Distance control of sound image using line array loudspeaker for three-dimensional audio visual system

[1]Yasushige Nakayama, [1]Hiroyuki Okubo, [1]Kimio Hamasaki, [1]Ando Akio and [2]Setsu Komiyama

[1]Science & Technical Research Lab. NHK, Japan, [2]Broadcast Engineering Department, NHK, Japan

E-Mail: nakayama.y-em@nhk.or.jp

Keywords: *Sound image distance control; Loudspeaker array; WFS; Mixed reality AV system.*

EXTENDED ABSTRACT

A loudspeaker array has been studied as a tool to control the distance of sound images in a 3D audio visual system.

The factors considered to affect the perception of sound image distance include intensity, spectrum, direct-to-reverberant energy ratio (D/R), and binaural difference. Recent studies suggest that in room environments, intensity and D/R are the primary acoustic cues to perceived distance. To control the distance, the above factors should be controlled as efficiently as possible.

A loudspeaker array such as the one shown in Figure 1, which consists of a large number of loudspeaker units (around 350 units!) and has input feeds through appropriate time delay processing, is able to control D/R by focusing the sound waves close to the listener. Subjective experiments using hundreds of loudspeaker units showed that the sound image is perceived to be near the focal point. However, 2-dimensional system has a disadvantage which is difficult to provide a wide listening area. Accordingly, a line array loudspeaker which is able to provide widely one would be necessary for a practical 3D audio-visual system for a two or more listeners.

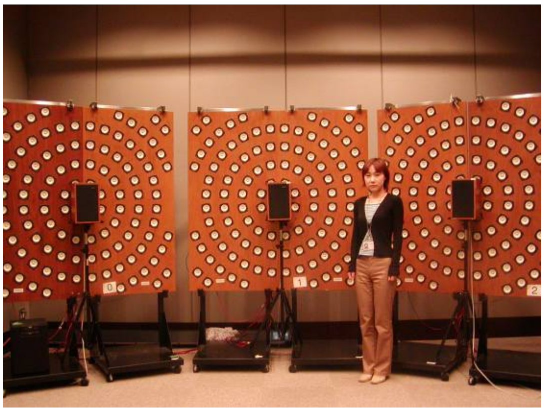


Figure 1. Two-dimensional loudspeaker arrays were used for conventional 3D audio visual experiments at NHK.

However, the line array system has a problem in that the effect of the sound image that is supposed to “jump out” at the listener tends to be smaller as the array loudspeaker size gets smaller.

We have developed a 3D audio visual system, called the “Mixed Reality Audio Visual reproduction system (MRAV)”, to facilitate studies on interactive 3D audio visual reproduction and to enable dynamic control of the sound field in synchronization with movement of the viewpoint in a virtual world. MRAV reproduces the sound image to create the virtual audio field including those distance by using line array loudspeaker systems (Figure 2).

Figure 2. A line array loudspeaker for MRAV

In this paper, four methods to put the sound image as far as possible from the listener were discussed in order to expand the range in which the perceived distance of the sound image can be controlled using line array loudspeakers.

This paper indicated that a method which reduces D/R by suppressing direct sound components without suppressing indirect sound components would be able to make the sound image seem very distant while keeping the change in tone quality small at the listening point. Moreover, D/R measurements showed that this method provided sound image control over a range of 2.5 m in equivalent distance to a real sound source.

1. INTRODUCTION

We have been studying 3D television and interactive virtual reality systems with the goal of producing super high presence programs for future broadcast systems. An example of such a service would be a virtual concert hall where any audience member can listen to a performance at favorite seat in the hall. In the last decade, VR (Virtual Reality) systems for representing 3D CG and 3D audio have been developed (e.g. Blauert et al). These systems try to achieve directional synchronization between the CG object and virtual sound source position and produce appropriate reverberant sound in the virtual space. To provide a sense of distance to the sound image, the attenuation of audio signal is only adjusted by controlling the sound power level in accordance with an inverse law. For an immersive sensation of reality, one also needs to reproduce sound images that move with the 3D visual images that "jump out" at the viewer. The creation of highly realistic programs may rest on the ability to control the listener's perceived distance of sound image as well as control of the perceived one's direction.

In such a way that the sense of distance of sound and an image is reproduced synchronously, Okubo et al. have developed a 3D audio visual system, called the "Mixed Reality Audio Visual reproduction system (MRAV)", to facilitate our studies on interactive 3D audio visual reproduction and to enable dynamic control of the sound field in synchronization with movement of the viewpoint in the virtual world. The MRAV provides listeners with 3D virtual space including the sense of sound distance by using the line array loudspeaker system.

Komiyama et al. proposed a sound image distance control system using array loudspeakers. The loudspeaker array, which consists of a large number of loudspeaker units on a 2-dimensional plane and input feeds through appropriate time delay processing, is able to control the direct-to-reverberant energy ratio (D/R) by focusing the sound waves close to the listener. Subjective experiments using hundreds of loudspeaker units showed that the sound image is perceived to be near the focal point. However, 2-dimensional system has a disadvantage which is difficult to provide a wide listening area. Accordingly, a line array loudspeaker which is able to provide widely one would be necessary for a practical 3D audio-visual system for a two or more listeners. But, the array system has a problem in that the effect of sound images that are supposed to "jump out" at the listener tends to be smaller as the array loudspeaker size gets smaller.

Previous studies (Komiyama et al) suggested that emphasizing the direct part of the sound helps to make the sound image jump out from the array loudspeaker. Moreover, they have indicated how many loudspeaker units are necessary to reproduce a sound image sufficient for 3D applications, and have identified the conditions regarding the number of loudspeaker units. These studies, however, have not investigated the method of naturally localizing a sound image at a distant point from the listener. Thus, in this study we have examined methods to localize a sound image far away from a listener by using a line array loudspeaker.

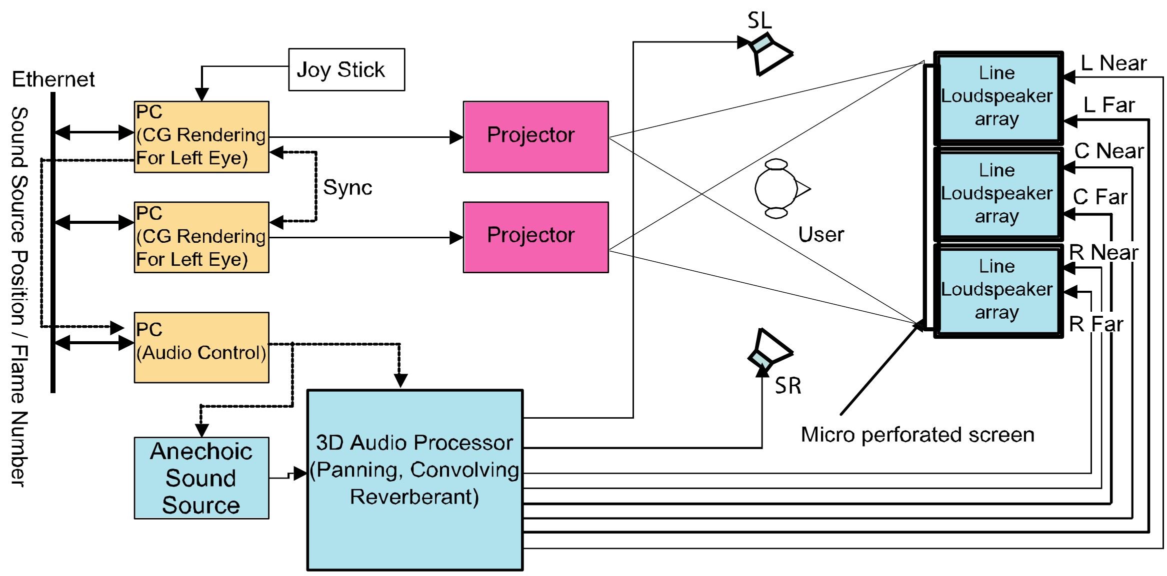


Figure 3. Outline of MRAV with line loudspeaker arrays (3-D audio visual experimental system of NHK).

2. PROPERTIES OF AUDITORY DISTANCE PERCEPTION WITH LOUDSPEAKER ARRAYS

2.1. Auditory Distance Perception

There are several signal properties that affect the human perception of auditory distance. Blauert discusses distance perception in terms of the human auditory system, while more recently, other researchers, e.g., Zahorik, Bronkhorst et al., and Nielsen, have experimentally examined the factors that affect distance perception. They have reported that sound waves radiating through space (e.g., across a room) from an audio source undergo multiple changes in their physical properties before reaching the listener, providing cues for judging the distance of the audio source. The following four physical factors have been proposed as cues for distance perception:

1) Intensity

2) Direct-to-reverberant energy ratio (D/R ratio)

3) Tone color

4) Binaural characteristics

It has also been confirmed that distance perception differs for different audio source. This phenomenon has been observed when judging the empirical distance of familiar audio sources such as speech. Zahorik showed experimentally that subjects make judgments not by applying these four factors uniformly, but by varying the weightings of these cues depending on the audio source. According to the report, the D/R ratio is more heavily weighted when judging the distance of a burst noise audio source compared with speech in a reverberant circumstance such as in an ordinary room.

2.2. Loudspeaker Array

Figure 4 shows the 2-dimensional loudspeaker array. The loudspeaker array can control all of the factors mentioned above. Below we discuss the technique of controlling D/R, which we found to be the property most affecting perceived distance.

To locally increase the sound pressure at a given point (focus) in the room, the signals that drive the loudspeaker array are delayed to compensate for the differences between the distance from the center of the loudspeaker array to the focus and the distance from each loudspeaker unit. When the listener is positioned slightly to the rear of that focus, the sound image is localized close to that

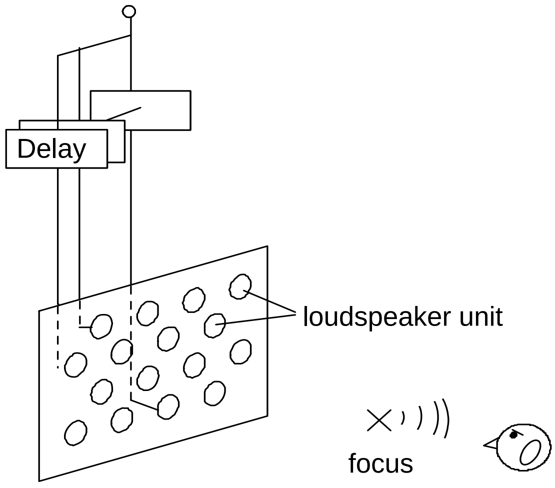
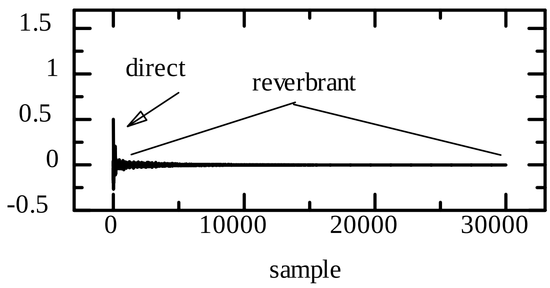
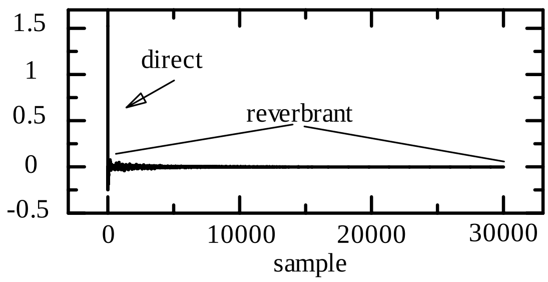


Figure 4. Loudspeaker array



(a) Loudspeaker placed 3 m from listening position.



(b) Loudspeaker array placed 3 m from listening position.

Figure 5. Impulse response of loudspeaker and loudspeaker array at listening position.

focus, so that the listener is given the impression that the sound is originating from that point.

Figure 5 shows an example of (a) the loudspeaker and (b) the loudspeaker array impulse response at the listening position, where the delay in the loudspeaker array was set for the sound image to be focused 50 cm from the listening position. As the figure shows, the loudspeaker array results in much higher sound pressure for the direct sound components at the listening position compared with the pressure for the loudspeaker. Given the sound reflections in a listening room, the direct sound components alone can be increased with little effect on indirect sound, and the energy ratio between the direct and indirect sound components (direct-to-reverberant energy ratio), which is a physical factor of distance perception, can be varied relatively. This makes it possible to control the listener's perception of the sound image's distance by using a loudspeaker array.

Komiyama et al. determined the number of loudspeakers necessary for producing a near sound image. Here, we considered that the loudspeaker is ideal monopole sound source. The wave front is premised on spreading in all the directions from the loudspeaker. When the interval between loudspeakers is larger than $\lambda/2$ (λ is wave length) and the listening position is very close to the focus, the equivalent monopole sound source distance, r_{eq} , can be written as follows:

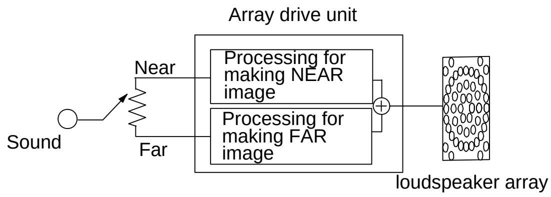
$$r_{eq} = \frac{d}{\sqrt{N}} \tag{1}$$

where d is the average distance between the loudspeaker array and its focus, and N is the number of loudspeakers. It follows that if the number of loudspeakers is 100 and the interval between them is $\lambda/2$ or more, the auditory distance will be close to d/10. When the number of loudspeakers decreases, the "jump-out" effect deteriorates.

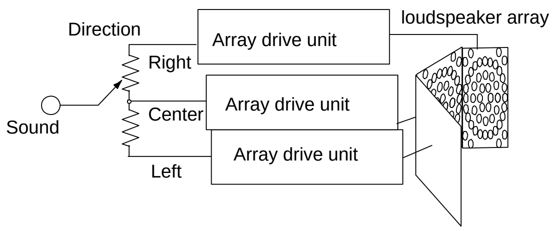
To continuously control distance by using a loudspeaker array, the position of the focus can be directly controlled in the same way as the actual sound source moves (We called "focus movement method"). In this method, increasing the number of sound sources being controlled also increases the calculations needed to control the delay, making real-time processing more difficult. There is also a tendency for continuity to be lost in the movement of the perceived sound image.

We attempted to devise an effective continuous real-time control over the listener's impression of distance by controlling the levels between the positions of a "near sound image" and a "far sound image" (We called "Amplitude zooming method"). This is equivalent to controlling the distance to the sound image between loudspeakers arranged in a straight line at different distances from the listening position, in the same way as the sound image direction can be controlled form side to side between two loudspeakers installed on the left and right sides of the room (amplitude panning method). To use the amplitude zooming method with the loudspeaker array, we first prepared a "near sound image" with the focus set very close to the listening position, as shown in Figure 4, and then prepared a "far sound image" located such that the differences between the paths from the loudspeakers would as far as possible not generate sound pressure peaks close to the listening position, thereby ensuring that the sound heard at the listening position would primarily be indirect sound (unlike the near sound image). Figure 6 (a) illustrates the distance control using Intensity zooming method.

Also, by arranging multiple loudspeaker arrays around the listening position and using intensity to control left-right directionality in the same way as a conventional stereophonic audio system, the sound image can be localized among the loudspeaker arrays. Using this method, each of the sound images generated at a given distance by the respective loudspeaker arrays behaves just as if there were a loudspeaker (virtual loudspeaker) at that distance, and the fusion of the left and right sound images between the virtual loudspeakers localizes the sound image between the loudspeaker arrays while giving the impression that the sound images are close by. Figure 6 (b) illustrates the direction and distance control method.



(a) Distance control of a sound image.



(b) Directional control of a sound image.

Figure 6. Sound image control using loudspeaker arrays.

An interesting issue is how a sound image is made to seem distant. For instance, if the number of speakers is limited to such as in a line array loudspeaker system, the controllable range of the sound image distance could be expanded in practical systems. In the following section, we discuss some techniques for localizing the sound image at a perceptually distant place.

3. METHODS FOR LOCALIZING SOUND IMAGE TO A DISTANT PLACE

As discussed previously, the number of loudspeakers is a determinant to the jump-out effect. To achieve this effect, the loudspeaker array controls the direct sound so that it is predominantly at a focal point near a listener. In contrast, if the indirect sound is controlled so that the focal point is away from the listener, the sound image will be perceived as far away. Below, we evaluate four techniques for a line loudspeaker arrays to make the indirect sound components dominant in the listening area.

(1) Addition of artificial reverberation (Type I)

Additional artificial reverberation is fed to the loudspeaker units. The listener gets the impression of a far sound image because the artificial indirect sound is dominant. However, the tone color is different from actual room reverberation. Additionally, since reverberation can be heard only from the direction of the loudspeaker array, a perceived location will feel unnatural. Nevertheless, this technique is useful for changing virtual sound environments as it can simulate concert halls, tunnels, rooms of various sizes, etc.

(2) Random time delay (Type II)

Sound signals with a random time delay are fed to loudspeaker units so that the direct sound won't be focused near the listener. The tone color difference between focused sound (near) and random time delay sound (far) is small, and desirable. However, if the time delay difference between each loudspeaker is small, the direct sound will become dominant at focal point, and thus the sound image does not get localized at a distant place. A suitable time delay differential between loudspeaker units is about 100 msec. If the time delay were longer than 100 msec, the phase difference between loudspeaker arrays becomes large and the phantom image defocuses.

(3) Negating direct components by phase inversion (Type III)

The reverse phase of each loudspeaker unit with a delay setup is saved and is focused so that only the direct sound is cancelled at the focal point. Consequently, the indirect sound of the room became dominant. A listener thus perceives the sound image to be distant. However, since the phase of the adjacent loudspeaker unit is reversed, low components of the radiation power of the whole line array loudspeaker are decreased. Therefore, the tone quality of the far sound image differs from the focused near sound image.

(4) Negating direct components by phase rotation (Type IV)

This technique, which was the subject of the experiment described in the next section, minimizes the tone quality difference between the near and far sound image by phase reversal. The phase difference of an adjacent unit is made as small as possible, and the phase control is performed so that the phase is reversed, and the direct sound near the focus is negated by minimizing the influence of the frequency characteristics of the radiation power of the whole line loudspeaker array. The direct sound can be cancelled, and at the same time, the tone quality difference between the near and far sound images can quickly be minimized.

Table 1 shows the summary of properties of above techniques to localizing sound image to a distance. The far sound image made by Type IV technique is most orientated to nature and a distant place in the above techniques. Accordingly, Type IV technique is employed for creating the far sound image using a line array loudspeaker.

Table 2. The summary of properties of four techniques to localizing sound image to a distance

Technique type	I	II	III	IV
Naturalness of indirect component	△	○	○	○
Tone quality	○	○	×	△
quality of latitude localized	○	×	○	○
domination of indirect component	○	×	○	○

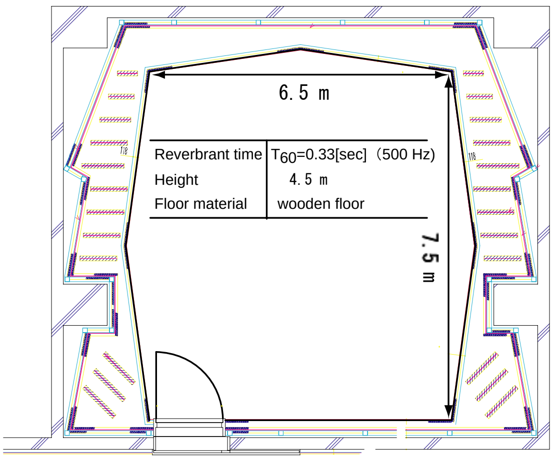
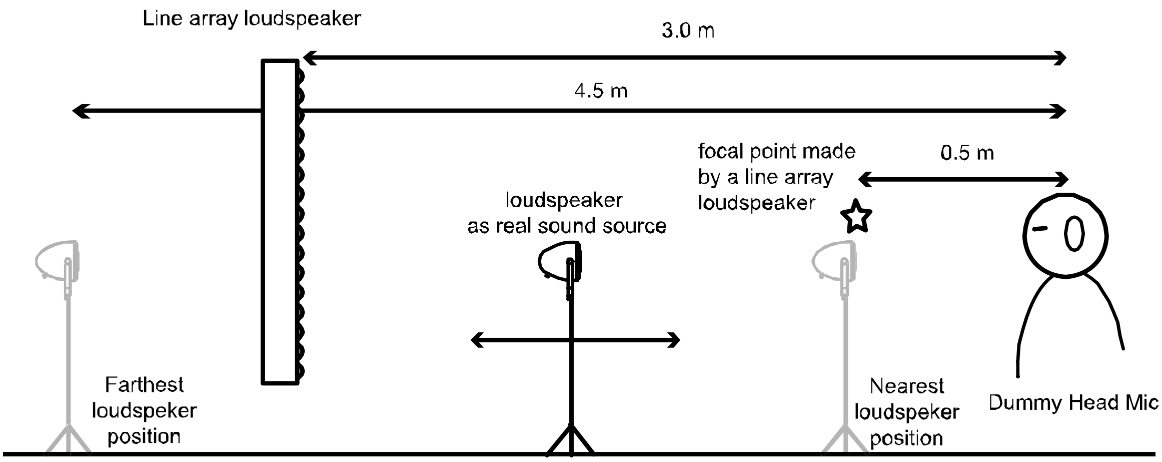


Figure 7. Experimental room



4. EXPERIMENT

4.1. Criterion of assessment for D/R

The measurements were made in our experimental room (Fig. 7). Figure 8 outlines the microphone and speaker arrangement. The D/R ratio is defined by the equation

$$D/R = 10\log_{10}\left(\int_0^{0.05} s^2(t)\,dt \middle/ \int_{0.05}^{\infty} s^2(t)\,dt\right) \text{ [dB]} \quad (2)$$

where s(t) is the Binaural Room Impulse Response (BRIR) as measured by a dummy head microphone (CORTEX).

4.2. D/R and distance to real sound source

D/R was measured when changing the distance of a real sound source's D/R and sound pressure, and it was measured when changing the distance to the real sound source. A loudspeaker (FOSTEX FX120) was used as the real sound source. The BRIR from the loudspeaker to the microphone was measured with the TSP signal, and D/R was obtained from the BRIR. The distance from the loudspeaker to microphone was changed in 0.5 m steps from 0.5 m to 4.5 m. The sound pressure at the listening position was measured using a sound-level meter (RION NA-27). The sound pressure was set to 80 dB when the dummy head was 0.5 m from the loudspeaker. Figure 9 shows the measured results. Since the sound pressure decreases linearly according to speaker distance, it turns out that the environment with direct sound is superior to the indirect sound one.

4.3. D/R and distance of sound image using line array loudspeaker

The D/R of the far and near sound images was measured in the same way as with the real sound source. The distance from the listening position to

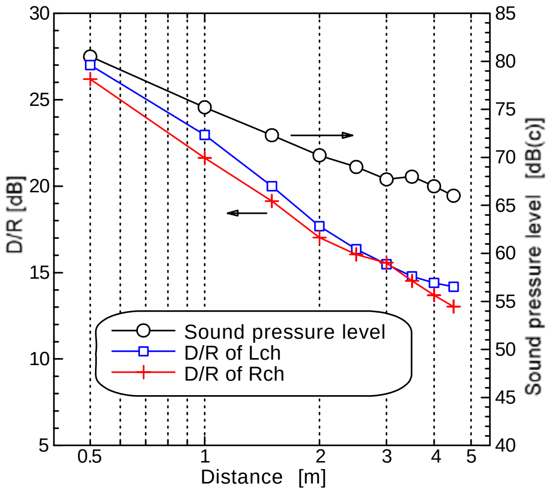


Figure 9. Sound pressure level and D/R for a real sound source.

the line loudspeaker array was set at 3 m. To make the near sound image, the sound signal was fed to each loudspeaker unit with a suitable delay so that the focus would be 0.5 m from the listening position. On the other hand, to make the far sound image, the phase was controlled such that the sound signals fed to each loudspeaker unit would be increased by 22.5 degrees from top to down. The distance to the sound image was continuously changed by using the gain factor of the near sound image and far sound image with sine and cosine as level control curves. Since only the sound of the near image exists when the gain factor for far image is 0 and the gain factor of near image is 1, listeners will hear the direct sound concentrated at the focus (sound image feels nearby). Moreover, when the gain factor for far image is 1 and the gain factor of near image is 0, the direct sound is negated at the focal point and listeners will hear mainly the indirect sound from the room (sound image feels far away). The BRIR was measured with the dummy head microphone using the TSP signal as in the case of the real sound source sound, and the D/R was calculated from it. Figure 10 shows the D/R of the sound image of the line array loudspeaker together with the gain factors.

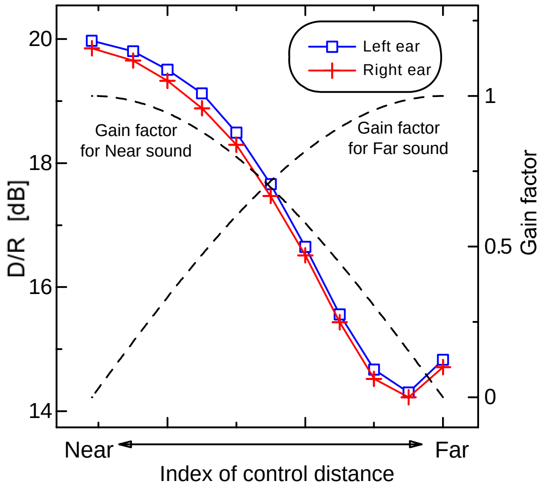


Figure 10. The D/R of line array loudspeaker and The gain factors of near and far sound images for distance control.

The results indicate that the D/R differences of near and far sound by using type IV technique was about 5 dB. Table 2 shows the equivalent sound image distance of the farthest and nearest sound images as computed from the D/R of the real sound source. It turns out that the range of image distance control for the line array loudspeaker is about 2.5 m. The listener can best hear the far sound image 1 m away from a line array loudspeaker (a maximum of 4.1 m in equivalent distance) rather than at the distance at which line array loudspeaker was installed (3 m).

Table 2. Comparing the sound images of real sound source and line array loudspeaker

	farthest sound image		nearest sound image	
	D/R	EQSD*	D/R	EQSD
L-ch	14.3 dB	4.1 m	19.9 dB	1.5 m
R-ch	14.2 dB	3.6 m	19.8 dB	1.3 m

EQSD*=EQuivalent Sound image Distance

5. SUMMARY

The control of the perceived distance to a sound image by using a line array loudspeaker was discussed. Four techniques for making a sound image seem to originate from a distant place were discussed. The technique of controlling the phase difference between loudspeaker unit as small as possible, setting up a focus near the listeners, was deemed best for both canceling the focused direct sound and making the tone quality change with the sound image small. Moreover, D/R measurements indicated that the line array loudspeakers could localize the sound image at 4 m (equivalent real sound source distance).

Further study is needed to determine an adequate gain factor curve of near and far sound for obtaining a D/R realistically reflecting to the change in a real sound source with distance.

6. REFERENCES

Blauert,J. Lehnert,H. Sahrhage J. and Strauss, H. (2000), An Interactive Virtual-Environment Generator for Psychoacoustic Research. I: Architecture and Implementation. ACUSTICA, Vol. 86, pp.94-102

Bronkhorst, A.W. Houtfast, T. Auditory distance perception in rooms, Nature,397,pp.517-520.

Funkhouser, T. Tsingos, N. Carlbom, Elko, I. G. Sondhi, M. West, J. Pingali, E. Min, G. P. and Ngan, A. (2004), A beam tracing method for interactive architectural acoustics. J. Acoust. Soc. Am., Vol.115, No.2, pp.739-756

Gilkey, R.H. Anderson, T.R. (1997), Binaural and Spatial hearing in Real and Virtual Enviroments. Library of Congress Cataloging-in-Publication Data,.

Komiyama, S. Nakayama, Y. Ono, K. Koizumi, S. (2003), A loudspeaker-array to control sound image distance. Acoust. Sci.& Tech.24,5,242-249.

Larson,P. Vaestfjaell, D. and Kleiner, M. (2002), Better Presence and Performance in Virtual Environments by Improved Binaural Sound Rendering. Proc. AES 22nd Conf.

Nielsen, S.H. (1993), Auditory distance perception in different rooms. J.Audio Eng.Soc.41, pp755-770.

Okubo, H. Nakayama, Y. Naito, Y. Ikenaga, T. Komiyama, S.(2005), An Application of Line-formed Loudspeaker Array System for Mixed Reality Audio-Visual Reproduction System. Preprints of AES 119th Convention, New York

Savioja, L. Huopaniemi, J. Lokki, T. and Vaeaenaenen, R. (1999), Creating Interactive Virtual Acoustic Environments. J. Audio Rng. Soc., Vol. 47, No. 9, pp. 675-705 (1999)

Tsingos,N. Gallo, E. and Drettakis, G. (2004), Perception Audio Rendering of Complex Virtual Environments. ACM Trans. Graphics, Vol.23, No.3, pp.249-258

Zahorik P. (2002), Direct-to-reverberant energy ratio sensitivity. J.Acoust.Soc.Am, 112(5), Pt.1, pp2110-2117.