

22.2 Multichannel Sound System for Ultra High-Definition TV

Kimio Hamasaki¹, Toshiyuki Nishiguchi¹, Reiko Okumura¹, Yasuhige Nakayama¹ and Akio Ando¹

¹ NHK Science & Technical Research Laboratories, Tokyo, 157-8510, Japan

ABSTRACT

The 22.2 multichannel sound system was developed for an Ultra High-Definition TV (UHDTV) system. It consists of three layers of loudspeakers: an upper layer with nine channels, a middle layer with ten channels, and a lower layer with three channels and two channels for LFE (Low Frequency Effects). Subjective evaluations were performed for comparing the impression of various spatial attributes using UHDTV contents with pictures in large room. These evaluations demonstrate that viewers have better impressions of various spatial attributes in a wider listening area with the 22.2 multichannel sound system than with other sound systems. This paper describes the 22.2 multichannel system for UHDTV, and discusses the advantages of the 22.2 multichannel sound. It also describes the sound recordings and productions by the 22.2 multichannel sound system.

1. Introduction

The high-definition television (HDTV) system used in digital broadcasting in Japan provides high-resolution images to viewers in their homes. Although HDTV produces more realistic images than conventional standard television, the sensation of presence and reality depend strongly on viewing angle. Research has found that the sensation of reality increases as the viewing angle increases, becoming saturated at a horizontal angle of around 100° [1].

An UHDTV system with 4000 scanning lines was developed to achieve this 100° horizontal viewing angle. The system has 16 times the resolution of HDTV, more than twice the resolution of a 70-mm film, and provides a viewing distance of 0.75 H (H: picture height). Table 1 compares the specifications of the UHDTV with 4k×8k pixels with those of the HDTV system [2].

A 5.1 surround sound system has also been used in digital broadcasting in Japan and currently provides two-dimensional spatial impression to viewers in their homes. However, several issues must be solved to create a very natural spatial impression by the conventional multichannel sound systems. Japan Broadcasting Corporation (NHK) has developed a 22.2 multichannel sound system for UHDTV to reproduce an immersive and natural three-dimensional sound field that creates a superior sense of presence and reality [3]. The 22.2 multichannel sound system was first made public in the UHDTV Theater at World EXPO 2005 in Japan [4].

	HDTV	UHDTV
Number of pixels	1080 × 1920	4320 × 7680
Viewing angle (pixel invisible)	30° horizontally	More than 100° horizontally
Relative resolution	Equivalent to 35-mm motion film	More than twice 70-mm motion film

Table 1: HDTV and UHDTV system specifications

NHK produced various ultrahigh-definition (UHD) video programs with the three-dimensional sound of a 22.2 multichannel audio system. Two programs were produced for World EXPO 2005 held in Aichi, Japan. Subjective evaluations of sound in these EXPO programs were conducted using the semantic differential (SD) method [5] to investigate the wideness of the optimum listening area on different multichannel sound systems including 2.0, 5.1 and 22.2 multichannel sound. About 50 subjects who were selected among professional sound engineers, students of sound engineering and music, gave subjective responses to the program. The stimuli were selected from UHD video content from World EXPO 2005 using 22.2 multichannel sound, and were reproduced in the theatrical environment of an UHDTV theater with a 450-inch screen. The subjects were asked to report their impressions of different attributes of spatial sound at different listening positions for three different sound systems: 2 channel stereo, 5.1 surround sound, and 22.2 multichannel sound.

The results of these subjective evaluations demonstrate that 22.2 multichannel sound systems provide a very wide listening area with exceptional spatial sound quality. This paper describes the details of these subjective experiments and discusses the improvement in spatial sound quality of a three-dimensional sound system as compared with conventional two-dimensional sound systems such as 5.1 surround sound.

With the development of 22.2 multichannel sound and an UHDTV system for future broadcasting, the possibility of live transmission has become important and must be demonstrated. Live transmission of 22.2 multichannel sound with UHD video was demonstrated through an optical fiber network in 2005. Uncompressed audio and video signals were transmitted at about 28 Mbit/s and 24 Gbit/s, respectively, over a distance of 260 km on an optical fiber network.

However, such huge data rates are not practical for broadcasting. Therefore, a live transmission demonstration with a new scheme was carried out to demonstrate the possibility of practically applying the system to future broadcasting. UHDTV was successfully transmitted at a data rate of about 640 Mbit/s using an existing IP optical network in 2006. The UHD video signal was encoded by the MPEG-2 coding scheme and compressed to less than 600 Mbit/s from its original data rate of 24 Gbit/s. A 32-channel digital audio signal sampled at 48 kHz with 24-bit sampling was embedded in the MPEG-2 streams, so the 22.2 multichannel sound of uncompressed digital audio was transmitted over the IP network from Tokyo to Osaka, about 500 km to the west. This transmission experiment was also the first time the three-dimensional sound of a large-scale musical TV program has been live mixed using a 22.2 multichannel sound system.

This paper also discusses various issues of sound design, sound capturing, and mixing of three-dimensional sound based on experimental productions of 22.2 multichannel sound. These are important because three-dimensional sound design needs new and different approaches to move beyond 5.1 surround sound.

2. 22.2 multichannel sound system

UHD video was developed to achieve the highest possible sensation of reality, so the sound system used with UHD video should contribute to this impression of presence. The sound field reproduced by conventional multichannel sound systems, such as 5.1 surround sound system, for movies is strongly related to the visual image on the screen. Therefore, multichannel sound systems for conventional movie theaters prioritize frontal sound reproduction over rear sound reproduction. However, the UHD video system provides an immersive sensation because of the wide viewing angle associated with an extra large screen. The multichannel sound system therefore has to provide a sound field that surrounds viewers with various sound sources.

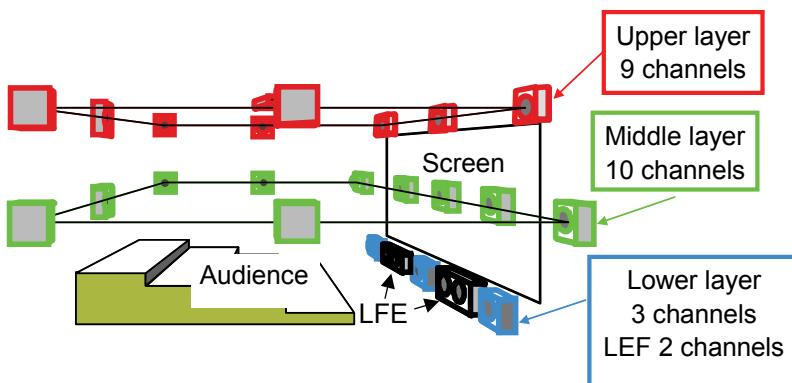


Figure 1 : 22.2 multichannel sound system

and vertical direction of the visual image and the sound match. Based on these requirements, the sound system of an UHDTV should:

- Stably localize frontal sounds over the entire screen area.
- Reproduce sound image in all directions around a viewer including elevation.
- Reproduce a three-dimensional spatial impression that augments the sense of reality.
- Create a wide listening area with exceptional sound quality.
- Be compatible with existing multichannel sound systems.

The 22.2 multichannel sound system was developed to fulfill these conditions. As shown in Figure 1, the system consists of loudspeakers with an upper layer of nine channels, a middle layer of ten channels, and a lower layer of three regular channels and two low frequency effects (LFE) channels. Table 2 and Figure 2 show the name and number of each channel in the 22.2 multichannel sound system, and Figure 3 shows the horizontal standard arrangement of each channel in a 22.2 multichannel sound system.

Previous research, which evaluated the sense of presence created by the sound of symphony music with no accompanying picture, showed that the 22.2 multichannel audio system can reproduce a greater sensation of presence over a wider listening area than a conventional multichannel audio system, and that the upper layer of loudspeakers is essential for reproducing a better sense of presence [3].

3. Subjective evaluations

Subjective evaluations were conducted to assess the effective wideness of listening areas of three different multichannel audio

The horizontal viewing angle of an UHD video is around 100° in the front row and around 40° in the back row in a typical UHDTV theater. Therefore, the horizontal and vertical angles of movements in the visual image are extremely wide in UHD video as compared to conventional movie theaters. Consequently, the ability to localize sound content over the screen images should be improved so that the horizontal

Channel No.	Channel name (Shortened form)	Channel name
1	M-F-L	Middle layer - Front - Left
2	M-F-L/C	Middle layer - Front - Left/Center
3	M-F-C	Middle layer - Front - Center
4	M-F-C/R	Middle layer - Front - Center/Right
5	M-F-R	Middle layer - Front - Right
6	M-S-L	Middle layer - Side - Left
7	M-S-R	Middle layer - Side - Right
8	M-B-L	Middle layer - Back - Left
9	M-B-C	Middle layer - Back - Center
10	M-B-R	Middle layer - Back - Right
11	U-F-L	Upper layer - Front - Left
12	U-F-C	Upper layer - Front - Center
13	U-F-R	Upper layer - Front - Right
14	U-S-L	Upper layer - Side - Left
15	U-S-C	Upper layer - Side - Center
16	U-S-R	Upper layer - Side - Right
17	U-B-L	Upper layer - Back - Left
18	U-B-C	Upper layer - Back - Center
19	U-B-R	Upper layer - Back - Right
20	L-F-L	Lower layer - Front - Left
21	L-F-C	Lower layer - Front - Center
22	L-F-R	Lower layer - Front - Right
23	LFE-L	LFE - Left
24	LFE-R	LFE - Right

Table 2: Channel name and assignment of 22.2 multichannel sound

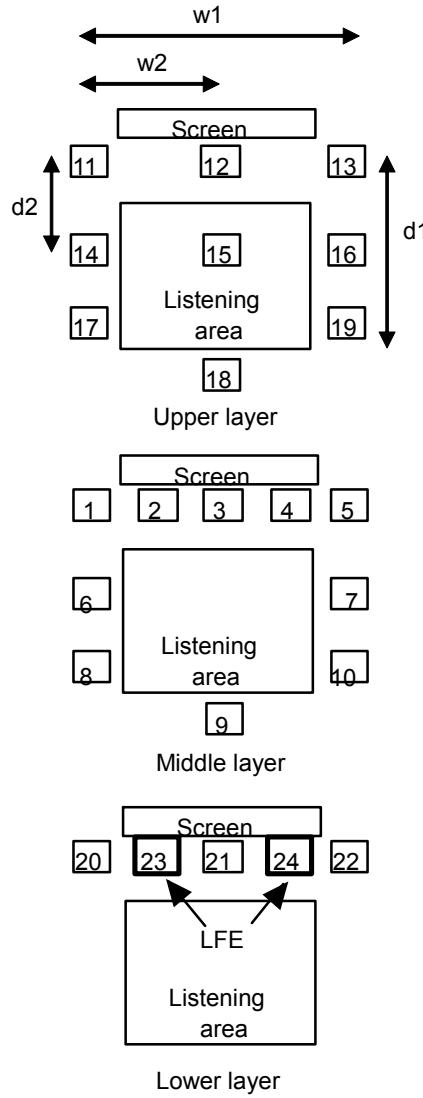


Figure 2: Loudspeaker arrangement for channels in a 22.2 multichannel sound system

EXPO 2005. Table 3 lists the sound content of each scene of the two programs (A and B). Programs A and B were screened in the first and second halves of the exposition, respectively. The two programs were originally produced using the 22.2 multichannel sound system, so the 5.1 surround sound and 2 channel stereo sound were produced by down mixing from the original 22.2 multichannel sound based on the down mixing equations shown in Figure 4. The coefficients (k , m , n) in the down mixing equations adjust the total sound level and balance the sound levels in the upper, middle, and lower layers. The k , m , and n during the subjective evaluations were set to 1.0 to maintain the same total loudness in the different multichannel sound systems.

As both programs (A and B) had 4 different parts, and each part had three different sound reproductions, there was a total of 24 stimuli (segments of content) in the experiment.

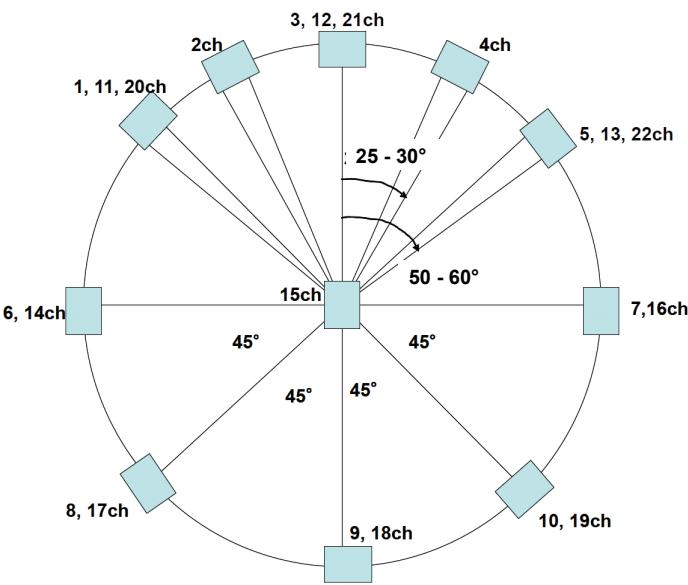


Figure 3: Standard horizontal arrangement of each channel in a 22.2 multichannel sound system

systems: 2 channel stereo, 5.1 surround sound, and 22.2 multichannel sound. The stimuli for the subjective evaluations were selected from the World EXPO 2005 programs. The sounds and video were screened in NHK Labs' UHDTV Theater (a 450-inch screen) for the theatrical environment evaluation. The subjects were asked to report their impressions of the sound provided by the different sound systems when shown with different images on the different screens. The subjects were asked to sit in different positions so differences in impressions based on position with regard to the screen could be measured.

3.1. Stimuli for evaluations

Two programs, each of which was about eight minutes long and divided into four scenes, were produced for World

	Program A	Program B
Scene 1	Short sound clip of three dimensional movement without video followed by music with female singers with video	Three dimensional movement of sound based on CG followed by symphony orchestra music
Scene 2	Music and surrounding sounds of shooting location	Symphony orchestra music followed by three dimensional sound of water gushing from a huge dam
Scene 3	Typical three dimensional sound scene for Japanese festival followed by music	Three dimensional sound creation of sound effects with symphony orchestra music
Scene 4	Music with children singing	Rocket launch

Table 3: Sound content in each Program of ultrahigh-definition TV for World EXPO 2005

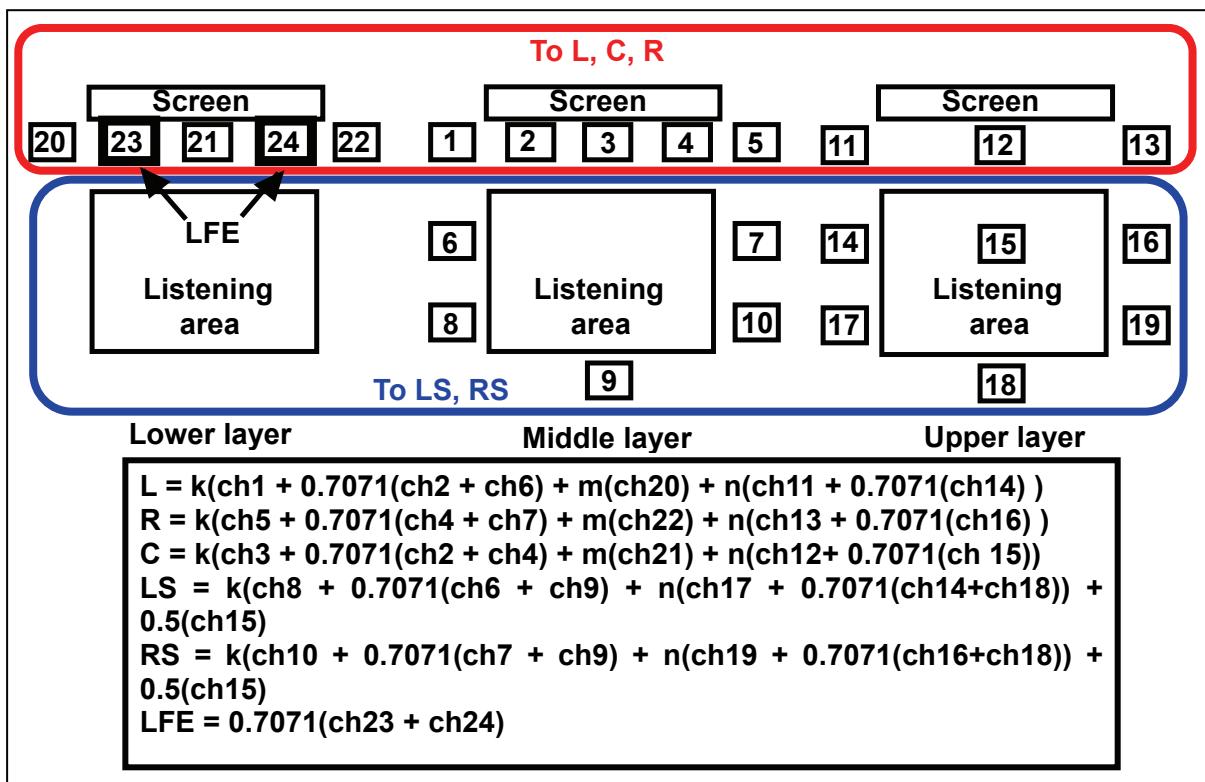


Figure 4: Down mixing sound from 22.2 to 5.1 and 2.0

3.2. Acoustical conditions during evaluations

Experiment was done in the UHDTV Theater. The reverberation time in the theater was 0.70 s at 500 Hz. The reference sound level for each channel was set to 85 dB(C) by the pink noise of -18 dBFS(RMS) in experiment. The reference sound level of LFE channel was set to 95 dB(C) by the pink noise of -18 dBFS(RMS) in the LFE frequency band.

The channels (speakers) in the 22.2 multichannel sound system were numbered as shown in Figure 2. The sound stimulus of the 5.1 surround sound was reproduced by channels 1, 3, 5, 8, 9, and 23, and the sound stimulus of the 2 channel stereo was reproduced by channels 1 and 5. The distance between

speakers 1 and 5, which was the same as the distance between speakers 11 and 13 and is indicated as w1 in Figure 2, was 11 m in the UHDTV Theater. The distance between speakers 1 and 3, which was the same as the distance between speakers 11 and 12 and is indicated as w2 in Figure 2, was 5.5 m. The distance between speakers 5 and 9, which was the same as the distance between speakers 13 and 19 and is indicated as d1 in Figure 2, was 9.5 m. The distance between speakers 1 and 6, which was the same as the distance between speakers 11 and 14 and is indicated as d2 in Figure 2, was 4.75 m. The average height at which the middle layer loudspeakers were positioned was 2 m and the average height at which the upper layer loudspeakers were positioned was 6.28 m in the UHDTV Theater.

From 3 to 10 subjects listened to the stimuli at once while sitting in positions spaced at 1 m intervals around the central listening position during screenings in the UHDTV Theater. The layout of these positions is shown in Figure 5.

Term	Scale	Term	Short description in figures
dynamic	1-2-3-4-5-6-7	static	dynamic
gaudy	1-2-3-4-5-6-7	plain	gaudy
powerful	1-2-3-4-5-6-7	weak	powerful
interesting	1-2-3-4-5-6-7	not interesting	interesting
delightful	1-2-3-4-5-6-7	dull	delightful
impressive	1-2-3-4-5-6-7	not impressive	impressive
distinct	1-2-3-4-5-6-7	indistinct	distinct
beautiful	1-2-3-4-5-6-7	dingy	beautiful
transparent	1-2-3-4-5-6-7	not transparent	transparent
clear	1-2-3-4-5-6-7	muddy	clear
bright	1-2-3-4-5-6-7	dark	bright
loud	1-2-3-4-5-6-7	soft	loud
unique	1-2-3-4-5-6-7	ordinary	unique
real	1-2-3-4-5-6-7	unreal	real
natural	1-2-3-4-5-6-7	unnatural	natural
presence	1-2-3-4-5-6-7	no presence	presence
deep	1-2-3-4-5-6-7	flat	depth
wide in left and right	1-2-3-4-5-6-7	narrow in left and right	width
wide in front and rear	1-2-3-4-5-6-7	narrow in front and rear	front/rear
wide in above and below	1-2-3-4-5-6-7	narrow in above and below	up/down
clear movement of sound	1-2-3-4-5-6-7	obscure movement of sound	movement
sound come from every direction	1-2-3-4-5-6-7	sound come from lopsided direction	direction
rich reverberant	1-2-3-4-5-6-7	poor reverberant	reverberant
rich envelopment	1-2-3-4-5-6-7	poor envelopment	envelopment

Table 4: Pairs of evaluation terms

3.3. Evaluation method

The SD method was used in this experiment. Subjects were asked to rate their impressions on a seven-point scale for the pairs of evaluation terms listed in Table 4. Each pair contained two terms

with opposite meanings, such as “dynamic” and “static”. Subjects were asked to select a score from 1 to 7, where 1 means very dynamic, 2 means fairly dynamic, 3 means slightly dynamic, 4 means neither dynamic nor static, 5 means slightly static, 6 means fairly static, and 7 means very static. Subjects rated each stimulus (segment of content) using the 24 evaluation pairs listed in Table 4. There were 53 subjects (28 university students of music or audio engineering and 25 audio professionals) in experiment.

3.4. Results of experiment

Figure 6 shows the total mean values for each pair of evaluation terms for each sound system to compare the basic features of the three different multichannel sound systems. Each mean value calculated from all results obtained in experiment is marked with a 95% confidence interval for different terms and for different sound systems. The horizontal axis represents the scale listed in Table 4, and the vertical axis contains each pair of evaluation terms.

Figure 6 shows that the 22.2 multichannel sound system for every evaluation term except “loud”, and was rated significantly better than 5.1 surround sound system for the terms “front/rear”, “up/down”, “movement”, “direction”, “reverberant”, and “envelopment” (short descriptions of terms are listed in Table 4). The results also show that the 5.1 surround sound system was rated significantly better than the 2 channel stereo system for every term except “loud”. The results suggest that there was no difference in the loudness of the different sound systems. The results also suggest that the 22.2 multichannel sound system provided a better three-dimensional spatial sound quality than either the 2 channel stereo system or the 5.1 surround sound system.

3.5. Discussion

Subjective evaluations using the semantic differential method showed that the 22.2 multichannel sound was better at 5% significance level for various sensations than the 2 channel stereo or the 5.1 surround sound system in the theatrical environment with a 450-inch UHDTV screen.

The other attributes such as reality, transparency, and gaudiness were not rated more highly for the 22.2 multichannel sound than for the 5.1 surround sound in the theatrical environment. This might be due to the influence of the picture on the audio impression, because the conditions in the theater had a much larger and higher resolution screen than that in the domestic environment.

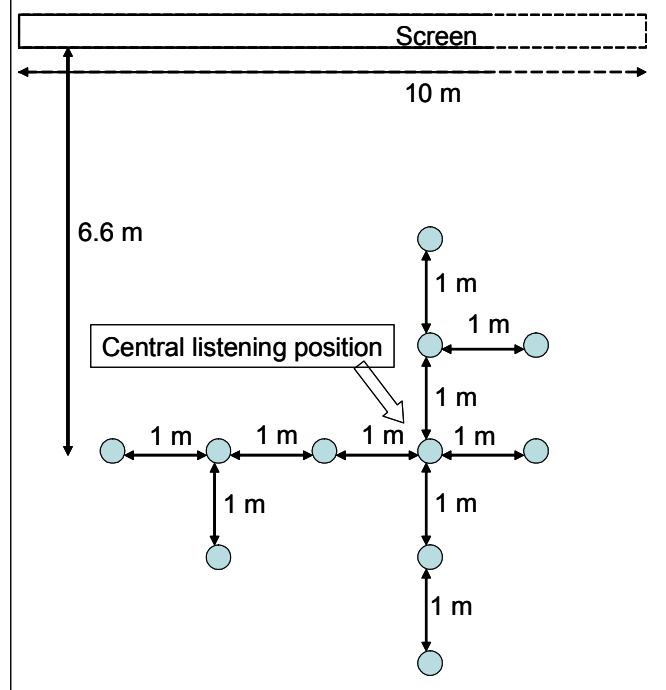


Figure 5: Listening positions in experiment

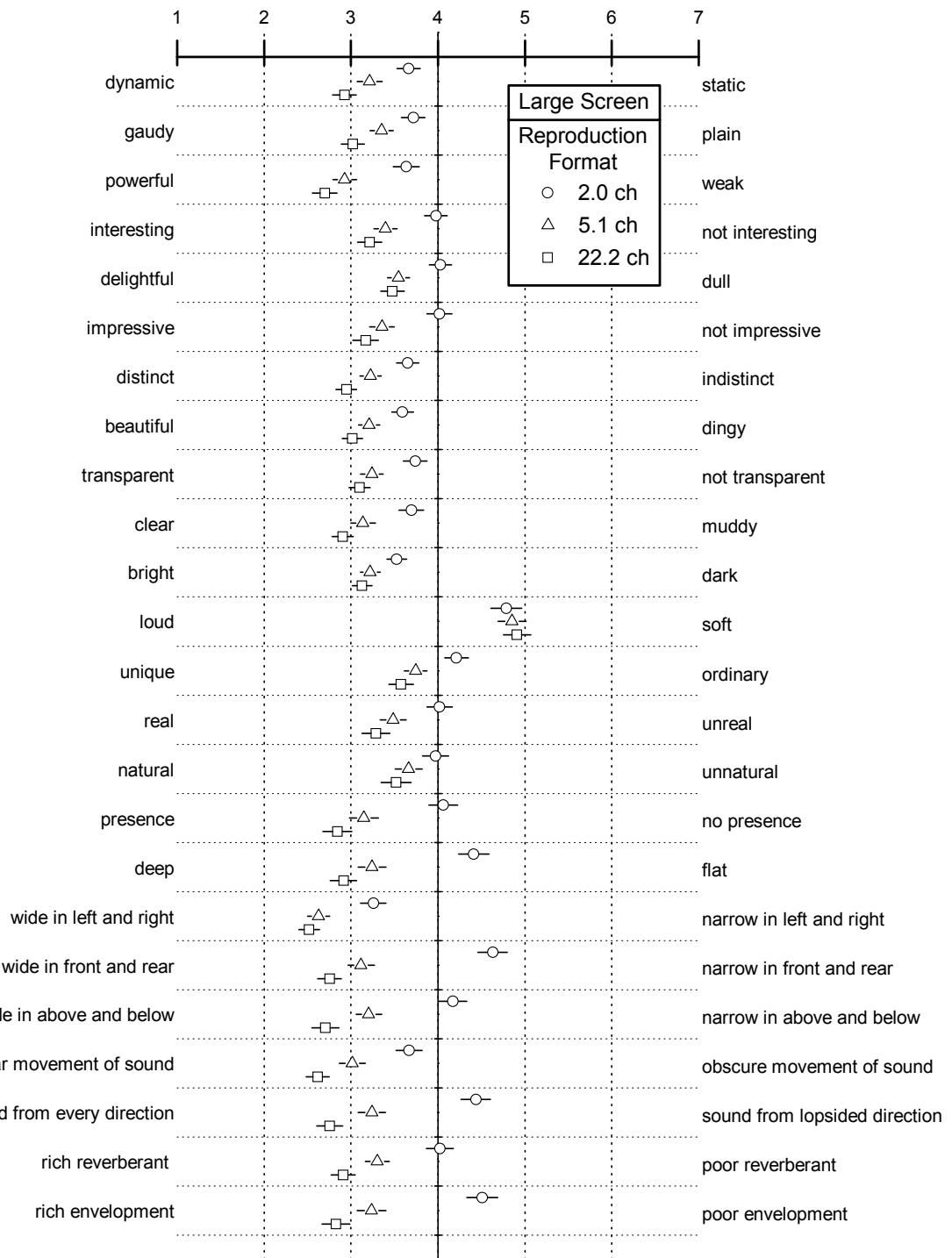


Figure 6: Evaluations of 2.0, 5.1, and 22.2 channel sound systems using a large screen

4. Sound design and mixing approach in the three-dimensional sound reproduction system

Mixing of music must be discussed from both acoustic and artistic viewpoints. Reproduction of the ambience of a concert hall in musical shows can be easily discussed from both viewpoints. However, mixing of pop music cannot be easily discussed from the viewpoint of acoustics or physics, and the sound design of TV programs using three-dimensional sound systems very much depends on personal artistic taste and the content of the program. In the early stages of development of 22.2 multichannel sound, several subjective evaluations were done to demonstrate the effectiveness of the 22.2 multichannel sound system in creating a three-dimensional sound field with music and other programming [3], [6], [7]. These evaluations were done using already mixed and produced program content, but no evaluations were done in terms of mixing. Although it is not easy to quantitatively analyze the mixing, it is very important to qualitatively analyze it to develop the production tools for three-dimensional sound and for scientific research of spatial audio.

4.1. Contribution of individual channels to 22.2 multichannel sound mixing

Magnitude estimation (ME) was used to investigate how the recording engineer and sound designer used each channel of the 22.2 multichannel sound system in their mixing. Six engineers and one sound designer who already had a lot of experience mixing or designing sound in 22.2 multichannel systems were asked to rate how necessary each channel was for mixing the 22.2 multichannel sound and to rate how much each channel contributes to overall mixing by rating each channel as compared to channel 3 (channel M-F-C) which was given a reference value of 100 points. The engineers and designer were free to score as they liked. If they felt that a channel was twice important comparing with channel M-F-C, they could give it 200 points.

Figure 7 shows the results of this assessment of a total mix of various programs. The score in the frames reflecting the placement of the channels (the channel number and name of each frame is shown in Figure 2 and Table 2) are geometrical means of the scores given by all subjects. Total mix means the final mixing of a program in which sound engineers and designers decide the final balance of the three stems (sound effects, music, and speech). Almost every channel except channel U-S-L, U-S-R, and U-B-C scored higher than 70 points. These results show that sound engineers and designers used every channel effectively.

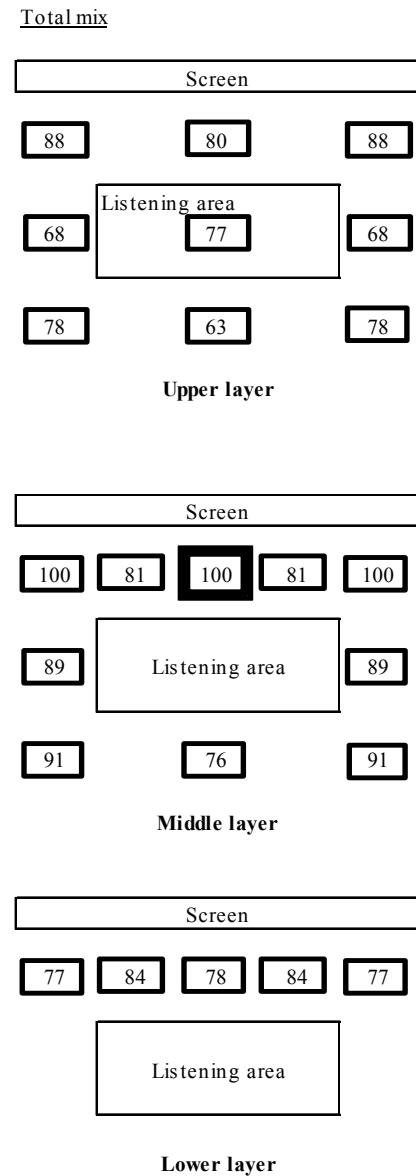


Figure 7: Need for each channel compared with M-F-C channel for total mix

4.2. Three-dimensional sound mixing

While several ultrahigh-definition TV programs have already been produced, sound engineers and designers still do not have adequate know-how and experience of three-dimensional sound. The conventional, current applications of layers of their mixing are enumerated below.

Upper layer

- Reverberation and ambience
- Sound localized above, such as loudspeakers hung in gymnasiums and airplanes and at fireworks shows
- Unusual sound such as meaningless sound

Middle layer

- Forming the basic sound field
- Reproducing envelopment

Lower layer

- Sounds of water such as the sea, rivers, and drops of water
- Sound on the ground in scenes with bird's-eye views

Sound engineers are also discussing several issues in three-dimensional sound mixing. The principal issues are listed below.

- Effective use of the upper and lower layer
- Three-dimensional movement of sound sources
- Creating a sense of elevation
- Interaction between immersive audio and visual cues

Basic features and advantages of three-dimensional sound with the 22.2 multichannel system are well understood and have been shown by several evaluations and various productions. However, further research and production are still necessary to develop three-dimensional sound production know-how. Three-dimensional sound production tools also must be developed.

5. Conclusion

Results obtained by subjective experiments clearly indicate that three-dimensional sound by a 22.2 multichannel sound system can produce better sensations of spatial sound quality, reality, and presence in a wider listening area than 5.1 and 2.0 multichannel sound systems. However, objective experiments using different stimuli will be necessary to support this conclusion, because the evaluation seems to depend on content. Further studies must also be done to investigate the interaction between audio and visual cues on immersive presence audio-visual systems.

Results obtained by magnitude estimation shows that each channel of a 22.2 multichannel sound system seemed to be used effectively by the sound engineers and sound designer to mix the three-dimensional sound. Further studies in terms of sound productions using 22.2 multichannel sound system is necessary to establish three-dimensional sound mixing know-how and for the development of advanced tools for three-dimensional sound production.

6. Acknowledgements

The authors thank the staff of Bose K.K., YAMAHA Corporation, and Fairlight Japan Inc. for their help in building the 22.2 multichannel sound system and producing the content. We also thank the staff of the NHK production operations center for their contribution in producing the demonstration content for the 22.2 multichannel sound system at World EXPO 2005 in Aichi, Japan.

7. References

- [1] T. Hatada et al., "Psychophysical analysis of the sensation of reality induced by a visual wide-field display", SMPTE Journal, Vol. 89, pp. 560-569 (1980)
- [2] F. Okano, M. Kanazawa, K. Mitani, K. Hamasaki, M. Sugawara, M. Seino, A. Mochimaru, and K. Doi, "Ultrahigh-definition Television System with 4000 Scanning Lines", presented at NAB 2004, Las Vegas, USA (2004)
- [3] K. Hamasaki, K. Hiyama, T. Nishiguchi, and K. Ono, "Advanced multichannel audio systems with superior impression of presence and reality", AES 116th Convention, Berlin, Germany, Convention paper 6053 (2004)
- [4] K. Hamasaki, K. Hiyama, and R. Okumura, "The 22.2 multichannel sound system and its application", AES 118th Convention, Barcelona, Spain, Convention paper 6406 (2005 May 28-31)
- [5] C.E. Osgood, G.J. Suci, and P.H. Tannenbaum, "The Measurement of Meaning", University of Illinois Press, Urbana (1957)
- [6] K. Hamasaki, T. Nishiguchi, K. Hiyama, and R. Okumura, "Effectiveness of height information for reproducing presence and reality in multichannel audio system", presented at AES 120th Convention, Paris, France, Convention paper 6679, 2006 May 20-23.
- [7] K. Hamasaki, T. Nishiguchi, K. Hiyama, and R. Okumura, "Influence of picture on impression of three-dimensional multichannel sound", presented at AES 28th International Conference, Piteå, Sweden, 2006 June 30-July 2.