# 22.2 Multichannel Audio Format Standardization Activity

The audio format for Super Hi-Vision, which is under research and development for high-presence broadcasting in the future, is 22.2 multichannel audio. The 22.2 multichannel format is a 3D sound play-back format that is able to reproduce sound impressions that are much closer to natural sound than the 5.1 multichannel audio format used in current digital broadcasts. Here, we present an overview of the 22.2 multichannel audio format and describe current activities toward international standardization.

## 1. Introduction

At NHK STRL, we are developing Super Hi-Vision[1] and conducting research toward the realization of high-presence broadcasting. Super Hi-Vision is an ultra-high-resolution video system with 4000 scaning lines and a viewing angle[*1] of 100°. The standard viewing distance for a video system is set such that the pixels are just smaller than a person with 20/20 vision can resolve. For a fixed screen size and aspect ratio, the viewing angle is determined by the standard viewing distance, and increasing the number of pixels increases the viewing angle. The sense of presence reaches saturation at a viewing angle of 100°[2], and the Super Hi-Vision system produces a viewing angle of 100° at the standard viewing distance. The goal of the 22.2 multichannel audio format[3] under research and development for use with Super Hi-Vision is to produce, record, transmit and play back high-presence audio-visual content together with ultra-high-resolution video. In order to make possible program production and broadcasting of 22.2 multichannel audio in many countries and enable international exchange of content, it is vital that compatibility be maintained for audio play-back formats, audio devices, interfaces, and production and playback environments. International standards are what guarantee this compatibility internationally, and standardization is the process for deciding on these standards. International standardization of the 22.2 multichannel audio format is currently in progress.

What follows in this article is an overview of the 22.2 multichannel audio format, an introduction to the standardization organizations related to acoustic and audio technologies, and a description of the standardization activities that we are participating in.

## 2. The 22.2 Multichannel Audio Format
### 2.1 Requirements for a High-Presence Audio Format

The requirements for a high-presence audio format for use with Super Hi-Vision are as follows:

**(1) Must be able to localize an audio image anywhere on the screen.**

In viewing environments with a wide field of view and large screen, it is important to be able to match the positions of the video and audio images[*2]. Thus, a format that is able to localize the sound image anywhere on the screen is required.

**(2) Must be able to reproduce sound coming from all directions surrounding the viewing position.**

Sound arrives from all directions in the actual acoustic field. Accordingly, to reproduce a high sense of presence, the audio format must be able to reproduce sound coming from all directions.

**(3) Must be able to reproduce a natural, high-quality 3D acoustic space.**

The format must be able to reproduce a 3D acoustic space with ample reverberation, as in a concert hall, as naturally and with as high quality as possible.

**(4) Must have an enlarged optimal listening zone.**

With multichannel audio reproduction, the sound reproduced near the center of the listening area are very faithfully to what content producer desired, where distances to each speaker are nearly equal. However, for environments with more than one viewer, the format must be able to reproduce audio with as high quality as possible over a wide area.

**(5) Must be compatible with existing multichannel audio formats.**

Newly developed multichannel audio formats must be compatible with multichannel audio formats that are already implemented and other new formats that have already been proposed. There are two aspects to this compatibility. The first is that content produced for other multichannel formats must be reproducible with the newly developed format. The other is that it must be possible to play back content produced for the newly developed format using other multichannel audio

---

[*1] The horizontal angle within the field of view occupied by the width of the screen.

[*2] The image of the sound source according to the sense of hearing.

formats by applying some sort of processing.

**(6) Must support live recording and live broadcasting.**

The format is intended for broadcast, so the recording and playback formats must be capable of live recording and live broadcasting. An important requirement of a broadcast audio format is that broadcast programming can be produced in the format, and that the creativity of engineers and directors involved in program production can be utilized to their full potential. Accordingly, it is essential that high-quality sound production and reproduction be equivalent or better than conventional broadcast sound formats, in addition to reproducing an acoustic space.

The 22.2 multichannel audio format satisfies these requirements. Figure 1 shows the positions of each channel with circles and accompanying channel labels. Channels correspond to the direction from which the sound arrives, and speakers are installed at the positions shown in the figure to reproduce the sound signal for the corresponding channel. The channels for the 22.2 multichannel audio format are arranged in three layers: upper, middle and lower. The upper layer consists of nine channels arranged at the height of the top of the screen or the ceiling. The middle layer consists of ten channels arranged at the height of the center of the screen or the viewer's ear level. The lower layer consists of three channels arranged at the height of the bottom of the screen or the floor. The lower level also includes two low-frequency effects (LFE) channels.

The 22.2 multichannel audio format normally requires 22 wide-band speakers and two LFE speakers, but it is also possible to use fewer speakers to reproduce sound arriving from the directions of channels shown in Figure 1. However, in such cases, the optimal listener area is smaller.

### 2.2 Features of the 22.2 Multichannel Audio Format

The channels of the middle layer reproduce the primary sound sources. Existing multichannel audio formats[*3] such as 5.1, 6.1 and 7.1 all have their primary channels in the middle layer. That means the 22.2 multichannel audio format can easily reproduce audio content produced for the existing multichannel audio formats.

There are three channels on each of the left and right sides of the listener (front, side, back), allowing motion of the sound image in the forward and backward directions. The side channels also enable natural, high-quality
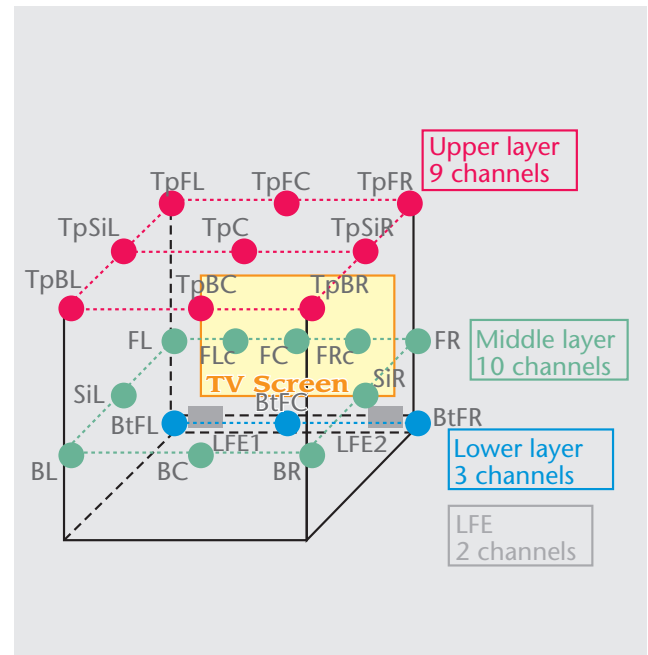
---

[*3] The 5.1 multichannel audio format has three channels in front of the viewer (left, center, right), and two behind (left, right). The 6.1 multichannel audio format has three channels in front of the viewer (left, center, right), and three behind (left, center, right). The 7.1 multichannel audio format has five channels in front of the viewer (left, between left and center, center, between center and right, right), and two behind (left, right).



**Figure 1: The 22.2 Multichannel Audio Format**

reproduction of an acoustic space. Reflections from the sides are very important for producing the sense of spatial broadening in musical performances in concert halls, or a sense of being enveloped in sound[4]. The side speakers are able to reproduce such reflected sounds, and thus, they help to give the listener the impression of the natural acoustic space of a concert hall.

The upper-layer channels can be used to localize the sound image anywhere above the viewer, or can be used in conjunction with the middle or lower level channels to produce motion of the sound image in the vertical direction. This means that the format is able to reproduce sound images with a vertical sense, which is not possible with the earlier multichannel audio formats. The upper-layer channels are also important for reproducing a good sense of enveloping sound over a wide listening area. In particular, a good acoustic space can be reproduced over a wide listening area by appropriately reproducing early reflections and late reverberation in the upper-layer channels[3].

Sound can be localized at any position on the screen using the three channels at the top of the screen (TpFL, TpFC, TpFR), the three channels at the bottom of the screen (BtFL, BtFC, BtFR), and the five forward channels at the middle level (FL, FLc, FC, FRc, FR). The two LFE channels positioned at the bottom of the screen also improve the spatial impression of the sense of breadth and being enveloped in sound[5].

## 3. Standardization Organizations for Acoustic and Audio Technologies

The representative organizations engaged in standardization of acoustic and audio technologies are listed in Figure 2. Research and development on the
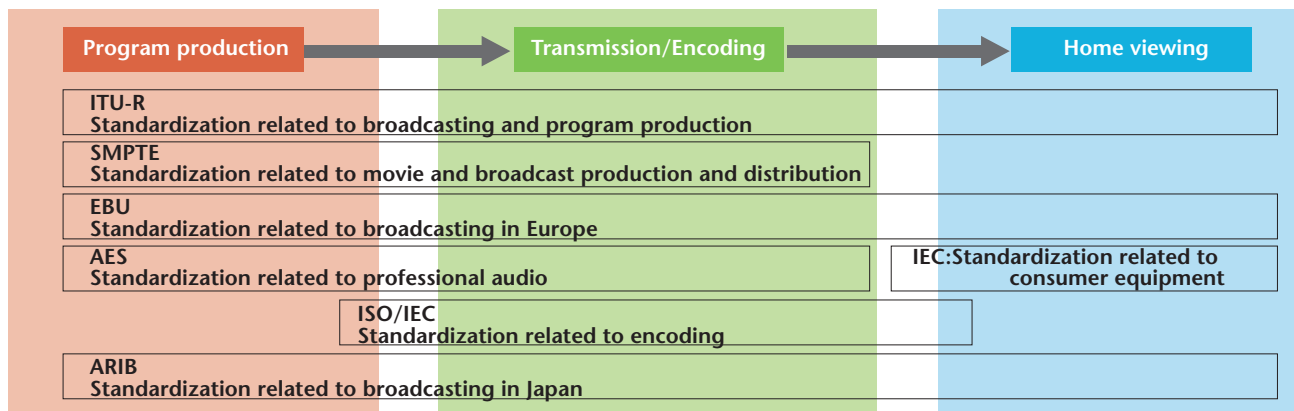
**Figure 2: Major Standardization Organizations for Acoustic and Audio Technologies**

22.2 multichannel audio format is being done with the objective of production, recording, transmission and play-back of high-presence audio-visual content, but Figure 2 shows the relationships among the various standardization organizations in terms of three processes related to actual broadcasting, namely program production, transmission and encoding, and home viewing. Below, we give simple explanations of the standardization organizations shown in Figure 2. Note that the ITU-R is an inter-governmental organization, while the others are private organizations.

**- ITU-R (International Telecommunication Union - Radiocommunication Sector)**

Conducts standardization of acoustic and audio technology related to broadcasting with radio waves and other media. It is an important international organization for broadcasters.

**- SMPTE (Society of Motion Picture and Television Engineers)**

Conducts standardization of acoustic and audio technology related to movies and television.

**- EBU (European Broadcasting Union)**

A union consisting of European broadcasters that engages in standardization related to broadcast audio technology.

**- AES (Audio Engineering Society)**

The only international organization related to professional audio technology. Standardization committees conduct a wide range of activities related to acoustic and audio capture, recording, transmission and reproduction. In particular, the digital audio interface standards from AES (E.g.: AES3-20096)) are important standards that are referenced by nearly all other digital-audio-related standards from other organizations.

**- IEC (International Electrotechnical Commission)**

The TC100 (Audio, Video, Multimedia systems and devices), technical committee of the IEC, conducts standardization related to digital audio and focused on consumer devices.

**- ISO (International Organization for Standardization)/IEC**

The ISO/IEC JTC1/SC29/WG11[*4], which is a joint organization between the ISO and IEC, conducts standardization of digital audio encoding under the popular name, Moving Picture Experts Group (MPEG).

**- ARIB (Association of Radio Industries and Businesses)**

Conducts investigations, research, development, and standardization related to efficient use of radio frequencies in the communications and broadcasting fields, as well as standardization of audio technology related to broadcasting in Japan.

## 4. Standardization Activities Related to the 22.2 Multichannel Audio Format

As shown in Figure 3, standardization work related to the 22.2 multichannel audio format is being done by multiple standardization organizations on various technical elements including transmission, encoding, and home viewing technologies. The EBU is not shown in Figure 3, NHK STRL is participating in an EBU audio technology standardization project and contributing to standardization of the Broadcast Wave Format (BWF)[*5], which may be related to the 22.2 format. Below, we introduce the standardization activities regarding the 22.2 multichannel audio format that are underway at each of the standardization organizations.

### 4.1 ITU-R

ITU-R is responsible for international standardization of broadcasting using the 22.2 multichannel audio format. Multichannel audio formats for digital broadcasting are regulated in recommendation BS.775-2[7]. Among these, the 5.1 multichannel audio format is in broad use. To begin research on audio formats exceeding the 5.1 multichannel audio format, a revised research question proposal was created, and Research question 135/6[8] has been approved. Research on multichannel audio formats with 3D speaker arrangements is being done on the basis

---

[*4] Joint Technical Committee 1/Subcommittee 29/Working Group 11

[*5] A standard audio and data exchange format that is an extension to the Microsoft WAV audio format and was established by the EBU. It is used at broadcast stations.
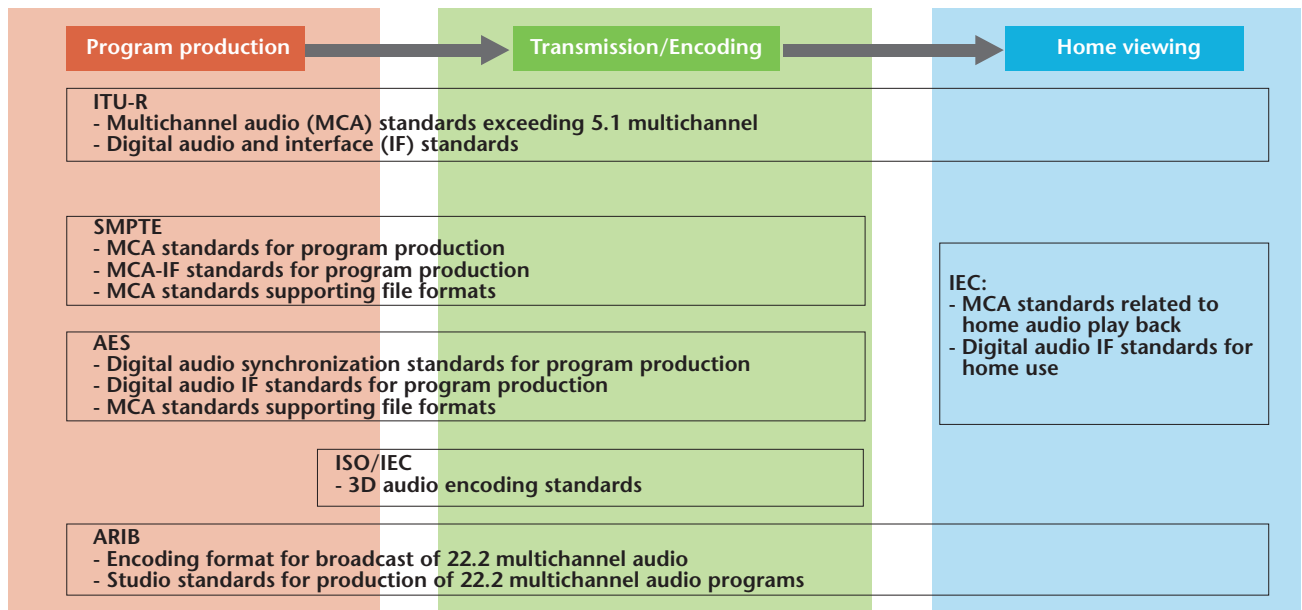
**Figure 3: Standardization of the 22.2 multichannel audio format**

of research question 135/6 on arrangements beyond those of the horizontal plane.

The BS.2159[9] report was approved in 2009. It describes the latest activity in multichannel audio formats, including the 22.2 multichannel audio format.

A new draft recommendation, "3D Multichannel Studio Sound Format Standard," was also tabled at the November 2009 meeting. This draft recommendation proposes studio standards for the 22.2 multichannel audio format as the highest-rank in a hierarchy of 3D audio formats and describes the three-layer speaker placement as well as the names for each of the audio channels and channel mappings. Currently, this draft recommendation is being studied by a rapporteur group of ITU-R WP6C.

### 4.2 SMPTE

SMPTE is conducting standardization related to program production using the 22.2 multichannel audio format. Standardization of Super Hi-Vision video and audio at SMPTE is being done in the form of Ultra-High Definition TV (UHDTV) video and audio. SMPTE is standardizing UHDTV1, with 3,840 x 2,160 pixels, and UHDTV2, with 7,680 x 4,320 pixels[10]. UHDTV2 has the same video resolution as Super Hi-Vision, and the video, audio, and interface formats are respectively standardized in ST2036-1-2009[10], ST2036-2-2008[11], and ST2036-3-2010[12].

The audio format described in ST2036-2-2008 includes digital audio sampling frequencies of 48 kHz or 96 kHz, bit lengths of 16, 20, or 24 bits, 24 channels, and no pre-emphasis[*6]. The 22.2 multichannel audio format is specified with the channel mappings and channel labels and names shown in Table 1. Note that the speaker arrangement in Figure 1 is indicated as a reference example.

ST299-1-2009[13] standardizes a digital audio signal

**Table 1: Channel mapping, labels, and names for the 22.2 multichannel audio, as standardized in SMPTE ST2036-2-2008**

| AES Pair No./Ch No. | Channel No. | Label | Name |
|---|---|---|---|
| 1/1 | 1 | FL | Front left |
| 1/2 | 2 | FR | Front right |
| 2/1 | 3 | FC | Front center |
| 2/2 | 4 | LFE1 | LFE-1 |
| 3/1 | 5 | BL | Back left |
| 3/2 | 6 | BR | Back right |
| 4/1 | 7 | FLc | Front left center |
| 4/2 | 8 | FRc | Front right center |
| 5/1 | 9 | BC | Back center |
| 5/2 | 10 | LFE2 | LFE-2 |
| 6/1 | 11 | SiL | Side left |
| 6/2 | 12 | SiR | Side right |
| 7/1 | 13 | TpFL | Top front left |
| 7/2 | 14 | TpFR | Top front right |
| 8/1 | 15 | TpFC | Top front center |
| 8/2 | 16 | TpC | Top center |
| 9/1 | 17 | TpBL | Top back left |
| 9/2 | 18 | TpBR | Top back right |
| 10/1 | 19 | TpSiL | Top side left |
| 10/2 | 20 | TpSiR | Top side right |
| 11/1 | 21 | TpBC | Top back center |
| 11/2 | 22 | BtFC | Bottom front center |
| 12/1 | 23 | BtFL | Bottom front left |
| 12/2 | 24 | BtFR | Bottom front right |

with a sampling rate of 96 kHz for transmission over a High Definition Serial Digital Interface (HD-SDI)*7. This standard revises ST-299m-2004, which had specified the transmission of a 48 kHz digital audio signal over HD-SDI, into one that supports a sampling rate of 96 kHz.

SMPTE is currently studying standards for including multichannel audio format channel labels and acoustic space definitions (speaker placements) as metadata in the Material Exchange Format (MXF), which is a standard file format for exchanging video and audio content.

### 4.3 AES

AES is conducting standardization on a digital audio interface that supports Super Hi-Vision. AES11, which regulates synchronization of video frame rates and digital audio, was recently revised as AES11-2009[14] to cover synchronization of digital audio sampling frequencies with all video frame rates specified in SMPTE ST2036-1-2009. AES also issued AES10-2008[15] in 2008, which is a revision of the Multichannel Audio Digital Interface (MADI) standard of 1991 for transmission of 24-channel digital audio signals over a single 75Ω coaxial cable.

AES is currently studying issues including metadata and new digital audio interfaces related to multichannel audio. AES is also collaborating with IEC, as described below.

### 4.4 IEC

The TC100 group at IEC is studying consumer-level digital audio interfaces for home play-back of the 22.2 multichannel audio format. In order to transmit the 22.2 multichannel audio format signal over a digital audio interface, channel mappings, labels, and names for its 24 channels must be standardized, as they are in SMPTE ST2036-2-2008. Accordingly, IEC is working on multichannel mapping standards supporting the 22.2 multichannel audio format and other multichannel audio formats. IEC plans to issue the new IEC62574 standard within 2011. IEC is also studying IEC60958-3, which is the AES3-2009 consumer specification, from the perspective of 22.2 multichannel audio format broadcasts and home play back.

### 4.5 ISO/IEC

ISO/IEC is conducting standardization on the 3D audio encoding used in the 22.2 multichannel audio format. The audio encoding format of current digital broadcasting, MPEG2-AAC, cannot be used to describe channel configurations for 3D audio formats. Thus, a revision to MPEG2-AAC was proposed, and the revised document was issued in 2009. As a result, it is now possible to encode 3D audio signals, including those of the 22.2 multichannel audio format, using MPEG2-AAC.

### 4.6 ARIB

ARIB is conducting standardization activities for broadcasting using the 22.2 multichannel audio format within Japan. In 2003, the 26th revision of the Ministry of Internal Affairs and Communications', "Standard digital broadcast format for standard television broadcasts," issued regulations stating that the maximum number of input audio channels shall be 22 channels and two low-frequency effect channels for Advanced BS digital broadcasting and advanced wide-band CS digital broadcasting. To support these changes, ARIB revised its ARIB STD-B3216, adding standards for an MPEG-2-ACC audio encoding format supporting audio modes up to 22.2 multichannel audio.

Also, in 2008, the "Ultra-high-resolution television studio equipment development group" was established to conduct standardization of studio equipment needed by broadcasters to realize broadcasting of television signals with resolutions of 1,080 lines or more. Under this group, the "Audio systems study work group" was established to study audio formats needed for ultra-high definition television studio equipment. The work group is studying draft studio specifications for 3D multichannel audio formats with three-layer speaker arrangements (upper, middle and lower layers), based on ARIB STD-B32. In the draft specifications, the 22.2 multichannel sound format is the audio format with the most channels. The main points of the specification are given below.

(1) Speaker positions in the three-layer structure (upper, middle, lower) and the permitted ranges for the installation angles. Figure 4 shows an example of speaker positions and installation angles for the 22.2 multichannel audio format.
(2) A hierarchical set of formats from 22.2 12.2, 10.2, 7.1 down to 6.1.
(3) The channel labels and mappings for the 22.2 multichannel audio format.
(4) The quality of the digital audio (sampling frequency, number of quantization bits)
(5) Studio-standard play-back levels.
(6) Formulas for conversion of the above hierarchical formats into 5.1 multichannel audio and two-channel audio.

### 5. Conclusion

This article gave overviews of the 22.2 multichannel audio format and the standardization organizations related to acoustic and audio technology, and it introduced standardization activities of these organizations for the 22.2 multichannel audio format.

Mr. Ben Burtt is a legendary specialist of sound design who was involved in the sound design for the

---

*6 An operation applied to the analog signal before it is converted into a digital signal. It effectively increases the signal-to-noise (SN) ratio by increasing the amplitude of specific ranges of high frequencies. This requires the amplitudes to be reduced by the amount of increase (de-emphasis) when the digital signal is converted back into analog.

*7 An interface for transmitting an uncompressed digital Hi-Vision video and 16-channel digital audio signal over a single cable. Regulated in SMPTE ST292M.
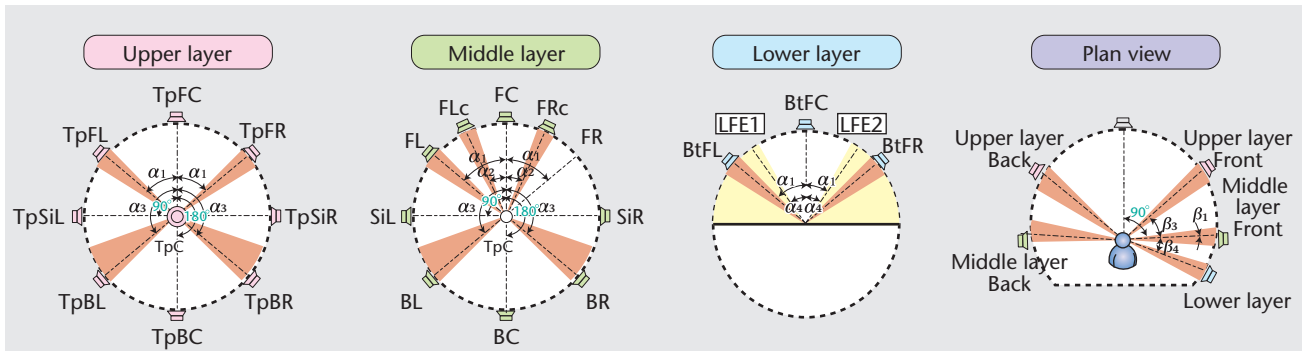
**Figure 4: Speaker positioning and angles for the three-layer (upper, middle, lower) structure for 22.2 multichannel audio ($\alpha_1$: 45°-60°, $\alpha_2$: $\alpha_1$/2, $\alpha_3$: 110°-135°, $\alpha_4$: 30°-90°, $\beta_1$: 0°-5°, $\beta_2$: 0°-15°, $\beta_3$: 30°-45°, $\beta_4$: 15°-25°)**

Star Wars and Indiana Jones series' of movies. He has been responsible for sound effects on 12 Academy Award nominated movies and has himself received four Oscars. He was invited to speak at the AES 129th Convention[17] held in San Francisco in November 2010 and had the following to say regarding the role of sound in the creation of video content:
- Sound gives CREDIBILITY to the image.
- Sound BINDS separate images together into a WHOLE.
- Sound creates a PLACE.
- Sound EXTENDS the place beyond frame.
- Sound enhances EMOTION.
- Sound enhances CHARACTER.
- Sound controls PACE.
- Sound makes TRANSITION.
- Sound COUNTERPOINTS images to give a separate message.
- Silence EMPOWERS the impact of sound.

A goal of research and development of the 22.2 multichannel audio format is to better realize these aspects in the production of high-presence audio-visual content. In the future, we will continue to pursue our goal of international standardization of the 22.2 multichannel audio format and enable those involved in broadcasting and content production around the world to use the effects of 3D sound to their maximum potential. We will also continue with our research and development to allow viewers to enjoy exciting content on the audio system of their choice and suited to their own viewing environment. **(Kimio Hamasaki)**

### References
1) M. Kanazawa, K. Mitani, K. Hamasaki, M. Sugawara, F. Okano, and K. Doi, M.Seino: "Ultrahigh-definition Video System with 4000 Scanning Lines," Proc. IBC (2003).
2) T. Hatada, H. Sakata and H. Kusaka: "Psychophysical Analysis of the Sensation of Reality Induced by a Visual Wide-field Display," SMPTE J., Vol.89, pp. 560-569 (1980).
3) K. Hamasaki, T. Nishiguchi, R. Okumura, Y. Nakayama, and A. Ando: "A 22.2 Multichannel Sound System for Ultra-High-Definition TV (UHDTV)," SMPTE Motion Imaging J., Vol. 117, No. 3, pp. 40-49 (2008).
4) D. Griesinger: "Spaciousness and Envelopment in Musical Acoustics," AES 101st Convention, 4401 (1996).
5) W. Martens, J. Braasch, W. Woszczyk: "Identification and Discrimination of Listener Envelopment Precepts Associated with Multiple Low-frequency Signals in Multichannel Sound Reproduction," AES 117th Convention, 6229 (2004).
6) AES 3-2009, "AES Standard for Digital Audio Engineering - Serial Transmission Format for Two-channel Linearly Represented Digital Audio Data," (2009).
7) Rec. ITU-R BS.775-2, "Multichannel Stereophonic Sound System with and without Accompanying Picture," (2006).
8) Question ITU-R 135/6, "System Parameters for Digital Sound Systems," (2010).
9) Report ITU-R BS.2159, "Multichannel Sound Technology in Home and Broadcasting Applications," (2009).
10) SMPTE ST2036-1-2009, "Ultra High Definition Television - Image Parameter Values for Program Production," (2009).
11) SMPTE ST2036-2-2008, "Ultra High Definition Television - Audio Characteristics and Audio Channel Mapping for Program Production," (2008).
12) SMPTE ST2036-3-2010, "Ultra High Definition Television - Mapping into Single-link or Multi-link 10 Gb/s Serial Signal/Data Interface," (2010).
13) SMPTE ST299-1-2009, "24-Bit Digital Audio Format for SMPTE 292 Bit-Serial Interface," (2009).
14) AES 11-2009, "AES Recommended Practice for Digital Audio Engineering -Synchronization of Digital Audio Equipment in Studio Operations," (2009).
15) AES10-2008, "AES Recommended Practice for Digital Audio Engineering ? Serial Multichannel Audio Digital Interface (MADI)," (2008).
16) ARIB STD-B32, "Video and audio encodings and multiplexing format standards for digital broadcasting," (2009) (Japanese).
17) Ben Burtt: "The Sound Behind the Image," The Richard C. Heyser Distinguished Lecturer for the 129th AES Convention (2010).