Interactive 3D Audio Video Reproduction System

- Audio video reproduction with an enhanced sensation of reality -

We are studying the elemental technologies of an interactive broadcasting service conveying a stronger sensation of reality than is possible with the present HDTV systems. Besides providing high-quality video and sound such as system will let a viewer effectively watch and listen from any position in reference to the presentation. For example, if a user has a favorite seat in an auditorium and is listening to a music program of a live concert in that auditorium, the service would configure the presentation so that the viewer could experience the performance as if he or she were actually sitting in that seat.

Our recent developments along this line include the mixed reality audio-visual reproduction system (MRAV), which realizes 3D audio reproduction linked to binocular 3D video and interactive viewing control (see Figure 1).

Hiroyuki Okubo,
Acoustics and Audio Signal Processing

I joined NHK Science & Technical Research Laboratories in 1992 and have researched method of sound field simulation, its reproduction, room acoustical measurement and evaluation of human spatial impressions. My recent research includes interactive 3D audio reproduction system. I am now engaged in the development of the Advanced Multi-channel Audio Systems with Superior Impression of Presence and Reality.

■ 3D video display

Our latest 3D video display program controls two workstations that respectively output video images to the left and right eyes (the viewer wears special glasses). These images are synchronized by ensuring that the workstations send the same video frame numbers. As illustrated in Figure 2, it is also capable of presenting 3D CG background images, actual-shot 3D HDTV images, and 3D moving objects* on the same screen.

(a) Composition of actual-shot binocular 3D HDTV video

(b) Composition of 3D moving objects

Figure 2: Example of 3D image using newly developed display program

■ 3D audio reproduction

3D audio is reproduced with three loudspeaker-arrays in front of the listener and two loudspeakers to the rear. The reproduction level and delay for each loudspeaker are controlled in an integrated manner with a 3D audio signal generator. To realize natural-sounding reverberation in a virtual space, the reflected sound from every direction is processed. The key to synchronizing sound with 3D video is the use system's loudspeaker arrays. These arrays can create sounds that appear to "pop out" right in front of the listener.

■ Interactive viewing control

Data for constructing a viewing point in the virtual space as designated by a viewer or data for localizing a sound source are sent to the 3D audio signal generator from the workstations via Ethernet. The synchronizations needed for matching the sound's direction, distance and timing are executed by the generator.

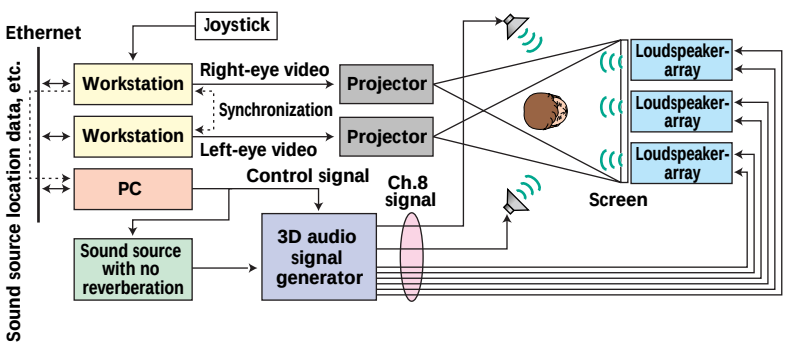


Figure 1: System overview

■ Screen development

To present clear sound from loudspeakers installed behind the screen, there are many small holes in the screen's surface. This feature, however, sometimes disperses light from the projectors, causing inaccurate right-eye and left-eye video data to be delivered to a viewer. A new screen that has both good sound transmittance and 3D video capabilities was fabricated to overcome this problem. This screen is permeable to sound with frequencies of up to nearly 16 kHz.

The system can be used for subjective evaluations of the synergy effect of binocular 3D video and 3D audio. The subjective experiments are being planned.

* Actually shot moving 3D images generated from multiple cameras surrounding the subject.

Image Extraction Technology Applied to Sports Programs

- Virtual display of offside line -

The virtual studio technology, which displays CG objects on the real image, is now widely applied to live sports programs, as an effective visual aid for viewers understanding the rules and plays of the game, such as a virtual first-down line in American football. Virtual objects are superimposed as if they were in the real 3D world by measuring pan and zoom operations of the on-air camera.

When using conventional techniques, however, it has been extremely difficult to present fast-changing information right after the play. In order to overcome this problem, we developed a computer-based video analyzer for sports images that is capable of detecting dynamic information in real-time (Figure 1).

This system is applicable to a diverse range of sports events by developing specific application software for each. The pitching trajectory renderer, for example, has already been frequently utilized in live baseball broadcasts and also in a softball educational program. Another application, described in detail in this article, can present the offside lines on the pitch in a soccer game.

■ Estimation of offside line position

Based on images acquired by two fixed sensor cameras (Figure 2), each covering half of the court, the application estimates the positions of the offside lines as follows:

(1) detection of individuals (players and referees) based on differences in color or texture from background turf,
(2) image-based positioning of each individual on the pitch,
(3) team identification from the uniform color, and
(4) estimation of offside line positions based on formation analysis.

■ Offside line rendering

Based on measured zoom, pan, and tilt values of an on-air camera, virtual offside lines are rendered by a perspective projection method[1]. A chroma-keyer[2] is also employed in order not to draw the virtual lines over players, but just to render on the green-colored turf area. The offside lines can be superimposed as if they were lying on the pitch as shown in Figure 3. The delay time for the computation that takes place before display is approximately 0.5 seconds.

In the Japanese live soccer program "J League" broadcast on BS1 on September 23, this virtual effect was used as slowed-motion replay clips immediately after offside scenes and as those for formation commentary.

We are now conducting field experiments to improve the accuracy of offside line estimation. New sports applications will also be developed besides baseball and soccer.

It is our hope that image extraction technologies will contribute to production of sports programs that are more enjoyable to our viewers.

[1] A geometric perspective in which the farther away an object is, the smaller the object is rendered, based on camera position and angle.
[2] A technique to compose an image in an area of a specific color, such as blue or green.

Toshihiko MISU,
Intelligent Information Processing

I joined NHK Science & Technical Research Laboratories in 1999 and have studied scene-adaptive image segmentation technologies for object-based video coding. My recent research interests include data-fusion approaches to moving object tracking and metadata extraction from video information. I am now engaged in the development of object-linked video effectors for sports scenes based on real-time image analysis and visualization.

Figure 1: Video analyzer

Figure 2: On-air camera and sensor cameras

Figure 3: Virtual display of offside line

Multi-channel Audio System

Super Hi-Vision is an audio video system that offers an unconventional sensation of reality and immersion. The goal of Super Hi-Vision is an advanced broadcasting medium far beyond the capabilities of HDTV. Our research has so far led us to construct a prototype three-dimensional audio system consisting of 22 loudspeaker channels and two LFE* loudspeaker channels for low-bass effects.

The optimum viewing distance for the Super Hi-Vision system (the distance from the screen) is equivalent to 0.75 times the screen height, meaning that the optimum seating area is extremely close to the screen. It is feasible to create an enhanced visual sensation for all the audience seats. To ensure that realistic audio is presented over the entire seating area, progress is being made on finding the best loudspeaker arrangement.

One significant feature of this 22.2 channel audio system is its 3-loudspeaker arrangement (see figure), which has loudspeakers placed at the audience's height, as well as ones above and below. Unlike the common audio systems for theaters which only install speakers at the audience's height, the addition of upper and lower level loudspeakers can bring out the following effects.

■ Conveying a perception of the vertical sound direction

The upper-level loudspeakers allow a listener to perceive sound arriving from above. For example, this will allow reproduction of scenes such as one in which an angel alights on a location above the seats and speaks to the audience, giving listeners the perception of sound in the vertical direction.

■ Synchronizing the video on the screen with the sound image

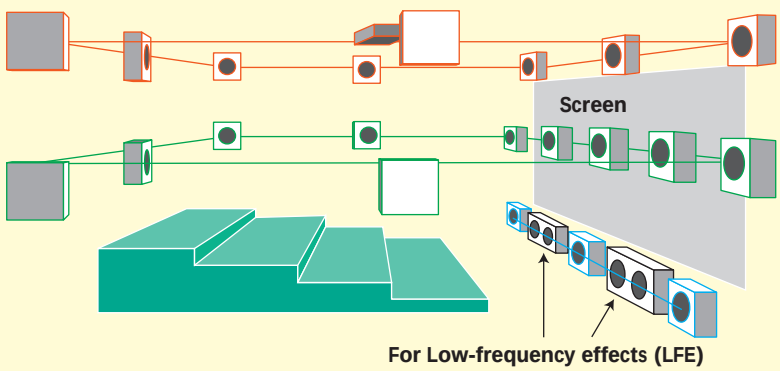
Super Hi-Vision has an extremely wide viewing field angle in the horizontal and vertical directions, in comparison with conventional HDTV and movies. This requires a highly accurate sound image localization that matches the sound image to the video data in a scene where a video cut moves in the up-down or right-left directions on the screen. The 22.2 channel system obtains sound image localization through the incorporation of a 5-channel middle level to localize over a wide horizontal viewing angle, and 3 higher and 3 lower level channels for localization over a large vertical viewing angle.

■ Maintaining the sensation of reality over a wide listening range

When reflected sound and reverberation are reproduced using only the middle level loudspeakers, the sound coming from loudspeakers installed on the side walls of the theatre sometimes becomes too loud, creating a sense of incongruity for listeners at seats near the wall. This issue was dealt with by making it so that only the upper loudspeakers reproduced reflected sound and reverberation, thus securing a good sound space impression throughout the seating area.

At present, we are preparing for an exhibition of "super-reality" audio based on the 22.2 channel audio system at Expo 2005 in Aichi, Japan. It will be at the "Super Hi-Vision Theater" in the Global House.

* LFE (Low Frequency Effect)



22.2 channel audio system loudspeaker arrangement

Koichiro HIYAMA,
Acoustics and Audio Signal Processing

Koichiro Hiyama received M.S. degree in acoustic design from Kyushu Institute of Design in 1994. And same year he joined NHK, Japan Broadcasting Corporation. In 1999 he joined NHK Science and Technical Research Laboratories and has worked in research on 3D audio systems. He is a member of the Audio Engineering Society (AES), the Acoustical Society of Japan.