International Telecommunication Union

ITU-R

Radiocommunication Sector of ITU

Report ITU-R BS.2159-9

(03/2022)

Multichannel sound technology in home and broadcasting applications

BS Series

Broadcasting service (sound)

International Telecommunication Union

Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from http://www.itu.int/ITU-R/go/patents/en where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Reports

(Also available online at http://www.itu.int/publ/R-REP/en)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	**Broadcasting service (sound)**
BT	Broadcasting service (television)
F	Fixed service
M	Mobile, radiodetermination, amateur and related satellite services
P	Radiowave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management

***Note**: This ITU-R Report was approved in English by the Study Group under the procedure detailed in Resolution ITU-R 1.*

Electronic Publication
Geneva, 2022

REPORT ITU-R BS.2159-9

Multichannel sound technology in home and broadcasting applications

(2009-2010-05/2011-10/2011-05/2012-11/2012-2013-2015-2019-2022)

TABLE OF CONTENTS

1 Introduction

ITU-R has developed Recommendation ITU-R BS.775 for the 3/2 multichannel stereophonic sound system (5.1 sound system) with and without accompanying picture. Multichannel stereo as well as 2-channel stereo audio services are widely used as part of digital broadcasting services. Recommendation ITU-R BS.775 specifies a hierarchy of compatible multichannel sound systems to enhance the directional stability of the frontal sound image and the sensation of spatial reality (ambience), and each loudspeaker is set at the same height as a listener's ears.

Some television applications with higher resolution imagery including ultra-high definition television and high-dynamic range television specified in Recommendations ITU-R BT.2020 and BT.2100, large screen digital imaginary (LSDI) application[1] and advanced immersive audio visual system including 360° images, both providing wider viewing angle, may need multichannel stereophonic sound systems that can reproduce the sound sources, which are localized at a higher position over the listener and a lower position below the screen, and vertical movements of the sound sources. Several

[1] LSDI is defined as a service whereby programmes are distributed in the form of digital signals, in real-time or non-real-time, for collective viewing in theatres or other group venues equipped with appropriate electronic projectors, to provide excellent presentation in terms of picture and sound quality, size of the presentation screen and presentation environment.

multichannel stereophonic sound systems are currently applied or studied for higher resolution imagery, and some of them have loudspeakers arranged above and below the viewer. There would be value in continued studies in this area for future broadcasting applications in order to evolve beyond the 5.1 sound system.

The advanced sound system is a system with a reproduction configuration beyond the 5.1 sound system specified in Recommendation ITU-R BS.775 or a system that can support channel-based, object-based, scene-based or their combination with audio-related metadata. A number of ITU-R Recommendations relevant to the advanced sound system have been developed:

– Recommendation ITU-R BS.1909 – The performance requirements for an advanced multichannel stereophonic sound system for use with or without accompanying picture
– Recommendation ITU-R BS.2051 – Advanced sound system for programme production
– Recommendation ITU-R BS.2127 – Audio Definition Model renderer for advanced sound systems.

ITU-R has also developed the following Recommendations of audio-related metadata:

– Recommendation ITU-R BS.2076 – Audio Definition Model
– Recommendation ITU-R BS.2094 – Common definitions for the Audio Definition Model
– Recommendation ITU-R BS.2125 – A serial presentation of the Audio Definition Model
– Recommendation ITU-R BS.2088 – Long-form file format for the international exchange of audio programme materials with metadata.

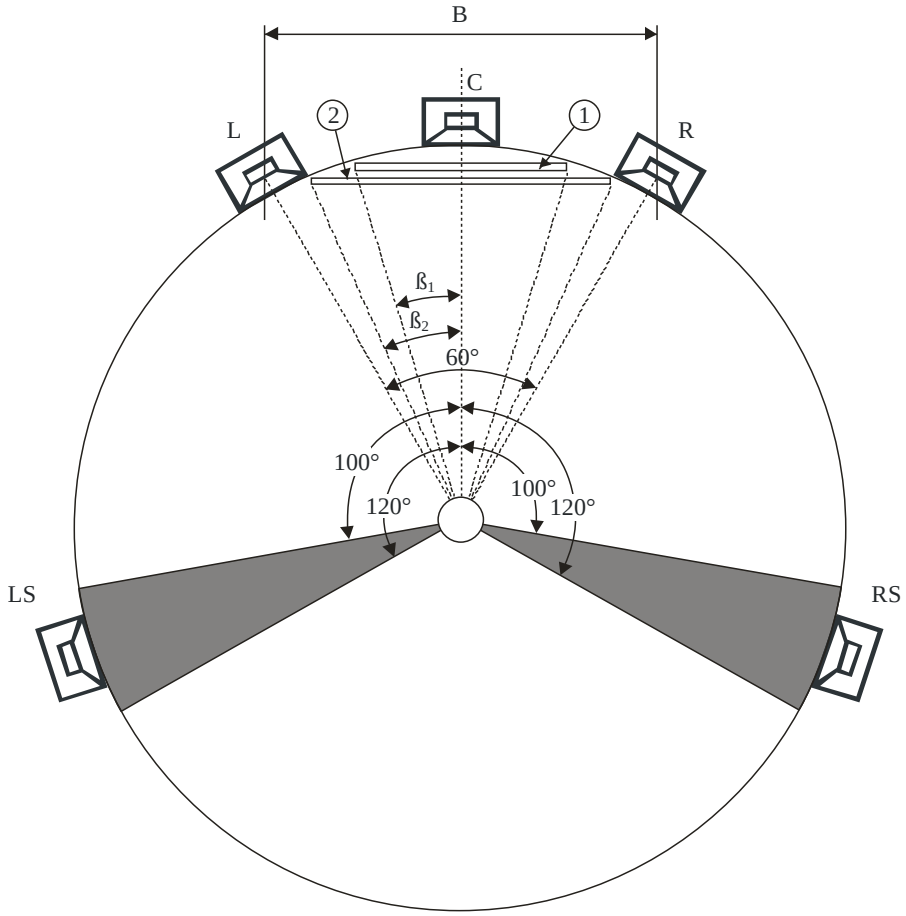
This Report contains information on the subject of multichannel sound technologies in home and broadcasting applications beyond the 5.1 sound system specified in Recommendation ITU-R BS.775.

2 5.1 multichannel sound system

The 5.1 channel sound system has been specified in Recommendation ITU-R BS.775. The system is widely used as a part of digital broadcasting services. It enhances the directional stability of the frontal sound image and the sensation of spatial reality (ambience). The reference loudspeaker arrangement is shown in Fig. 1, in which each loudspeaker is set at the same height as a listener's ears.

FIGURE 1

Reference loudspeaker arrangement with loudspeakers L/C/R an LS/RS



Screen 1 HDTV - Reference distance = 3 $H(2ß_1 = 33°)$

Screen 2 = 2 $H(2ß_2 = 48°)$

H: height of screen

B: loudspeaker base width

Loudspeaker	Horizontal angle from centre (degrees)	Height (meters)	Inclination (degrees)
C	0	1.2	0
L, R	30	1.2	0
LS, RS	100 ... 120	≥ 1.2	0 ... 15 down

Report BS.2159-01

3 Basic requirements of multichannel sound systems beyond the 5.1 sound system

The following requirements are related to the multichannel sound system beyond the current 5.1 channel sound system specified in Recommendation ITU-R BS.775.

1) The directional stability of the frontal sound image should be maintained over the entire higher resolution imagery area. Coincidence of position between sound image and video image also should be maintained over the wide imagery area.
2) The sound image should be reproduced in all directions around the listener, including elevation.
3) The sensation of three-dimensional spatial impression that augments a sense of reality should be significantly enhanced. This may be achieved by the use of side and/or back, top and/or bottom loudspeakers.
4) Exceptional sound quality should be maintained over wider listening area than that provided by current 5.1 channel sound system.
5) Compatibility with the current 5.1 channel sound system specified in Recommendation ITU-R BS.775 should be ensured to an acceptable degree.
6) Live recording, mixing and transmission should be possible.

The reasoning for the basic requirements for advanced multichannel sound systems is provided below:

3.1 Basic requirements of the sound image

– The directional stability of the frontal sound image should be maintained over the entire higher resolution imagery area. Coincidence of position between sound image and video image also should be maintained over the wide imagery area.
– The sound image should be reproduced in all directions around the listener, including in the elevation.

Reason:

The following requirements are defined in Recommendation ITU-R BS.775 for the 5.1 channel sound system:

– *The directional stability of the frontal sound image shall be maintained within reasonable limits over a listening area larger than that provided by a conventional two-channel stereophony.*

The following requirement is also defined:

– *It is not required that the side/rear loudspeakers should be capable of the prescribed image locations outside the range of the front loudspeakers.*

For advanced multichannel sound systems beyond the 5.1 channel sound system, the reproduction of the sound images should be improved from the following two aspects:

− The directional stability of the sound image come from all horizontal directions, i.e. the front/back and left/right directions, should be maintained within reasonable limits over the listening area.
− The sound image included in the elevation directions should also be reproduced.

Therefore, the aforementioned **basic requirement 2)** is defined.

In addition, considering advanced multichannel sound systems applied for television applications, which have high-resolution imagery with a horizontally and vertically wide field of view, the coincidence of the position between sound images and video images is needed over the entire imagery area. Therefore, the aforementioned **basic requirement 1)** is defined.

3.2 Basic requirement of sensation of a spatial impression

– The sensation of three-dimensional spatial impression that augments a sense of reality should be significantly enhanced. This may be achieved by the use of side and/or back, top and/or bottom loudspeakers.

Reason:

The following requirement is defined in Recommendation ITU-R BS.775 for the 5.1 channel sound system:

– *The sensation of spatial reality (ambience) shall be significantly enhanced over that provided by a conventional two-channel stereophony. This shall be achieved by the use of side and/or rear loudspeakers.*

Because each loudspeaker of the 5.1 channel sound system is set at the same height as the listener's ears, the sensation of spatial reality is fundamentally limited to the horizontal plane.

For advanced multichannel sound systems beyond the 5.1 channel sound system, the sensation of spatial impression should be enhanced around the listener, including in the upward/downward elevation sensation, reverberation and ambience. Therefore, the aforementioned **basic requirement 3)** is defined.

3.3 Basic requirement of listening area

– Exceptional sound quality should be maintained over wider listening area than that provided by current 5.1 channel sound system.

Reason:

The following requirement is defined in Recommendation ITU-R BS.775 for the 5.1 channel sound system:

– *The directional stability of the frontal sound image shall be maintained within reasonable limits over a listening area larger than that provided by a conventional two-channel stereophony.*

As frontal two (left and right) loudspeakers are placed for the conventional 2-channel sound system, the listening area of the 5.1 channel sound system should be considered only for the frontal sound image by comparing it to that of a conventional 2-channel stereophony.

To extend the basic requirement of the 5.1 channel sound system, the listening area of advanced multichannel sound systems should be enlarged by comparing them to the 5.1 channel sound system. Therefore, the aforementioned **basic requirement 4)** is defined.

3.4 Basic requirement of compatibility with existing sound systems

– Compatibility with the current 5.1 channel sound system specified in Recommendation ITU-R BS.775 should be ensured to an acceptable degree.

Reason:

The following requirement is defined in Recommendation ITU-R BS.775 for the 5.1 channel sound system:

– *Downward compatibility with sound systems providing lower number of channels (down to stereophonic and monophonic sound systems) shall be maintained.*

The aforementioned compatibility means that, for example, the down-mixed stereophonic or monophonic sound quality from 5.1 channel sound signals should be maintained to an acceptable degree. To extend the basic requirement of the 5.1 channel sound system, the down-mixed 5.1 channel or 2-channel sound quality from advanced multichannel sound signals should be maintained to an acceptable degree for advanced multichannel sound systems.

Additionally, the compatibility should be considered from the view of programme production facilities and exploiting the expertise of sound mixing engineer. Even in future broadcasting services, every programme will not likely be produced by the advanced multichannel sound format. In other words conventional sound formats, such as mono, 2-channel stereo, or 5.1 channel sound format may be operated even in the future broadcasting depending on the programme genre or other service requirements. Thus, broadcasters would prefer to be able to produce various sound programme formats even in a single production studio. As a result, channel compatibility with the 5.1 channel sound system and conventional 2-channel sound system should be considered to an acceptable degree. It also takes advantage of sound mixing engineer's know-how, cultivated by the 5.1 channel sound production. Therefore, the aforementioned **basic requirement 5)** is defined.

3.5 Basic requirement of live broadcasting

– Live recording, mixing and transmission should be possible.

Reason:

The following requirement is defined in Recommendation ITU-R BS.775 for the 5.1 channel sound system:

– *Real-time mixing for live broadcast shall be practicable.*

The live broadcast is the most essential factor for broadcasting services. Therefore, the aforementioned **basic requirement 6)** is defined.

4 Multichannel sound systems beyond the 5.1 sound system under development for broadcasting applications

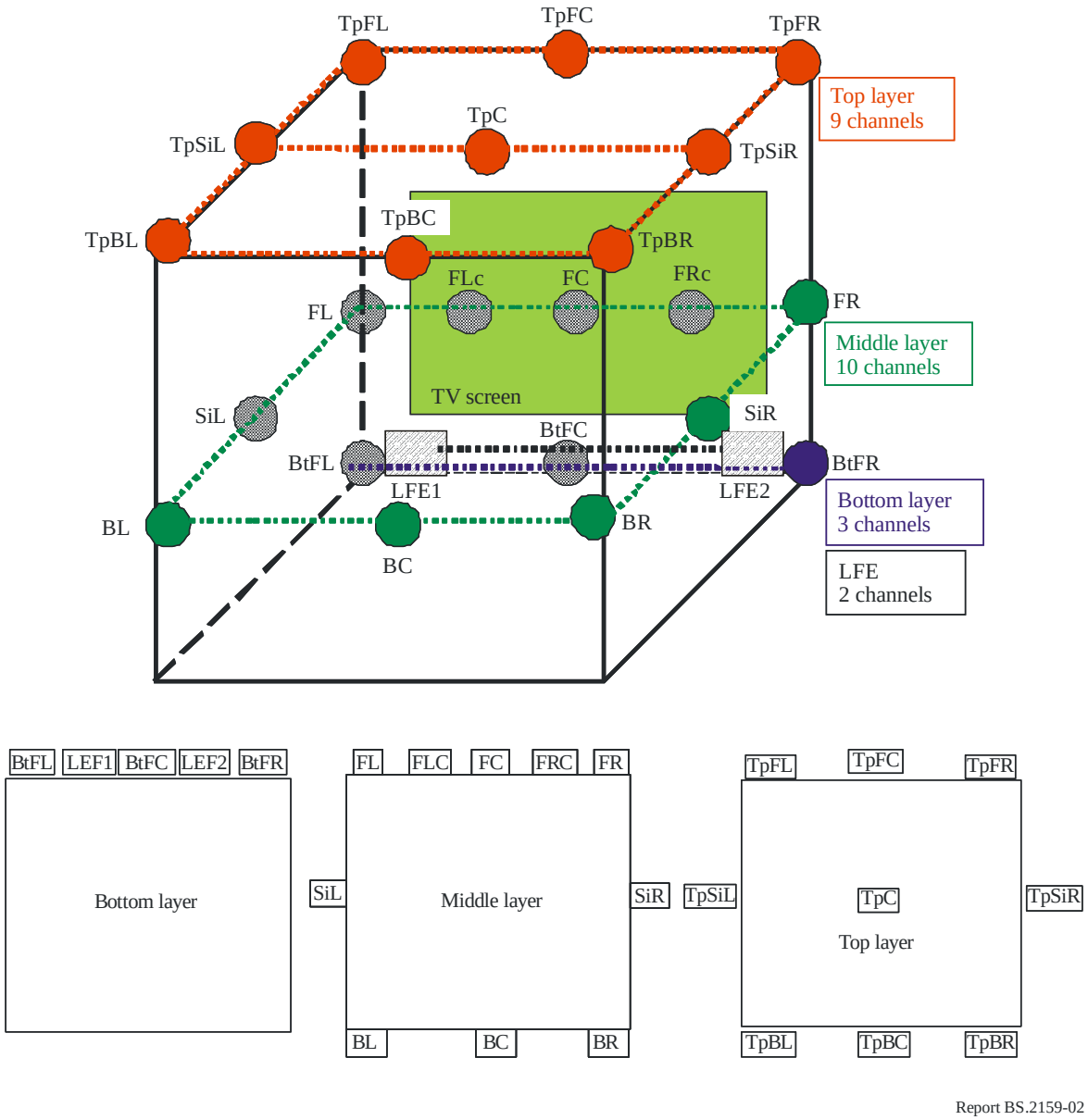
Several multichannel sound systems have been studied to improve the spatial impression of sound. The following systems seem to have the capability for practical use.

4.1 22.2 multichannel sound system (system H specified in Recommendation ITU-R BS.2051)

The 22.2 multichannel sound system was developed by NHK (Japan Broadcasting Corporation). It has nine channels at the top layer, ten channels at the middle layer, three channels at the bottom layer and two low-frequency effects (LFE) channels. This system is suited to wide screens, such as 100 inch FPD displays, because it can two-dimensionally localise a sound image over the entire screen by using three bottom channels, five middle channels and three top channels around the screen.

FIGURE 2

The 22.2 multichannel sound system



Report BS.2159-02

This system has common channels at each three layers with other multichannel systems so that its audio can be easily down-mixed to other multichannel sound systems and has compatibility with every multichannel sound system.

The audio characteristics and audio channel mapping for UHDTV programme production for the 22.2 multichannel sound format have been specified as system H in Recommendation ITU-R BS.2051 and standardised by SMPTE (SMPTE 2036-2), as mentioned in § 8.1.1.

Several sound reproduction systems based on the 22.2 multichannel sound have been developed and exhibited at more than a dozen international expositions and exhibitions, such as World Exposition at Aichi (Japan), NAB Show in Las Vegas (United States of America), IBC in Amsterdam (Netherlands), CEATEC Exhibition in Tokyo (Japan), Broadcast Asia in Singapore, and Grand Exposition for Yokohama's 150th Year. The advanced satellite digital broadcasting of UHDTV with 22.2 multichannel sound started in Japan in December 2018.

There are also permanent installations which can reproduce 22.2 multichannel sound. They are:

– One theatrical demonstration room with hundreds of seats in NHK Fureai Hall at Tokyo.

– One theatrical production/demonstration room with hundreds of seats, and one home theatre demonstration room in NHK Science & Technology Research Laboratories.
– One laboratory installation in Fraunhofer IIS.
– One laboratory installation in McGill University CIRMMT.
– One laboratory installation at Qualcomm Technologies, Inc, San Diego, USA.

A large theatrical sound system with a 600 inch screen at World Exposition at Aichi, Japan in 2005 is shown in Fig. 3.

FIGURE 3

Theatrical 22.2 multichannel sound system at World Exposition held at Aichi, Japan in 2005

Report BS.2159-03

Post production rooms for 22.2 multichannel sound have also been installed as follows:

– Two mixing rooms in NHK broadcasting centre at Tokyo, one of which is applicable to live production.
– Two pre-production rooms, a production studio to receive contributions and a monitoring room for emission, in NHK broadcasting centre at Tokyo.
– One OB van.
– A mixing console for 22.2 multichannel sound is being installed in an 8K television studio (under construction).

Home sound systems for 22.2 multichannel sound have also been developed. Figure 4 shows the living room with 22.2 multichannel sound, where 24 loudspeakers are installed into the walls and ceiling.

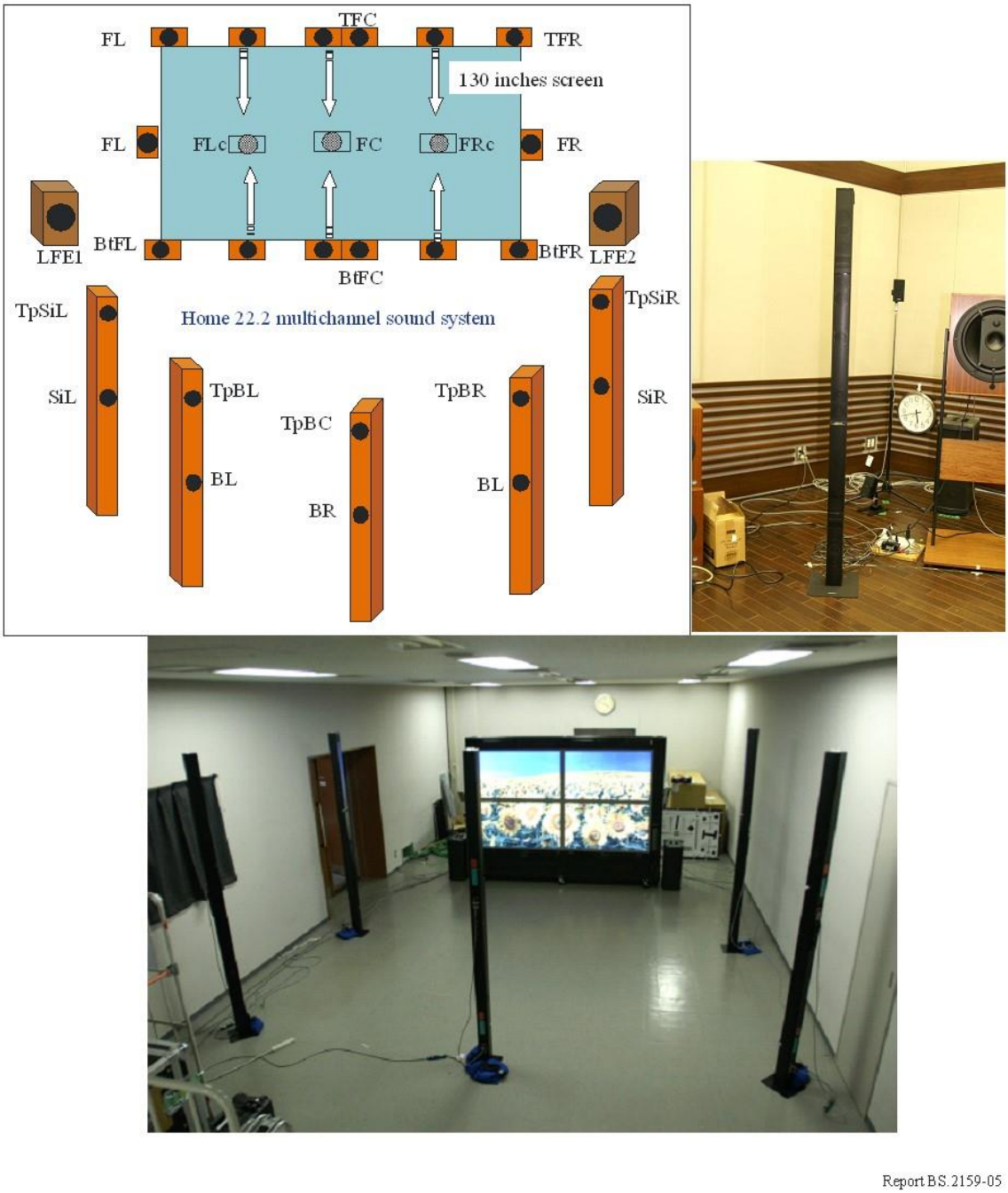
FIGURE 4

Living room with 22.2 multichannel sound

A tallboy type loudspeaker has been developed to reproduce three vertical channels (i.e. top, middle and bottom channels) by a single loudspeaker. These loudspeakers are used for the home 22.2 multichannel sound system with UHDTV FPD on which compact loudspeaker units are rigged up to reproduce frontal sound channels as Fig. 5.

FIGURE 5

Home 22.2 multichannel sound system using tallboy type loudspeakers



Report BS.2159-05

A headphone processor to provide 22.2 multichannel sound has been developed; it is shown in Fig. 6. This processor enables listeners to enjoy an accurate immersive 3D sound with ordinary headphones. Because the headphone processor can reproduce the 22.2 multichannel sound without the use of loudspeakers, TV programmes with 3D sound can be efficiently produced on location in places such as an OB van.

FIGURE 6

22.2 multichannel sound headphone processor

Report BS.2159-06

4.2 10.2 surround sound system (Type A)

4.2.1 Background

The Immersive Audio Laboratory is a part of the Integrated Media Systems Center at the University of Southern California and its practitioners have worked since the mid-1990s in the development of multichannel sound, especially 10.2-channel sound. This sound system is a logical extension of Recommendation ITU-R BS.775 and its 5.1-channel layout. Although called 10.2 as shorthand, it actually employs 14 electrical channels, explained below. 10.2 describes the number of loudspeaker locations, since some loudspeaker channels can be combined into one physical location.

4.2.2 Highlights

There are eight permanent installations of 10.2 channel sound as of February 2010. They are:

- Two cinemas with hundreds of seats each. Note that these use a variant on the basic system, designed specifically for cinemas (typ. > 2 500 cu. m.), where surround arrays are used for left, right, and rear surround, along with point sources for left and right surround.
- One home theatre demonstration room operating in an audio-video store. In operation for many years and used virtually continuously to demonstrate the advantages of more sound channels to the public.
- One high-power installation at USC's Institute for Creative Technologies, funded by DARPA.
- Two laboratory installations in Ronald Tutor Hall at USC.
- One installation at Inha University, Incheon, Kore.
- One installation in a private home.

In addition, more than a dozen temporary exhibitions have been made on four continents (North America, South America, Europe, and Asia).

There are more than 20 items of produced programme material. Since 10.2-channel sound is a playback platform, not a recording/playback system, a wide variety of methods of recording have been employed, from classic ones, to completely constructed spaces using advanced digital signal processing algorithms.

The system is scalable from very small listening rooms to cinemas. Changes are made in the physical system to accommodate the range of conditions encountered and its calibration, and the focus of the work has been in deriving the maximum interchangeability among the various size installations. There is no recalibration or mixing necessary to scale from the smallest to the largest space.

The loudspeaker layout was chosen considering physical acoustics of spaces to be reproduced; psychoacoustics of multichannel listening; and the desires of composers, sound designers, and other interested parties. Publications are available detailing these choices.

4.2.3 The loudspeaker channel layout

The loudspeaker channel layout starts with standard 5.1:

- L –30° in plan view, approximately 0° in elevation (raised slightly for line-of-sight in multi-row listening for direct path sound, or the L screen channel in cinemas which are 2/3 of the way up the height of the motion-picture screen to the high-frequency section for instance). Reference point is the centre of the listening area.
- R +30° in plan, same elevation and reference position as L.
- C 0° in plan (straight ahead), same elevation and reference position as L.
- LS direct –110°±10° in plan, same elevation and reference position as L.
- RS direct +110°±10° in plan, same elevation and reference position as L.

To which are added the following:

- Left Wide (LW) –60° in plan, same elevation and reference position as L.
- Right Wide (RW) +60° in plan, same elevation and reference position as L.
- LS diffuse. For “small” rooms of a typical room volume of 85 cu. m: typically a dipole type loudspeaker radiation pattern (low bass excepted) at –110°±10° in plan, elevated above the LS direct loudspeaker. For “large” rooms (cinemas) which are typically >1 000 cu. m: typically a surround array composed of four to twelve loudspeakers laid out for uniform sound level coverage of the listening area.
- RS diffuse. For “small” rooms as above: typically a dipole-type loudspeaker radiation pattern (low bass excepted) at +110°±10° in plan, elevated above the RS direct loudspeaker. For “large” rooms (cinemas) as above: typically a surround array composed of four to twelve loudspeakers laid out for uniform sound level coverage of the listening area.
- Back Surround (BS): For “small” rooms as above: +180° in plan, same elevation and reference position as L.
- Left Height (LH): –45° in plan and elevated +45° (or whatever is practical) above the listening plane.
- Right Height (RH): +45° in plan and elevated +45° (or whatever is practical) above the listening plane.
- L Sub: Systems employ bass management. Bass below the operating frequency range of all of the left channel loudspeakers (L, LH, LW, LS direct, LS diffuse) and C are added together at equal level and L LFE is added in at +10 dB in-band gain. Typical crossover frequency is 25-50 Hz. Typical L LFE low pass filter frequency (brick wall) is 120 Hz. The combined signals are sent to one or more subs located left of the listener. In cinemas they may be in the left front corner. In small rooms they may be on the left side of the room.

– R Sub: Bass below the operating frequency range of all of the right channel loudspeakers (R, RH, RW, RS direct, RS diffuse) and BS are added together at equal level and R LFE is added in at +10 dB in-band gain. Typical crossover frequency is 25-50 Hz. Typical R LFE low pass filter frequency (brick wall) is 120 Hz. The combined signals are sent to one or more subs located right of the listener. In cinemas they may be in the right front corner. In small rooms they may be on the right side of the room.

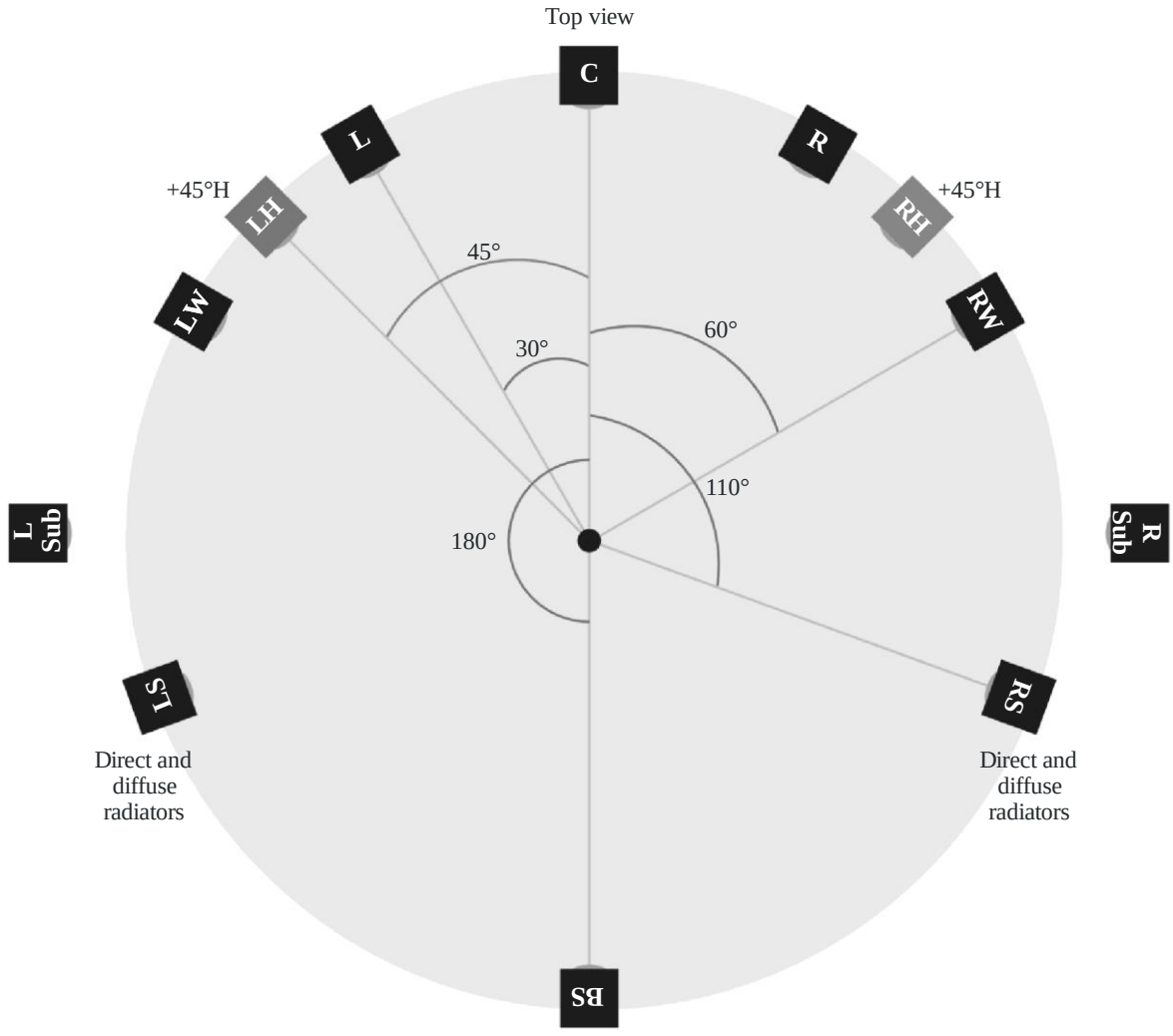
The consolidated positions of Left Surround direct and diffuse radiators, and Right Surround direct and diffuse radiators (applicable in "small" rooms), result in 10.2 total speaker locations. The system thus requires 14 electrical channels. Additionally, two channels of a sixteen-channel layout are reserved for Hearing Impaired and Visually Impaired descriptive service channels.

4.2.4 Standardization

By following the outlined speaker locations and sound calibration methods, 20 installations have been made to sound as similar as possible. 10.2 is recognized as a format in Apple Quicktime and in SMPTE Digital Cinema standards. It has been implemented by one audio workstation manufacturer, and another is expected to join.

FIGURE 7

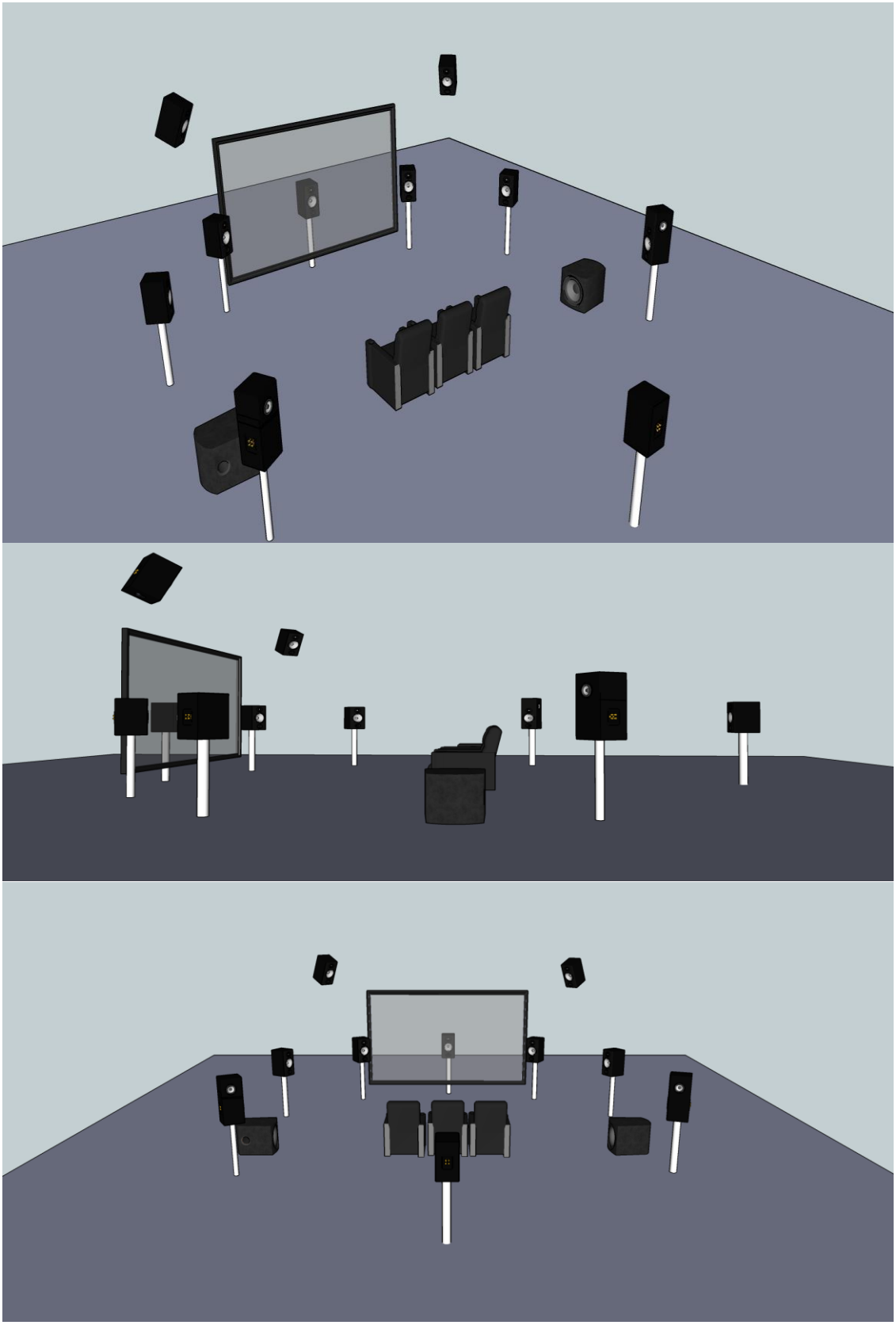
Diagram of speaker layout for 10.2



Report BS.2159-07

FIGURE 8

Typical “small room” installation schematic



Report BS.2159-08

4.3 10.2 channel sound system (Type B)

4.3.1 Background

In the Republic of Korea, a 10.2 channel audio system was developed and this multichannel system was standardized as "Audio Signal Formats for Ultra High Definition (UHD) Digital TV" in the Republic of Korea, TTAK.KO-07.0098 in 2011. This standard has been developed based on 3/4 loudspeaker arrangement of Recommendation ITU-R BS.775 for backward compatibility with the conventional system. The specific information about this system is described below. It is somewhat different from 10.2 surround sound system (Type A) of § 4.2.

4.3.2 The loudspeaker channel layout

Firstly, the terms for this layout are defined. The loudspeaker layout is composed of three heights, layers:

– middle layer: the height which is an ear position of listener;

– top layer: the height which is a position over the listener's head;

– bottom layer: the height which is a position under the listener's leg.

The 10.2 channel loudspeakers are defined as below.

TABLE 1

Channel definition of 10.2 channel

Channel	Label	Definition
L	Left channel/signal/speaker	Front left position on middle layer
R	Right channel/signal/speaker	Front right position on middle layer
LB	Left Back channel/signal/speaker	Rear left position on middle layer
RB	Right Back channel/signal/speaker	Rear right position on middle layer
C	Centre channel/signal/speaker	Front centre position on middle layer
LFE1	Left Low Frequency Effect channel/signal/speaker	Left side on bottom layer
LS	Left Side channel/signal/speaker	Left position on middle layer
RS	Right Side channel/signal/speaker	Right position on middle layer
LH	Left Height channel/signal/speaker	Front left position on top layer, elevated
RH	Right Height channel/signal/speaker	Front right position on top layer, elevated
CH	Centre Height channel/signal/speaker	Rear centre position on top layer, elevated
LFE2	Right Low Frequency Effect channel/signal/speaker	Right side on bottom layer

Then the 10.2 channel loudspeakers are arranged as below.

FIGURE 9

The 10.2 channel loudspeaker layout

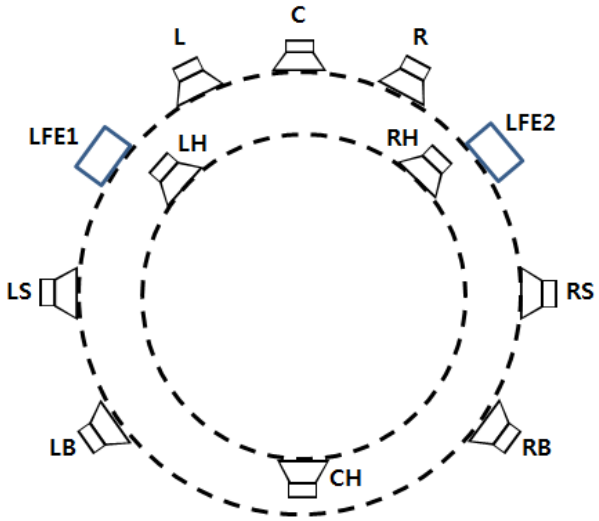


TABLE 2

Channel arrangements of 10.2 channel

Channel	Azimuth	Remark
C	0°	–
L, R	±30°	left and right each
LS, LB, RS, RB	±60° ~ 150°	at left and right, two channels each
CH	90° ~ 135°H	in that range
LH, RH	±30° ~ 45° & 30° ~ 45°H	horizontally and vertically each

The loudspeaker channel layout starts with the 5.1 and 3/4 loudspeaker arrangement of Recommendation ITU-R BS.775:

– L–30° in middle layer, 0° in elevation. Reference point is the centre of the listening area.

– R+30° in middle layer, same elevation and reference position as L.

– C0° in middle layer, same elevation and reference position as L.

– LS and LB-60~–150° in middle layer, same elevation and reference position as L.

– RS and RB+60~+150° in middle layer, same elevation and reference position as L.

To which are added the following:

– Left Height (LH)-30~–45° in middle layer with +30~+45° elevated. Reference point is the ear level of listener and this channel positioned in top layer.

– Right Height (RH)+30~+45° in middle layer with +30~+45° elevated and reference position as LH.

– Centre Height (CH)+90~+135° elevated and reference position as LH.

– LFE1: Systems employ bass management. Bass below the operating frequency range of all of the left channel loudspeakers (L, LS, LB, LH) C, and CH are added together at equal level; and

– LFE2: Bass below the operating frequency range of all of the right channel loudspeakers (R, RS, RB, RH), C and CH are added together at equal level.

So the resulting arrangement is depicted in Figs 10 and 11 below.

FIGURE 10

Middle layer and top layer of 10.2ch loudspeaker layout

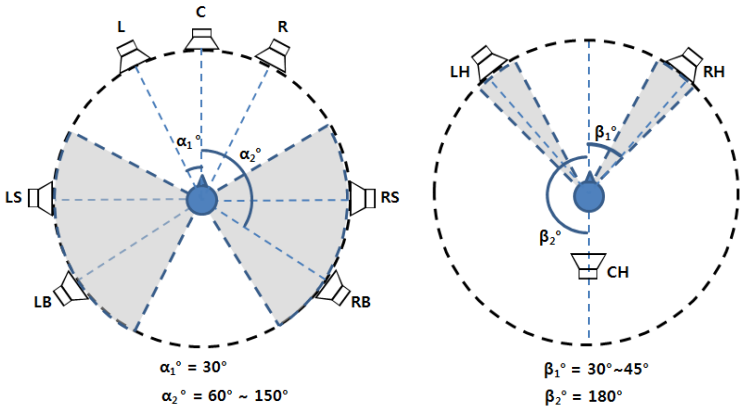
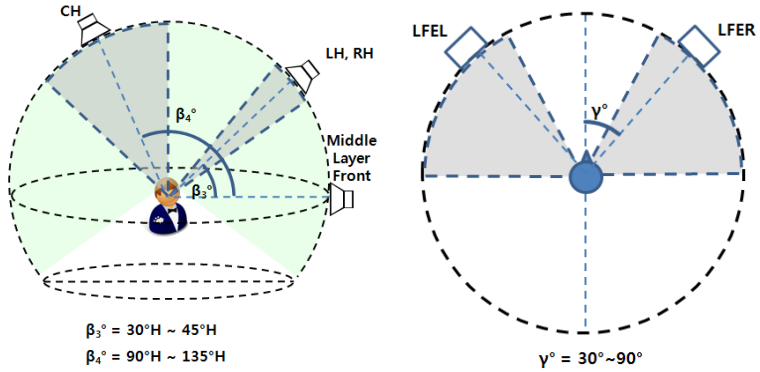


FIGURE 11

Top layer and bottom layer of 10.2ch loudspeaker layout



4.3.3 Recommended arrangement of 10.2 channel

The outlined speaker location as the recommended arrangement of 10.2 channel audio system is as follows:

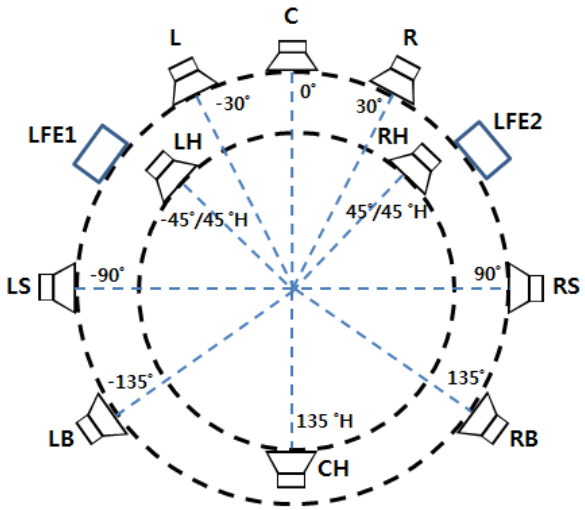
TABLE 3

Specific channel arrangements of 10.2 channel

Channel	Azimuth
LS	–90°
RS	90°
LB	–135°
RB	135°
LH	–45°/45°H
RH	45°/45°H
CH	135°H

FIGURE 12

Specific channel arrangements of 10.2 channel



4.4 Wave-field-synthesis

Wave-field-synthesis (WFS) was invented by the Delft University of Technology, Netherlands in 1989. In the European project CARROUSO components for the complete chain including recording, coding, transmission, decoding, and sound reproduction were developed. Since then WFS has been refined to deliver truly immersive sound. Application areas include cinema (with a priority on combination with 3D video), theme parks, virtual reality (VR) installations (in combination with 3D audio) and, in the long run, home theatres. In February 2003 the first cinema using this system started daily operation (Ilmenau, Germany). In 2004 the first WFS system was installed in a sound stage in Studio City, CA. Since 2008, the Chinese 6 Theatre and the Museum of Tolerance in Los Angeles have been equipped with WFS sound systems. These systems are also used in themed environments. Commercial examples of IOSONO GmbH (a spinoff of Fraunhofer IDMT) include the installation in the 4D cinema at the Bavaria Filmstadt (Munich), the Odysseum Science Adventure Park (Cologne), and the “Haunted Mansion” at Disney World (Orlando). Virtual reality installations at the University of Surrey and the Technical University of Ilmenau use WFS with two loudspeaker arrays in the front to enable the proper reproduction of elevation. These two systems also use stereoscopic video projection. An extension of WFS with additional loudspeakers above the listeners was presented at “the 2008 Expo” in Saragossa.

FIGURE 13

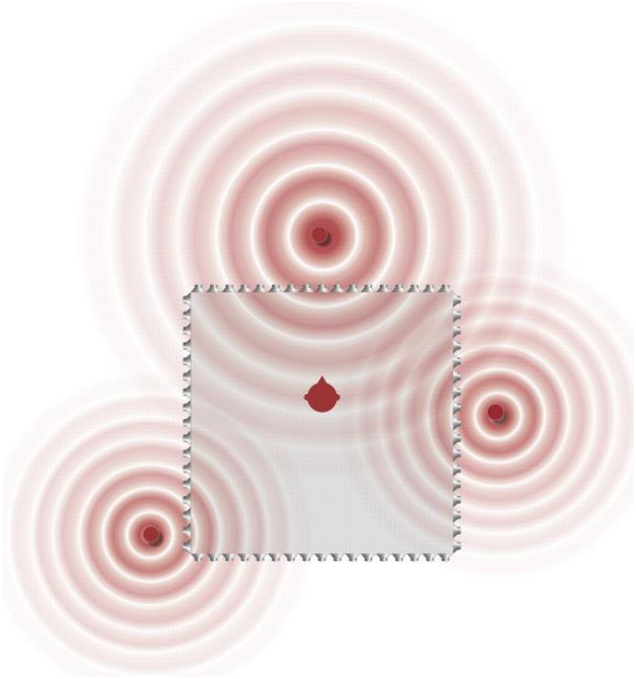
WFS sound system in the German cinema "Linden lichtspiele ilmenau"

Report BS.2159-13

WFS is an object-oriented approach to accurately recreate a replication of a sound field using the theory of waves and of the generation of wave fronts. This concept is best explained by the well-known Huygens principle: points on a wave front serve as individual point sources of spherical secondary waves. This principle is applied in acoustics by using a large number of small and closely spaced loudspeakers (loudspeaker arrays). The driving signal is calculated for each of the loudspeakers in real time at the reproduction site. The number of loudspeakers is independent from the number of transmission channels and only related to the size of the reproduction room. Loudspeaker arrays controlled by WFS reproduce wave fields that originate from any combination of (virtual) sound sources like an acoustic hologram. When manipulated properly, the system recreates wave fronts approaching perfect temporal, spectral and spatial properties throughout the listening room.

FIGURE 14

Working principle of WFS



Report BS.2159-14

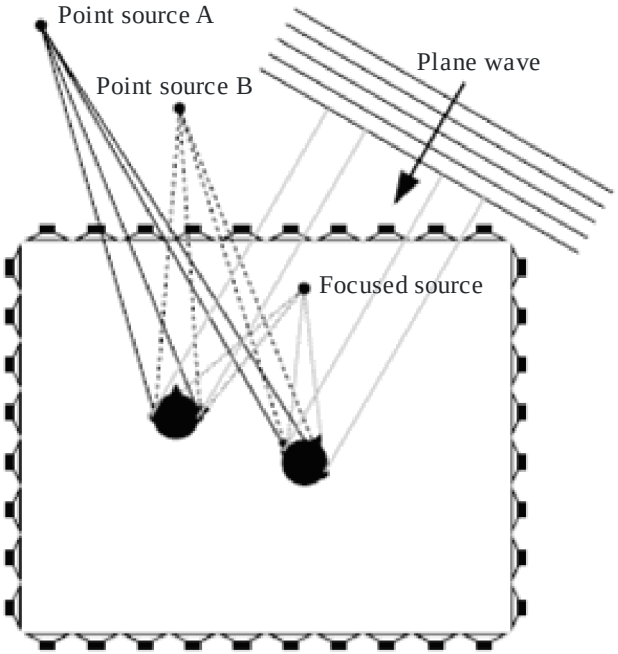
Three representations of sound sources are possible in WFS (see Fig. 15). In the first two, virtual sound sources can be placed behind the loudspeaker arrays (so-called point sources) as well as in front of loudspeaker arrays (so-called focused sources). In the case of sound sources in front of the array, the array radiates convergent waves toward a focus point, from which divergent waves propagate into the listening area. The third type is the so-called plane waves. Plane waves come from the same angular direction for all positions in the listening space.

The other commonly used channel-oriented sound reproduction approaches require a well-defined loudspeaker setup, i.e. the number and positions of the loudspeakers are predefined. In the mastering process the target setup must be determined and the loudspeaker signals prepared in a way that allows them to perfectly fit the assumed setup. This implies that it is difficult to feed the generated signals into another sound system.

This problem can be solved by the object-oriented sound reproduction paradigm, which was develop for WFS but which is not restricted to it. In this method, the audio content is represented as audio objects containing the pure audio content together with metadata describing the position of the object in real time along with the properties of the audio object like directivity. On the rendering site the driving signal for each individual loudspeaker is calculated taking into account its exact position in the reproduction room. Besides the positioning of direct sound, a position-dependent calculation of early reflections and diffuse reverberations is possible, which enables the generation of realistic but also artificial spatial environments. Through the availability of the direct sound of each source and a parametric description of the properties of the room, an optimal reproduction can be adapted to the given spatial environment. This can be a WFS setup of any size (and number of loudspeakers) but also an arbitrary loudspeaker configuration. Increasing the number of loudspeakers increases the size of the sweet spot and makes the sound sources more stable. This results in an increased freedom when deciding which loudspeaker setup to install, because the actual loudspeaker signals are calculated at the reproduction site through a process called rendering.

FIGURE 15

Reproduction of point sources, focused sources and plane waves



Report BS.2159-15

WFS overcomes the restrictions of a sweet-spot and enable the location of sound objects at any position outside and inside the reproduction room without problems of phase or sound coloration. All formats mentioned in §§ 4.1, 4.2 and 5 can be reproduced using WFS by the concept of virtual loudspeakers enabling an enlarged sweet-spot for any content already produced.

4.4.1 Object-based multichannel audio system

This system was developed based on the principles of wave-field synthesis, originally invented by TU Delft and explored in the European project CARROUSO[2]. In CARROUSO the MPEG-4 BIFS was used to represent the audio data. For commercial applications this very flexible format proved to be too complex (in terms of storage requirements and computing power) and therefore a less expensive version of the file format had to be developed. With the intention of keeping the perceptual properties of wave-field synthesis, a system for flexible 3D speaker layouts was developed in Germany[3].

A complete production and reproduction chain based on the object-based paradigm is available. More than 20 commercial and demonstration installations of systems exist worldwide (e.g. Chinese 1 Multiplex Theater and Chinese 6 Multiplex Theater in Hollywood, Los Angeles). Virtual reality installations at the University of Surrey and the Technical University of Ilmenau use WFS with two loudspeaker arrays in the front to enable the proper reproduction of elevation. These two systems also use stereoscopic video projection. An extension of WFS with additional loudspeakers above the

---

[2] Partners in CARROUSO: Fraunhofer IDMT (Federal Republic of Germany), IRT (Federal Republic of Germany), University of Erlangen (Federal Republic of Germany), France Telecom (France), IRCAM (France), Thales (France), TU Delft (Netherlands), Aristotle University of Thessaloniki (Greece), EPFL (Switzerland), Studer (Switzerland).

[3] By IOSONO GmbH Erfurt (Federal Republic of Germany) and Fraunhofer IDMT Ilmenau (Federal Republic of Germany).

listeners was presented at “the 2008 Expo” in Saragossa. A 3D setup with two layers of loudspeakers was shown at Prolight + Sound 2011 in Frankfurt, Germany. A few installations are shown here to illustrate the diversity that can be realized using the object-based audio system. Content can be exchanged between all these systems.

FIGURE 16

Installation with 64 loudspeakers at the Chinese Multiplex Theater, Hollywood

Report BS.2159-16

FIGURE 17

Setup with 2 flexible layers of 34 loudspeakers presented at a trade show

Report BS.2159-17

FIGURE 18

Wave-field synthesis based setup at Peltz Theatre, Beverly Hills

Report BS.2159-18

A headphone processor to process object-based scene description for dynamic binaural headphone reproduction has been developed. The headphone processor can be used to simulate several loudspeaker layouts to monitor the auditory scene as it would be rendered in a real loudspeaker setup.

4.5 Object-based audio formats

The other commonly used channel-oriented sound reproduction approaches require a well-defined loudspeaker setup, i.e. the number and positions of the loudspeakers are predefined. In the mastering process the target setup must be determined and the loudspeaker signals prepared in a way that allows them to perfectly fit the target setup. This implies that it is difficult to feed the generated signals into another sound system. This problem can be solved by the object-based sound reproduction paradigm, which was develop for WFS but which is not restricted to it.

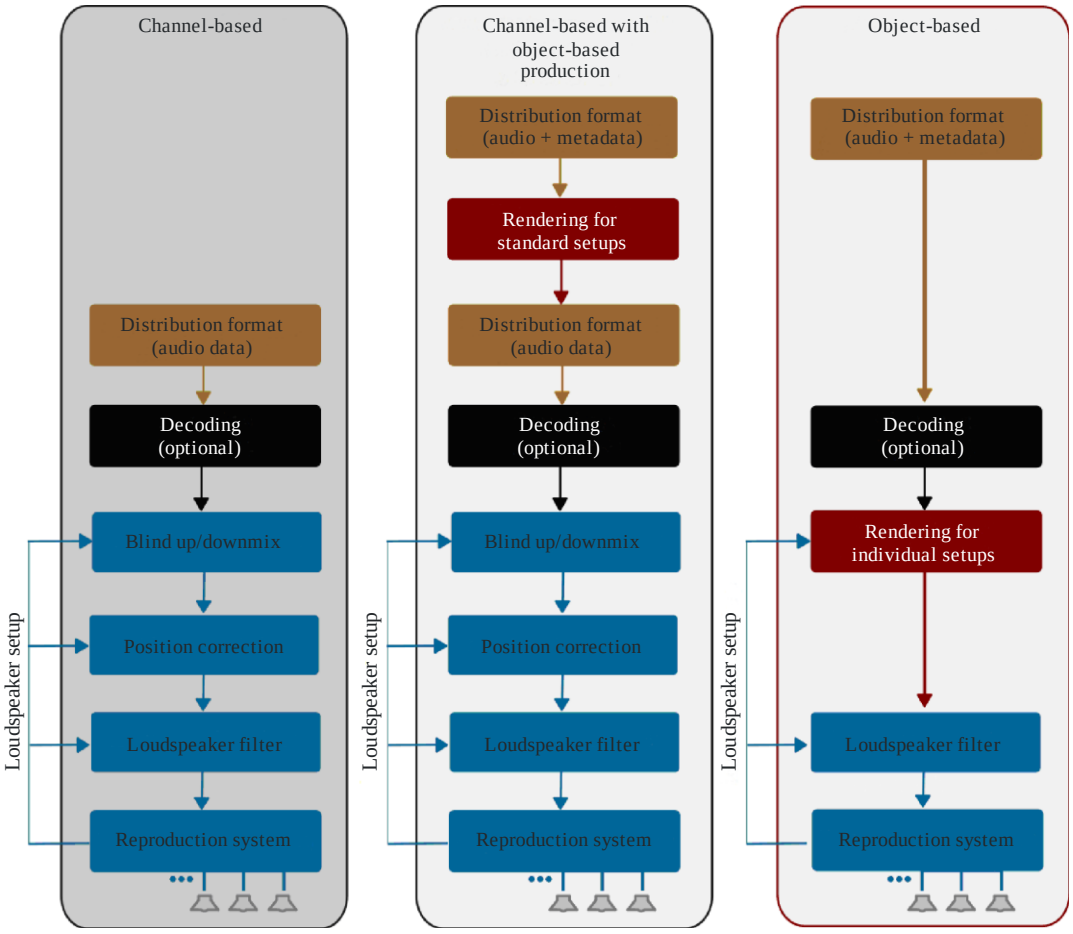
In an object-oriented system the audio content is created independently of any specific loudspeaker layout. The audio content is represented as audio objects containing the pure audio content, together with metadata describing the position of the audio object along with the properties of the audio object such as directivity in real time.

On the rendering site the driving signal for each individual loudspeaker is calculated, taking into account its exact position in the reproduction room. Such representations can be rendered in real time to loudspeaker setups from 5 to more than 500 speakers. The setups do not have to be regular or in a specific layout but standard layouts can easily be supported (as shown in Fig. 19). Furthermore, the auditory scene can be scaled to the current screen size and size of the audience area in a reproduction venue. In that way the content can be transferred between different cinemas as well as to domestic-size screens. Due to the adaptive rendering, loudspeakers do not have to be placed in a specific relationship to the screen. The setup of a system in a home environment becomes flexible and acceptable. This results in an increased freedom when deciding which loudspeaker setup to install, because the actual loudspeaker signals are calculated at the reproduction site through a process called rendering.

WFS overcomes the restrictions of a sweet-spot and enables the location of sound objects at any position outside and inside the reproduction room without problems of phase or sound coloration, if an appropriate number of loudspeakers are installed. All current or future multichannel formats can be reproduced using WFS by the concept of virtual loudspeakers enabling an enlarged sweet-spot for any content already produced.

FIGURE 19

Comparison between channel based and object-based production system



Report BS.2159-19

4.5.1 Rendering and reproduction of object-based audio

Depending on the specified speaker setups the algorithm scales the reproduction of an object-based scene. If only a few loudspeakers are available, a rendering with comparable quality to any multichannel format is the result. On the other end, if a wave-field synthesis loudspeaker setup is available, wave-field synthesis is used for the rendering process. Due to its flexibility loudspeaker signals for multichannel layouts like 22.2, 10.2, 9.1 or 5.1 can be rendered in real time directly using the production or reproduction tools. Using the spatial audio processor a rendering with a specific adaptation to a venue is possible and loudspeaker setups from 5 to 500 speakers are possible. Such a system can reproduce different source types which are known from wave-field synthesis. Point sources enable the perception of a fixed source position for the whole audience area. Plane waves enable the perception of a fixed source direction for the whole audience area. Depending on the number of loudspeakers the focusing can be used to create a source position between loudspeaker and listener.

4.6 Hybrid channel/object-based system

4.6.1 Introduction

Recently there has been considerable interest in alternative spatial audio description methods in the audio industry. The developers of this hybrid channel/object based system had long recognized the potential benefits of moving beyond “speaker feeds” as a means for distributing spatial audio.

At a high level, there are three main spatial audio description formats:

– Speaker feed – the audio is described as signals intended for loudspeakers at nominal speaker positions. Binaural audio is a special case where the speakers are located at the left and right ears.

– Model- or Object-based description – the audio is described in terms of a sequence of audio events at specified positions.

– Sound field description – describes the acoustic sound field, not a set of sound sources (e.g. objects or speakers). For example, an acoustic sound field can be described within a region using spherical harmonics.

The **speaker-feed format** is the most common because it is simple and effective. If the playback system is known in advance, mixing, monitoring and distributing a speaker feed description that identically matches the target configuration provides the highest fidelity. However, in most cases the playback system is not known and can only be assumed to conform to a general standard e.g. stereo, 5.1. Deviation from nominal speaker placement results in distortions of the spatial information; however timbre is generally well preserved. For content where spatial accuracy is not critical, the speaker-feed format is effective. There is a large body of excellent stereo and multi-channel audio programmes that support this statement.

The **object-based description** is the most adaptable because it makes no assumptions about the rendering technology and is therefore most easily applied to any rendering technology. This adaptability allows the listener the freedom to select a playback configuration that suits their individual needs or budget – with the audio rendered specifically for their chosen configuration. The model-based description efficiently captures high resolution spatial information and enables accurate and lifelike reproduction that is particularly effective for discrete audio images. The object-based model includes much information beyond position, including size.

This system combines these two scene description methods.

Hybrid system

A hybrid channel- and object-based audio system provides all the benefits of a traditional speaker-based format:

– high timbre control and fidelity,

– direct control of speaker signals when desired,

– efficient transmission of dense audio ambiences and textures,

– traditional authoring options that allow mixers to make use of their experience and expertise, while incorporating the new capabilities at their own pace,

and extends the capabilities to include the following benefits

– more immersion and envelopment,

– increased spatial resolution, e.g. an audio object can be dynamically assigned to any one or more loudspeakers within a traditional surround array,

– ability to effectively bring sound images off screen,

– single inventory distribution compatible with effective adaption to alternative rendering modes including 5.1 and 7.1,

– familiar surround mixing paradigm. The front end of the mixing process is identical to existing tools. The rendering step is delayed until after distribution.

This system has been introduced into the cinema marketplace under the trade name “Dolby Atmos”.

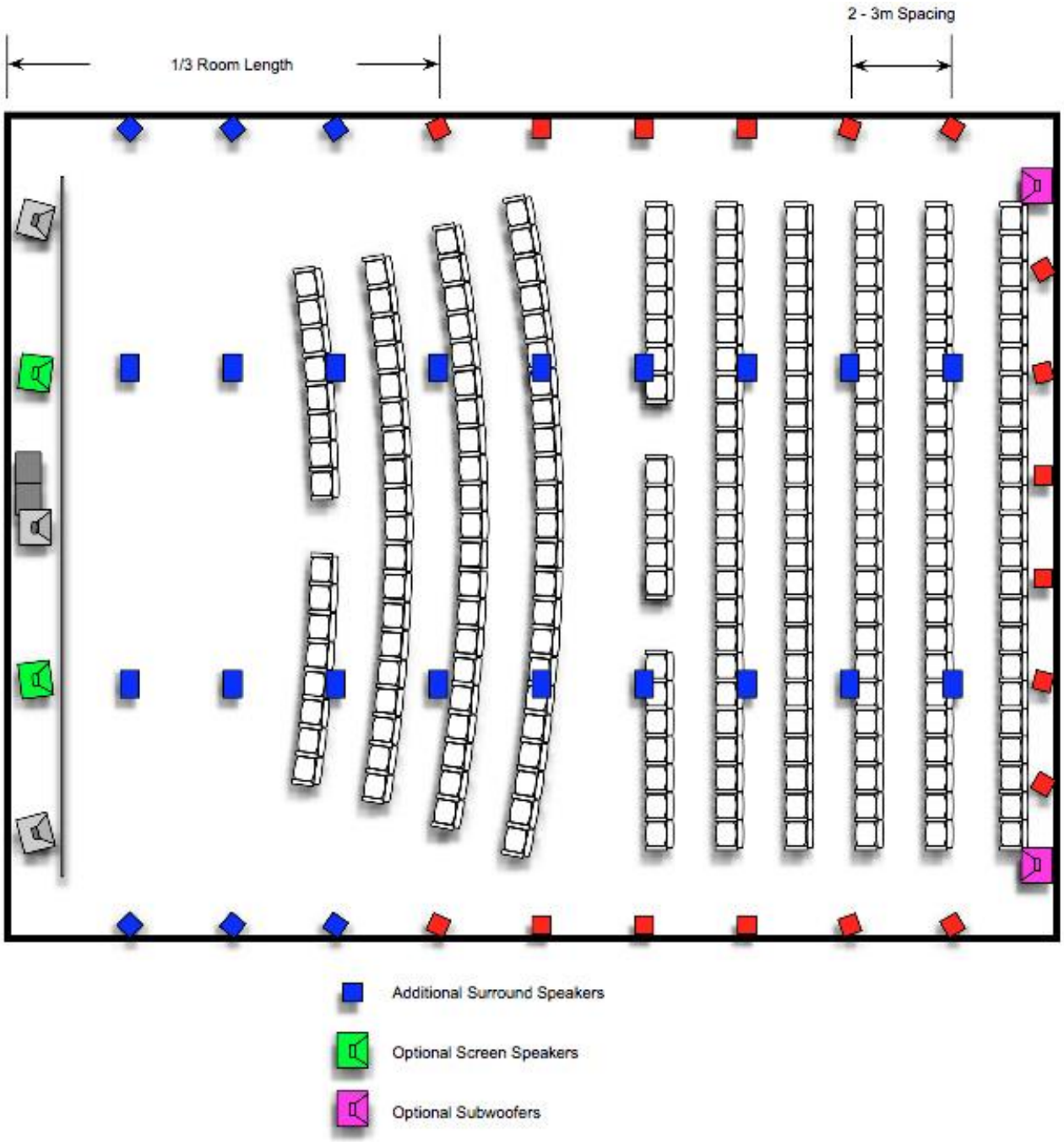
4.6.2 System design in theatres and LSDI venues

The recommended layout of speakers for this hybrid system remains compatible with existing cinema systems and LSDI venues, which is important so as not to compromise the playback of existing 5.1 and 7.1 channel-based formats. In the same way that the intent of the content creator must be preserved with the introduction of this system, the intent of mixers of 7.1 and 5.1 surround content must equally be respected. This includes not changing the positions of existing primary front channels in an effort to heighten or accentuate the introduction of new speaker locations. This hybrid format is capable of being accurately rendered in the cinema to speaker configurations such as 7.1, allowing the format (and associated benefits) to be used in existing venues with no change to amplifiers or loudspeakers.

Different speaker locations can differ in effectiveness depending on the room design, and therefore the industry appears to agree that there is not an ideal number or placement of channels. As a result, this hybrid format is adaptable and able to play back accurately in a variety of rooms, whether they have a limited number of playback channels or many channels with highly flexible configurations. Figure 20 shows a diagram of suggested speaker locations in a typical auditorium. The reference position referred to in the document corresponds to a position two-thirds of the distance back from the screen to the rear wall, on the centre line of the screen.

FIGURE 20

Recommended speaker locations



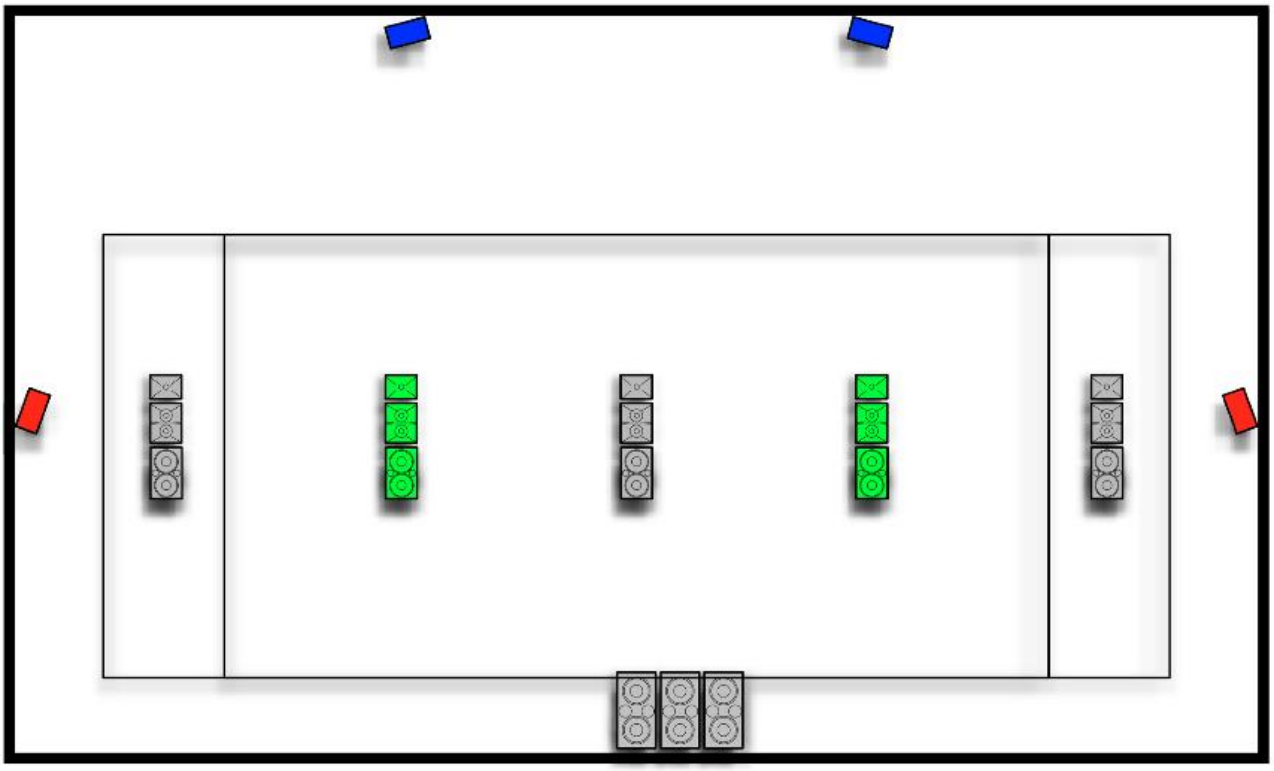
Screen speakers

The developers have studied the perception of a higher speaker density (both vertical and horizontal) in the screen plane. It was found that additional speakers behind the screen, such as Left Centre (Lc) and Right Centre (Rc) screen speakers (in the locations of Left Extra and Right Extra channels in 70 mm film formats), can be beneficial in creating smoother pans across the screen, while additional layers of vertical speakers provide little benefit in particular in the context of stadium seating configurations. Consequently, for cinemas and LSDI venues it is recommended to install these additional speakers, particularly in auditoria with screens greater than 12 m (40 ft) wide. All screen speakers should be angled such that they are aimed toward the reference position. The recommended placement of the subwoofer behind the screen remains unchanged, including maintaining asymmetric cabinet placement, with respect to the centre of the room, to prevent stimulation of standing waves.

Figure 21 shows a diagram of suggested speaker locations at the screen. For home viewing applications, screen height speakers have been introduced in the past (e.g. in the Dolby ProLogic 2z format) as they offer a convenient trade-off between addition of some height dimension into the mix and ease of installation.

FIGURE 21

Recommended speaker locations (screen, side surrounds, and top surrounds)



Surround speakers

Ideally, surround speakers should be specified to handle an increased SPL for each individual speaker, and also with wider frequency response and the ability to provide uniform coverage throughout the seating area where possible.

As a rule of thumb for an average-sized theatre, the spacing of surround speakers should be between 2 and 3 m (6' 6" and 9' 9"), with Left and Right Surround speakers placed symmetrically. However, the spacing of surround speakers is most effectively considered as angles subtended from a given listener between adjacent speakers, as opposed to using absolute distances between speakers.

For optimal playback throughout the auditorium, the angular distance between adjacent speakers should be 30° or less, referenced from each of the four corners of the prime listening area. Good results can be achieved with spacing up to 50°. For each surround zone, the speakers should maintain equal linear spacing adjacent to the seating area where possible. The linear spacing beyond the listening area, such as between the front row and the screen, can be slightly larger.

Side surrounds

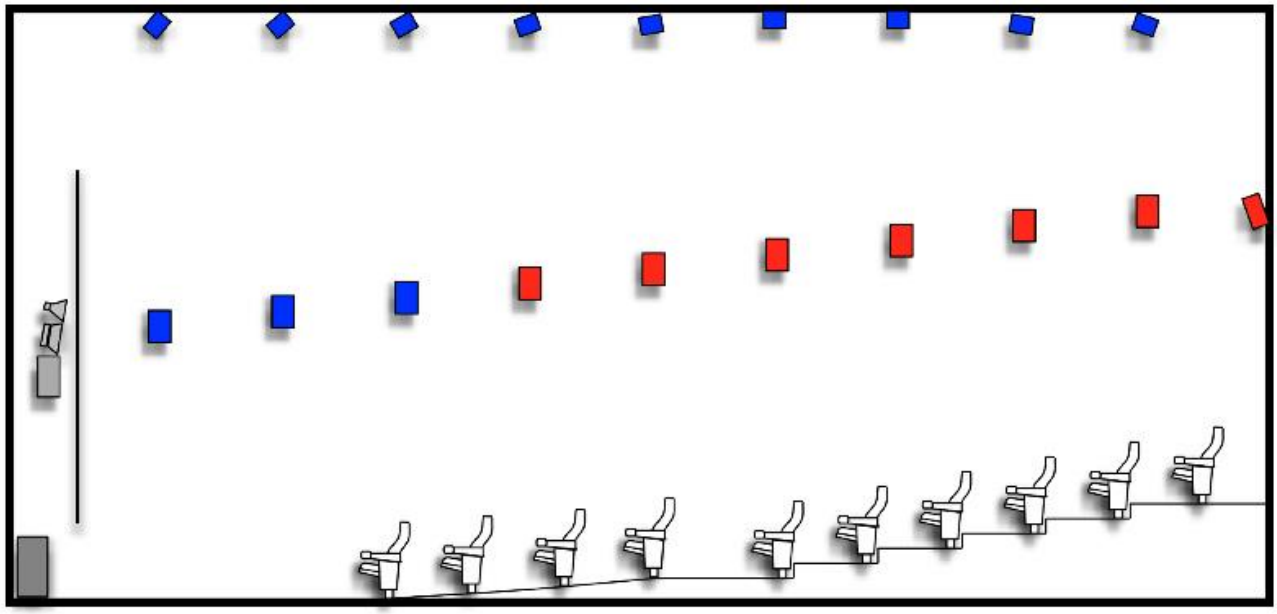
Additional side surround speakers should be mounted closer to the screen than the currently recommended practice of starting approximately one-third of the distance to the back of the auditorium. These speakers are not used as side surrounds during playback of 7.1 or 5.1 soundtracks, but they will enable smooth transition and improved timbre matching when panning objects from the screen speakers to the surround zones.

To maximize the impression of space, the surround arrays should be placed as low as is practical, subject to the following constraints: the vertical placement of surround speakers at the front of the array should be reasonably close to the height of the screen-speaker acoustic centre, and high enough to maintain good coverage across the seating area according to the directivity of the speaker. The vertical placement of the surround speakers should be such that they form a straight line from front to back, and (typically) slanted upward so the relative elevation of surround speakers above the listeners is maintained toward the back of the cinema as the seating elevation increases, as shown in Fig. 22. In practice, this can be achieved most simply by choosing the elevation for the front-most and rear-most side surround speakers, and placing the remaining speakers in a line between these points.

The distance between side surround speakers should be determined based on the guiding principles at the start of this section.

FIGURE 22

Recommended side wall and ceiling speaker locations



Rear surrounds

The number of rear surround speakers, and the distance between them, should be determined based on the same guiding principles as for the side surrounds. The back-wall speakers should have approximately the same linear spacing as the side surrounds adjacent to the seating area, although it may be necessary to slightly increase the density of back surrounds in order to meet the angular requirements. Such an increase in density can also be an advantage for power handling of the left and right rear surround zones, which are typically half the length of the side surround zones.

Top surrounds

Overhead (or top surround) speakers should be in two arrays from the screen to the back wall, nominally in alignment with the Lc and Rc screen channels of a typical auditorium, where the screen width is effectively the width of the theatre and the screen top is near the ceiling. They should always be placed symmetrically with respect to the centre of the screen. The top surrounds should have the same design characteristics as the side surrounds to maintain timbre matching.

The number and spacing of the top surround speakers should be based on the position of the side surround speakers. However, the spacing of top surround speakers is less critical than for side surrounds, and so it is acceptable for the number and front-back position to vary relative to the side surrounds if necessary. The top surround array should also extend to the screen in the same manner as the side surrounds. For home installations, screen height speakers when present can therefore be considered as being part of the top surround zones.

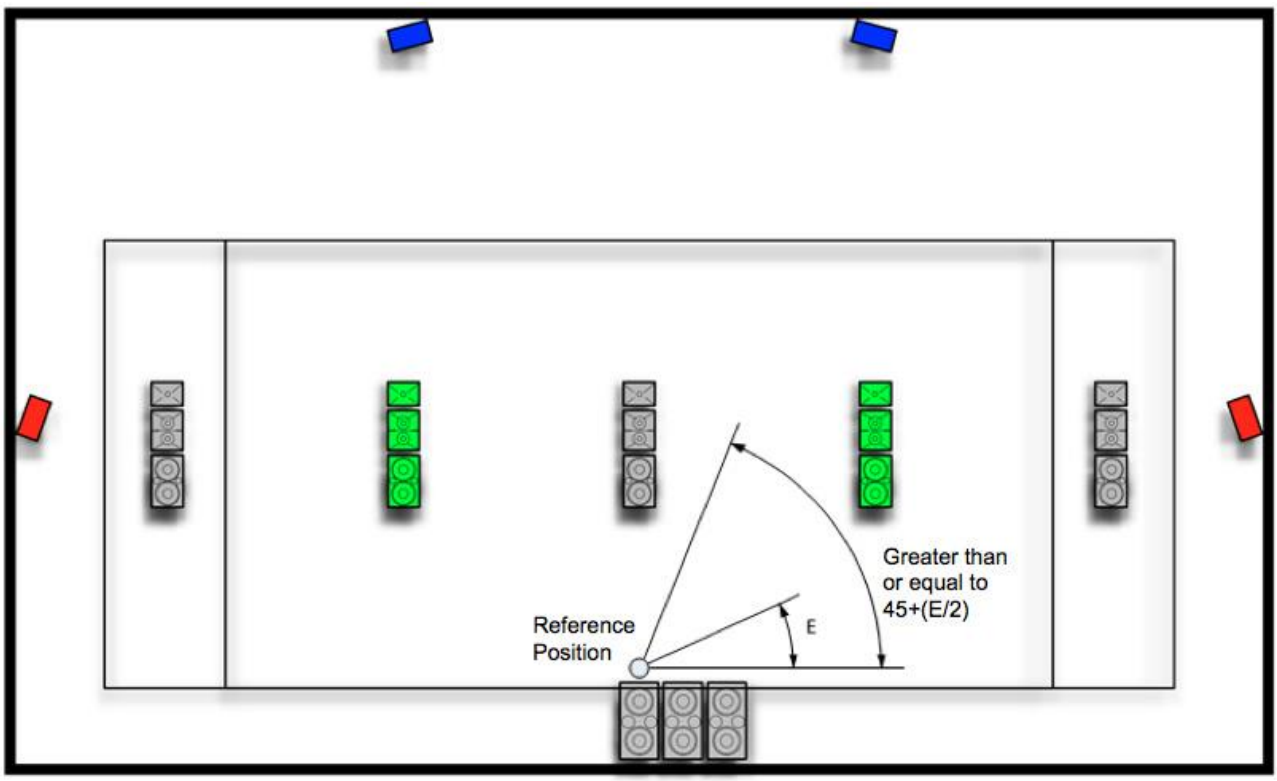
The lateral position of the arrays should be chosen to optimize spatial immersion and uniformity across the listening area. As stated earlier, placing the top surround speaker arrays in alignment with Lc and Rc screen channels will generally give good results. For rooms where the seating area is significantly wider than the screen, or the top surrounds are mounted significantly higher than the level of the top of the screen, it is desirable to have the overhead arrays more widely spaced.

The minimum width is the Lc and Rc spacing. The maximum width should be determined based on elevation angles as follows.

Let E be the elevation angle of the nearest side surround, measured from a reference position in the middle of the seating area (typically 15 to 25 degrees). The elevation angle of the corresponding top surround array should be greater than or equal to 45° plus half of angle E, as shown in Fig. 23. For example, if E is 20°, then the elevation angle of the top surround array should be greater than or equal to 55°.

FIGURE 23

Example of top surround lateral position



Individually addressable surround speakers

The full channel set is as follows:

9.1 channel: Left, Right, Centre, LFE, Left side surround, Right side surround, Left rear surround, Right rear surround, Left top surround and Right top surround.

While there are now 10 channels, there are many more speakers. For example, Fig. 24 shows 42 surround loudspeakers within the six surround zones. The channel renderer will group the speakers from these six zones into arrays, and send to each of these arrays the matching channel signals.

However, using the object-based representation, an audio object can be precisely positioned in the room. The object renderer will determine the best speaker, or speakers, to play back the object audio stream. The speaker(s) used could be within an array, or span multiple speakers across arrays.

Bass management

A significant shortcoming of traditional large venue surround sound is the lack of "full range" surrounds. Specifically, while typical screen channels have a frequency response extending down to 40 Hz and lower, surround speakers often begin to roll off at 100 Hz. If a full range sound is panned off the screen, the timbre will shift dramatically as a result of the lack of low frequency capability of the surrounds. As a result of this timbre shift (as well as the lack of spatial resolution) mixers hesitate to bring sound objects off the screen. To address this issue, the concept of a left/right "surround direct" loudspeaker pair has been standardized in SMPTE 428-3, and can also be found in some Imax configurations, where the Ls and Rs loudspeaker arrays are replaced by a pair of full range speakers.

The recommendation for this hybrid system is to include a pair of surround subwoofers. The channel and object renderers redirect low frequency content from the left and right arrays (side, rear and top) to the subwoofers, taking advantage of the limited directionality of low frequency sound. In effect, every surround loudspeaker becomes a surround direct loudspeaker. The appropriate crossover frequency is established during installation based on the capabilities of the surround loudspeakers. Bass redirection is optional. The goal is full range surrounds. If the surround loudspeakers have sufficient low frequency extension, bass redirection is not needed.

5 Multichannel sound systems in use for home audio release media

The following multichannel sound systems are used in home audio entertainment.

5.1 DVD audio

DVD audio is a digital format for delivering exceptionally high-fidelity audio content on a DVD. It offers many channel configurations of audio channels, ranging from mono to 5.1-channel surround sound, at various sampling frequencies and bit resolution per sample (from compact disc 44.1 kHz/16 bits up to 192 kHz/24 bits). Compared with the CD format, the much higher capacity DVD format enables the inclusion of considerably more music (with respect to total running time and quantity of songs) and/or far higher audio quality (reflected by higher sampling frequencies and greater bit resolution per sample, and/or additional channels for spatial sound reproduction).

Audio is stored on the disc in linear pulse code modulation (PCM) format, which is either uncompressed or losslessly compressed with Meridian Lossless Packing (MLP). The maximum permissible total bit rate is 9.6 Mbit/s. In uncompressed modes, it is possible to get up to 96 kHz/16 bits or 48 kHz/24 bits in 5.1-channel surround sound. To store 5.1-channel surround sound tracks in 88.2 kHz/20 bits, 88.2 kHz/24 bits, 96 kHz/20 bits or 96 kHz/24 bits MLP encoding is mandatory.

5.2 SACD

Super Audio CD (SACD) is a read-only optical audio disc format that provides higher fidelity digital audio reproduction. SACD audio is stored in a format called Direct Stream Digital (DSD), which differs from the conventional PCM used by compact disc or conventional computer audio systems. DSD is 1-bit and has a sampling frequency of 2.8224 MHz. This gives the format a greater dynamic

range and wider frequency response than that of the CD. The system is capable of delivering a dynamic range of 120 dB from 20 Hz to 20 kHz and an extended frequency response up to 100 kHz.

SACD supports up to six channels at full bandwidth. In its current form the SACD standard does not precisely specify how the channels shall be used.

222 sound currently uses SACD to provide 2 + 2 + 2 sound contents consist of 6 channels including 4 channels (front left, front right, rear left and rear right) and 2 height channels (top front left and top front right).

5.3 BD

BD is an optical disc format. The format was developed to enable recording, rewriting and playback of high-definition (HD) video, as well as storing large amounts of data. BD pre-recorded application format (BD-ROM) is designed not only for pre-packaged HD movie content but also as a key component of a consumer HD platform. The BD platform is designed to provide access to HD content throughout the home via HD digital broadcast recording and HD playback functions.

One of the key features offered by BD-ROM is:

– Industry standard high definition video and surround sound audio:
  - MPEG-2, MPEG-4 AVC, and SMPTE VC-1 video formats;
  - LPCM as well as Dolby Digital, Dolby Digital Plus, Dolby Lossless, DTS digital surround, and DTS-HD audio formats.

BD-ROM supports six types of audio stream formats ranging from a low bit rate to high audio quality, as shown in Table 4.

TABLE 4

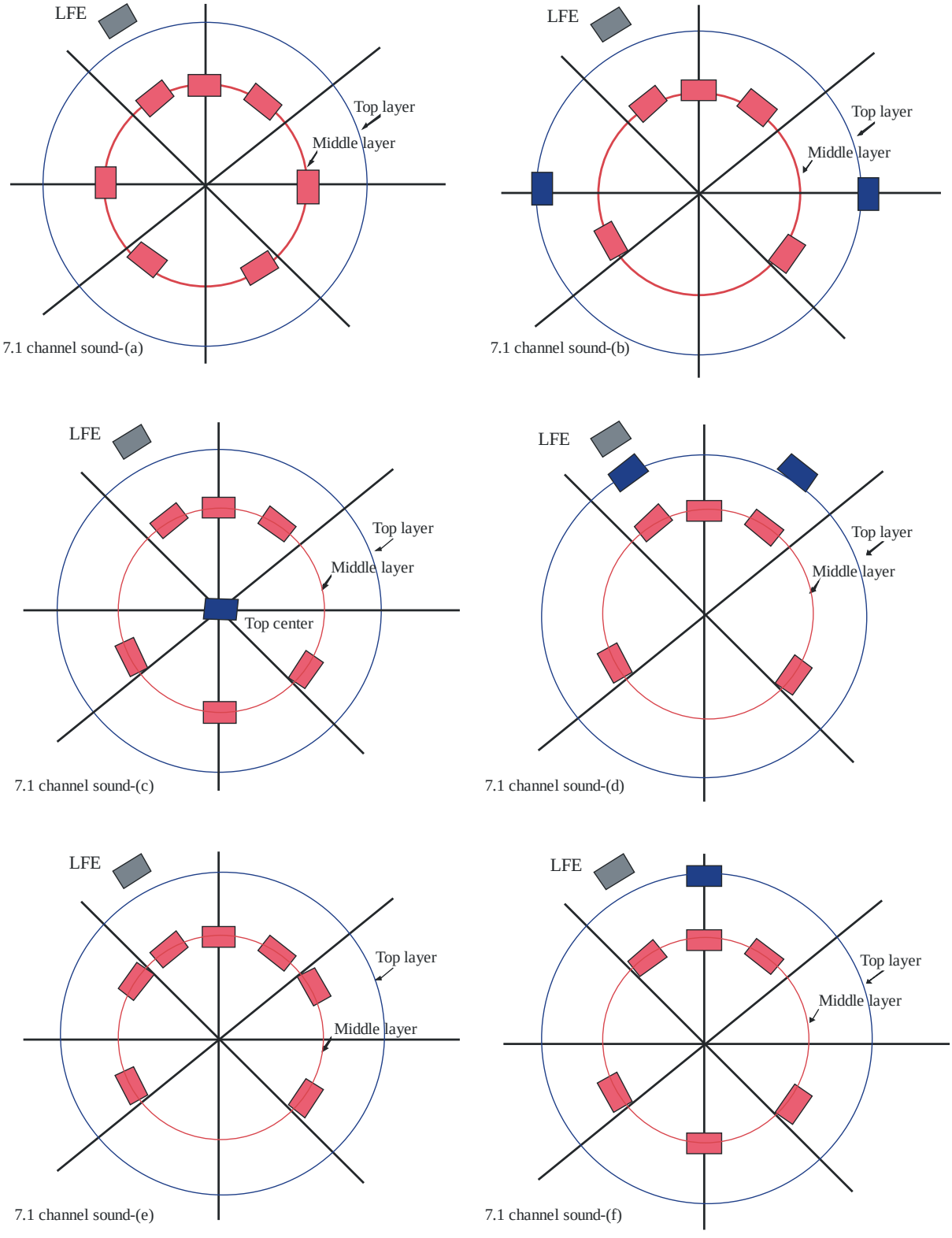
Specification of BD-ROM audio streams

CODEC	LPCM	Dolby Digital	Dolby Digital Plus	Dolby lossless	DTS digital surround	DTS-HD
Max. bit rate	27.648 Mbit/s	640 kbit/s	4.736 Mbit/s	18.64 Mbit/s	1.524 Mbit/s	24.5 Mbit/s
Max.ch	8(48 kHz, 96 kHz), 6(192 kHz)	5.1	7.1	8(48 kHz, 96 kHz), 6(192 kHz)	5.1	8(48 kHz, 96 kHz), 6(192 kHz)
bits/sample	16, 20, 24	16-24	16-24	16-24	16, 20, 24	16-24
Sampling frequency	48 kHz, 96 kHz, 192 kHz	48 kHz	48 kHz	48 kHz, 96 kHz, 192 kHz	48 kHz	48 kHz, 96 kHz, 192 kHz

Whilst 7.1 channel sound is available in Dolby Digital Plus and DTS-HD, several channel mappings are proposed in terms of 7.1 channel sound as shown in Fig. 24. The proposed mappings consist of two layers of loudspeaker positions, middle and top layer. The middle layer is basically at the same height with the listener's ear and the top layer is at a higher position such as at ceiling level.

FIGURE 24

Examples of loudspeaker mapping of 7.1 channel sound



Report BS.2159-24

6 Multichannel sound programme production in studio for home audio

6.1 Production of 5.1, 6.1 and 7.1 channels

Many countries are currently producing 5.1 channel sound programmes for broadcasting and audio and video releases. Production of 6.1 channel and 7.1 channel sound programmes is also increasing for audio and video releases. Several microphone techniques had been already proposed by many sound engineers and audio researchers for 5.1 channel sound recording. As described above, 7.1 channel sound is functional with the loudspeakers at a higher position. Several issues regarding how to use height channel properly or effectively were discussed in various workshops.

6.2 Production of 22.2 multichannel sound

6.2.1 Principles of three-dimensional sound mixing

NHK has already produced several UHDTV programmes with 3D sound using the 22.2 multichannel sound mixing system. Sound engineers and designers have been developing know-how and experience in the 3D sound field. The current, conventional applications of layers on 22.2 multichannel sound used for mixing are enumerated below.

Top layer

– Reverberation and ambience.

– Sound localised above, such as loudspeakers hung in gymnasiums and airplanes and at fireworks shows.

– Unusual sound, such as meaningless sound.

Middle layer

– Basic sound field formation.

– Envelopment reproduction.

Bottom 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 have also been discussing several issues in 3D sound mixing. The principal issues are as follows.

– Effective use of the top and bottom layers.

– 3D movement of sound images.

– Creating a sense of elevation.

– Interaction between immersive audio and visual cues.

6.2.2 22.2 multichannel sound post-production system

A 22.2 multichannel sound post-production system has been developed for producing 3D sound. The system has been installed in production rooms as shown in Fig. 25. The production system currently has the following features:

– Digital audio workstation including plug-in of sound effects.

– Sound mixing console with 3D pan on each channel strip as shown in Fig. 26.

– 3D reverberator with 22 directional impulse responses.

– Loudness meter specified in Recommendation ITU-R BS.1770-4.

– 3D audio signal compressor on 24-channel master bus.
– Down-mixing function to produce stereo and 5.1 multichannel sound simultaneously.
– Applicable to over 1 000 sound tracks.

FIGURE 25

22.2 multichannel sound production room

FIGURE 26

22.2 multichannel sound mixing console

Report BS.2159-25

6.2.3 Examples of live mixing of three-dimensional sound live

6.2.3.1 A large-scale musical TV programme at NHK Hall

The 22.2 multichannel sound of a large-scale musical TV programme at NHK Hall was live mixed using about 150 sound input feeds. Multiple microphones were arranged in the manner of standard pop recording, basically as a “close setup near the sound sources”, so the 22.2 multichannel sound mixing was also done with the conventional pop music mixing technique, i.e. the multi-microphone recording technique. The major difference in microphone arrangement and mixing between 5.1 channel sound and 22.2 multichannel sound is in how the ambience of a concert hall is recorded and mixed. It is important to reproduce the acoustical impression of the huge dimensions of the NHK

Hall, which has a 4 000-seat capacity and the impression of being surrounded by an enormous audience. Spatial sound reproduction advantages of the 22.2 multichannel system include the improvement of the listener's sense of envelopment and the enlargement of the listening area with exceptional sound quality. For the achievement of these new features of spatial reproduction with a 22.2 multichannel sound system, the following concept was planned as shown in Fig. 27.

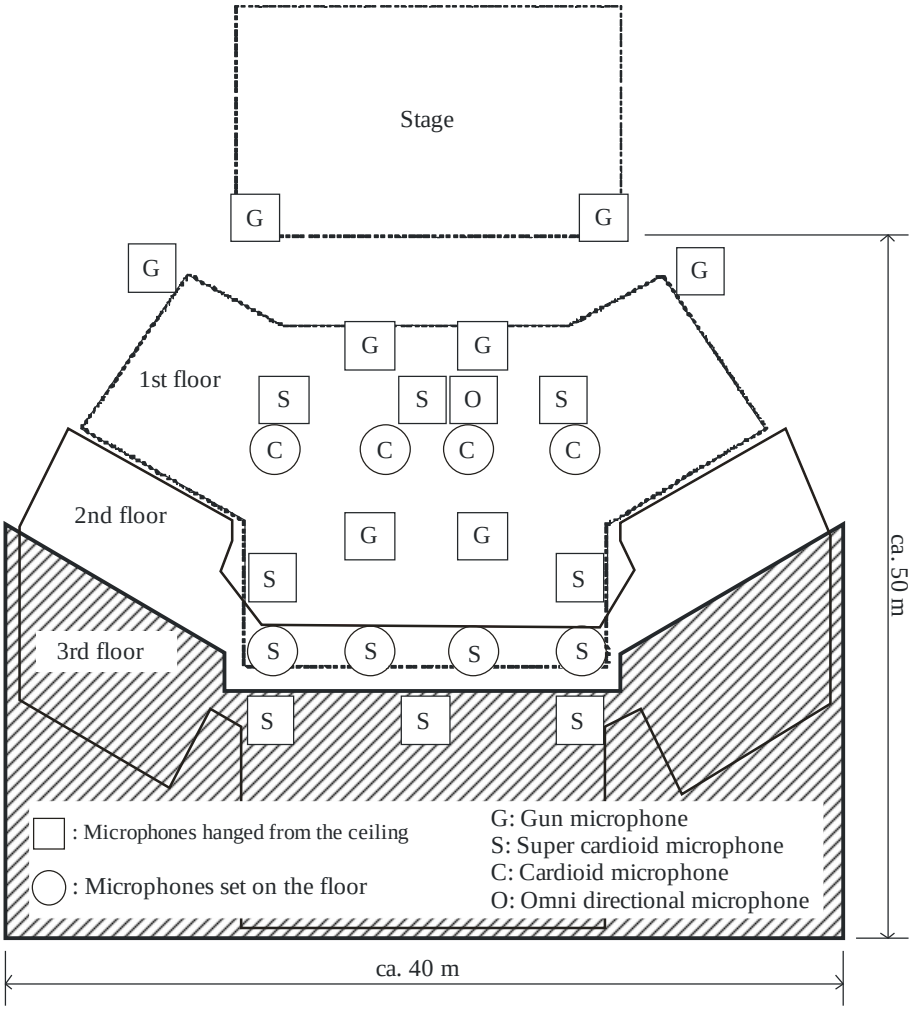
– Reflection and reverberation in the auditorium of NHK Hall, which are captured by microphones hung from the ceiling, are reproduced by the top layer loudspeakers of the 22.2 multichannel sound system to widen the listening area and create a sense of the listener being enveloped.

– The sounds of the audience, such as applause and shouts of encouragement, which are captured by several microphones set close to the audience, are reproduced by the middle layer loudspeakers to give the viewers a good sense of presence, as if they were sitting in the audience in NHK Hall.

– As the sound of musical instruments and vocals are reproduced by the sound reinforcement (SR) loudspeaker system in NHK Hall, reproduced sound reflected by the wall, ceiling, and floor of the hall is captured by the ambience microphones and reproduced by the top and middle layer loudspeakers to give the viewers the same sense of presence.

FIGURE 27

Arrangement of microphone for live mixing of a large-scale musical UHDTV programme by the 22.2 multichannel sound at NHK Hall



Report BS.2159-26

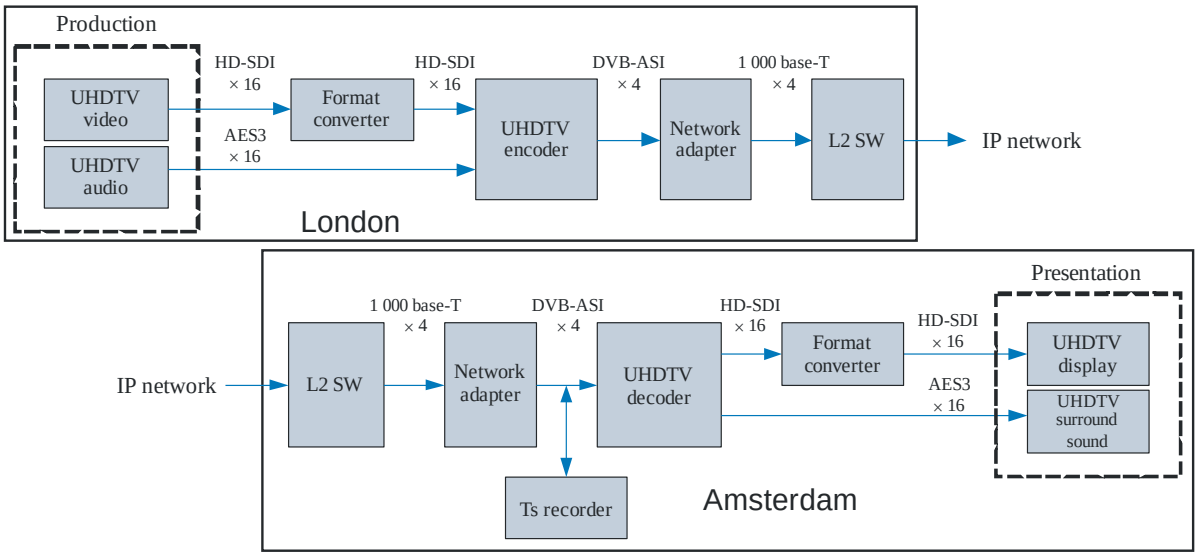
6.2.3.2 Emulated live news reports demonstrated at IBC2008

Ultra high definition television (UHDTV) and 22.2 multichannel sound technologies were demonstrated at IBC2008 by the international collaborative group called the Broadcast Technology Futures group (BTF), which included international live contribution link over an ultra-broadband IP network. The outline is depicted in Fig. 28.

UHDTV live pictures and sound captured in central London were carried to Amsterdam over an ultra-broadband IP network. In order to demonstrate the live nature of the link, the scenario set up was to emulate live news reports from London to Amsterdam with two-way interaction between a reporter in London and a presenter in the theatre in Amsterdam.

FIGURE 28

UHDTV and 22.2 multichannel sound IP transmission system at IBC2008



Report BS.2159-27

Sound acquisition system adopted in London was a microphone array with 15.2 system rather than a full-blown 22.2 system due to limitation on number of channels in mixing desk. This meant that there would be a middle layer containing eight of the ten specified 22.2 microphone complement, the top layer would be reduced to four microphones from nine, and the lower layer would have the full complement of five microphones, of which two were LFE channels. The microphone array is shown in Fig. 29. The total of 18 audio channels, including the 15.2 channels and one commentary channel, were sent to Amsterdam to be reproduced for the 22.2 multichannel system in the viewing theatre there. The 3D surround sound quality reproduced by 22.2 multichannel sound in Amsterdam was completely convincing; ambient sounds of London were reproduced effectively, even the sounds of airplanes and helicopters flying overhead sounded as if they were flying over the theatre.

FIGURE 29

Microphone array

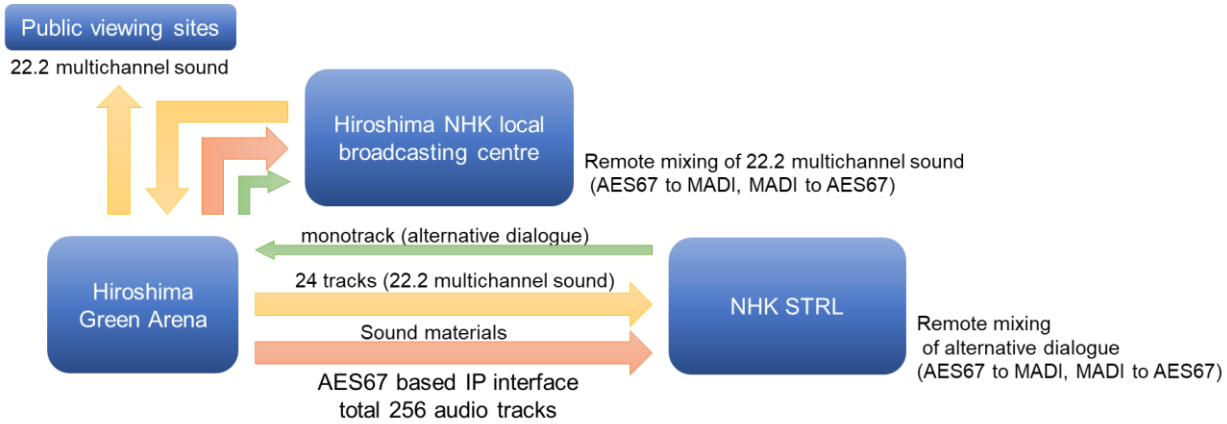
Report BS.2159-28

6.2.3.3 Remote production using IP interfaces

A remote production has been conducted for live production of the ISU Grand Prix of Figure Skating 2018/2019 NHK Trophy, as shown in Fig. 30. A total of 256 audio tracks including 22.2 multichannel sound were transported over AES67-based IP interfaces that can convey up to 64 audio channels converted from a MADI (multichannel audio digital interface) in a single stream. Recording was operated in the arena and all the recorded sound and stereo mix were transported to a local broadcasting centre and NHK STRL (Science and Technology Research Laboratories). A 22.2 multichannel sound mix was produced at the local broadcasting centre and an alternative dialogue for multilingual services was produced using stereo sound down mixed from 22.2 multichannel sound at NHK STRL. The total delay when sending an alternative dialogue from STRL to the local broadcasting centre was 21.2 ms when 24-bit audio signal with 48 kHz sampling and a packet time of 1 ms was used.

FIGURE 30

Remote production of 22.2 multichannel sound



6.3 Production of 10.2 multichannel sound (Type A)

Programme material has been originally recorded for the 10.2 multichannel sound format, and some has been repurposed from other multichannel material. What is standardized is the playback platform including environment. No standardization of recording technique is required or desirable. A range of methods of recording were used, including adding microphones to more conventional 5.1-channel recording, layouts of microphones that mimic loudspeaker locations, and more complex pop style mixing wherein a large number of source microphone channels, up to 48 in several cases, are remixed to the 10.2 loudspeaker format.

There are more than 20 items of produced programme material. Since 10.2-channel sound is a playback platform, not a recording/playback system, a wide variety of methods of recording have been employed, from classic ones, to completely constructed spaces using advanced digital signal processing algorithms.

6.4 Object-based post-production system

6.4.1 Object-based sound system with ADM metadata

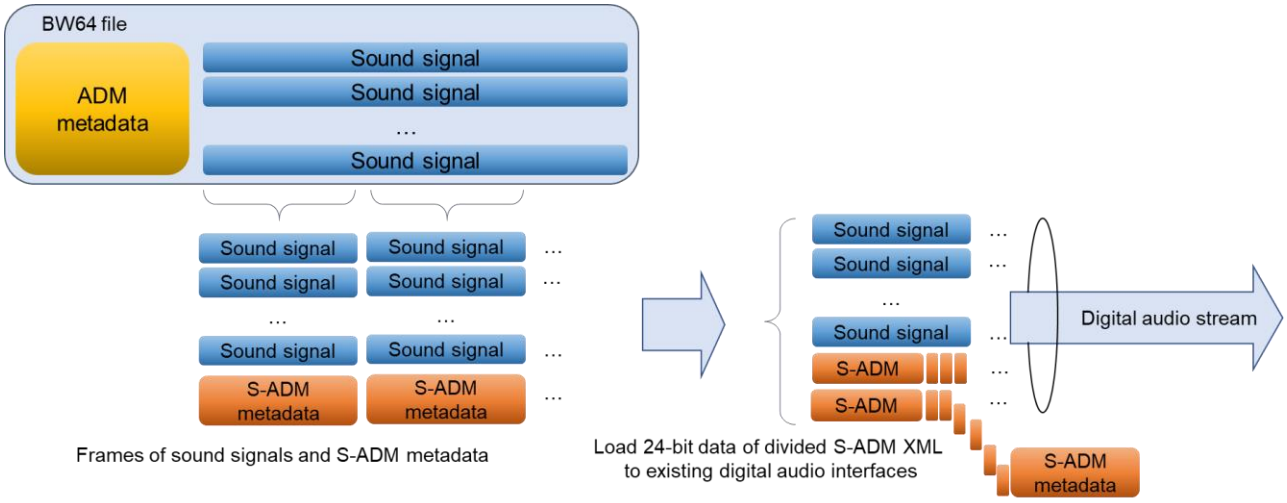
6.4.1.1 Interface to transport serial ADM

Object-based sound systems require audio-related metadata including gains and positions of audio tracks and a group of audio tracks to be reproduced simultaneously. The Audio Definition Model (ADM) and its serialised presentation (S-ADM) have been standardised for a production metadata set as Recommendations ITU-R BS.2076 and ITU-R BS.2125, respectively.

NHK has developed production tools of object-based sound using ADM and S-ADM metadata to study the specifications of next generation audio services. ADM metadata is stored in the axml chunk of a BW64 file and an S-ADM stream is transported by MADI (AES3-based multichannel audio digital interfaces). An authoring tool generates an S-ADM stream from an ADM XML code. S-ADM frames are synchronised with associated audio frames over MADI as shown in Fig. 31. The authoring tool can add S-ADM metadata to an audio signal stream when the tool receives an audio signal over MADI as output from a mixing console. The tool is applicable to live production with only static metadata even when an existing channel-based mixing console is used.

FIGURE 31

Transport of Serial ADM (S-ADM) over audio digital interface



6.4.1.2 Live mixing console supporting Serial ADM (S-ADM)

6.4.1.2.1 Specifications of the console

NHK has developed a live mixing console that supports S-ADM, as shown in Fig. 32. The console supports the creation of audio objects using up to 22.2 multichannel audio (sound system H specified in Recommendation ITU-R BS.2051) and S-ADM, and the transmission of the S-ADM stream. It can output a stream of advanced audio content including up to 60 audio tracks for audio objects such as dialogue and ambient sound, and up to four audio tracks for S-ADM metadata via the serial multichannel audio digital interface known as MADI specified in Recommendation ITU-R BS.1873. The console generates S-ADM metadata based on prepared ADM metadata by itself and enables broadcasters to produce immersive and personalised audio contents even in a live production.

FIGURE 32

Live mixing console supporting S-ADM

TABLE 5

Specifications of live mixing console

Sampling rate (kHz)	48
Bit depth (bit)	24
Mixing engine	TAMURA NT900c
Audio signal inputs (channel)	Up to 192
Audio metadata input format	ADM
Audio signal outputs for transmission (channel)	Up to 64 including S-ADM via MADI
Audio signal outputs for monitoring (channel)	Up to 24 (22.2 multichannel audio supported)
Audio metadata signal output format	S-ADM
S-ADM transmission frequency (fps (1/ms))	20 (1/50), 50 (1/20), 59.94 (1/16.68), 60 (1/16.67)
Number of S-ADM transmission tracks	1 (MADI Ch16 or Ch64), 2 (MADI Ch63–64), 4(MADI Ch61–64)
Coding format of S-ADM	gzip
Largest format of audio object	22.2 multichannel audio
Number of audio objects	60 in mono-channel equivalent

Figure 33 shows a diagram of the console. Traditional audio consoles output final mixed channel-based audio signals of stereo, the 3/2 multichannel sound system or so on. The console, on the other hand, does not output final mixed audio signals but various audio objects of diverse formats. The mixing engine of the console mixes various input audio signals such as microphones and audio from recorders into some audio group buses[4] such as dialogue and ambience groups, and outputs audio signals from group buses as audio objects.

Input ADM metadata is converted into various frames of S-ADM and inserted into audio matrices in the form of audio signals. Various audio objects and S-ADM are arranged on audio tracks in order and are output via one MADI as the main output of the console. For monitoring, the main output is distributed and input into the S-ADM renderer. The rendering algorithm complies with Recommendation ITU-R BS.2127.

Figure 34 shows the user interface of the S-ADM renderer. The user can set the configurations of input and output signals in the left panel. “S-ADM metadata track” sets audio channels of MADI that convey the S-ADM stream. “Speaker layout” indicates the reproduction sound system, such as sound systems A (stereo, 0+2+0), B (5.1, 0+5+0) and so on. “Output Matrix” specifies the relationship between audio channels of MADI and loudspeaker labels. The user can also control user interaction parameters in the right panel. Combinations of audioProgramme, audioObject and audioComplementaryObject are freely selected by clicking texts on the user interface. If the metadata allows, the user can change playback levels of audioObjects. “gain” and “mute” are related to the “gainInteract” and “onOffInteract” elements of audioObject, respectively.

[4] An audio bus is a signal path that can combine multiple audio signal paths.

FIGURE 33

Diagram of live mixing console

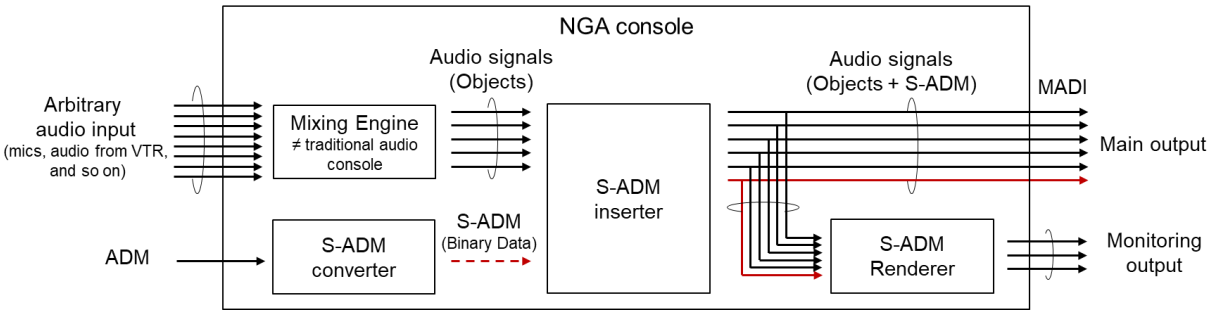
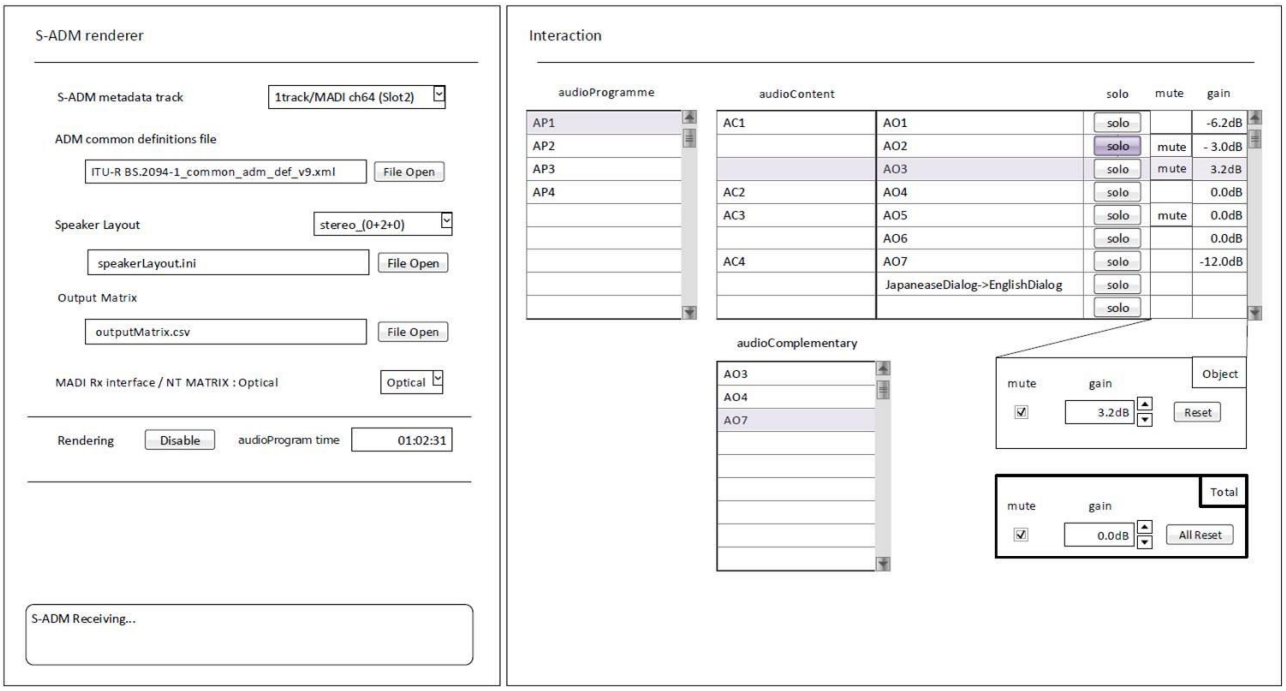


FIGURE 34

User interface of S-ADM renderer



6.4.1.2.2 Connection to MPEG-H 3DA encoder

An audio encoder and decoder pair complying with MPEG-H 3DA baseline profile level 4 was developed and connected to the console for verification. Figures 35 and 36 show the MPEG-H 3DA encoder and decoder, and an overview of the verification system, respectively.

FIGURE 35

MPEG-H 3DA encoder and decoder supporting baseline profile level 4

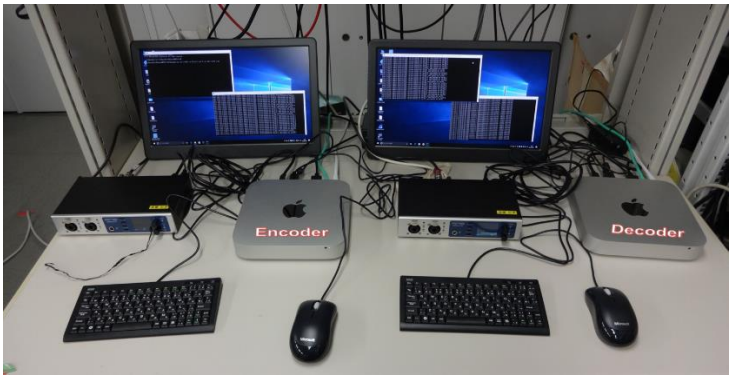
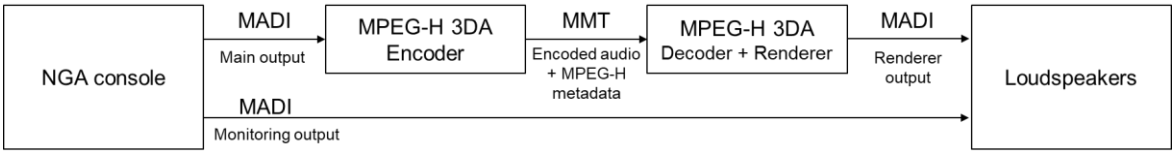


FIGURE 36

Verification system



Test contents using S-ADM were used in this verification test. The functionalities described below were confirmed by monitoring the S-ADM renderer in the console and the MPEG-H 3DA renderer in the MPEG-H 3DA decoder.

Functionality for immersive audio:

– Reproduction by different loudspeaker layouts: format conversion from 22.2 multichannel audio (sound system H) to 7.1.4 (sound system J), 5.1.4 (sound system D), 5.1 (sound system B) and stereo (sound system A).

Functionalities for personalisation:

– Switching objects of object-based audio for multilingual services: switching dialogue objects with those of other languages.

– Switching objects of channel-based audio for Home & Away in sports programmes: switching not only dialogue objects but also background sound objects. Audiences can feel they are together with supporters, cheering on their own team, the home team or the away team.

– Adjusting playback levels of individual objects for dialogue enhancement: changing the level balance between dialogue objects and background sound objects.

6.4.1.3 Application of Serial ADM metadata

The S-ADM can be used to transport ADM metadata to an ADM-originated renderer, an encoder for emission and a converter for IBB systems in production environments as shown in Fig. 37.

In production and monitoring

The authoring tool can transport up to 124 audio signals and S-ADM metadata. The size of the metadata is less than four audio tracks in a frame. The monitoring renderer supports 128 input audio tracks including the S-ADM metadata field and 32 output audio tracks, and can reconstruct the original ADM from the S-ADM stream.

In emission

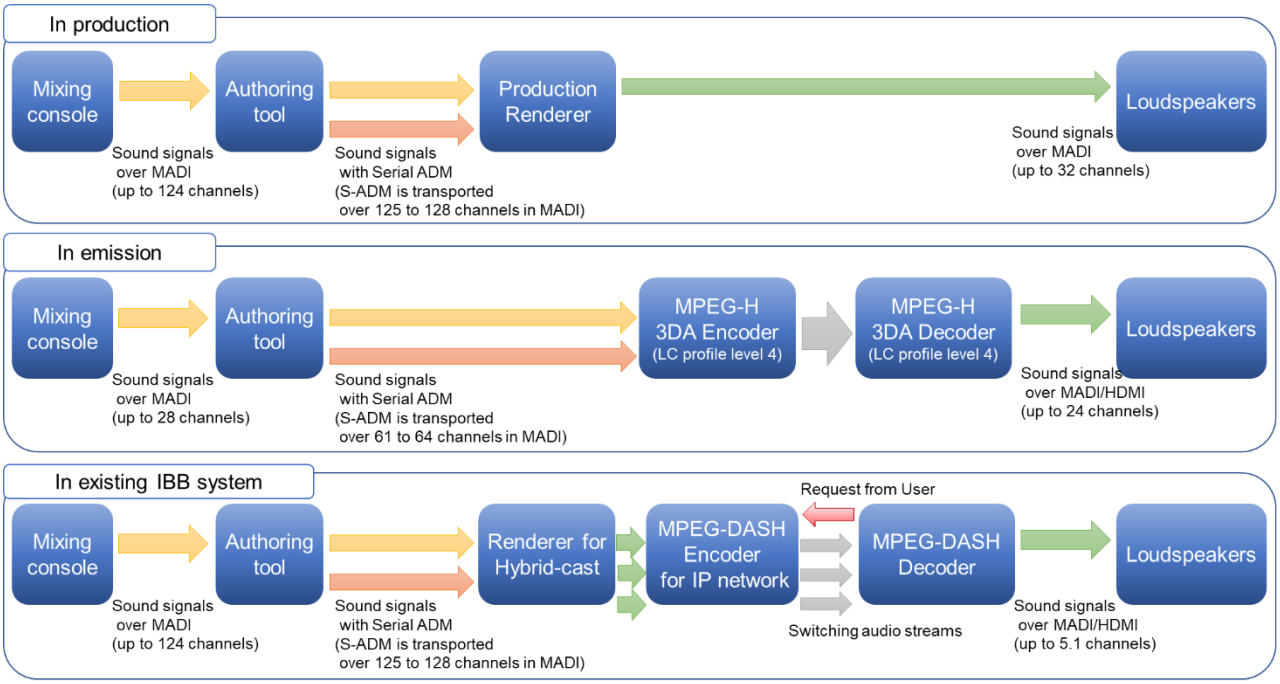
The authoring tool can be connected to a real-time encoder of MEPG-H 3DA LC profile level 4. The system can support up to 28 input audio tracks, that is, 22.2 multichannel sound with four dialogues and S-ADM. The system has been implemented experimentally to study the specifications of next generation terrestrial broadcast services.

In IBB systems

A live audio programme may be produced using S-ADM with different service parameters (e.g. multilingual, dialogue enhancement, different mixing balances, different loudspeaker layouts), from which a user can select his/her preferred one on the receiver. An IBB system may be used to deliver such a programme using MPEG-DASH.

FIGURE 37

Application of Serial ADM (S-ADM)



6.4.1.4 Example of object-based audio programme with static metadata

Object-based sound systems can adapt to listening environments including the loudspeaker layout and listener's preferences by using multiple stems, for instance, multilingual dialogues, multi-viewing background sound and sound effects. An object-based test programme was produced to verify the behaviour of the authoring tool and the description method of ADM metadata.

A total of 28 programmes of a baseball game were created using 13 objects, including seven types of dialogue, three types of background sound (neutral, home and away) and three types of sound effect, and 121 audio tracks in a BW64 file as shown in Table 6. The dialogue object is a single audio track and a type of 'matrix' from a single track to the front two, three or four channels. The background sound and sound effects are 22.2 or 11.1 multichannel sound and a type of 'DirectSpeakers'. The number of programmes and objects depends on the variation of the dialogue, the background sound and the sound effects but the number of audio tracks depends on not only the variation of the content but also the channel configurations for audio objects. The profile and level of ADM metadata should be defined by a combination of both the variation of the programme and the loudspeaker layout.

TABLE 6

Example of object-based sound programme of a baseball game

Tracks	Description	Objects	Format	Programmes
0001	Neutral narration in Japanese (mono)	AO_1001	matrix (1)	APR_1001, APR_1011
		AO_1011	matrix (2)	APR_1021, APR_1031
0002	Neutral narration in English (mono)	AO_1002	matrix (1)	APR_1002, APR_1012
		AO_1012	matrix (2)	APR_1021, APR_1031
0003	Home-side narration in Japanese (mono)	AO_1003	matrix (1)	APR_1003, APR_1013
		AO_1013	matrix (2)	APR_1021, APR_1031
0004	Visitor-side narration in Japanese (mono)	AO_1004	matrix (1)	APR_1004, APR_1014
		AO_1014	matrix (2)	APR_1021, APR_1031
0005	Comedy-like narration in Japanese (mono)	AO_1005	matrix (1)	APR_1005, APR_1015
		AO_1015	matrix (2)	APR_1025, APR_1035
0006	Professional description (mono)	AO_1006	matrix (1)	APR_1006, APR_1015
		AO_1016	matrix (2)	APR_1026, APR_1035
0007	Easy description (mono)	AO_1007	matrix (1)	APR_1007, APR_1017
		AO_1017	matrix (2)	APR_1027, APR_1037
0008 to 0031	Neutral background sound (22.2)	AO_1101	22.2 (4)	APR_1001, APR_1002, APR_1005, APR_1006, APR_1007 APR_1011, APR_1012, APR_1015, APR_1016, APR_1017
		AO_1201	matrix (3)	APR_1021, APR_1022, APR_1025, APR_1026, APR_1027 APR_1031, APR_1032, APR_1035, APR_1036, APR_1037
0032 to 0055	Home-side background sound	AO_1103	22.2 (4)	APR_1003, APR_1013
		AO_1203	matrix (3)	APR_1024, APR_1034
0056 to 0079	Visitor-side background sound	AO_1104	22.2 (4)	APR_1004, APR_1014
		AO_1204	matrix (3)	APR_1024, APR_1034
0080 to 0099	Sound effects (video game)	AO_1108	22.0 (5)	APR_1011, APR_1012, APR_1013, APR_1014, APR_1015, APR_1016, APR_1017
		AO_1208	matrix (3)	APR_1021, APR_1022, APR_1023, APR_1024, APR_1025, APR_1026, APR_1027 APR_1031, APR_1032, APR_1033, APR_1034, APR_1035, APR_1036, APR_1037

TABLE 6 (*end*)

Tracks	Description	Objects	Format	Programmes
0100 to 0110	Sound effects (with friends)	AO_1109	11.0 (6)	APR_1021, APR_1022, APR_1023, APR_1024, APR_1025, APR_1026, APR_1027
0111 to 0121	Sound effects (with audience of PV)	AO_110a	11.0 (6)	APR_1031, APR_1032, APR_1033, APR_1034, APR_1035, APR_1036, APR_1037

(1) Matrix from mono to front four loudspeakers (M+000, M+060, M-060, U+000).

(2) Matrix from mono to front three loudspeakers (M+000, M+060, M-060).

(3) Matrix from 22.2 to front two loudspeakers (M+060, M-060).

(4) Sound system H specified in Recommendation ITU-R BS.2051.

(5) Sound system H specified in Recommendation ITU-R BS.2051 without two LFEs.

(6) Sound system J specified in Recommendation ITU-R BS.2051 without LFE.

6.4.1.5 Example of object-based audio programme with dynamic metadata

6.4.1.5.1 Overview of content

NHK and Fuji Television Network produced an experimental music video with object-based audio programme. The goal of this work was to gather experiences to produce highly immersive audio content and to establish a workflow.

The content is originally composed music with the following characteristics:

– Interactive audio contents with fixed solo part and four selectable accompanied audio object groups;

– Combination of music instruments at fixed positions and dynamically moving sound effects;

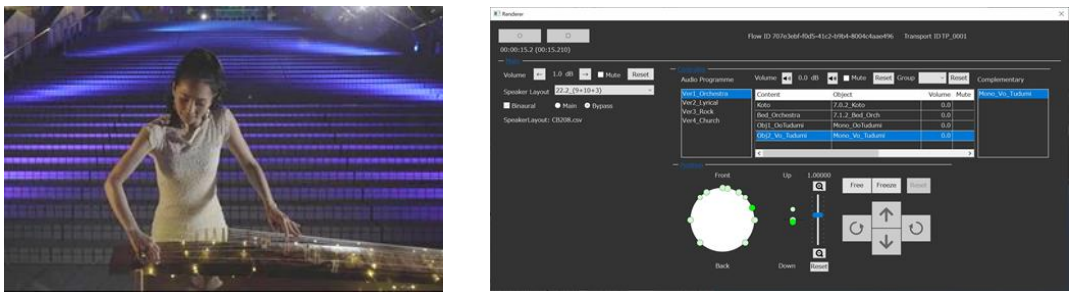
– Playback environment compliant with 7.1.4 multichannel audio (sound system J in Recommendation ITU-R BS.2051);

– 56 audio tracks, which is the maximum number for baseline profile level 4 of MPEG-H 3DA and playable by the ITU-R ADM renderer specified in Recommendation ITU-R BS.2127.

Figure 37 shows a scene of the content and the user interface for the renderer.

FIGURE 37

Scene of music video and user interface for renderer

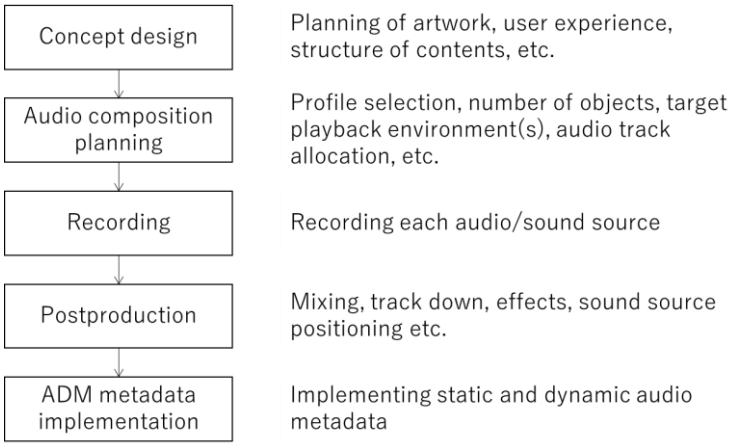


6.4.1.5.2 Workflow and production

The production of object-based audio content needs to be thoughtfully planned right down the line for well-performed implementation. Figure 38 shows the workflow of this content creation.

FIGURE 38

Workflow of object-based audio programme production



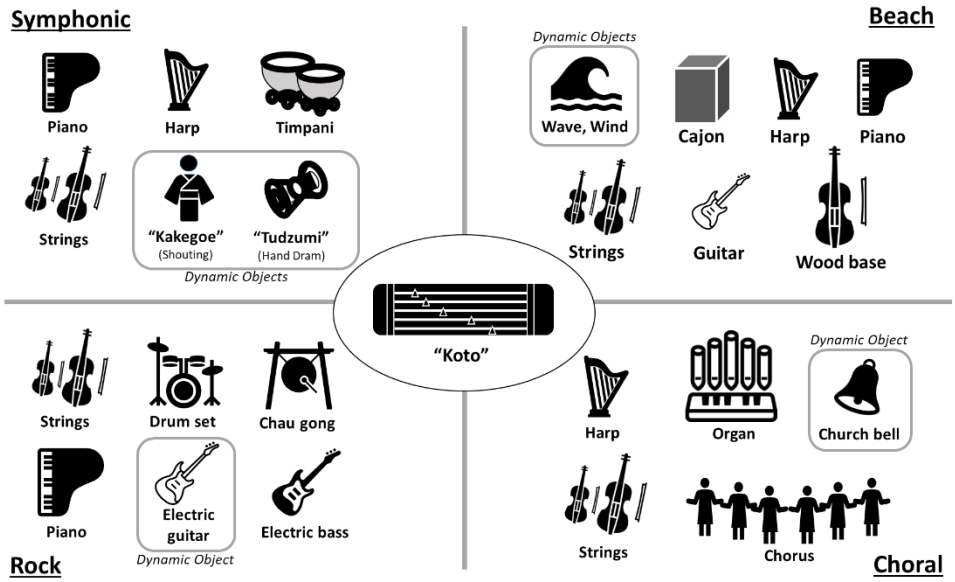
6.4.1.5.2.1 Concept design

The purpose of this content was to invent a new experience of music video in which a cantus with a koto, a Japanese traditional harp, leads four types of accompaniment, with sound effects corresponding to different ambiences: symphonic, beach, rock and choral. Figure 39 shows the variations of accompaniment and composition of the musical instruments. Some objects are dynamically moved for enhanced immersion.

Users can switch the accompaniments and move dynamic objects arbitrarily.

FIGURE 39

Variations of accompaniment and composition of the musical instruments



6.4.1.5.2.2 Audio composition planning

Before studio production, composition planning for object-based audio programme production is essential considering the constraint of the target codec profile possible in plural presentations combining multiple objects flexibly.

Baseline profile level 4 of MPEG-H 3DA was selected, in which 28 audio tracks can be used for simultaneous playback within the maximum number of tracks of 56. Table 7 shows the audio composition of this programme.

TABLE 7

Audio object composition and number of tracks

Programme variation	**Cantus**	**Audio objects of accompaniments and sound effects (tracks, format)**	**Subtotal audio tracks**
Symphonic	Koto (9, channel-based signal: 7.0.2)	Symphonic accompaniments (10, 7.1.2) **Hand drum** (1, mono object) **'Kakegoe' (shouting)** (1, mono object)	21 (9+10+1+1)
Beach		Band set (10, 7.1.2) **Sound effect 1(wave)** (1, mono object) **Sound effect 2 (wind)** (1, mono object)	21 (9+10+1+1)
Rock		Band set (10, 7.1.2) **Electric guitar** (1, mono object)	20 (9+10+1)
Choral		Accompaniments (10, 7.1.2) Surround supplement (2, two-channel object for Left/Right top back)	21 (9+10+2)
Total audio tracks			56

Note: Boldface indicates tracks of the dynamic audio object

6.4.1.5.2.3 Recording and postproduction

The recording of object-based audio programmes is not much different from that of channel-based audio programmes. Each instrument in the experimental programme was recorded individually. The cantus with the koto was recorded in a surround format with multiple microphones to express the sense of presence and rich ambience. Accompaniment instruments were recorded to address the specific sound source position to the audience. Figure 40 shows the recording site.

FIGURE 40

Recording setup of koto


Postproduction requires two steps to finish the objects as follows:

– General procedures for surround audio post-production (mixing, panning, dubbing, acoustic effects);

– Allocation and adjustment of objects to match the 3D virtual space and record the position data of dynamically moving audio objects.

Since the textures of listening experiences are different among the different combinations of objects, quality control is part of our future study.

6.4.1.5.2.4 ADM metadata implementation

Completed audio packages need to be implemented with audio-related metadata such as ADM metadata. Static ADM metadata for channel-based objects and content information were manually written. On the other hand, dynamic metadata for moving audio objects was created by Pro Tools (Digital Audio Workstation, DAW) and the MPEG-H Authoring plug-in (plug-in software for Pro Tools) supporting ADM.

Owing to the limited number of tracks on the bus of the DAW, it was not immediately possible to generate completed ADM for the programme that used the 56 tracks in Table 7. The plug-in generates position metadata in each of 1024 samples, which are not synchronised with the audio frame in S-ADM. ADM metadata implementation had the following three steps:

– Generate and output ADM including dynamic metadata using the DAW and plug-in for the four variations of the audio programme listed in Table 7;

– Manually merge the four sets of ADM into one ADM and integrate multiple audio blocks to a single audio block;

– Regenerate the S-ADM stream from the merged ADM using the authoring tool described in § 6.4.1.1.

Future developments will include expansion of the number of tracks for the bus and I/O of DAWs and plug-ins, and conversion tools for ADM including dynamic metadata.

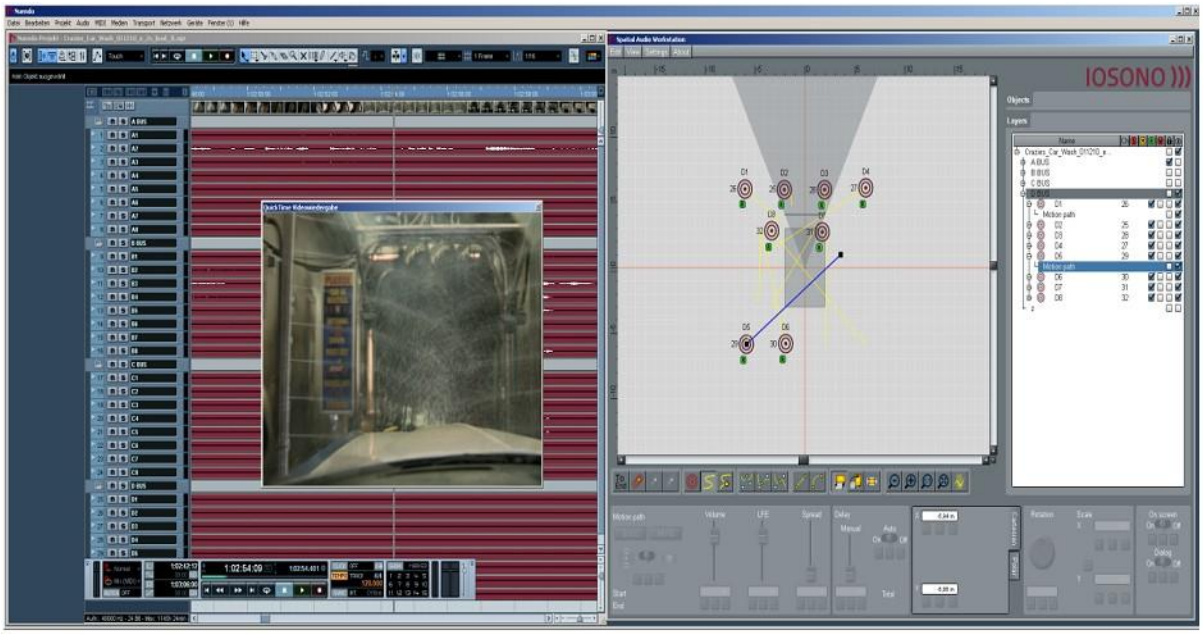
6.4.2 Object-based post-production system with individual audio object

The creation of an object-based sound scene involves associating spatial information with the sound signals comprising the scene. The IOSONO Spatial Audio Workstation (SAW) is a tool for object-oriented production, editing and mastering of auditory scenes for reproduction in different environments. This plug-in for a digital audio workstation enables the direct monitoring of the object-based scene in all multichannel layouts. In combination with external rendering, the production can be performed directly in a flexible speaker layout or mixing stage. It is currently realized as a plug-in

for the digital audio workstation Steinberg Nuendo. While Nuendo enables the editing and post production of audio streams, the SAW plug-in enables the sound engineer to create advanced sound source movements and complex audio scenes based on the audio material loaded into a Nuendo session (Fig. 41).

FIGURE 41

Spatial audio workstation used for a motion picture sound track production



Report BS.2159-29

Using the SAW, sound objects are positioned on the scene like marking points on a map. In addition, the SAW allows the definition of motion trajectories for the sound objects. The mixer can assign a discrete position to each sound object for its x, y and z coordinates. For moving sound objects, the position information is accompanied by a timestamp (SMPTE time code). The user has full control over the sound objects and the motion lines. Moreover the plug-in offers a wide range of functions for sound objects and motion lines, e.g. move, rotate, scale and group. The SAW is equipped with a graphical user interface that allows the mixer to easily assign a discrete position to each sound object in the listeners' space. This gives the sound engineer an intuitive view compared to traditional channel-oriented loudspeaker panning techniques. With this tool, even live mixing is possible.

The output of the object-based production tool can be directly feed to any multichannel formats without additional processing. Sound engineers can switch the output format whenever they like without changing anything in the production. Combined with an external rendering, any reproduction system, including wave-field synthesis, can be used with the same content file.

FIGURE 42

Spatial audio workstation and WFS speaker array installed at Todd-AO Stage 2 used by Rick Kline

Report BS.2159-30

Several productions have been performed in Todd-AO. Besides a number of trailers and demos, a complete motion picture sound-track has been produced using the spatial audio workstation (Fig. 39).

6.5 Production of cinematic hybrid content

High quality content authoring is getting increasingly complex, time-consuming and expensive as content creators strive to get more from surround sound. New mixing technology should enable new creative options, but it should also integrate into existing post production workflows without adding excessive time, therefore cost, to the process. The hybrid model of channels and objects allows most sound design, editing, pre-mixing, and final mixing to be performed in the same manner as they are today.

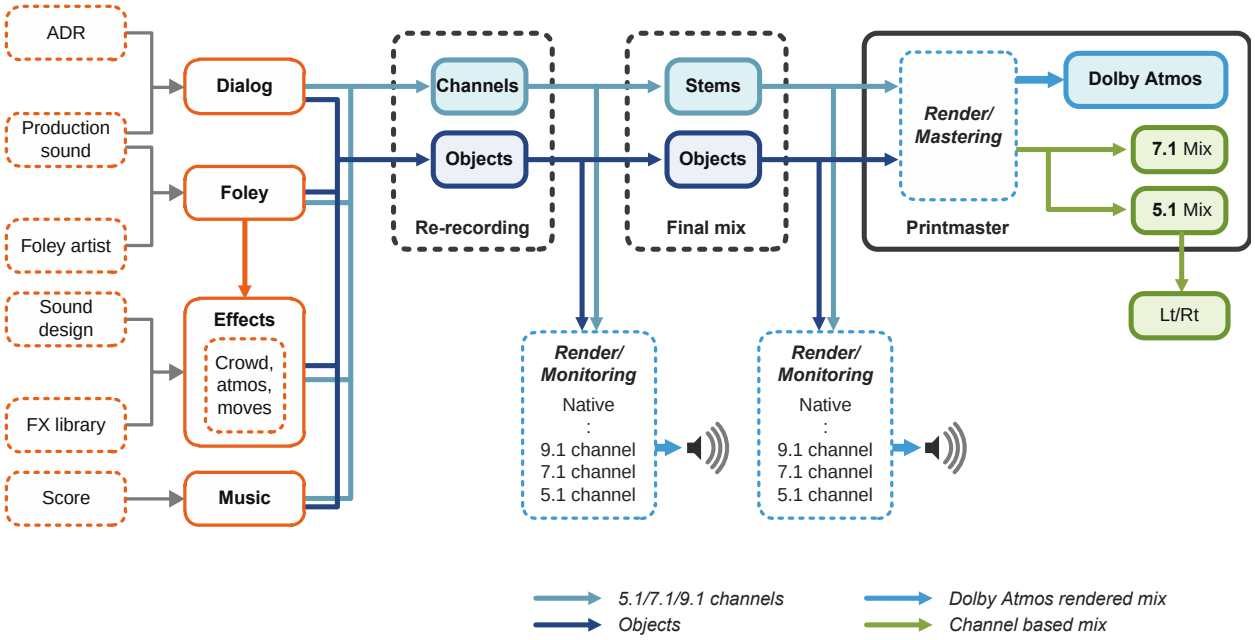
Plug-in applications for digital audio workstations allow existing panning techniques within sound design and editing to remain unchanged. In this way, it is possible to lay down both channels and objects within the workstation in 5.1-equipped editing rooms.

Object audio and metadata is recorded in the session in preparation for the pre- and final-mix stages in the dubbing theatre.

Metadata is integrated into the dubbing theatre's console surface, allowing the channel strips' faders, panning and audio processing to work with channels, channel sets ("stems") and audio objects. The metadata can be edited using either the console surface or the workstation user interface, and the sound is monitored using a reference rendering and mastering.

FIGURE 43

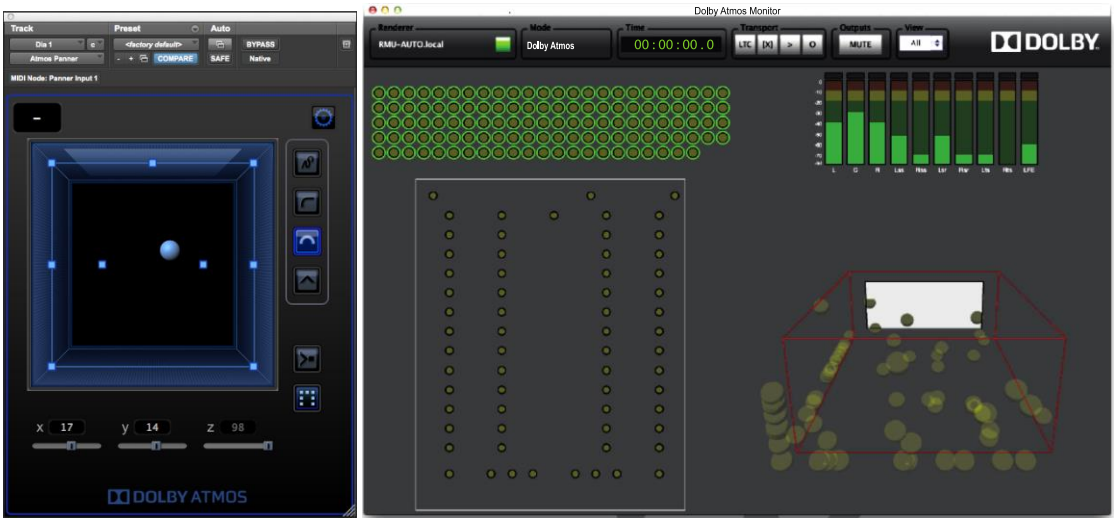
Authoring workflow, showing combination of channels and objects



A specific panning tool has been developed to allow the mixers to freely move sound sources in 3D space while taking advantage of all the speakers present in the environment. Figure 44 illustrates the panner plugin UI for Protools™. The interface is similar to common panning tools existing today but extends the creative palette of the mixer by introducing elevation controls for the sound objects. Sound objects can therefore be freely moved in the 3D space of the room and either panned between several loudspeakers or snapped to a single speaker closest to the intended location. In addition to the object location, the perceived width of the object can also be controlled. Several elevation constraints can also be used so that the objects elevation is automatically adjusted as they cross the room using different profiles (sphere, wedge, etc.). Finally, zone exclusion controls (e.g. no side wall, no back wall, etc.) can be enabled allowing the mixer to finely control the set of speakers involved in rendering a particular pan.

FIGURE 44

(a) Panner plugin UI for Protools™. (b) The monitoring application



During authoring, the state of the entire mix can be monitored using a separate application (Fig. 44b). This representation shows object activity status and input levels and 9.1 output metering information. The representation also includes a 3D spatial display of the room and object positions, as well as a speaker view showing current loudspeaker configuration and output levels.

The channel and object audio data and associated metadata is recorded during the mastering session to create a final master, which includes a hybrid mix and any other rendered deliverables (such as a 7.1 or 5.1 mix). This final master file is wrapped using industry-standard MXF wrapping procedures, hashed and optionally encrypted in order to ensure integrity of the audio content.

As of 3 October 2013, 35 post-production facilities around the world are equipped with mixing/encoding equipment for this hybrid format and 75 titles have been mixed in the format. More than 200 commercial theatres worldwide are equipped to playback this format.

6.6 3D Virtual Microphone Systems (VMS)

In an effort to develop ways of achieving an enveloping sound experience, the Rai Research Centre implemented two experimental Virtual Microphone Systems (VMS), one of them based on a commercial spherical 3D VMS probe, the other based on a planar 2D array probe. Both microphone systems are based on the Ambisonic theory, fall in the category of Object-Oriented Audio Systems and provide many valuable benefits in the production of audio programmes.

6.6.1 The spherical microphonic probe

The commercial spherical microphonic probe consists of a sphere of 8.4 cm in diameter; with 32 microphonic capsules positioned on its surface; each capsule has a diameter of 14 mm.

The 32 capsules are used to analyse the molecular behaviour in the space around them; thus permitting to synthesize up to seven virtual microphones and to also “zoom” one of those microphones on any sound source.

The connection between the microphone probe and its digital audio interface only requires an Ethernet cable, category 5 or higher, which carried the power supply in addition to the audio signals from the capsules.

The digital audio interface signals are fed to a computer that performs the calculations required to process the information from the probe. The computer processing software allows the operator to synthesize up to seven virtual microphones, choosing their characteristics, their spatial positions and their directivity while maintaining control of the seventh microphone through his computer mouse. This can be done in real time through a specially designed man-machine interface and is facilitated by a displays that shows the image of the stage as shot by a service camera and also shows coloured circles indicating the positions of the virtual microphones on the stage (see Fig. 45).

FIGURE 45

Example of the use of a 3D VMS system in a theatre

The outputs of the virtual microphones can be fed directly from the operator's computer to an audio mixer or to a recording system.

6.6.2 Typical uses of the spherical microphonic probe

A typical use of the 3D VMS probe is in a theatre or a concert hall. After positioning the probe, the operator can choose the spatial positions and the directivity characteristics of up to seven virtual microphones. Furthermore, the operator can control the seventh virtual microphone during the event, e.g. to point it to a specific musical instrument or to follow a performer that moves on the stage.

Another possible use of a 3D VMS probe is during the shooting of a sporting event, e.g. in a football stadium, where the operator can obtain a surround sound environment and he may even be able to capture the impact of a kick on the football.

Similarly, it is possible to follow a cycling race with the seventh virtual microphone, obtaining a fade-up/fade-down effect of the cheering crowd that watches the race alongside the road.

Such spatial effects can also be used to advantage to increase the impression of "space and presence" experienced by the audience of a sound broadcast, allowing listeners to identify the position of the anchor-man, the performers and the musicians in a 360° imaginary space.

This possibility has been successfully tested by positioning a 3D VMS microphonic probe in the centre of the stage of Teatro Regio in Turin, Italy, for the sound coverage of some lyric operas. One such test concerned the live coverage of Gaetano Donizetti's "Lucia di Lammermoor", when all the performers on the stage were picked up by a single 3D VMS probe.

Other tests were performed in the Vatican during some events held in the Paolo VI conference hall and in St. Peter's Basilica.

6.6.3 Typical uses of the planar array probe

A further interesting application is related to the use of the 2D VMS Planar Array probe briefly mentioned in the introduction to this Report.

The planar array probe is intended to only pick up a frontal event; in other words, it is not possible to process the probe outputs to synthesize virtual microphones positioned behind the event.

This probe does not allow to obtain a 'surround' effect, but it allows to increase the directivity of its virtual microphones, from order 6 Ambisonic (which is typical of the 3D VMS probe) to order 10 Ambisonic.

Positioning two such probes, one in front of the scene and one behind the scene in a television studio it can be expected to obtain full audio coverage with 14 virtual microphones, even when the performers turn their shoulders to the front of the scene. This feature would likely allow to dispense from the use of radio microphones, which are known to often pose some problems in operation.

7 Quality performance of the multichannel sound systems

7.1 22.2 multichannel sound system

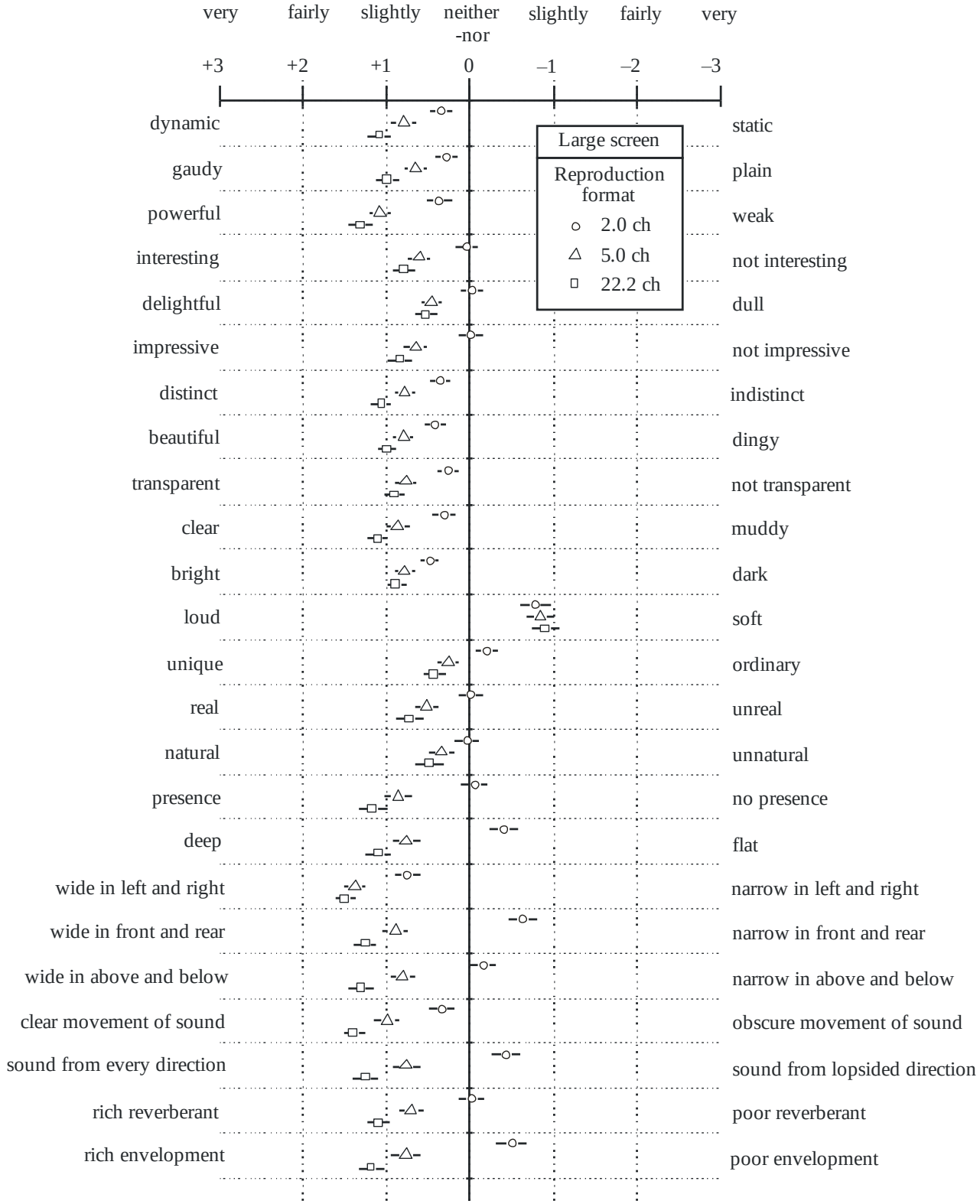
Subjective evaluations were conducted to assess the performance of three different multichannel audio systems: two-channel stereo, 5.1 channel sound, and 22.2 multichannel sound. The stimuli for the subjective evaluations were selected from the World Expo 2005 programmes. The sounds and video were screened at two locations: NHK Lab's UHDTV theatre (a 450-in. screen that is 14 m long, 15 m wide, and 10 m high) and NHK Lab's small post-production studio (a 50-in. screen that is 8 m long, 7.5 m wide, and 4 m high). The subjects were asked to report their impressions of the sound provided by the different sound systems when shown with different images on the screen. They were also asked to sit in different positions so that differences in impressions based on position with regard to the screen could be discussed.

The semantic difference evaluation method was used in this experiment. Subjects were asked to rate their impressions on a 7-point scale for the pairs of evaluation terms. Each pair contained two terms with opposite meanings, such as "dynamic" and "static." Subjects were asked to select a score from 3 to −3, in which 3 meant very dynamic, 2 meant fairly dynamic, 1 meant slightly dynamic, 0 meant neither dynamic nor static, −1 meant slightly static, −2 meant fairly static, and −3 meant very static. They rated each stimulus (segment of content) using the 24 evaluation pairs. There were 53 subjects (28 university students of music or audio engineering and 25 audio professionals) for this experiment.

Figures 43 and 44 show the total mean values of all the results from the 24 pairs of evaluation terms for each sound system in the large theatre and the small studio, respectively. Each mean value is marked with a 95% confidence interval for different terms and for different sound systems. The horizontal axis represents the scale of evaluation, and the vertical axis contains each pair of evaluation terms. Both figures show that the 22.2 multichannel sound system was rated significantly better (larger value) than the two-channel stereo system, for every evaluation term except "loud". Figure 46 shows that, in the large theatre, the 22.2 multichannel sound system was rated significantly better than the 5.1 channel sound system for the terms "gaudy", "distinct", "wide in front and rear", "wide in above and below", "clear movement of sound", "sound from every direction", "rich reverberant" and "rich envelopment". Figure 40 shows that, in the small studio, the 22.2 multichannel sound system was rated significantly better than the 5.1 channel sound system for every evaluation term except "loud", "dynamic", "gaudy" and "natural". The results also show that the 5.1 channel sound system was rated significantly better than the two-channel stereo system for every term except "loud". The results suggest that there was no difference in the loudness between each system. They also suggest that the 22.2 multichannel sound system provided a better 3D spatial sound quality than both the two-channel stereo and the 5.1 surround sound systems in both a large theatre and a small studio. Furthermore, the difference of rate between the 22.2 multichannel sound system and the 5.1 channel sound system or two-channel stereo system is basically bigger in a small studio than in a large theatre. Therefore, in a small studio, the 22.2 multichannel sound system may provide a better 3D spatial sound quality than both other sound systems than in a large theatre.

FIGURE 46

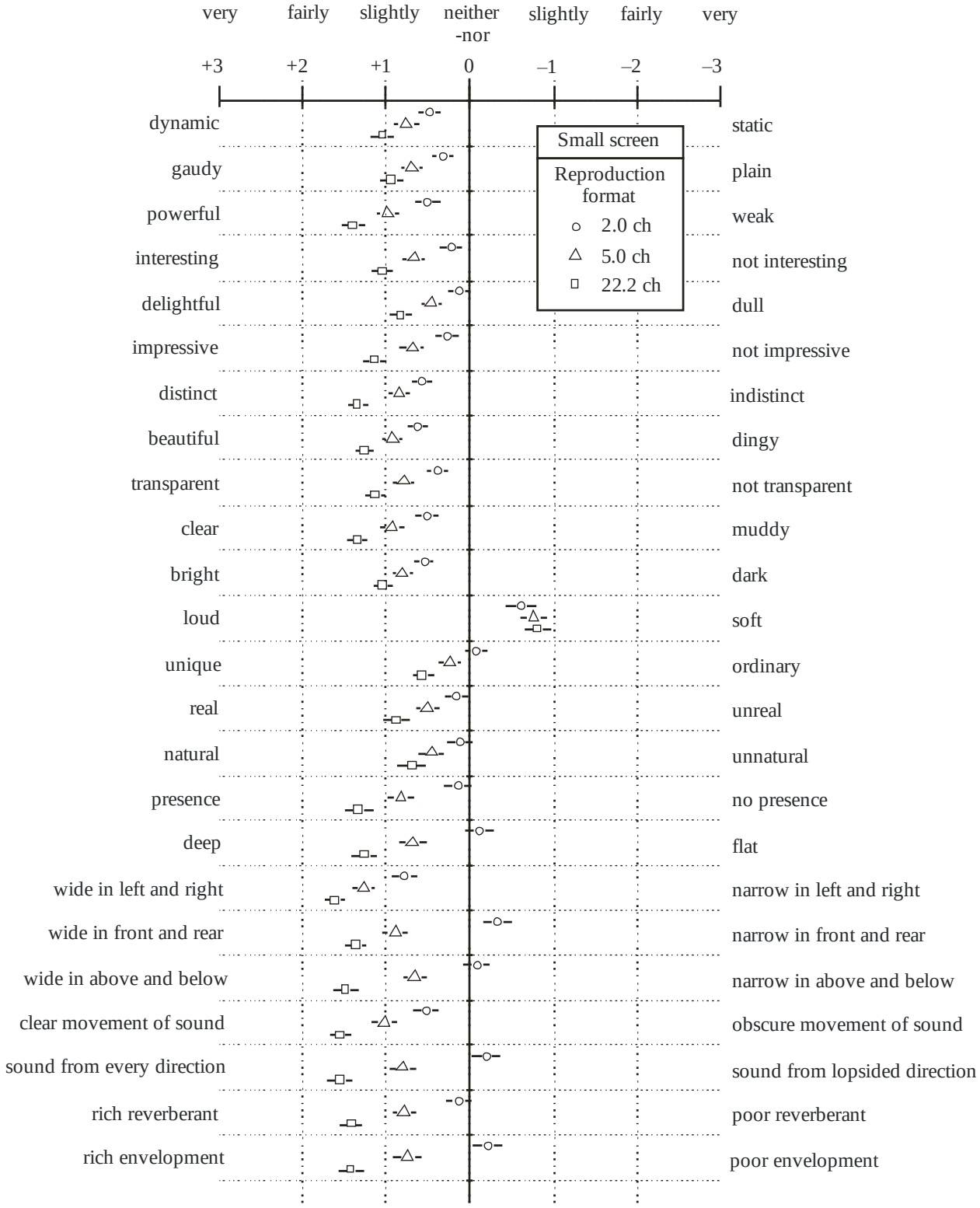
Results of subjective evaluation comparing three different sound systems with a large screen in a large theatre



Report BS.2159-34

FIGURE 47

Results of subjective evaluation comparing three different sound systems with a small screen in a small studio



Report BS.2159-35

7.2 10.2 channel sound system (Type B)

This provides subjective evidence of the 10.2 channel layout's performance.

7.2.1 Evaluation of directional audio quality

Because the loudspeakers of Top layer have an important role to make ambience, reverberation with direct sound.

To find a number of loudspeakers on Top layer, subjective evaluation is achieved.

– Layout of experiments:

  Middle layer was 3/4 channel of Recommendation ITU-R BS.775 without subwoofer in all case.

  NHK's Top 9 channel (Fig. 48(a)) was reference and Top 4 channel (Fig. 48(b)), Top 3 channel (Fig. 48(c)), Top 2 channel (Fig. 48(d)), Top 0 channel (Fig. 48(e)) layout were tested.

– Sound source of experiments: moving helicopter sound with VBAP rendered:

  Directions were circle and straight 0°/90°/180°. (Fig. 48(f)).

  Speed was 90°/s, 180°/s.

– Method of experiments: MUSHRA test about perception of directional difference:

  Imperceptible (100-80).

  Perceptible, but not annoying (80-60).

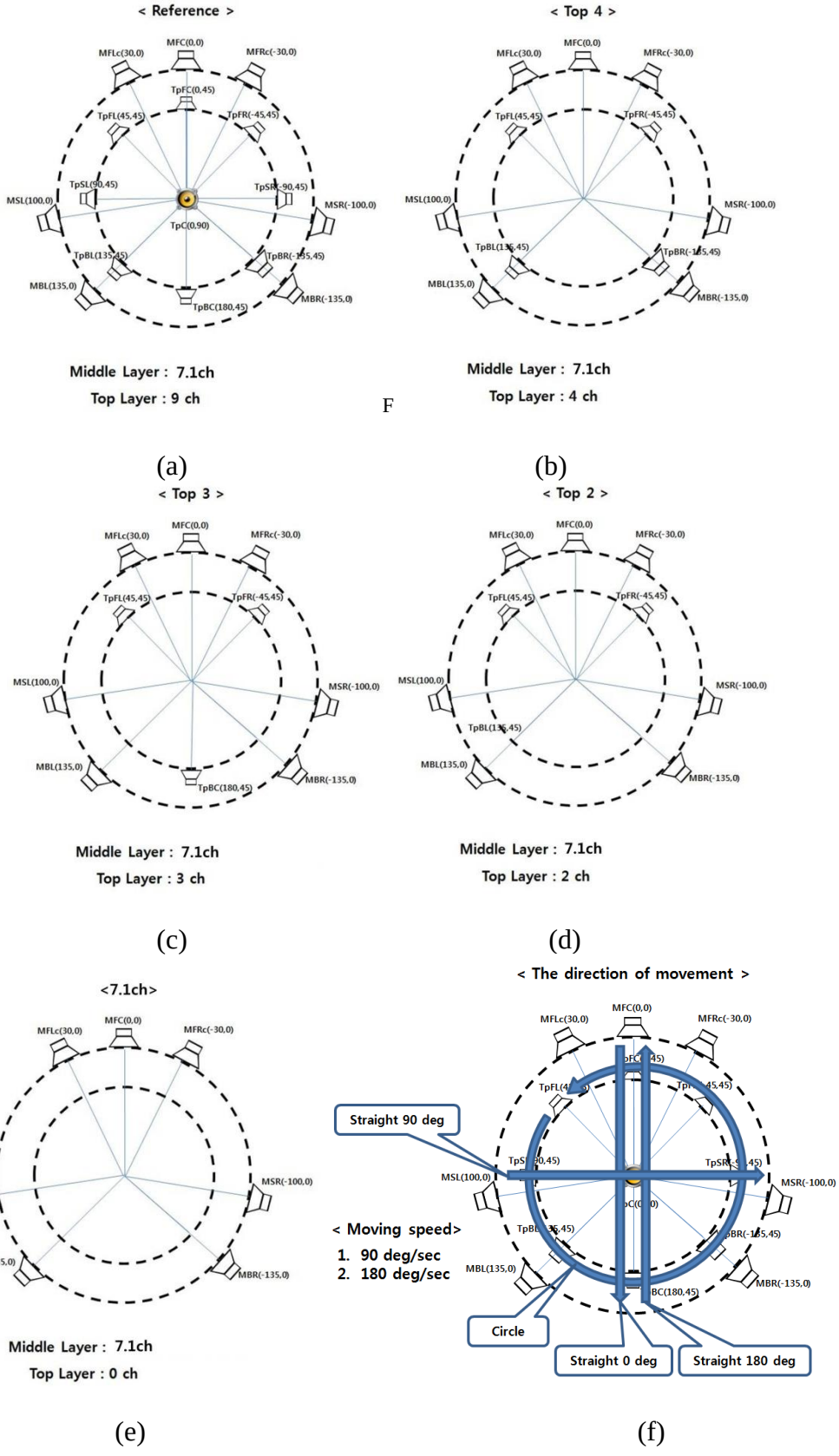
  Slightly annoying (60-40).

  Annoying (40-20).

  Very annoying (20-0).

FIGURE 48

The layout of directional audio quality evaluation

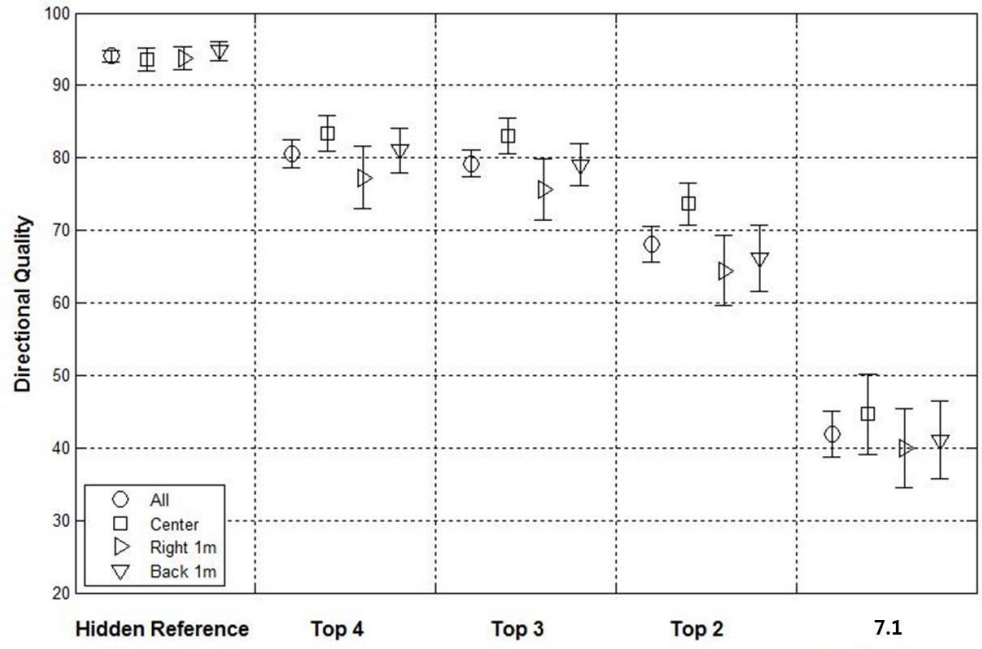


F

– Result of experiments: imperceptible at Top 3 channel.

FIGURE 49

The result of directional audio quality evaluation



7.2.2 Evaluation of overall audio quality

Using several audio contents, evaluation of overall audio quality by layouts is executed.

– Layout of experiments:

  NHK 22.2 (Fig. 50(a)), 11.2 (Fig. 50(b)), 10.2 (Fig. 50(c)), 7.1 (Fig. 50(d)).

– Contents of experiments:

  Movie mixing 1, 2, 3.

  Music mixing.

  DirAC (Directional Audio Coding, Fraunhofer & HUT) processed B-format live recording 1, 2.

– Method of experiments: MUSHRA test about perception of overall quality difference:

  Imperceptible (100-80).

  Perceptible, but not annoying (80-60).

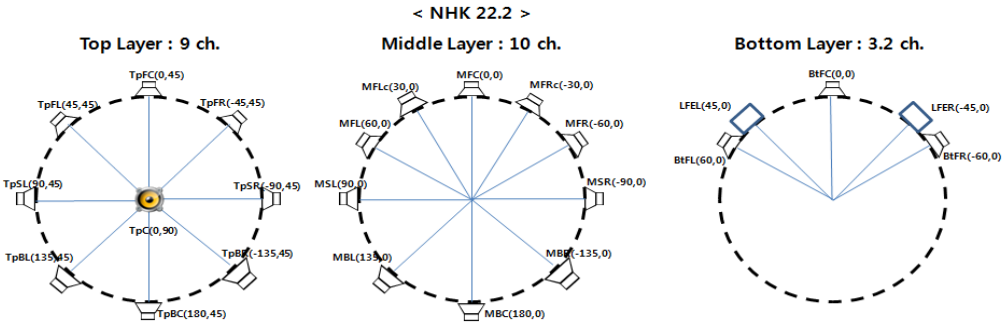
  Slightly annoying (60-40).

  Annoying (40-20).

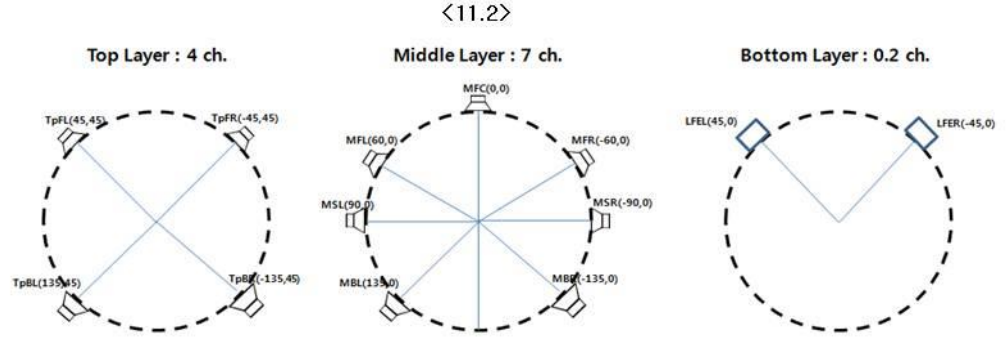
  Very annoying (20-0).

FIGURE 50

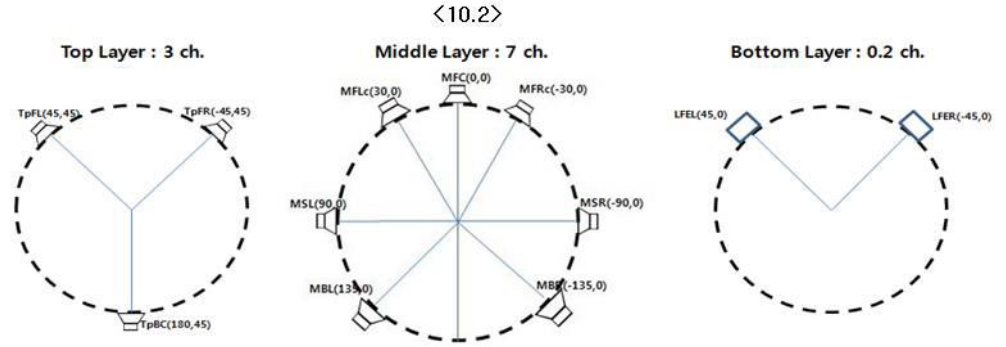
The layout of overall audio quality evaluation



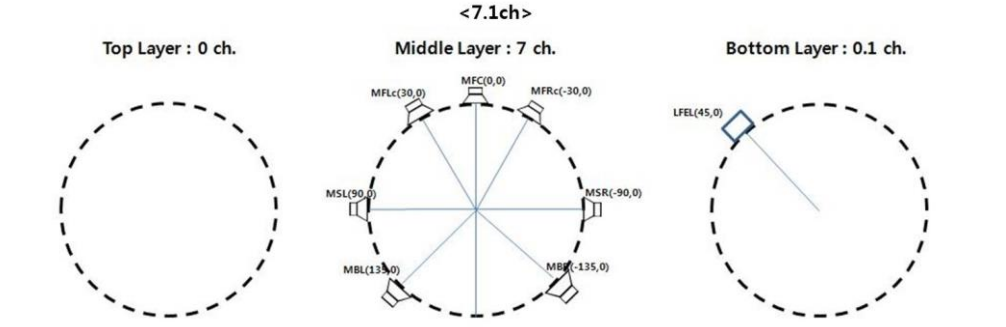
(a)



(b)



(c)



(d)

– Result of experiments:

The smaller the number of Top loudspeakers, the lower overall quality is.

But it is imperceptible at Top 3 channel in all cases.

FIGURE 51

The result of overall audio quality evaluation

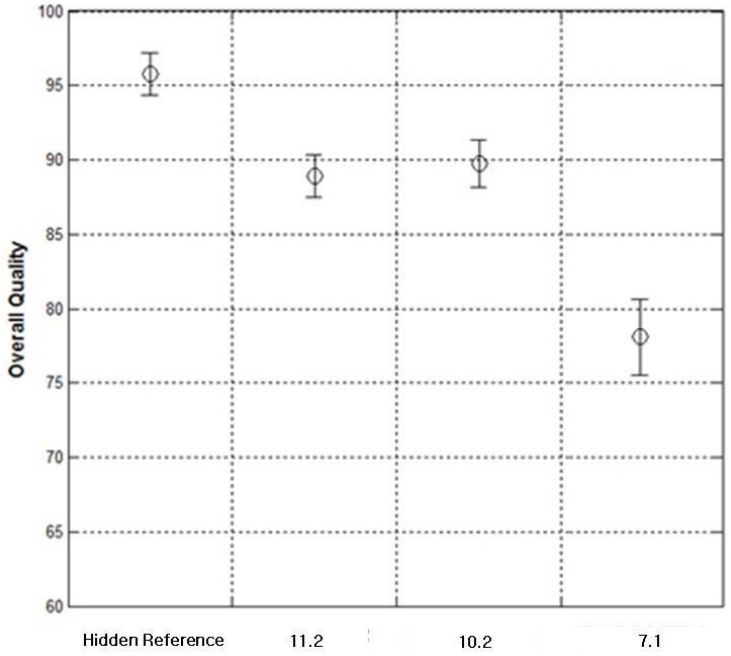
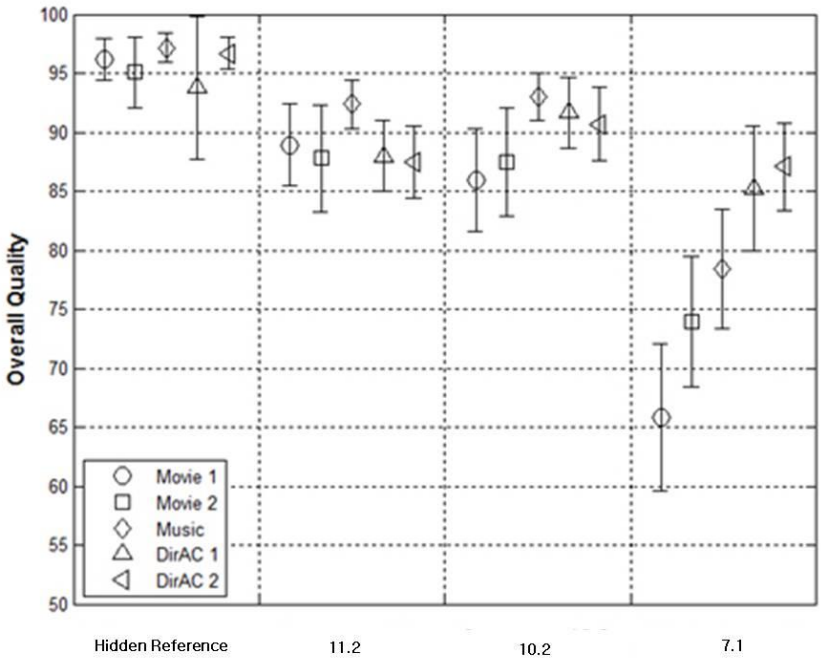


FIGURE 52

The result of overall audio quality evaluation by contents



7.3 Investigations into optimal speaker configurations for the hybrid object/channel system

The details on array count, position and density were determined through informal listening experiments. More formal testing paradigms would be necessary to extend the reported results to more general applications for determining optimal overhead speaker positioning of Advanced Multichannel Audio Systems. The current results were used solely to inform design decisions associated with the system described in this local text. Therefore, a limited methodological description of this listening is provided.

Listening environment

A commercial theatre was used for this listening With 205 seats, the room was representative of a medium sized theatre in a modern cineplex. This room had partial stadium seat configuration with a length of 56', a width of 47' and a maximum height of 21'. For testing additional sound, absorption material was installed to bring the reverb time down to RT60 approximately 350 ms for 500 – 10,000, rising to 580 ms at 120 Hz and dropping to 260 ms at 16 000 Hz. Three top surround arrays were installed on the ceiling as represented in Fig. 53.

Listeners and stimuli

Approximately 10 listeners, four listening positions, and multiple spectrally complex stimuli and moving sources were used to evaluate potential speaker configurations with attention to listener envelopment and spatial imaging. All listeners were presented audio (double blind) from a set of alternative speaker configurations derived as a subset of the speakers available. Listeners were instructed to grade each configuration on multiple parameters related to overall quality including general preference and listener envelopment. Only preference results are reported in this Report.

Results

Results from the Array Count investigation are shown in Figs 54 and 55. It was predicted that a listener's experience of the multichannel audio system would improve in a strictly monotonically manner with the introduction of an increasing number of speakers – limited only by the size of the listening venue. Listeners took part in two separate sections of assessment. Figure 51 shows the results for a test measuring listener preference for Ambient signals rendered in five different configurations.

A second test measured listener preference for panned, point sources. Results averaged over all listeners, listening positions and stimuli are shown in Fig. 55.

These experiments indicate the following: Two or three top surround arrays are significantly better than one or none. A third, centre array does not improve, and may even decrease envelopment in the case of ambient signals. A single centre array, while useful for some pan trajectories, does not seem to justify the additional install efforts. On this basis it was determined that the optimal speaker and channel recommendation for spatial audio reproduction is as shown in Fig. 20.

FIGURE 53

Experimental speaker installation

Top down view; screen is to the left. The 19 blue squares near the centreline of the room represent ceiling mounted speakers

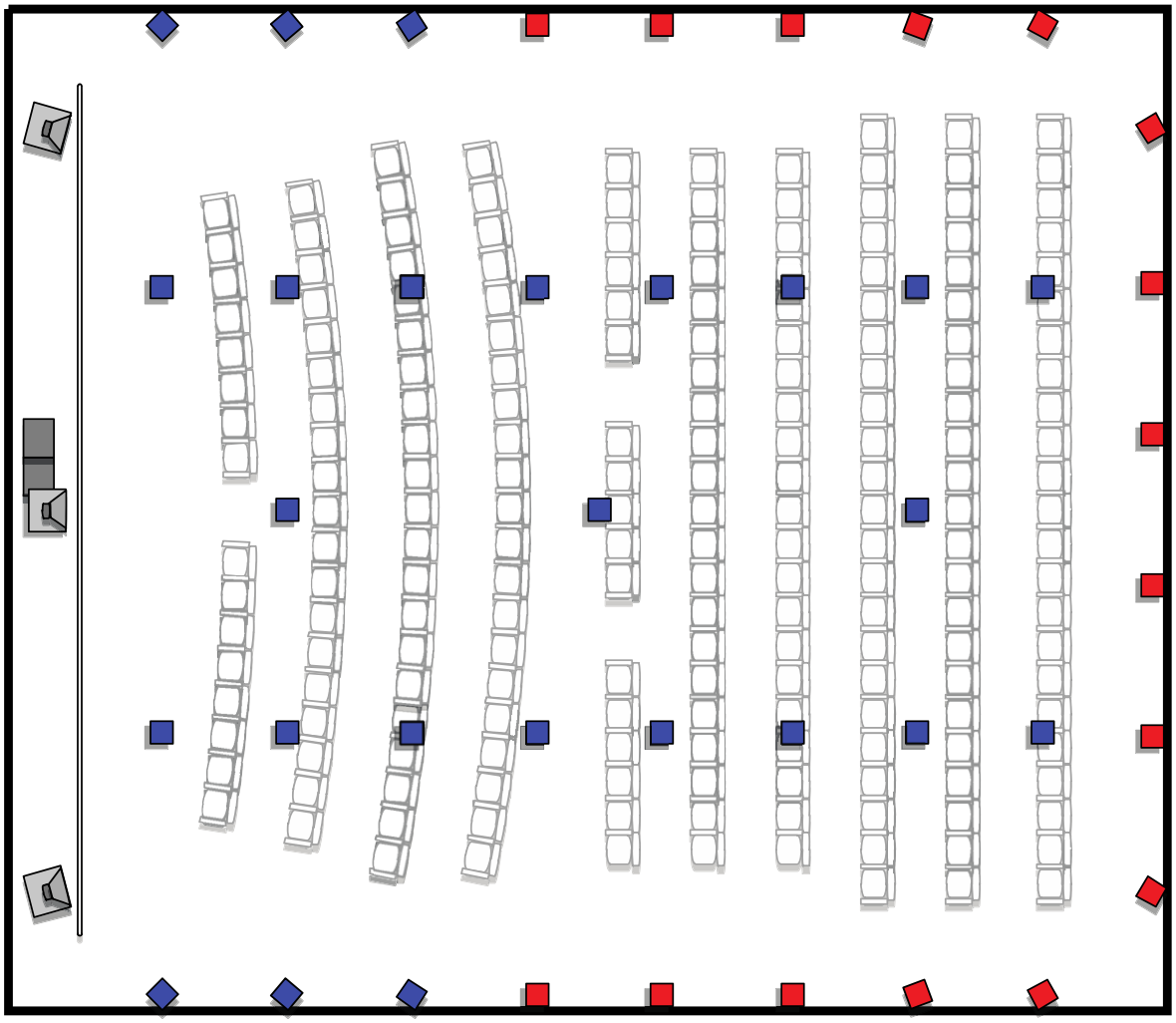


FIGURE 54

Listener preference for ambient sounds rendered in five different channel configurations: mono surround (ms), stereo surround (ss), ss + mono top surround, ss + stereo top surround, and ss + 3 top surround arrays. Stereo surround (ss) was the reference

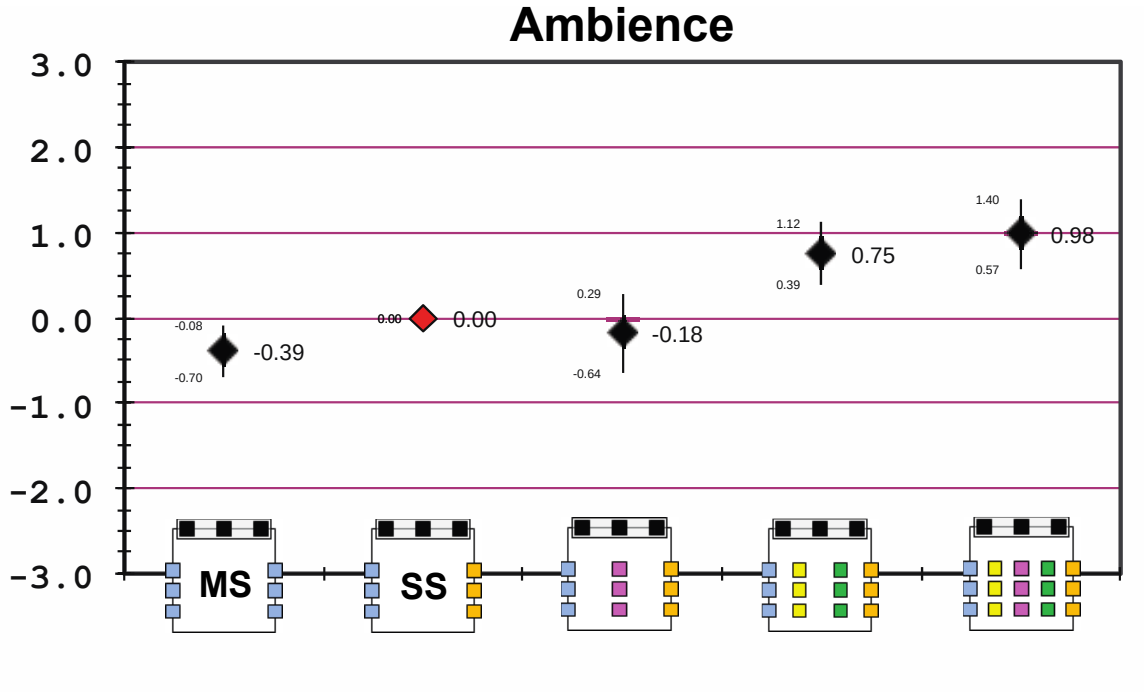
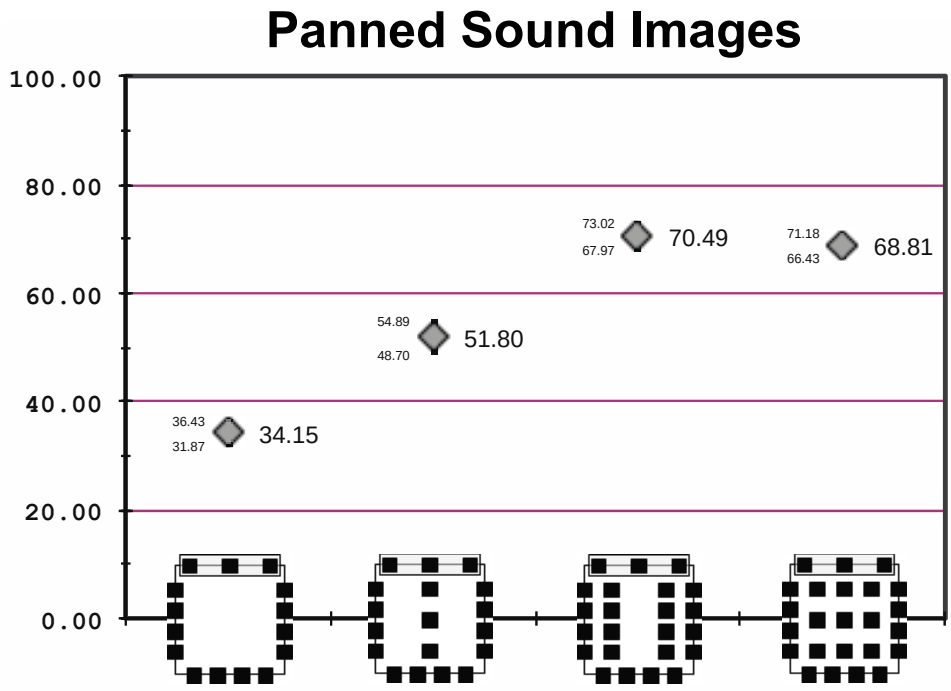


FIGURE 55

Listener preference for panned point-source sounds rendered using four different speaker configurations: no top surround speakers, one top surround array, two top surround arrays, and three top surround arrays. In this test the reference was a text description of the intended pan trajectory, but no audio reference



7.4 Further studies on quality performance relevant to multichannel sound systems

7.4.1 Elevation perception of phantom sound images in the frontal hemisphere

7.4.1.1 System configuration for subjective evaluation experiments

Recommendation ITU-R BS.1909 (Annex 1, § 1.3) recommends the following sound quality requirement:

– For applications with accompanying picture, the directional stability of the frontal sound image should be maintained over the entire area of high-resolution large-screen digital imagery. The coincidence of position between sound images and video images should also be maintained over a wide image and listening area.

Subjective evaluation experiments were conducted with a vertical loudspeaker configuration to investigate the number of audio channels required to reproduce vertically stable frontal sound images.

7.4.1.2 Subjective evaluation experiments

Preliminary subjective evaluation

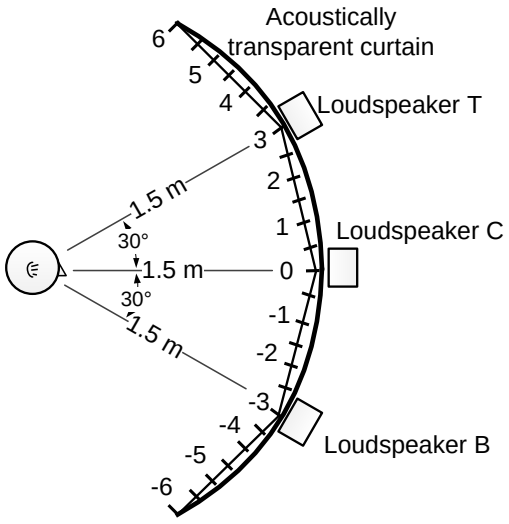
In a preliminary subjective evaluation, while a part of listeners reported that phantom sound images were unperceivable for a two-channel loudspeaker configuration with 60° vertical angle intervals, all listeners reported that phantom sound images were perceivable for a three-channel loudspeaker configuration.

Elevation perception of phantom sound images in the frontal direction (medial plane)

The experiments were conducted in an anechoic room. Three loudspeakers were placed at 30° intervals in a semicircle with a 1.5 m radius in the median plane (Fig. 56). To exclude the visual effect of loudspeakers, an acoustically transparent curtain was installed between the loudspeakers and listeners. Labels numbered from –6 to 6 were placed on the curtain to indicate the elevation angle for the listeners to respond in accordance with their perceived sound image. Listeners were positioned one at a time at the centre of the semicircle and asked to look at the label marked 0 directly in front of them during the evaluation.

FIGURE 56

Loudspeaker configuration

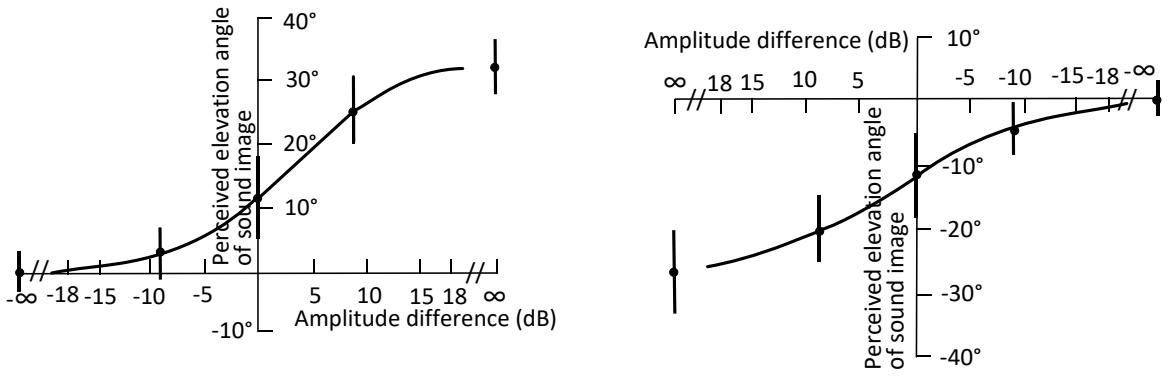


The stimuli comprised white noise with 1 s duration and 50 ms rise and fall time. The stimuli were reproduced by two loudspeakers (T and C or C and B) with amplitude difference. Five amplitude difference conditions were evaluated. The sound pressure level at the listening position was adjusted to 60 dBSPL for all conditions. Listeners were asked to respond the elevation angle that they felt corresponded to their perceived sound image. Five listeners with normal hearing participated in this experiment. Each listener assessed 15 trials for each condition, for a total of 75 trials.

As the distribution of the listeners' responses for perceived sound image did not significantly differ for each condition, average and standard deviation values were calculated from the data obtained for all 75 data. Figure 57 shows the experimental results obtained for the elevation angles of the perceived sound images. The horizontal axis indicates the amplitude difference between two loudspeakers, while the vertical axis indicates the overall mean elevation angle and standard deviation of perceived sound images. The results show that perceived sound images can be controlled by the pair-wise amplitude panning method over a 60° the vertical viewing angle.

FIGURE 57

Results for elevation angles of perceived sound images



(a) Loudspeakers T and C (b) Loudspeakers C and B

7.4.2 Elevation perception of phantom images in the frontal side direction

Subjective evaluation experiments were conducted to verify the vertical stability of reproduced sound images from the frontal side direction. The stimuli and other listening conditions were the same as those for the experiment described in § 7.4.1 with the exception of the azimuth angle of the loudspeakers configuration (Fig. 58). The experiments were conducted for the frontal right and upper side directions. The vertical semicircles were set up with azimuth directions of 30 and 60 degrees.

Figure 59 shows the experimental results obtained for the elevation angles of the perceived sound images. The horizontal axis indicates the azimuth angle of the loudspeaker configuration, while the vertical axis indicates the overall mean elevation angle of the perceived sound images. The curved lines indicate the results for each amplitude difference condition. The standard deviations (not shown in the Figure) were similar to those for the frontal direction experiment. The results show that perceived sound images can be controlled by the pair-wise amplitude panning method over a vertical direction of up to 60 degrees. A vertical angle of 60 degrees corresponds to the optimum vertical viewing angle for UHDTV.

FIGURE 58

Loudspeaker configuration

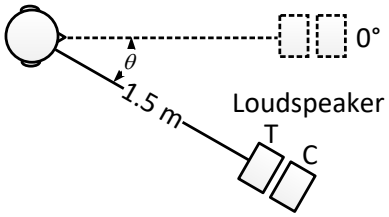
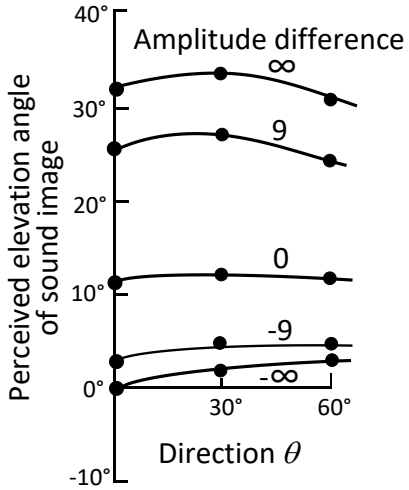


FIGURE 59

Results for perceived elevation angles of sound images



Conclusion

The results show that multichannel loudspeaker configurations with at least three vertically layered channels can reproduce directionally stable frontal sound images, which is desired for UHDTV applications whose optimal vertical viewing angle is relatively large.

7.4.3 Sensation of listener's envelopment[5] in an upper hemispherical sound field

System configuration for subjective evaluation experiments

Recommendation ITU-R BS.1909 recommends the following sound quality requirement:

– The sensation of a three-dimensional spatial impression that augments a sense of reality, which is related to ambience and envelopment, should be significantly enhanced over established sound formats in Recommendation ITU-R BS.775.

A subjective evaluation experiment was conducted on loudspeaker configurations with height channels to investigate the degree to which the number of sound channels and height channels enhances the sensation of LEV.

---

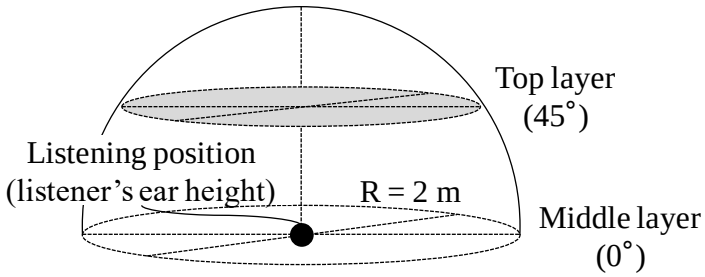
[5] As used in this Report, a "listener's envelopment is defined as the listener's impression of being in a spatial sound field in which he feels completely and uniformly surrounded by those sounds that are supposed to come from un-localisable sources. Examples of enveloping sound fields include concert hall reverberation, the wind in forest trees, falling rain, air conditioning noise". This definition is not to be directly referenced for use or extrapolation to the perceptual description of the performance of Advanced Multichannel Audio systems.

Subjective evaluation experiment (1)

The experiment was conducted in a listening room with a reverberation time of 0.18 s at 500 Hz. A total of 48 loudspeakers were placed around a listener in an upper hemispherical plane with two vertical layers, a middle layer, and a top layer, as shown in Fig. 60. Twenty-four of the loudspeakers were placed at a 15° aperture at the same height as the listener's ears. The other 24 were also placed at a 15° aperture, but above ear height with a 45 degrees elevation angle. The distance between the listener and each loudspeaker was 2 m.

FIGURE 60

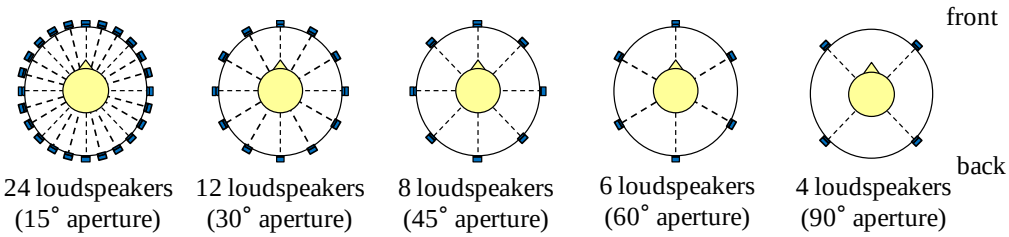
Loudspeaker layout in upper hemispherical plane



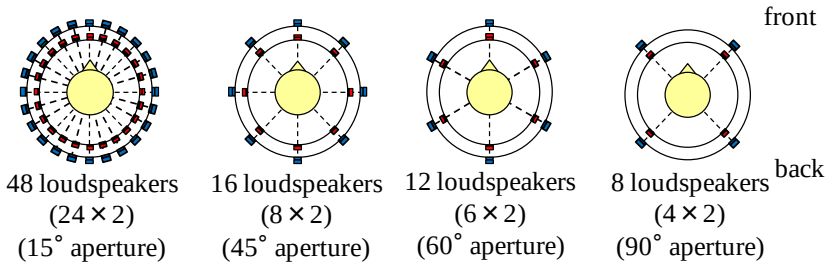
Nine test conditions for loudspeaker configurations were evaluated, five using a single (middle) layer and four using both the middle and top layers. These conditions are shown in detail in Fig. 61.

FIGURE 61

Loudspeaker configurations for subjective test



(a) Single layer (middle layer)



(b) Two layers (middle layer and top layer)

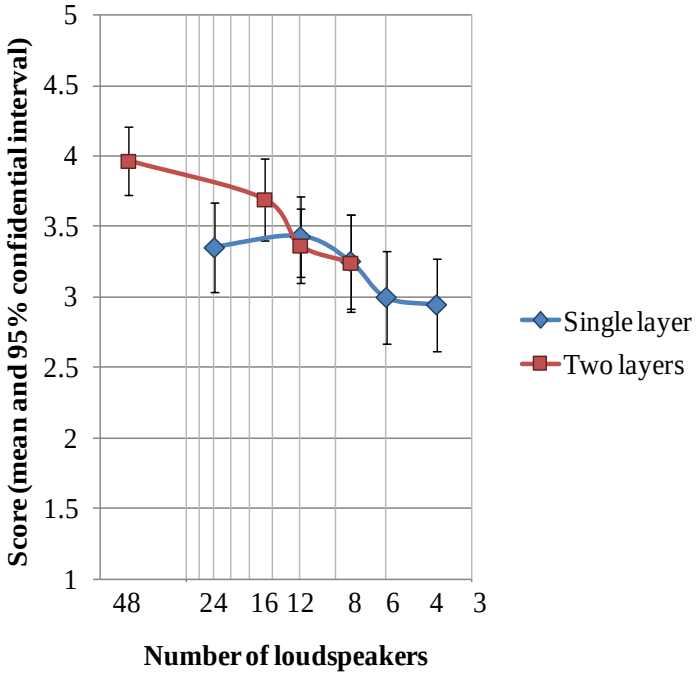
The stimuli comprised uncorrelated white noise with 6 s duration. The sound pressure level for each loudspeaker configuration was adjusted to 71.6 dBA ± 0.5 dBA at the listening point. For the conditions when both the middle and top layers were used, the sound pressure level at the listening point for the top layer was adjusted to −1.5 dB lower than that for the middle layer because the surface area of the top layer's spherical plane is $1/\sqrt{2}$ times that of the middle layer.

A total of 40 listeners with normal hearing, aged in their 20 s to 40 s, were asked to assess the LEV of white noise reproduced by the loudspeakers on a continuous five-grade quality scale, ranging from 5.0 (“Very much”) to 1.0 (“Not at all”). The listeners were instructed to consider LEV as being indicative of sounds with spatial homogeneity, i.e. sounds coming evenly from directions that were spatially connected. Each listener was positioned at the centre of the loudspeaker arrangement. The listeners were asked not to move during the evaluation, but their heads were not fixed to the chair.

As the distribution of the listeners’ grades did not significantly differ for each condition, scores (mean and 95% confidential interval) were calculated from the data obtained for all listeners. Figure 62 shows the experimental results obtained for the sensation of LEV. The horizontal axis indicates the total number of sound channels including height channels, while the vertical axis indicates the overall scores (mean and 95% confidential interval). The results show that the sensation of LEV was enhanced as the number of sound channels was increased for both the conditions in which a single (middle) layer was used and those in which the middle and top layers were used. When either eight or 12 sound channels were used, the sensation of LEV was perceived to be almost the same for the single-layer and two-layer cases. Although the sensation of LEV became saturated in the single-layer case even if the number of sound channels was increased to 24, in the two-layer case it continued to be enhanced as the number of sound channels was increased.

FIGURE 62

Results for sensation of LEV



The scores obtained for the sensation of LEV for 5.1 channel sound were similar to those obtained for 6-channel sound in a previous study [2] conducted by Hiyama *et al*. This shows that the sensation of LEV for an 8 × 2 channel configuration is significantly enhanced over that for the established sound formats in Recommendation ITU-R BS.775.

References

[1] Oode S. *et al.*, "Three-Dimensional Loudspeaker Arrangement for Creating Sound Envelopment", IEICE Technical Report, EA2012-46 (2012) (in Japanese).

[2] Hiyama K. *et al.*, "The minimum number of loudspeakers and its arrangement for reproducing the spatial impression of diffuse sound field," Proc. AES 113th Convention, 1-12 (2002).

7.4.4 Localization of phantom sound images in the elevation direction in an upper hemispherical sound field

System configuration for subjective evaluation experiments

Recommendation ITU-R BS.1909 recommends the following sound quality requirement:

– The sound image should be reproduced in all directions around the listener, including the elevation direction, within reasonable limits of stability.

A subjective evaluation experiment was conducted on loudspeaker configurations with height channels to investigate the localization and localization uncertainty of phantom sound images in the elevation direction generated by two loudspeakers located above the listener.

Previous studies of directional perception of phantom sound images in the horizontal plane (middle layer)

A number of previous studies have researched the correlation between the localization and localization uncertainty of sound image and inter-channel parameter relationships, such as pair-wise amplitude differences and/or pair-wise time differences. The studies have included investigations on a conventional 2-channel stereo configuration (with a loudspeaker aperture of 60 degrees) as well as on other apertures and lateral displacements using a so-called quadraphonic loudspeaker configuration (with a 90 degrees aperture for each loudspeaker) [1,2,3], a 6-channel "all round effect" loudspeaker configuration (with a 60 degrees aperture for each loudspeaker) [3] and the 5.1 channel stereophonic loudspeaker configuration specified in Recommendation ITU-R BS.775.

Studies done by Ratliff [1], Nakabayashi [2], and Theile [3] have shown that with a quadraphonic loudspeaker configuration it is not possible to achieve a uniform distribution of phantom sound images lateral to the listener, that even small amplitude differences between the front and back loudspeakers lead to large angle changes, and that localization jumps here and there between the loudspeakers at the front and at the back. Therefore, Theile [3] concluded that the right and left lateral directions must be represented through actual sources to achieve an "all round effect". In this case, if an aperture of up to 60 degrees for a loudspeaker pair is allowed, the 6-loudspeaker configuration shown in Fig. 60 results.

Martin and Woszczyk *et al.* [4] carried out 5.1 channel stereophonic sound listening tests to investigate the localization and localization uncertainty of phantom sound images. The images were produced using pair-wise panning with various amplitude differences or time differences. The results showed that for front sound images, when using front pairs of adjacent loudspeakers, pair-wise amplitude panning is a reasonably reliable method for producing predictable phantom image locations and produces smaller degrees of localization uncertainty. For side sound images, however, it produces lateral phantom images that are at best unstable, if indeed achievable. For back sound images, it produces localization uncertainties in the back pair of loudspeakers similar to those observed in the side pair.

Following up on these studies, a preliminary subjective test on the localization and localization uncertainty of phantom sound images in the horizontal plane (middle layer) was carried out prior to conducting a subjective test on phantom sound images in the elevation direction (top layer). The latter test was conducted using the subjective listening test method described below in § 3.3. The results

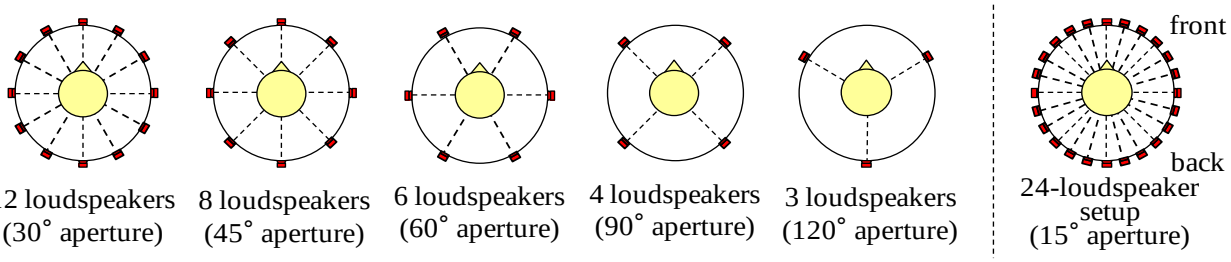
obtained for the directional localization uncertainty of phantom sound images were fairly consistent with those obtained in previous studies.

Subjective listening test method

The experiment was conducted in the same listening room with the same loudspeaker setup as those used in the LEV experiment (Fig. 59). Twenty-four loudspeakers were placed at a 15-degree aperture above ear height with a 45-degree elevation angle. We used certain portions of the 24 loudspeakers to evaluate five test conditions for loudspeaker configurations. These conditions are shown in detail in Fig. 63.

FIGURE 63

Loudspeaker configurations for subjective test and 24-loudspeaker setup



The stimuli comprised white noise with 1 s duration and 50 ms rise and fall time. The stimuli were reproduced by two adjacent loudspeakers with amplitude difference for each loudspeaker configuration. The pair-wise amplitude panning method of tangent law was used to derive the gains of the two adjacent loudspeakers. This law is shown in equation (1).

$$\frac{\tan\theta_T}{\tan\theta_0} = \frac{g_1 - g_2}{g_1 + g_2} \tag{1}$$

where:

θ_0 : half the angle between the two loudspeakers

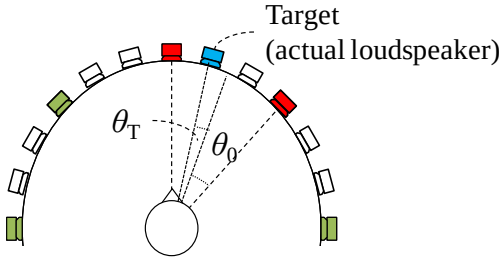
θ_T : desired angle of the phantom source

g_1, g_2 : the gains of the two adjacent loudspeakers with $g_1, g_2 \in [0,1]$.

The desired angle of the phantom sound images was set to the angle of the actual loudspeaker setup, so that 13 amplitude difference conditions in a semicircle were evaluated from 0 degree (front) to 180 degrees (back) with 15 degrees intervals. The sound pressure level at the listening position was adjusted to 71.5 dBA ± 0.5 dBA at the listening point.

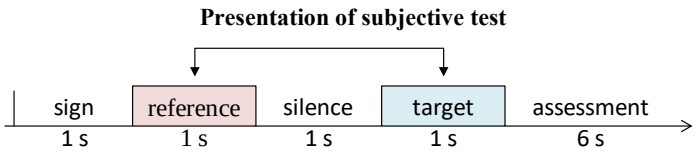
FIGURE 64

Pair-wise amplitude panning method



A total of 40 listeners with normal hearing, aged in their 20s to 40s, heard two white noises reproduced by the loudspeakers, and were asked to assess the "certainty of arrival direction" of the noises on a continuous five-grade quality scale, ranging from 5.0 ("Very much") to 1.0 ("Very little"). The listeners were instructed to consider "certainty of arrival direction" as being indicative of the difference in arrival direction between the two noises. One white noise was reproduced by an actual loudspeaker (reference), and the other was reproduced by two adjacent loudspeakers (target: phantom sound images). Each listener was positioned at the centre of the loudspeaker arrangement. The listeners were asked not to move during the evaluation, but their heads were not fixed to the chair.

FIGURE 65



Directional perception of phantom sound images in the upper plane (top layer)

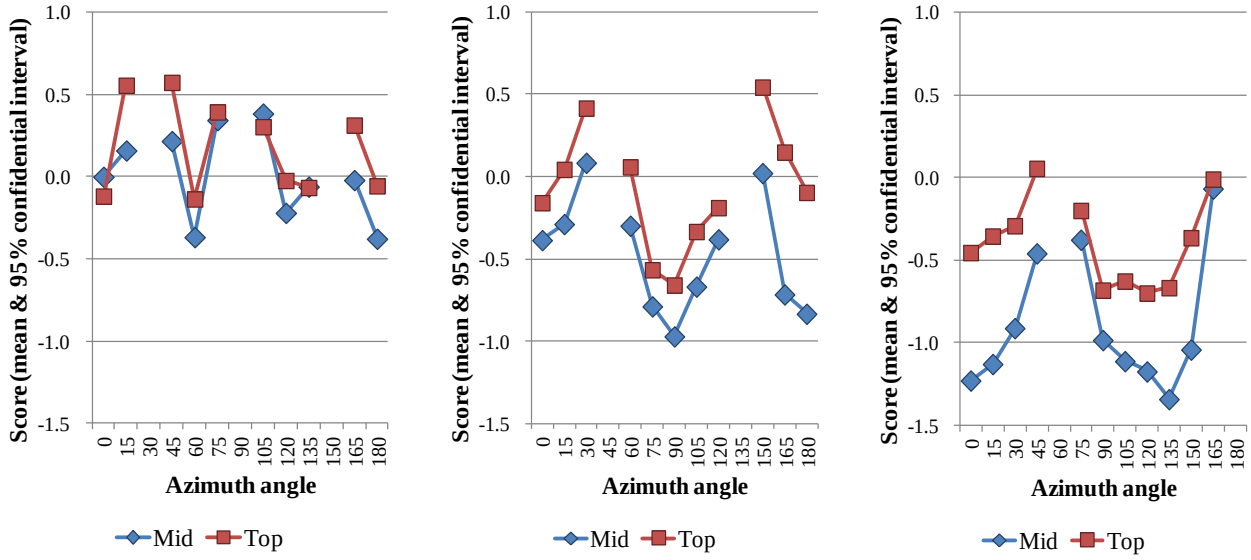
A subjective listening test was carried out in order to investigate the localization and localization uncertainty of phantom sound images above the listener using loudspeakers in the top layer. The test method was described in § 3. As the distribution of the listeners' responses for perceived sound image did not significantly differ for each condition, scores (mean and 95% confidential interval) were calculated from the data obtained for all listeners.

Figure 66 shows the experimental results obtained for the directional localization uncertainty of phantom sound images in the top layer as well as in the middle layer (horizontal plane). The horizontal axis indicates the desired azimuth angle of the phantom sound images, while the vertical axis indicates the overall mean of localization uncertainty of the images. The scores were normalized by the grades for the reference condition. The reference condition was set as the phantom sound image uncertainty in the straight-ahead direction for a conventional two-channel stereo configuration. These Figures show only the results for phantom sound images reproduced by the two adjacent loudspeakers. Stable sound images reproduced through actual sound sources (not shown in the Figure) were clearly perceived at the loudspeakers' locations even if the listeners were out of the centre listening position.

As the Figure shows, a similar tendency was obtained for the directional localization uncertainty of the phantom sound images in the top and middle layers. It also shows that the degree of localization uncertainty was smaller in the top layer than in the middle layer when the number of loudspeakers was decreased from six to three.

FIGURE 66

Results for localization uncertainty of phantom sound images in top and middle layers

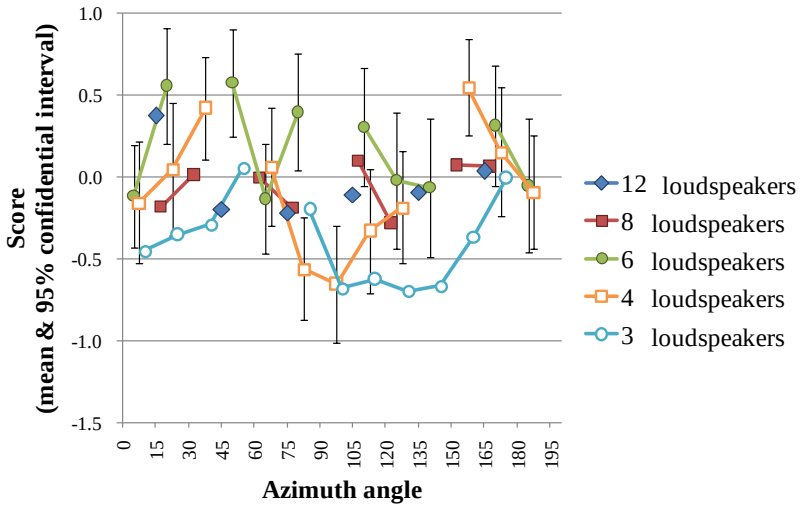


(a) 6 loudspeakers (b) 4 loudspeakers (c) 3 loudspeakers

Figure 67 shows the experimental results obtained for the localization uncertainty of the phantom sound images in the top layer. It includes 95% confidential interval values for 6 and 4 loudspeakers, and other confidential interval values obtained (not shown in the Figure) were similar to these values. As the Figure shows, the localization uncertainty scores for 12, 8 and 6 loudspeakers were around 0.0 or above for "all round directions", that is, the localization uncertainty is equivalent to that of the reference condition, which is phantom sound localization in the straight-ahead direction for a conventional two-channel stereo configuration. The scores for three loudspeakers were around −0.5 for some directions except those close to the actual loudspeaker directions. Those for four loudspeakers were less than −0.5 for the right and left lateral (90 degrees) directions.

FIGURE 67

Results for localization uncertainty of phantom sound images in top layer



Outline

According to the results obtained in previous studies and in the preliminary subjective test, in order to reproduce sound images in all round directions at ear height with good stability, sound channels at the side left/right positions and the back centre position are desirable in addition to the 5.1 channel loudspeaker configuration specified in Recommendation ITU-R BS.775. To reproduce sound images in all round directions over ear height with good stability, at least a 6 channel loudspeaker configuration is suitable at the centre position. It is assumed that an 8 channel loudspeaker configuration is preferable to maintain the directional stability of the front and back centre sound images over a wide viewing/listening area.

References

[1] Ratliff P. A., "Properties of Hearing Related to Quadrophonic Reproduction," BBC R&D 38 (1974).

[2] Nakabayashi, K., "Sound localization on the horizontal plane," The Journal of the Acoustic Society of Japan, Vol. 30, No. 3 (1974) (in Japanese).

[3] Theile, G. and Plenge, G., "Localization of Lateral Phantom Sources," Journal of the Audio Engineering Society, Vol. 25, No. 4 (1977).

[4] Martin, G. Woszczyk, W., Corey, J. and Quesnel, R., "Sound Source Localization in a Five-Channel Surround Sound Reproduction System," Proc. AES 107th Convention, Convention Paper 4994, pp.1-16, (1999).

[5] Oode S. *et al.*, "Sound Localization achieved by Loudspeakers Arranged in Three Dimensions," 2012 ASJ Autumn Meeting Report (in Japanese).

7.4.5 Influence of listening position on directional perception of frontal sound images in the frontal hemi-sphere

One of the required features of the advanced multichannel sound systems is a capability of reproducing the frontal sound image with directional stability over the entire area of high-resolution large-screen digital imagery and over a wide listening area.

This section describes the subjective evaluation on the influence of listening position and loudspeaker configurations with height channels on directional perception of frontal sound images.

System configuration for subjective evaluation experiments

Recommendation ITU-R BS.1909 recommends the following sound quality requirement:

1-3) For applications with accompanying picture, the directional stability of the frontal sound image should be maintained over the entire area of high-resolution large-screen digital imagery. The coincidence of position between sound images and video images should also be maintained over a wide image and listening area.

A subjective evaluation experiment was conducted on loudspeaker configurations with height channels to investigate the number of audio channels required to reproduce directionally stable frontal sound images over a wide listening area.

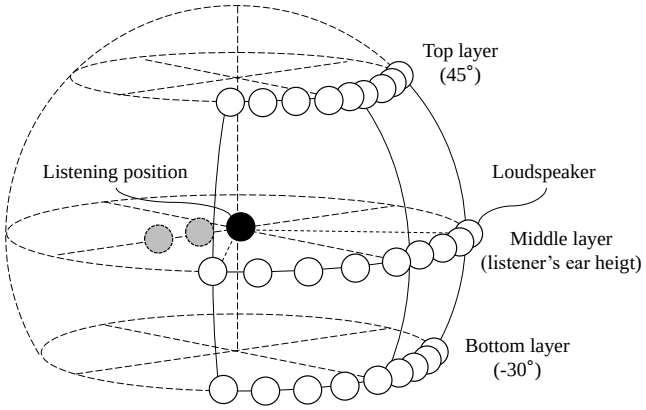
Subjective listening test method

The experiment was conducted in a listening room with a reverberation time of 0.18 s at 500 Hz. Loudspeakers were placed in front of a listener in an hemispherical plane with three vertical layers (a middle layer: at the same height as the listener's ears, a top layer: above ear height with a

45 degrees elevation angle, and bottom layer: below ear height with a 30 degrees elevation angle) as shown in Fig. 68.

FIGURE 68

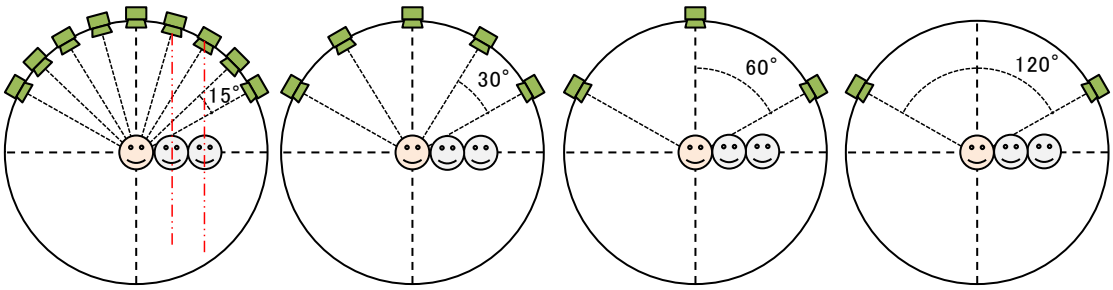
Loudspeaker layout in frontal hemispherical plane



Nine loudspeakers were placed at a 15° interval for each layer as shown in Fig. 69(a). The distance between a listener and each loudspeaker was 2 m. To exclude the visual influence on loudspeakers, an acoustically transparent but visually opaque curtain was installed between the loudspeakers and a listener. Labels were set to indicate the azimuth angle at each 1 degree interval to respond the perceived azimuth angle.

FIGURE 69

Loudspeaker configurations for subjective test



(a) 15° interval (b) 30° interval (c) 60° interval (d) 120° interval

Certain portions of the nine loudspeakers were used to evaluate four test conditions for loudspeaker configurations. These conditions are shown in detail in Fig. 66. For each loudspeaker configuration seventeen desired azimuth angle conditions of the sound images were set from –60° (left) to 60° (right) with 7.5° intervals by two adjacent loudspeakers.

The stimuli comprised white noise with 1 s duration and 50 ms rise and fall time. The stimuli were reproduced by two adjacent loudspeakers with amplitude difference for each loudspeaker configuration. The pair-wise amplitude panning method of tangent law was used to derive the gains of the two adjacent loudspeakers. The sound pressure level was adjusted to 71.5 dBA ± 0.5 dBA at the centre listening position. The stimuli were presented twice for the consistency of the listener's response.

A total of 20 listeners with normal hearing, aged in the 20s to 30s, were asked to respond the azimuth angle that corresponds to their perceived sound image. Each listener evaluated at three listening positions: (1) in front of the centre loudspeaker (centre listening position), (2) in front of the 15 degrees loudspeaker, and (3) in front of the 30 degrees loudspeaker. Each listener was asked to look at the label 0 degree marker of the centre loudspeaker position during the evaluation. Each listener evaluated 204 trials in total, which consists of four loudspeaker configurations, seventeen desired azimuth angle conditions and three listening positions.

Results of directional perception of frontal phantom sound images and listening position

As the distribution of the listeners' responses for perceived sound image did not significantly differ with each condition, scores (mean and 95% confidential interval: CI) were calculated from the data obtained from all listeners. Figure 70 shows the experimental results of azimuth angles of perceived sound image. The horizontal axis indicates the desired azimuth angle of sound image by the pair-wise amplitude panning method at the centre listening position, while the vertical axis indicates the overall mean azimuth angle of perceived sound images. As the results obtained from the position in front of the 15 degrees loudspeaker were similar to those of the position in front of the 30 degrees loudspeaker, the latter results are not shown. The 95% CIs are not shown to simplify these Figures.

7.4.6 Summary of studies on loudspeaker configurations to meet the requirements

The following series of studies have already been included in this Report:

– Elevation perception of phantom sound images in the frontal hemisphere, which is desired for UHDTV applications.

– The sensation of "listener's envelopment (LEV)", which is one of the primary features of a three-dimensional spatial impression.

– Localization and localization uncertainty of phantom sound images in the elevation direction generated by two loudspeakers located above the listener.

The study on the influence of listening position on directional perception of frontal sound images was described in § 7.4.1. The results show that a loudspeaker configuration with the centre channel is desired to maintain the directional stability of the front sound images over a wide listening area not only for the middle layer (the same height as the listener's ears) but also for the top and the bottom layers (the above or below the listener's ears). A loudspeaker configuration with five front channels for middle layer seems to be preferable.

Based on the results of the subjective evaluation, a summary on loudspeaker configurations with respect to the requirements is given in Table 8.

TABLE 8

Loudspeaker configurations to meet the requirements

Listening position	Quality requirements	Essential loudspeaker configuration to meet the requirements	
		Loudspeaker interval	**Number of loudspeakers**
Centre listening position	Localization of phantom sound images in all directions	Azimuth directional perception: 60° interval (middle layer and top layer)	Middle layer: 6 loudspeakers Top layer: 6 loudspeakers
		Elevation directional perception: 45° intervals	Middle layer and top layer and just above the listener
	Sensation of a three-dimensional spatial impression	Listener's envelopment (LEV) over horizontal plane: 45° interval	Middle layer: 8 loudspeakers Top layer: 8 loudspeakers
		LEV over vertical plane: 45° interval	Middle layer and top layer and just above the listener
	Directional stability of the frontal sound image over the entire image area(1)	Azimuth directional perception: 60° interval	Middle layer: 3 loudspeakers Top layer: 3 loudspeakers Bottom layer: 3 loudspeakers
		Elevation directional perception: 30° interval	Three layers: middle, top and bottom layer
Wide listening area(2)		Azimuth directional perception: 30° interval (maximum error is 10° or less)	Middle layer: 5 loudspeakers Top layer: 5 loudspeakers Bottom layer: 5 loudspeakers
		Middle layer: 30° interval top and bottom layer: 60° interval (maximum error is 20° or less)	Middle layer: 5 loudspeakers Top layer: 3 loudspeakers Bottom layer: 3 loudspeakers

(1) 7 680 × 4 320 Image system: horizontal viewing angle 96°, vertical viewing angle 54°.

(2) Listen at out of centre listening position.

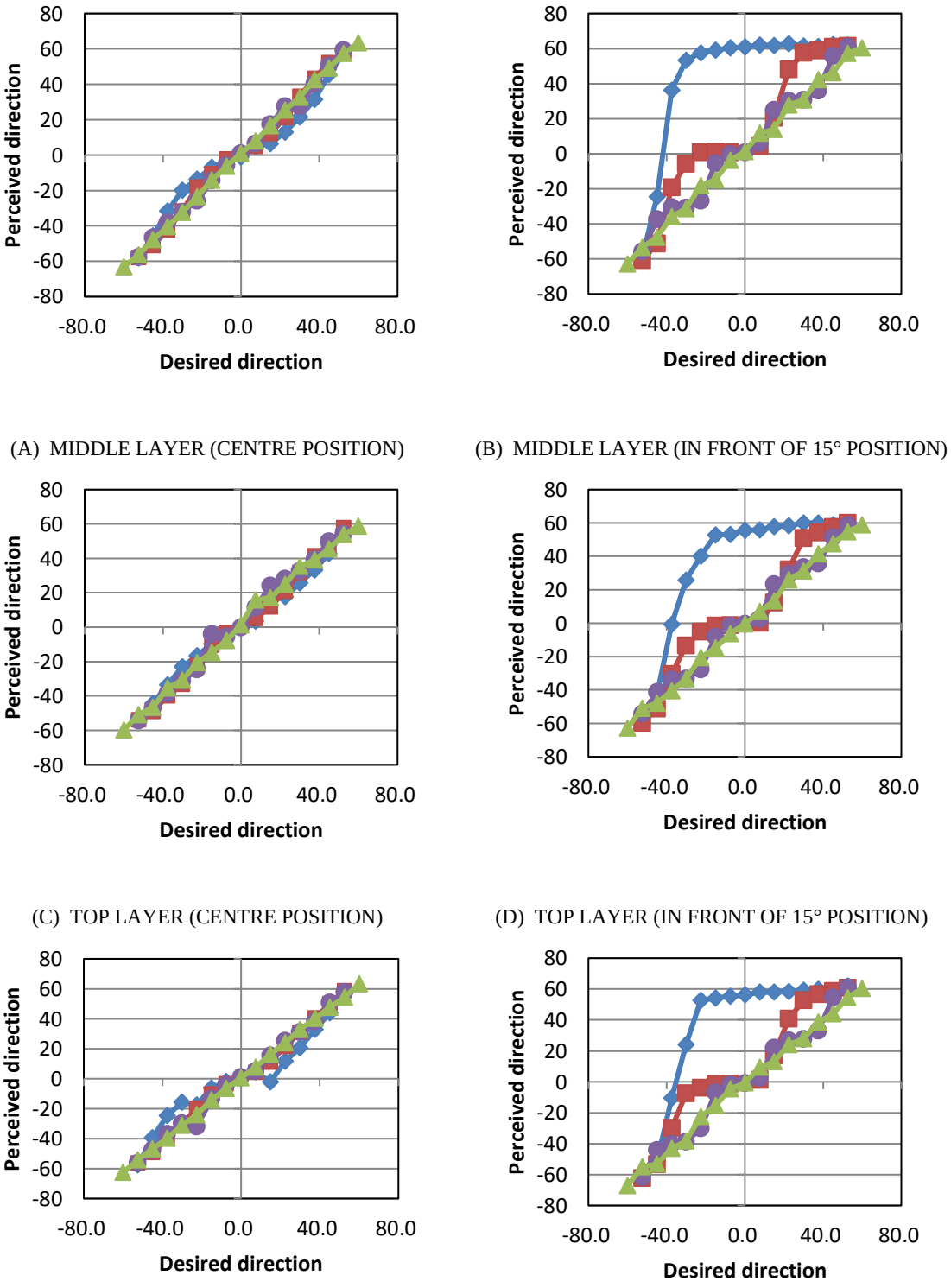
The results show that perceived sound images can be well controlled by the pair-wise amplitude panning method over the frontal horizontal angle of 120 degrees, for all the loudspeaker configurations and for all the vertical layers (middle, top and bottom layers), when listening at the centre listening position, as shown in Fig. 67(a), Fig. 67(c) and Fig. 67(e). However, some listeners reported that phantom sound images were unperceivable for the two-channel loudspeaker configuration with 120 degrees interval, even when listening at the centre listening position.

Perceived azimuth angles of sound images were considerably deviated from the desired azimuth angles for all the vertical layers in case of the 120 degrees interval loudspeaker configuration, when listening at the positions except for the centre position, as shown in Fig. 67(b), Fig. 67(d) and Fig. 67(f). For the 60 degrees interval loudspeaker configuration, differences between the perceived azimuth angle and the desired azimuth angle of sound images are drastically alleviated compared to the 120° interval loudspeaker configuration for all the three vertical layers. For the 30 degrees interval or 15 degrees interval loudspeaker configurations, the sound images were perceived well with sufficient accuracy to the desired azimuth angles for all the three vertical layers, even when listening at the positions except for the centre position. As regards 95% CIs (not shown in the Figures), while some CIs were up to 20 degrees for some of the test conditions of a 120 degrees interval loudspeaker configuration, most CIs were about 2-5 degrees.

The results were fairly consistent with those obtained in the previous study [1] on the influence of listening position on frontal phantom sound images localization of middle layer.

FIGURE 70

Results of evaluation



Outline

Some listeners reported that phantom sound images were unperceivable for the two-channel loudspeaker configuration with 120 degrees interval, even when listening at the centre listening position. In addition, differences between the perceived azimuth angle of sound image and the desired azimuth angle for the 60 degrees interval loudspeaker configuration are drastically alleviated compared to the 120° aperture loudspeaker configuration for all the three vertical layers, even when listening at the positions except for the centre position. Therefore a loudspeaker configuration with centre channel is desired to maintain the directional stability of the front sound images over the entire area of high-resolution large-screen digital imagery and over a wide listening area.

As shown in Fig. 70(b), Fig. 70(d) and Fig. 70(f), even when listening at the positions except for the centre position, the loudspeaker configuration with 30 degrees interval can control the perceived azimuth angle of sound images by the pair-wise amplitude panning method over the entire area with the horizontal viewing angle of 120 degrees for each of the three vertical layers. In case of loudspeaker configurations with 60 degrees interval, a similar tendency of deviation between the desired and perceived azimuth angle was obtained between the middle, top and bottom layers, but the degree of deviation of the top and bottom layer was smaller than that of the middle layer. An additional two channel configuration for middle layer, with five front channels in total, seems to be preferable, as listeners are sensitive to frontal sound images localization of middle layer than that of the top and the bottom layers, and as essential sound elements are apt to be presented around the middle layer.

8 Relevant documents concerning the multichannel sound systems developed by organizations outside ITU

8.1 SMPTE

8.1.1 SMPTE 2036-2-2008, “Ultra High Definition Television – Audio characteristics and audio channel mapping for programme production”

SMPTE 2036 Ultra High Definition Television (UHDTV) suite of documents is in multiple parts:

– Part 1: Image parameter values for programme production.

– Part 2: Audio characteristics and audio channel mapping for programme production.

SMPTE Standard 2036-2-2008 is Part 2 of SMPTE 2036 and describes the audio characteristics and audio channel mapping for programme production. This document specifies the characteristics of digital audio for UHDTV programme production and distribution, and also defines the mapping and labelling of 22.2 multichannel audio for UHDTV programme production.

The audio specifications are as follows:

1) Digital signal characteristics

UHDTV audio shall support a channel count of 24 full-bandwidth channels.

NOTE 1 – The two LFE channels are transported as full-bandwidth channels.

2) Channel mapping and channel labelling of 22.2 multichannel audio

TABLE 9

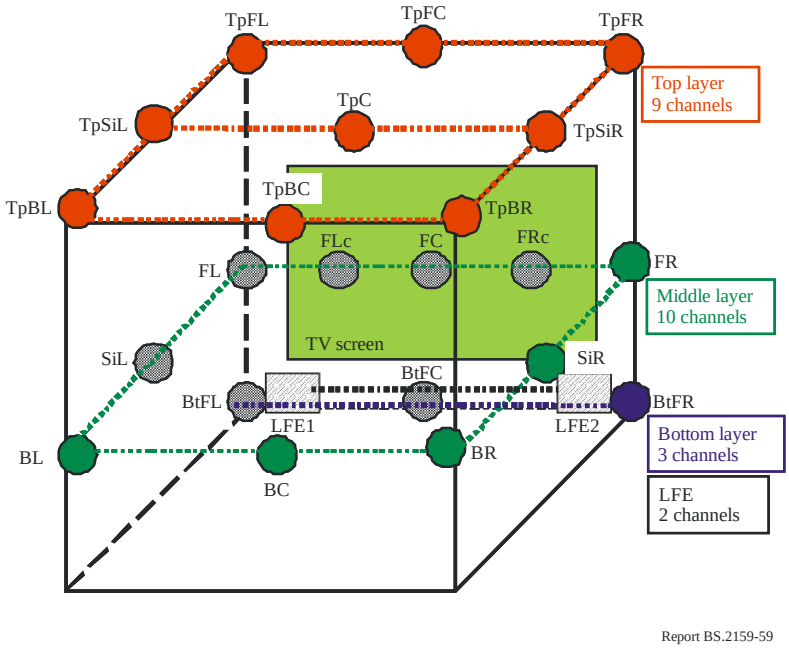
Channel maps and labels of 22.2 multichannel audio

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 centre
2/2	4	LFE1	LFE-1
3/1	5	BL	Back left
3/2	6	BR	Back right
4/1	7	FLc	Front left centre
4/2	8	FRc	Front right centre
5/1	9	BC	Back centre
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 centre
8/2	16	TpC	Top centre
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 centre
11/2	22	BtFC	Bottom front centre
12/1	23	BtFL	Bottom front left
12/2	24	BtFR	Bottom front right

3) Loudspeaker layout (informative)

Figure 71 illustrates the loudspeaker layout of a 22.2 multichannel sound system.

FIGURE 71



8.2 IEC

8.2.1 IEC 62574 “Audio, video and multimedia systems – General channel assignment of multichannel audio”

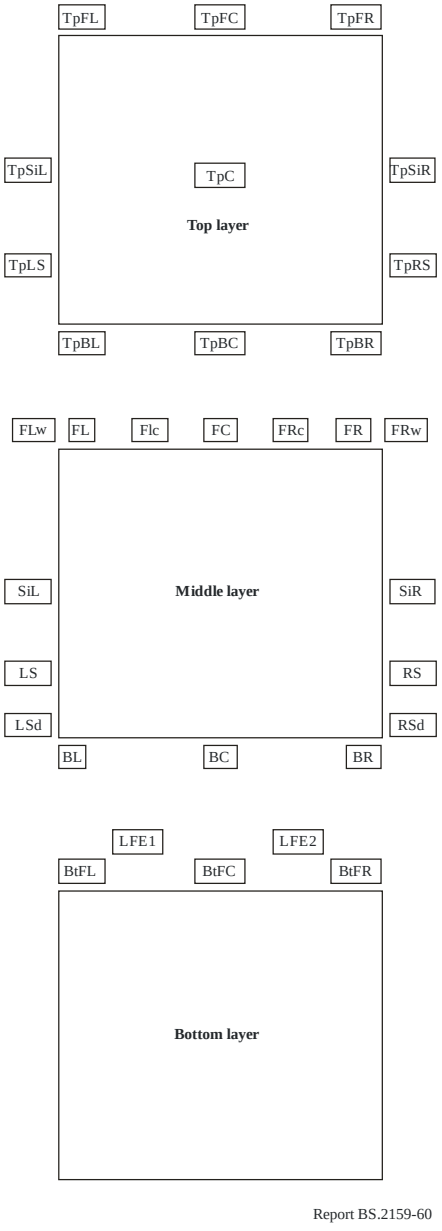
IEC 62574 “Audio, video and multimedia systems – General channel assignment of multichannel audio” specifies one general multichannel assignment. The general channel assignment as a channel mapping and labelling provides the unified usage of channel assignments for source devices, digital audio interfaces and sink devices. This standard excludes the specification of the exact position of each loudspeaker. It is aimed at consumer applications, but is not targeted for theatrical environments. Up to 32 labels for loudspeaker positions are specified, which can be used for all current multichannel formats.

Outline

IEC 62574 defines three layers, in each of which channels are assigned as depicted in Fig. 72. These are assigned channels and their labels, but they do not define exact loudspeaker positions.

FIGURE 72

Layers and channel assignment



Report BS.2159-60

Table 10 shows the channel maps and labels of the general channel assignment. Each channel label has an ID name as an abbreviation for the label.

TABLE 10

General channel assignment table

Channel number	Channel label ID name	Full name of ID
1/2	FL/FR	Front Left/Front Right
3/4	FC/LFE1	Front Centre/Low Frequency Effects-1
5/6	BL/BR	Back Left/Back Right
7/8	FLc/FRc	Front Left centre/Front Right centre
9/10	BC/LFE2	Back Centre/Low Frequency Effects-2
11/12	Sil/SiR	Side Left/Side Right
13/14	TpFL/TpFR	Top Front Left/Top Front Right
15/16	TpFC/TpC	Top Front Centre/Top Centre
17/18	TpBL/TpBR	Top Back Left/Top Back Right
19/20	TpSiL/TpSiR	Top Side Left/Top Side Right
21/22	TpBC/BtFC	Top Back Centre/Bottom Front Centre
23/24	BtFL/BtFR	Bottom Front Left/Bottom Front Right
25/26	FLw/FRw	Front Left wide/Front Right wide
27/28	LS/RS	Left Surround/Right Surround
29/30	LSd/RSd	Left Surround direct/Right Surround direct
31/32	TpLS/TpRS	Top Left Surround/Top Right Surround

8.3 MPEG (ISO/IEC JTC 1/SC 29/WG 11)

8.3.1 MPEG-2 AAC (ISO/IEC 13818-7) and MPEG-4 AAC (ISO/IEC 14496-3)

MPEG-2 and MPEG-4 Advanced Audio Coding (AAC), which have been standardized by ISO and IEC, specify a low-bit-rate audio coding scheme, which is able to include up to 48 audio channels in one stream. The technical corrigendum ISO/IEC 13818-7:2006/COR 1 on AAC has been approved to add the channel mapping for application-specific channel configurations. The audio coding signal of 3D sound systems including 22.2 multichannel audio can be transmitted by applying the application-specific channel configuration. The amendment ISO/IEC 14496-3:2009/ Amd 4:2013 has been approved to support the channel configuration of 22.2 multichannel audio as shown in Table 11.

8.4 EBU

8.4.1 EBU TECH 3306-2007, “RF64: An extended File Format for Audio”

The RF64 file format fulfils the longer-term need for multichannel sound in broadcasting and archiving. An RF64 file has additions to the basic Microsoft RIFF/WAVE specification to allow for either, or both:

– more than 4 Gbyte file sizes when needed;

– a maximum of 18 surround channels, stereo down-mix channel and bitstream signals with non-PCM coded data. This specification is based on the *Microsoft Wave Format Extensible* for multichannel parameters.

TABLE 11

Audio syntactic elements and channel alignment for an application-specific 22.2 channel configuration

Number of channels	Audio syntactic elements, listed in order received	Channel to speaker mapping
22+2	single_channel_element, channel_pair_element, channel_pair_element, channel_pair_element, channel_pair_element, single_channel_element, lfe_element, lfe_element, single_channel_element, channel_pair_element, channel_pair_element, single_channel_element, channel_pair_element, single_channel_element single_channel_element, channel_pair_element	centre front speaker, left, right front centre speakers, left, right front speakers, left, right side speakers, left, right back speakers, back centre speaker, left front low frequency effects speaker, right front low frequency effects speaker, top centre front speaker, top left, right front speakers, top left, right side speakers, centre of the room ceiling speaker, top left, right back speakers, top centre back speaker, bottom centre front speaker, bottom left, right front speakers

The file format is designed to be a compatible extension to the Microsoft RIFF/WAVE format and to the BWF format and its supplements and additional chunks. It extends the maximum size capabilities of the RIFF/WAVE and BWF format allowing for multichannel sound in broadcasting and audio archiving.

RF64 can be used in the entire programme chain from capture to editing and play out and for short or long term archiving of multichannel files.

An RF64 file with a bext chunk becomes an MBWF (multichannel BWF) file.

The following are specifications about audio channels:

1. Definition of a new format, RF64.

The wave format extensible channel mask contains 18 “#define” settings specifying different loudspeaker positions (or channel allocations).

Microsoft Wave Format Extensible Channel Mask

```
#define SPEAKER_FRONT_LEFT              0x00000001
#define SPEAKER_FRONT_RIGHT             0x00000002
#define SPEAKER_FRONT_CENTRE            0x00000004
#define SPEAKER_LOW_FREQUENCY           0x00000008
#define SPEAKER_BACK_LEFT               0x00000010
#define SPEAKER_BACK_RIGHT              0x00000020
#define SPEAKER_FRONT_LEFT_OF_CENTRE    0x00000040
#define SPEAKER_FRONT_RIGHT_OF_CENTRE   0x00000080
#define SPEAKER_BACK_CENTRE             0x00000100
#define SPEAKER_SIDE_LEFT               0x00000200
#define SPEAKER_SIDE_RIGHT              0x00000400
#define SPEAKER_TOP_CENTRE              0x00000800
#define SPEAKER_TOP_FRONT_LEFT          0x00001000
#define SPEAKER_TOP_FRONT_CENTRE        0x00002000
#define SPEAKER_TOP_FRONT_RIGHT         0x00004000
#define SPEAKER_TOP_BACK_LEFT           0x00008000
#define SPEAKER_TOP_BACK_CENTRE         0x00010000
#define SPEAKER_TOP_BACK_RIGHT          0x00020000
```

8.5 Japan

8.5.1 Advancement of satellite digital broadcasting

The advanced satellite digital broadcasting of UHDTV (4K and 8K) has been in operation in Japan since December 2018. Owing to the utilisation of state-of-the-art technologies, the advanced satellite broadcasting systems enable the effective use of the spectrum and the introduction of new services. The technical specifications were standardised by the Ministry of Internal Affairs and Communications of Japan and the Association of Radio Industries and Businesses (ARIB) of Japan in 2014. ARIB Standard STD-B59 specifies loudspeaker configurations for 22.2 multichannel sound (system H specified in Recommendation ITU-R BS.2051) and 7.1 multichannel sound (system C specified in Recommendation ITU-R BS.2051), and ARIB Standard STD-B32 Part 2 provides specifications for audio coding including audio source formats and audio coding methods.

The audio coding for the advanced satellite broadcasting system has a number of new features, including transmission of the downmixing coefficients for each programme, dialogue level control and dialogue replacement, and lossless transmission. The audio coding specifications, the downmixing coefficients, and the metadata related to dialogue level control and dialogue replacement are described in Tables 12, 13 and 14, respectively.

TABLE 12

Audio coding for advanced satellite digital broadcasting

Audio codec	Audio mode	Maximum bit rate (kbit/s)	Sampling (kHz)	Bit depth (bits)	Stream
MPEG-4 AAC (AAC profile, object-type LC)	Stereo	256	48	24	LATM/ LOAS
	5.1 (3/2.1) (1)	480			
	7.1 (2) (2/0/0+3/0/2.1)	640			
	22.2 (3) (3/3/3+5/2/3+3/0/0.2)	1 920			
MPEG-4 ALS (ALS simple profile)	Stereo (4)	2 429			
	5.1 (4) (3/2.1)	7 341			

(1) Audio mode of 5.1 multichannel sound with MPEG-4 AAC may accompany simulcast of stereo sound with MPEG-4 AAC.

(2) Audio mode of 7.1 multichannel sound with MPEG-4 AAC requires simulcast of stereo sound with MPEG-4 AAC and may also accompany simulcast of 5.1 multichannel sound with MPEG-4 AAC.

(3) Audio mode of 22.2 multichannel sound with MPEG-4 AAC requires simulcast of stereo sound with MPEG-4 AAC and may also accompany simulcast of 5.1 multichannel sound with MPEG-4 AAC.

(4) Audio mode of stereo or 5.1 multichannel sound with MPEG-4 ALS requires simulcast of stereo or 5.1 multichannel sound with MPEG-4 AAC and may also accompany simulcast of 7.1 or 22.2 multichannel sound with MPEG-4 AAC.

TABLE 13

Default values of downmixing coefficients to 3/2 multichannel sound system (MPEG-4 AAC)

System B (0+5+0.1)	System C (2+5+0.1)	System H (9+10+3.2)
C’	C	FC + (FLc + FRc)×c + (TpFC + TpC×f + BtFC)×e
L’	L×a + Lv×b	FL + FLc×c + SiL×d + (TpFL + TpSiL×d + BtFL)×e
R’	R×a + Rv×b	FR + FRc×c + SiR×d + (TpFR + TpSiR×d + BtFR)×e
Ls’	Ls	BL + BC×g + SiL×d + (TpBL + TpBC×g + TpSiL×d + TpC×f)×e
Rs’	Rs	BR + BC×g + SiR×d + (TpBR + TpBC×g + TpSiR×d + TpC×f)×e
LFE’	LFE	(LFE1 + LFE2)×l

a: −3 dB, b: −3 dB, c: −4.5 dB, d: −4.5 dB, e: 0 dB, f: −6 dB, g: −3 dB, l: −3 dB

TABLE 14

Additional ancillary data related to dialogue level control and dialogue replacement (MPEG-4 AAC)

Descriptor	Explanation
ext_dialogue_status	Existence of dialogue channels.
num_dialogue_chans	Number of main dialogue channels.
num_additional_lang_chans	Number of alternative dialogue channels.
dialogue_src_index[i]	Index of dialogue channels.
dialogue_main_lang_comment_bytes	Byte count of characters indicating content of main dialogue.
dialogue_main_lang_comment_data	Byte data of characters indicating content of main dialogue.
dialogue_main_lang_code	Language code of main dialogue.
dialogue_additional_lang_code[i]	Language code of *i*th alternative dialogue.
dialogue_additional_lang_comment_bytes[i]	Byte count of characters indicating content of *i*th alternative dialogue.
dialogue_additional_lang_comment_data[i]	Byte data of characters indicating content of *i*th alternative dialogue.
dialogue_gain_index[i]	Gain of alternative dialogue channels. (0000: 0 dB, 0001: −1 dB, 0010: −2 dB, ..., 1110: –14 dB, 1111: −∞ dB)
sn_dialogue_plus_index	Maximum gain of dialogue channels in receiver. (000: 0 dB, 001: +3 dB, 010: +6 dB, ..., 110: +18 dB, 111: +∞ dB)
sn_dialogue_minus_index	Minimum gain of dialogue channels in receiver. (000: 0 dB, 001: −3 dB, 010: −6 dB, ..., 110: −18 dB, 111: −∞ dB)
additional_dialogue_data_sync	Data stream element in which alternative dialogue data is packed.
additional_dialogue_index	Index of alternative dialogue channels corresponding to the “i” of dialogue_additional_lang_code[i].