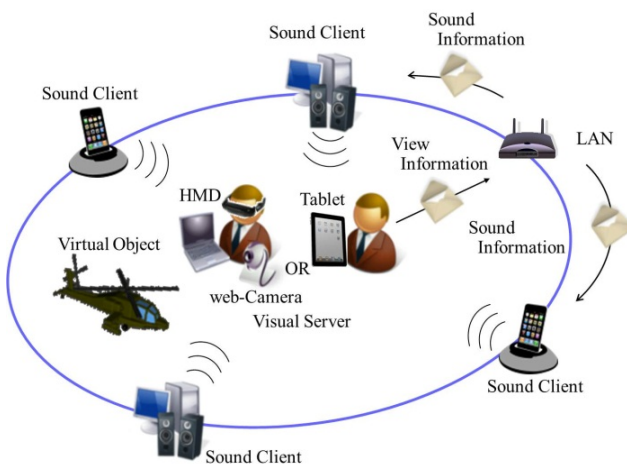

Fig. 1. Overview of the 3D audiovisual display system

and a camera or a tablet for visual information. Each device is connected to a wireless LAN and they are divided into one server and multiple sound clients. In this system, the visual server simulates the virtual space that contains spatial visual and audio data, which can be designed freely by a user. After the simulation, the visual server presents a spatial view to the HMD or tablet display. At the same time, the server transmits the information related to the virtual space to the sound clients via UDP multicast. After the clients receive the information, they implement the spatial sound via each speaker.

In our previous system, an HMD was used to present spatial views. However, with the development of mobile technology, we can now use a tablet to present spatial views as well (Figure 2). To present spatial views to multiple users, a visual client was developed as well as a sound client.

In this system, the visual server and client detect user postures via the gyroscope in the tablet, which is reflected by the viewing angle in the virtual space when presenting the spatial view. The visual server and client also present a mixed view, which comprises the virtual and real space for a user using the camera to generate the visual display. This visual display system is implemented on the same platform as the sound client (Apple iOS); therefore, it is possible to integrate all of the audiovisual display functions into one application. This integration greatly improves the portability and usability of the system, which facilitates the distribution of applications.
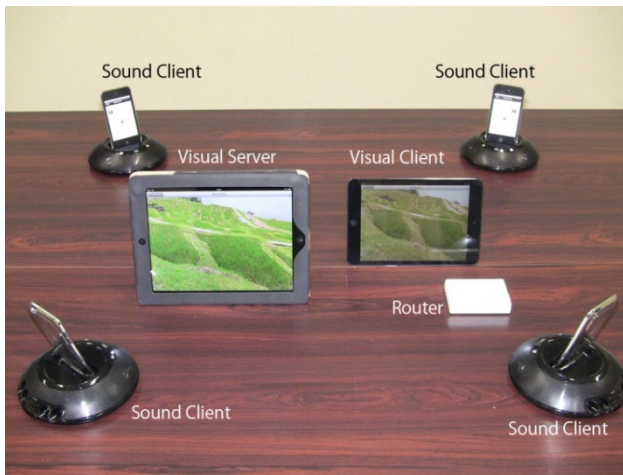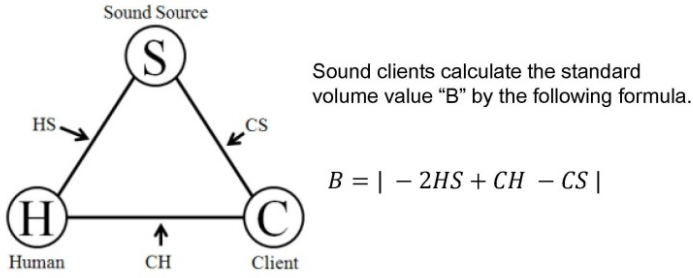

Fig. 2. Image showing the 3D audiovisual display

2.3   Calculation of Playback Sound

Our system reproduces the spatial sound using the sound clients to calculate the distance decrement of sound in three positions (the position of the listener, the sound source, and the sound client, as shown in Figure 3) [7]. Using these three positions,

Sound Source

S

HS

CS

H   C

Human   CH   Client

Sound clients calculate the standard volume value "B" by the following formula.

$$B = |-2HS + CH - CS|$$

The reproduction volume is calculated in the following formula.

$$(1 < B)$$
$$Volume[dB] = Max\ volume[dB] - 20\log(B) - 10$$
$$(0 \leq B \leq 1)$$
$$Volume[dB] = Max\ volume[dB] - B * 10$$

Fig. 3. Calculation of the sound volume

the system calculates the volume in real time and reproduces a spatial sound based on the volume difference between each client.

In our present system, if a user inputs the room information data into the system, the sound clients calculate the reverberation and overlays this information onto the

The position of mirror sound is calculated in the following formula.

$$\overrightarrow{CA} \times \overrightarrow{CB} = (n_x, n_y, n_z)$$

$$M = \frac{n_x(c_x - o_x) + n_y(c_y - o_y) + n_z(c_z - o_z)}{n_x{}^2 + n_y{}^2 + n_z{}^2}$$

$$o'_x = 2n_x M + o_x$$
$$o'_y = 2n_y M + o_y \qquad \ldots (1)$$
$$o'_z = 2n_z M + o_z$$

$O(o_x, o_y, o_z)$

A

C

B

$O'(o'_x, o'_y, o'_z)$

By using the following formula, reverberation time is determined.

$$T = 0.049\frac{V}{S\alpha} \qquad \ldots (2)$$

When the time is t, sound pressure level is represented in the following formula.

$$Volume = -4.024\frac{S\alpha}{V}t + Max \qquad \ldots (3)$$

T : Reverberation time
V : Volume
Sα : Absorption
Max : Initial sound pressure

Fig. 4. Calculation of reverberation

reproduced sound. Figure 4 shows the calculation method. To generate early reflections, this system calculates the mirror image of a sound source (formula (1)). To generate late reverberations, this system calculates the linear equation for attenuation from the early reflections and the approximate reverberation time (formulae (2) and (3)).

In this system, the server only sends position data to the sound clients and it is not concerned with phase variation. On the client side, they prepare playback buffers for early reflections, which are calculated from the room data, and they also prepare 10 playback buffers for late reverberations. After the sound clients receive the data related to the sound source, they calculate the sound volume and delay using the formulae in Figure 3. After the calculation, the clients play the sound sources with the appropriate delay timing and volume, so the correct reverberation is reproduced.

2.4   Synchronization between Sound Clients

In our previous system, the playback timing between clients could lead to the loss of synchronization. This was because the system controlled multiple mobile devices via UDP multicast so the packet was sometimes lost among other network traffic. Therefore, we implemented a synchronization process by calculating the network delay using the Network Time Protocol (NTP) method.

In this method, the server transmits test data to clients and records the sending time. Next, the clients receive the data and record the receiving time, before the clients send a response to the server and record the sending time. Finally, the server receives the response from the clients and records the receiving time. After this transmission, the server checks all of the times recorded and calculates the delays for all clients. The server then sends each delay time to the clients. Following this procedure, synchronization between clients is achieved if clients reproduce the sound source based on the ordinary signal from the server while they control the playback timing based on the delay time.

2.5   Coordination between Virtual and Real Space

In the previous system, a mixed view of virtual and real space was presented via an HMD where a magnetic sensor on the HMD was used to calculate the position and posture of the camera. As a result, the two coordinates were often misaligned and viewing was uncomfortable. In the present system, Parallel Tracking and Mapping (PTAM [10]), which is a marker-free augmented reality technique, is used to coordinate virtual and real space.

PTAM is a method for estimating the real camera position and posture. This method uses one camera and no markers to calculate the 3D map using feature points in real space. Figure 5 shows the coordination process between virtual and real space using PTAM. In this process, the Web camera captures two key frames and initializes the 3D map by calculating the visual features in real space. Next, the visual tracking and updating of the 3D map is executed as parallel thread processes. Visual tracking is achieved by calculating the feature points of the captured images and comparing the
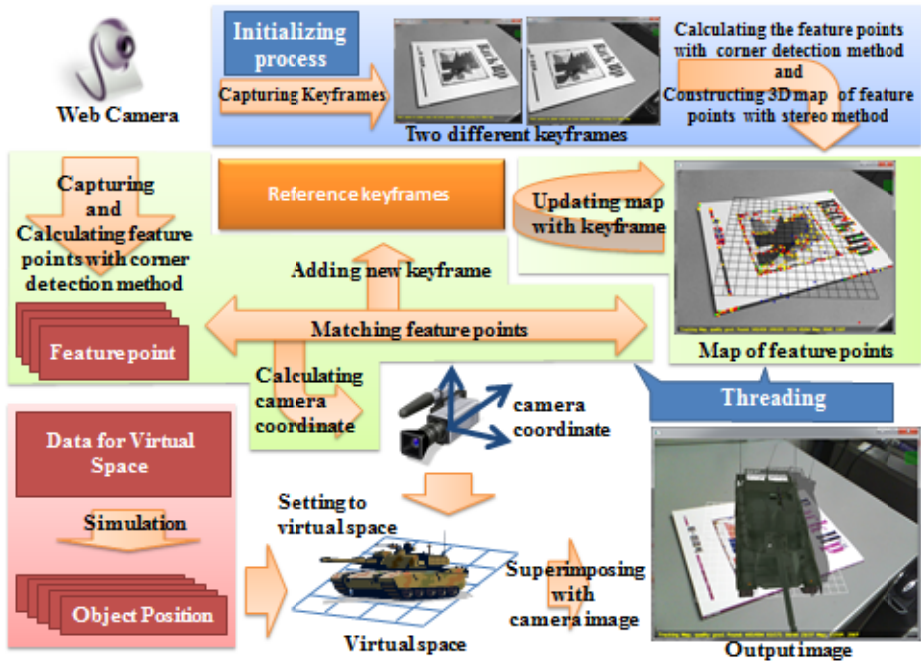
Fig. 5. Calculation of the coordination between virtual and real worlds

points to stored points in the 3D map in real time. At the same time, the system calculates the camera coordinates in real space.

The 3D map is updated by recalculating the map using newly stored key frames at a regular interval. In addition to these threading processes, our system calculates the positions of virtual objects and animates them in the virtual space using another threading process.

Finally, the position of the real camera is transferred to the coordinates of the virtual space and the real view captured by the Web camera is overlaid onto the virtual space. As a result, the final view is presented to the user. In this way, the visual difference between the virtual and real space can be reduced, so a user feels as though a virtual object is placed in the real space.

Our system is implemented at the software level. Therefore, our system can be built at low cost using popular mobile devices. The improvements introduced in this study imply that our system has greater portability, effective expression, and stability.

3   Web-Based Design System for Spatial Audio Content

3.1   Background

Despite the availability of free content on the Web, highly realistic content is not widespread. This is possibly because it is difficult for average users to create content

for these systems because this requires specialized knowledge and tools. To address this problem, it is necessary to remove these obstacles [8–9]. In our previous study, we proposed a Web-based design system to allow users to create and share spatial audio content [6]. However, the limitations of the client side technologies (JavaScript, WebGL, and so on) meant that the usability of this system was relatively low. However, recent progress in these technologies has increased the options during implementation. In this study, we improved the usability of our previous system using these technologies.

3.2   System Overview

Figure 6 shows an overview of our proposed system. This system comprises a "design system," "download system," and auditory displays. The "design system" is implemented as a rich internet application and users do not need any software other than the browser. With the "download system," users can search, preview, and download a designed sound space. To enjoy our sound space fully, users can apply our spatial audiovisual display system as explained in Section 2. However, if users do not apply this display, they can still enjoy the sound space via conventional 2 ch speakers on the preview page of a browser. This system can also export the designed spatial sound to a 5.1 ch surround sound format.
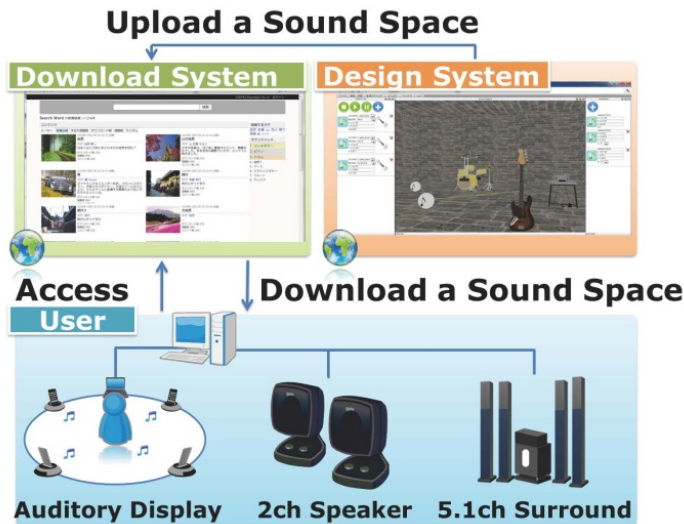

Fig. 6. Overview of the Web-based system

3.3   Designing the Interface

In current tools for designing the sound space, users generally have to coordinate the sound volume of fixed speakers. In our design system, however, users do not have to worry about the playback environment. Instead, they design the sound space by

positioning the sound sources and animating them on the browser. Figure 7 shows the window used for designing the sound space. Figure 7-1 shows the "object list window" used to add a sound object and for changing basic object properties (position, audio, and model data). The buttons at the top of window are used to "play," "stop," and "pause," as well as add a sound object. The properties of the

Fig. 7. System for designing the spatial sound content

sound object are located on right-hand side of the musical note icon. With the pull-down menu, users can select the audio data (recorded by users with an ordinary microphone or preregistered data) and the 3DCG model data. A button for the animation mode and a delete button are located on the right-hand side of the properties.

Figure 7-2 shows the "sound space window," which represents the 3DCG space from the user's viewpoint. Sound objects are represented as 3DCG models ("COLLADA" format data are imported as external data) where the movement paths are represented as lines and base points (white balls). To move a sound object, users manipulate the mouse cursor on the object by dragging (up, down, left, and right) or rotating the mouse wheel (forward and backward). Figure 7-3 shows the "animation window" known as the "object list window." In this window, users can specify the animation of sound objects using the button to add movement paths and a text field to add the movement speed.

Using these windows, users can design the sound space content via simple operations. After the design process, the contents are exported as XML data, and users can preview and download the file.

In the "download system," users can search, preview, and download the designed spatial audio content. At the top of the download system page, users can search for content using certain fields (such as "creator's name," "creator's comment," and "users' comments for evaluation"). Figure 8 shows the results of a search. On this

page, users click to preview the content, and the preview page and the design page are the same. On this page, users can listen to the designed sound via combined 2 ch speakers. If the users want to arrange the sound space and the permission of the contents is off, they can change the properties of the sound space on this page instantly.

Fig. 8. Download system

Thus, our Web-based design system does not require specialized tools, which allows users to design a sound space intuitively. A combination of the spatial audiovisual display and this Web system is expected to allow general users to enjoy highly realistic content in a more relaxed manner.

4   Conclusion and Future Work

In this study, we improved our spatial audiovisual display system and developed a Web-based system for designing and sharing sound space content.

In current spatial audiovisual display systems, it is impossible to manipulate virtual (or real) objects in real (or virtual) space. However, it is important to consider this virtual–real interaction to achieve hyper-reality. Therefore, in future work, we will introduce a depth sensor such as Microsoft Kinect to produce an interface that facilitates interaction between virtual and real spaces.

In the Web-based design system, we plan to develop a community system to enhance communication and active interaction among users. We will also develop a system to coordinate the sound space with omnidirectional video or 3DCG to produce more attractive content in future studies.

References

1. Lee, N.: IOSONO. ACM Computers in Entertainment 2(3), 3 (2004)
2. Hughes, C.E., Stapleton, C.B., Hughes, D.E., Smith, E.: Mixed reality in education, entertainment and training: an interdisciplinary approach. IEEE Computer Graphics and Applications 25(6), 24–30 (2005)
3. Defanti, T.A., Dawe, G., Sandin, D.J., Schulze, J.P., Otto, P., Girado, J., Kuester, F., Smarr, L., Rao, R.: The StarCAVE, a third-generation CAVE and virtual reality OptIPortal. Future Generation Computer Systems 25(2), 169–178 (2009)
4. Ito, T.: Future television - Super Hi-vision and beyond. In: Proceedings of IEEE Asian Solid-State Circuit Conference, pp. 5–8 (2010)
5. Takahashi, K., Yamamoto, T.: 3D audio-visual display using mobile devices. ACM SIGGRAPH 2010 Posters (52) (2010)
6. Matsuda, S., Yamamoto, T.: A Web system for creating and sharing 3D auditory contents. ACM SIGGRAPH 2010 Posters (80) (2010)
7. Takahashi, K., Ikeda, S., Yamamoto, T.: Light aural display using network connected multiple computers. In: Proceedings of HCI International 2009 - Posters, pp. 401–405. Springer (2009)
8. Hughes, D.E.: Defining an audio pipeline for mixed reality. in Proceedings of HCI International 2005 (2005)
9. Hughes, D.E.: Integrating and delivering sound using motion capture and multi-tiered speaker placement. In: Shumaker, R. (ed.) VMR 2009. LNCS, vol. 5622, pp. 179–185. Springer, Heidelberg (2009)
10. Georg, K., David, M.: Parallel tracking and mapping for small AR workspaces. In: International Symposium on Mixed and Augmented Reality, pp. 225–234 (2007)