International Telecommunication Union

ITU-R

Radiocommunication Sector of ITU

Report ITU-R BS.2159-5

(11/2012)

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 Annex 1 of 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, 2013

REPORT ITU-R BS.2159-5

Multichannel sound technology in home and broadcasting applications

(2009-2010-05/2011-10/2011-05/2012-11/2012)

TABLE OF CONTENTS

Page

1 Introduction

ITU-R has developed Recommendation ITU-R BS.775 for multichannel stereophonic 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 and large screen digital imaginary (LSDI) application, 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 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.

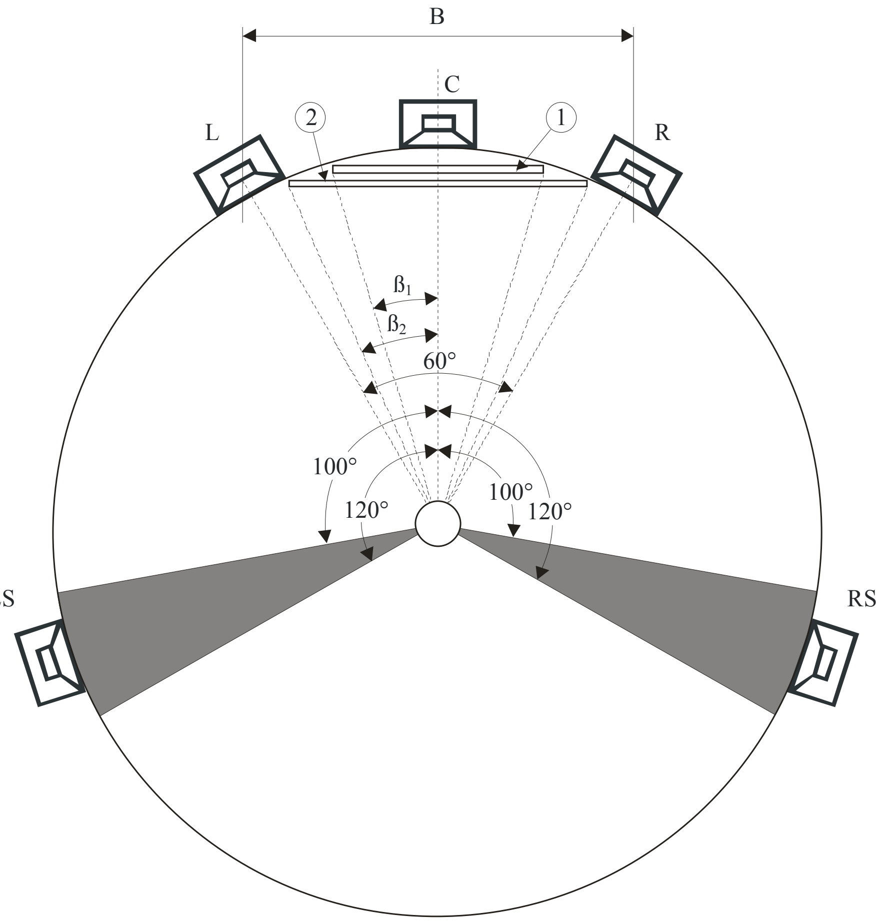
There would be value in continued studies in this area for future broadcasting applications in order to evolve beyond the current 5.1 channel sound system. This Report contains information on the subject of – Multichannel sound technology in home and broadcasting applications, beyond the current 5.1 channel 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

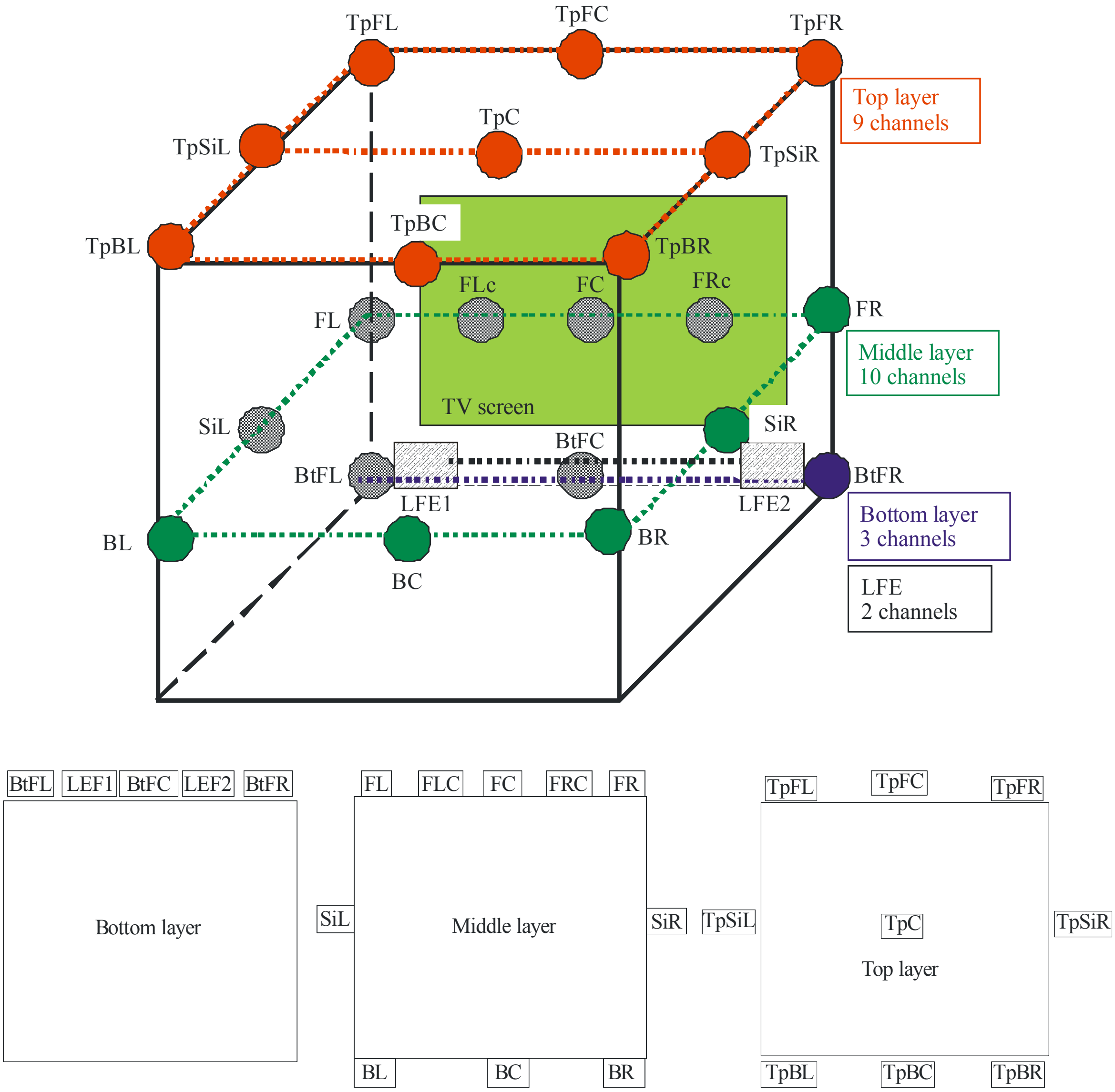
Several multichannel sound systems have been studied to improve the spatial impression of sound. The following systems seem to have the capability for practical use.

4.1 22.2 multichannel sound system

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 display, because it can localize two-dimensionally 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.

Audio characteristics and audio channel mapping for UHDTV programme production for the 22.2 multichannel sound format has been standardized 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.

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 production/demonstration room in NHK Science & Technology Research Laboratories.
- One laboratory installations in Fraunhofer IIS.
- One laboratory installations in McGill University CIRMMT.

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

Home sound systems for 22.2 multichannel sound have also been developed. Figure 4 shows the home 22.2 sound system using multiple compact loudspeakers.

FIGURE 4

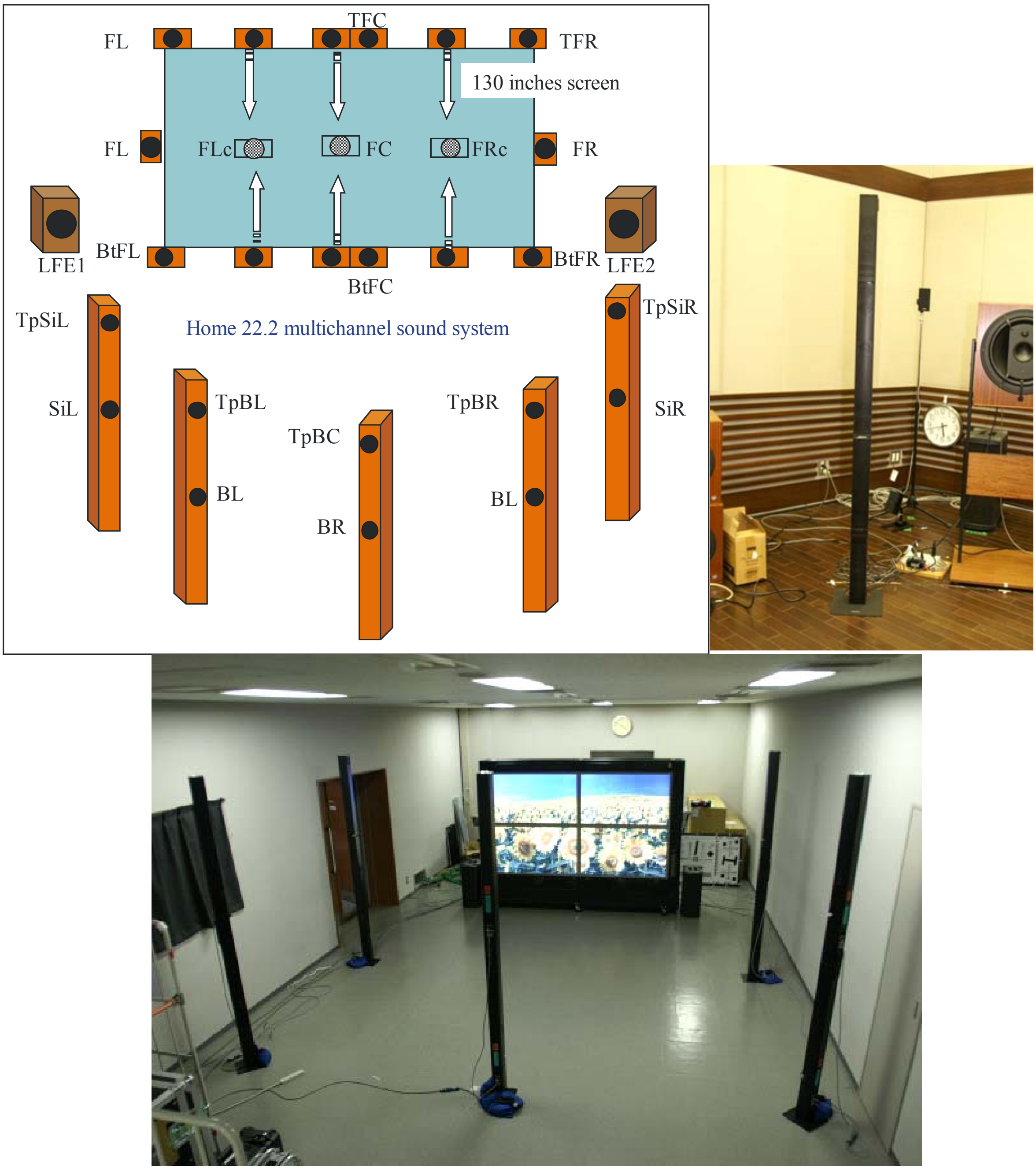
Home 22.2 multichannel sound system using multiple compact loudspeakers

Report BS.2159-04

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 a broadcast 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. > 2500 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, Korea.
- 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.

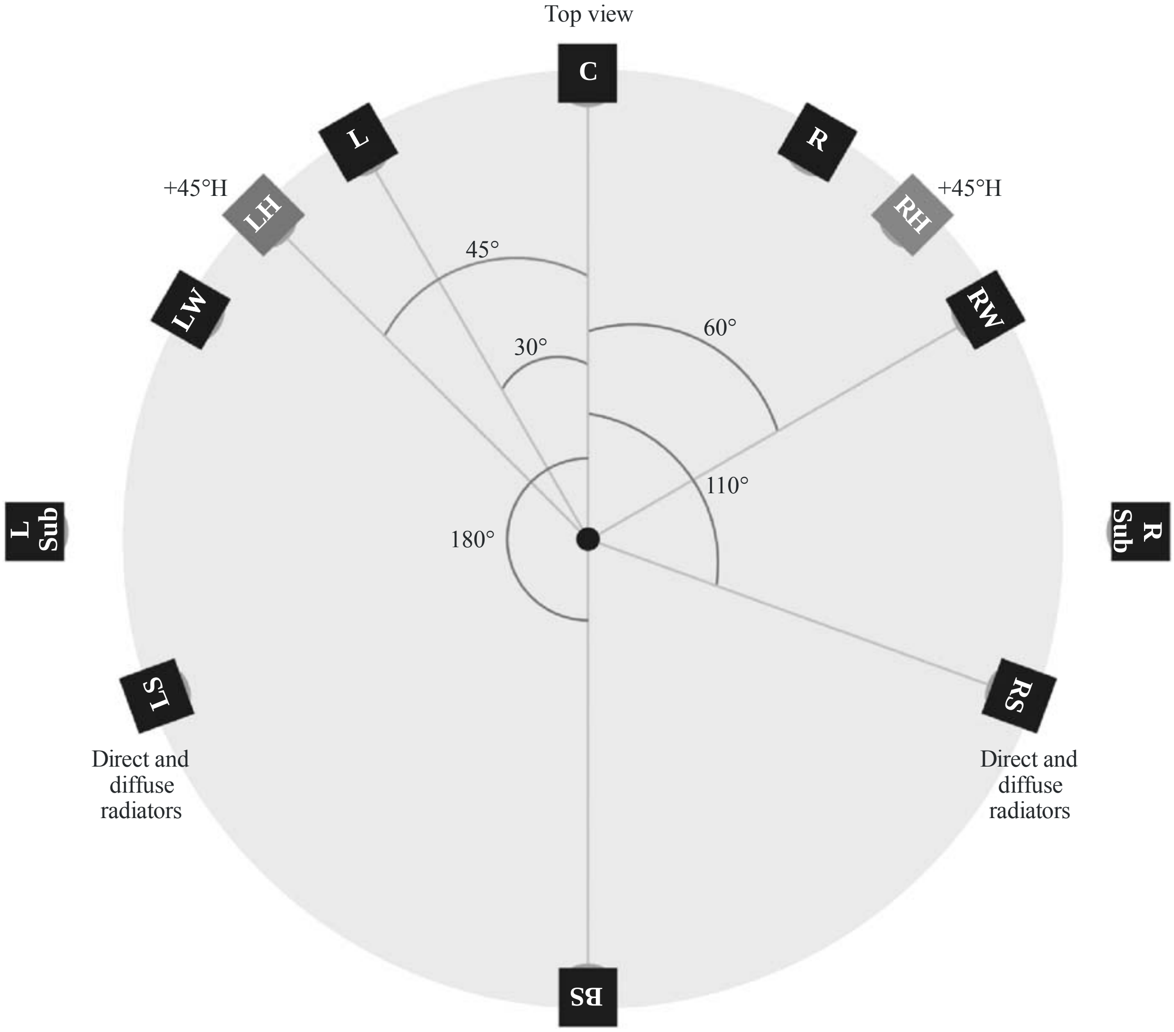
The consolidated positions of Left Surround direct and diffuse radiators, and Right Surround direct and diffuse radiators (applicable in "small" rooms), result in 10.2 total speaker locations. The system thus requires 14 electrical channels. Additionally, two channels of a sixteen-channel layout are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

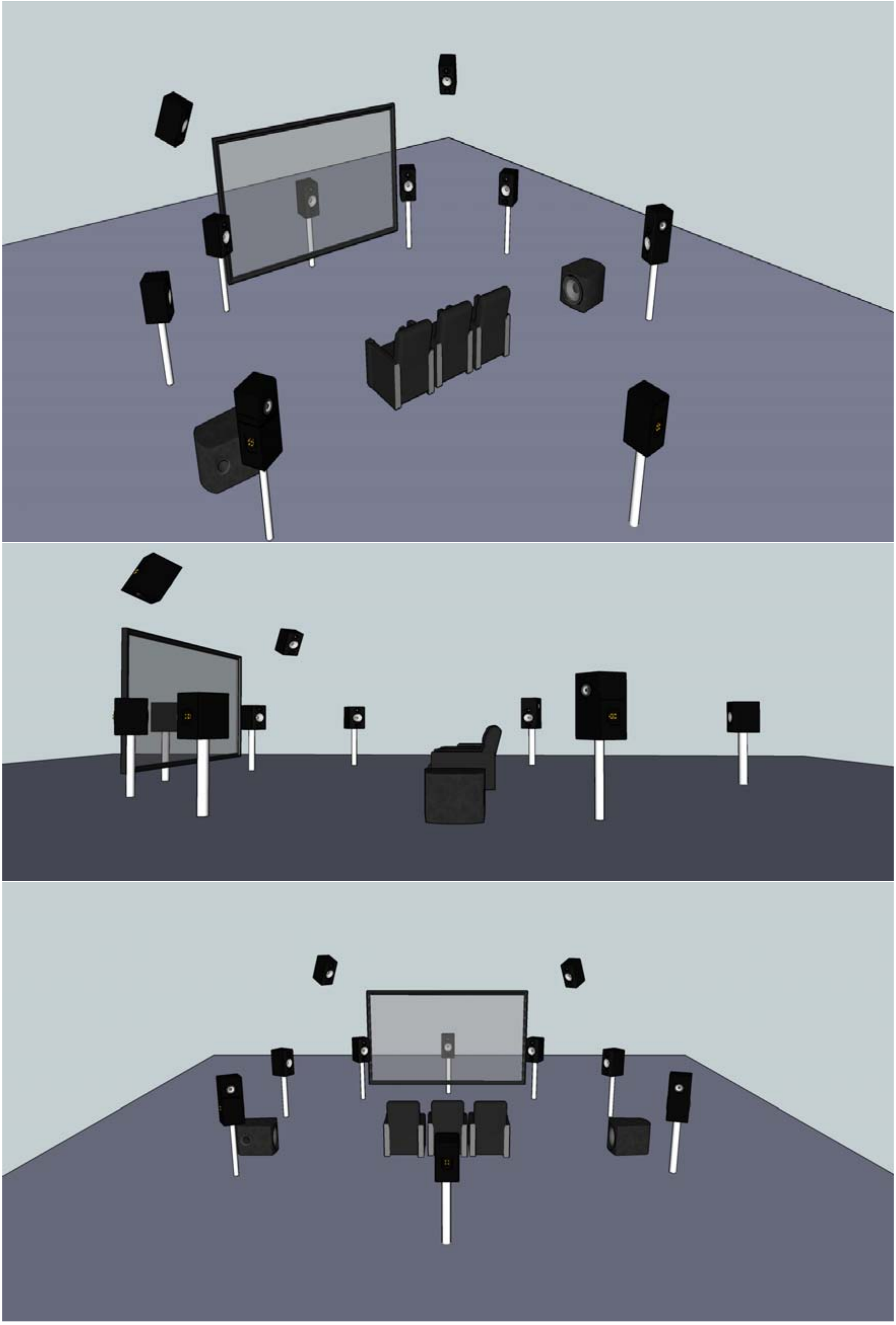
Diagram of speaker layout for 10.2



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

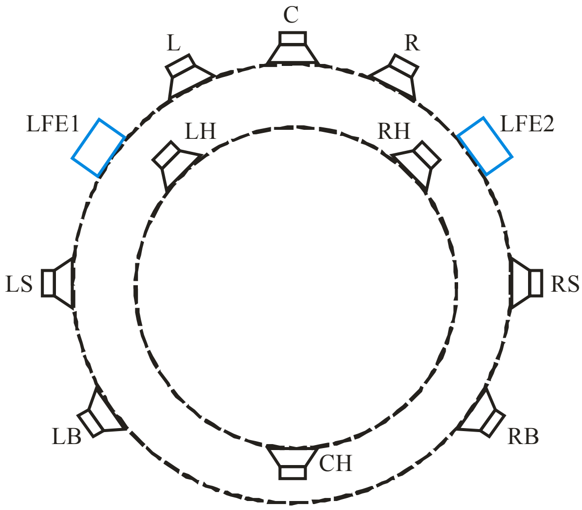
Channel definition of 10.2 channel

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



Report BS.2159-09

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:

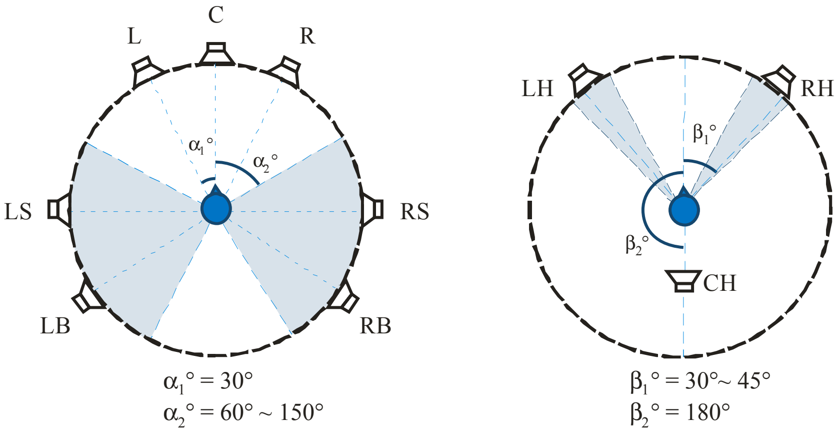
– Left Height (LH)-30~-45° in middle layer with +30~+45° elevated. Reference point is the ear level of listener and this channel positioned in top layer.
– Right Height (RH)+30~+45° in middle layer with +30~+45° elevated and reference position as LH.
– Centre Height (CH)+90~+135° elevated and reference 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

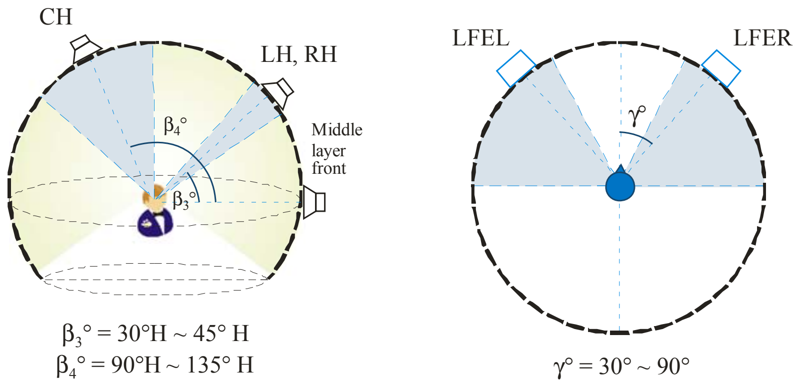
Middle layer and top layer of 10.2ch loudspeaker layout



Report BS.2159-10

FIGURE 11

Top layer and bottom layer of 10.2ch loudspeaker layout



Report BS.2159-11

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:

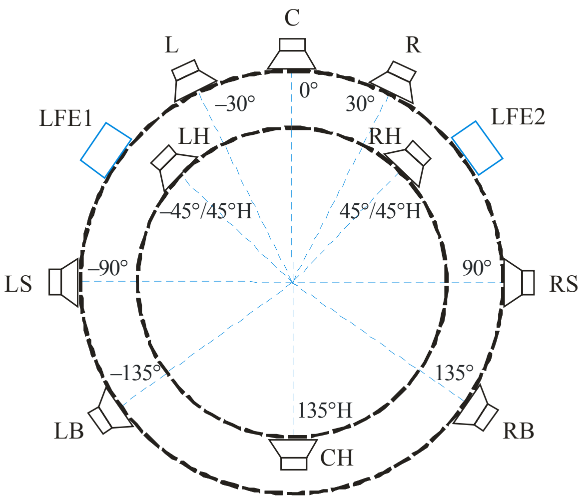
TABLE 3

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



Report BS.2159-12

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

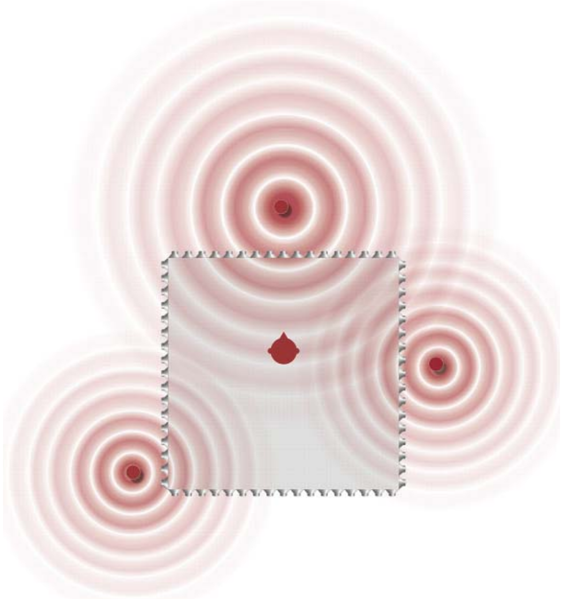
WFS sound system in the German cinema "Linden lichtspiele ilmenau"

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

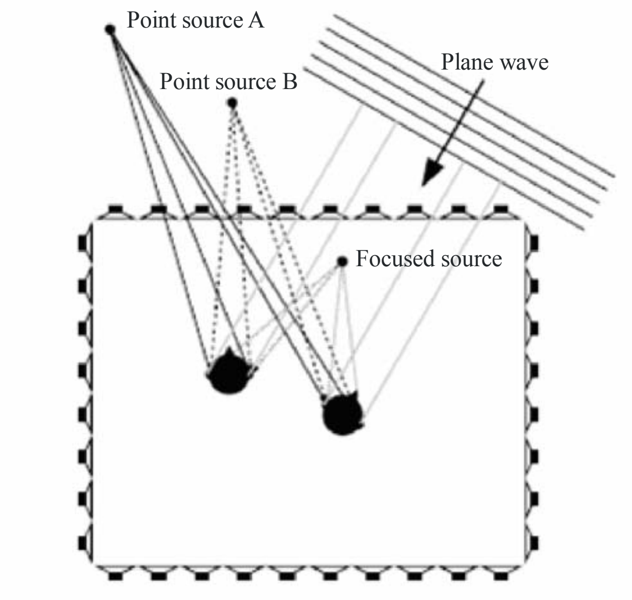
Three representations of sound sources are possible in WFS (see Fig. 15). In the first two, virtual sound sources can be placed behind the loudspeaker arrays (so-called point sources) as well as in front of loudspeaker arrays (so-called focused sources). In the case of sound sources in front of the array, 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[1]. 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[2].

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 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

---

[1] 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).

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

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.

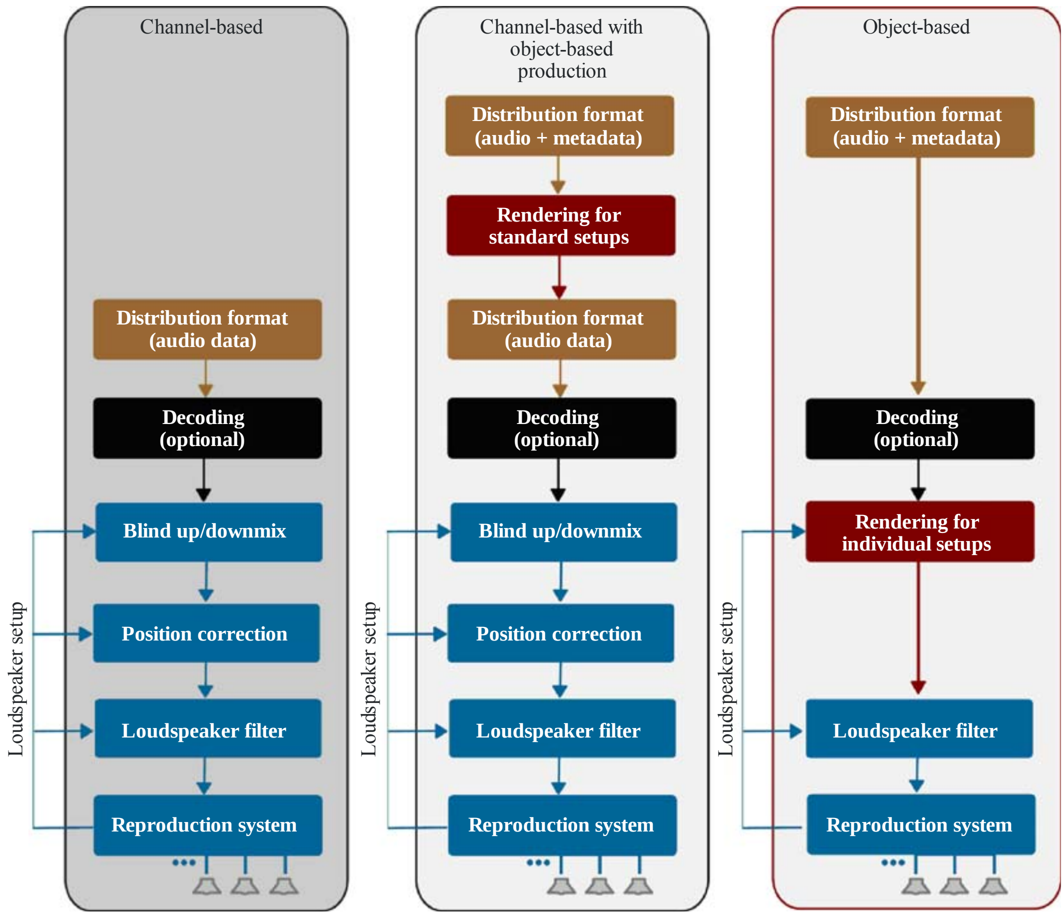
In an object-oriented system the audio content is created independently of any specific loudspeaker layout. The audio content is represented as audio objects containing the pure audio content, together with metadata describing the position of the audio object along with the properties of the audio object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

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

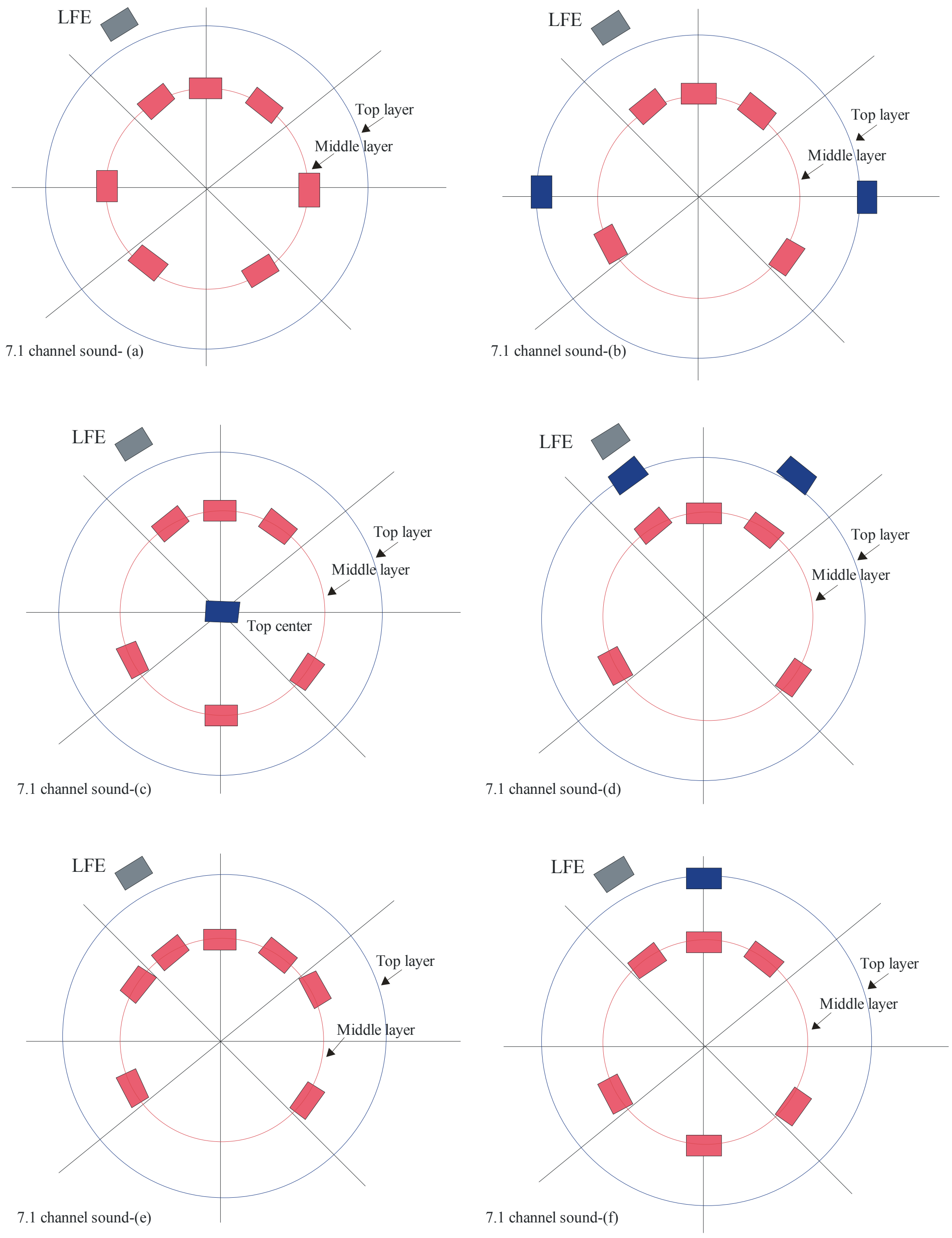
Specification of BD-ROM audio streams

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. 20. 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 20

Examples of loudspeaker mapping of 7.1 channel sound



Report BS.2159-20

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 3-D 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 localized 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 layer.

– 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 developing for producing 3D sound. This system currently includes a Digital Audio Workstation and a sound mixing console with 3D pan on each channel strip, 3D audio signal compressor on 24-channel master bus and a down-mixing function. It can mix over 1 000 sound tracks to produce 22.2 multichannel sound.

FIGURE 21

22.2 multichannel sound mixing console

Report BS.2159-21

6.2.3 Examples of three-dimensional sound live mixing

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. 22.

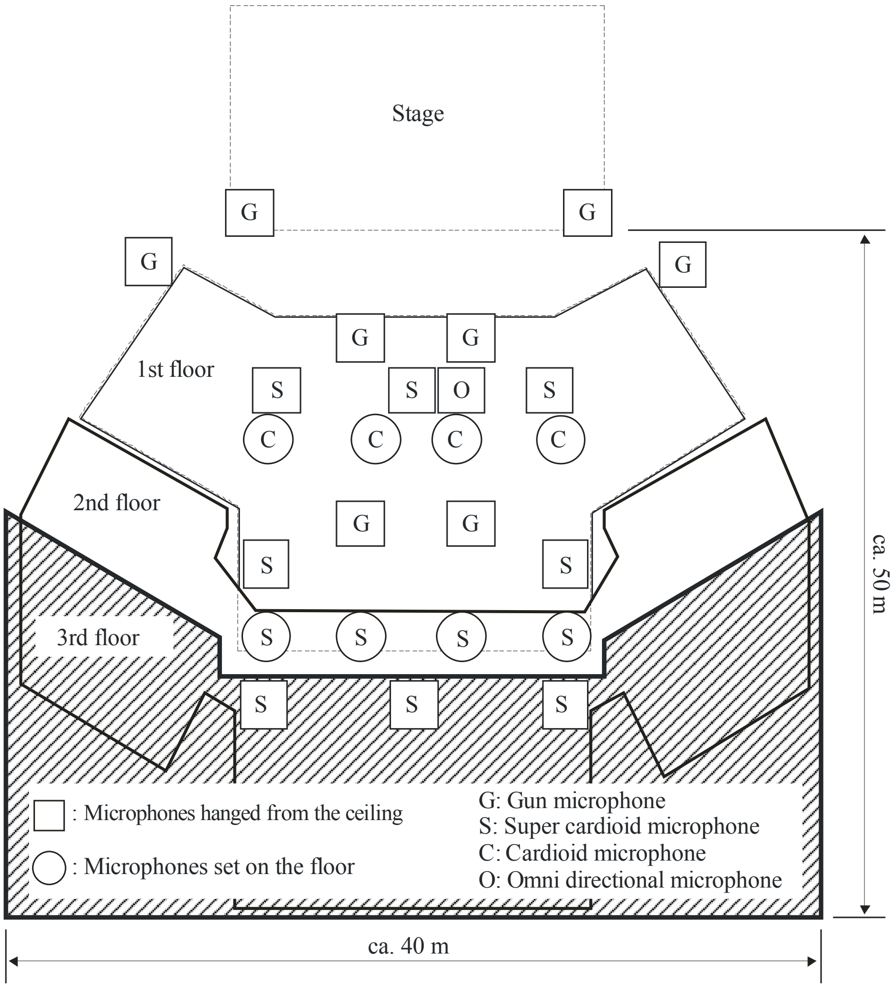
– Reflection and reverberation in the auditorium of NHK Hall, which are captured by microphones hung from the ceiling, are reproduced by the top layer loudspeakers of the 22.2 multichannel sound system to widen the listening area and create a sense of the listener being enveloped.

– 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 22

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



Report BS.2159-22

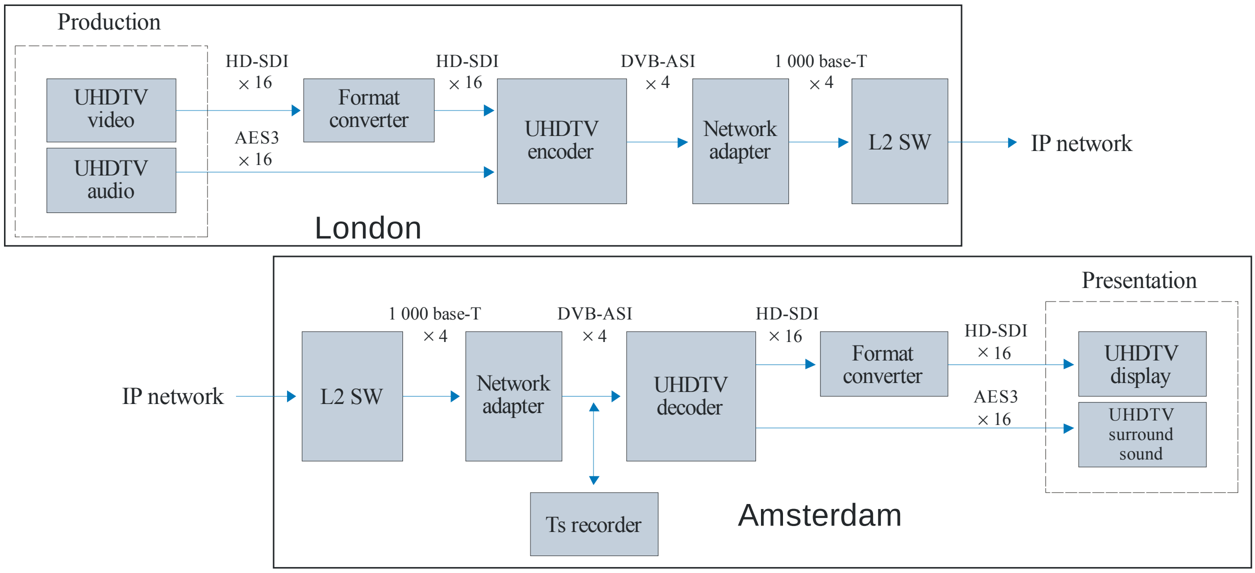
6.2.3.2 Emulated live news reports demonstrated at IBC2008

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. 23.

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 23

UHDTV and 22.2 multichannel sound IP transmission system at IBC2008



Report BS.2159-23

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. 24. 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.

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.

FIGURE 24

Microphone array

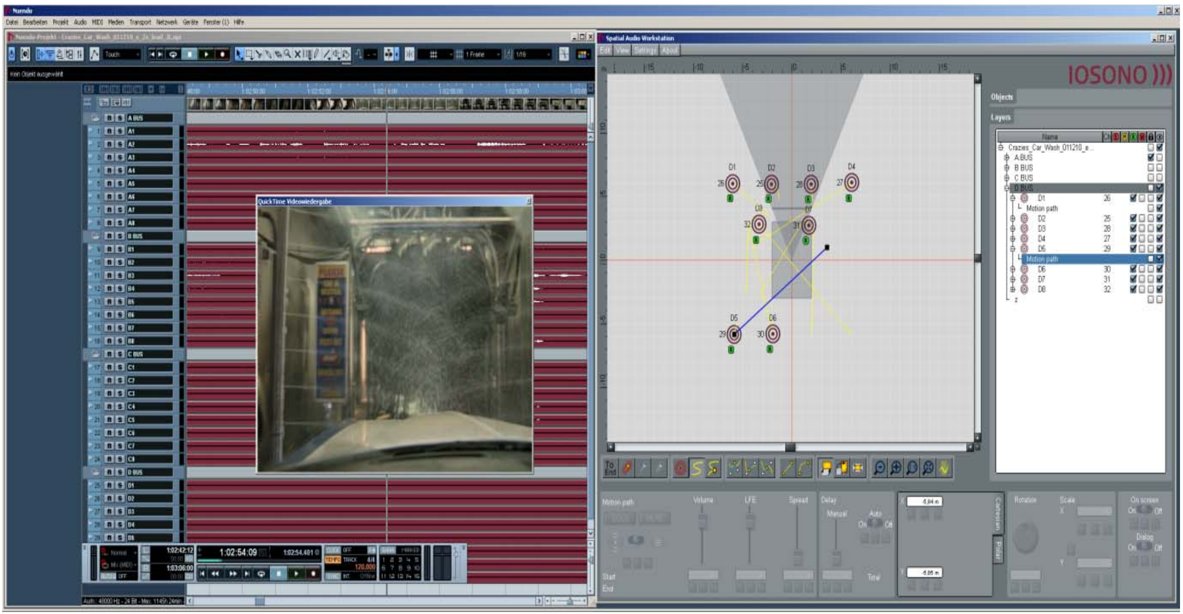
Report BS.2159-24

6.4 Object-based post-production system

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. 25).

FIGURE 25

Spatial audio workstation used for a motion picture sound track production



Report BS.2159-25

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 26

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

Report BS.2159-26

Several productions have been performed in Todd-AO (Fig. 26). Besides a number of trailers and demos, a complete motion picture sound-track has been produced using the spatial audio workstation (Fig. 26).

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 32 and 33 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 32 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 33 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.

NOTE 1 – Technical information on subjective evaluation results of elevation perception of phantom sound images is provided as Attachment 1.

NOTE 2 – Technical information on subjective evaluation results of the sensation of "listener's envelopment (LEV)", which is one of the primary features of a three-dimensional spatial impression, is provided as Attachment 2.

NOTE 3 – Technical information on subjective evaluation results of the localization and localization uncertainty of phantom sound images in the elevation direction generated by two loudspeakers located above the listener is provided as Attachment 3.

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. 27(a)) was reference and Top 4 channel (Fig. 27(b)), Top 3 channel (Fig. 27(c)), Top 2 channel (Fig. 27(d)), Top 0 channel (Fig. 27(e)) layout were tested.

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

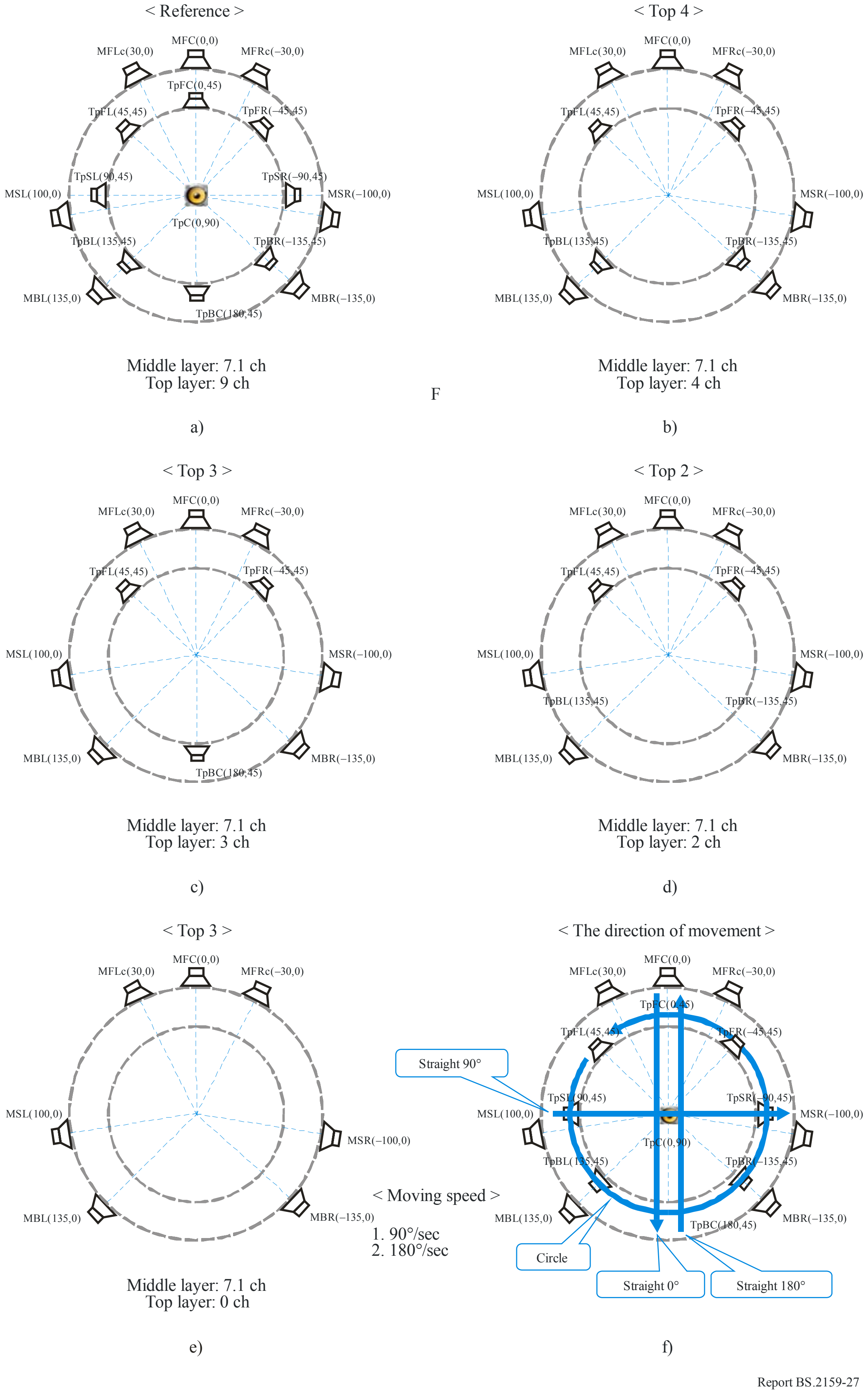
  Directions were circle and straight 0°/90°/180° (Fig. 27(f)).

  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 27

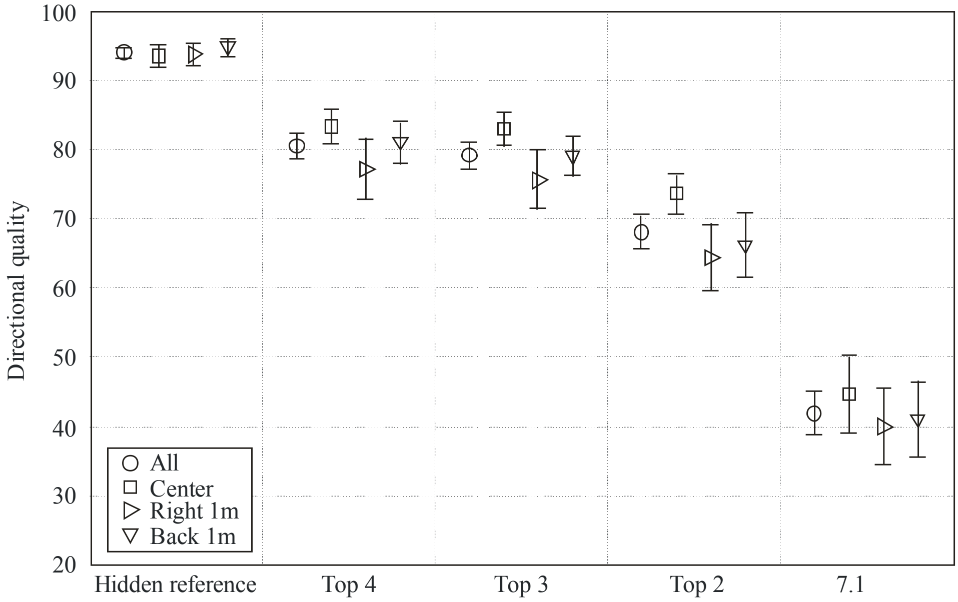
The layout of directional audio quality evaluation



– Result of experiments: imperceptible at Top 3 channel.

FIGURE 28

The result of directional audio quality evaluation



Report BS.2159-28

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. 29(a)), 11.2 (Fig. 29(b)), 10.2 (Fig. 29(c)), 7.1 (Fig. 29(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).

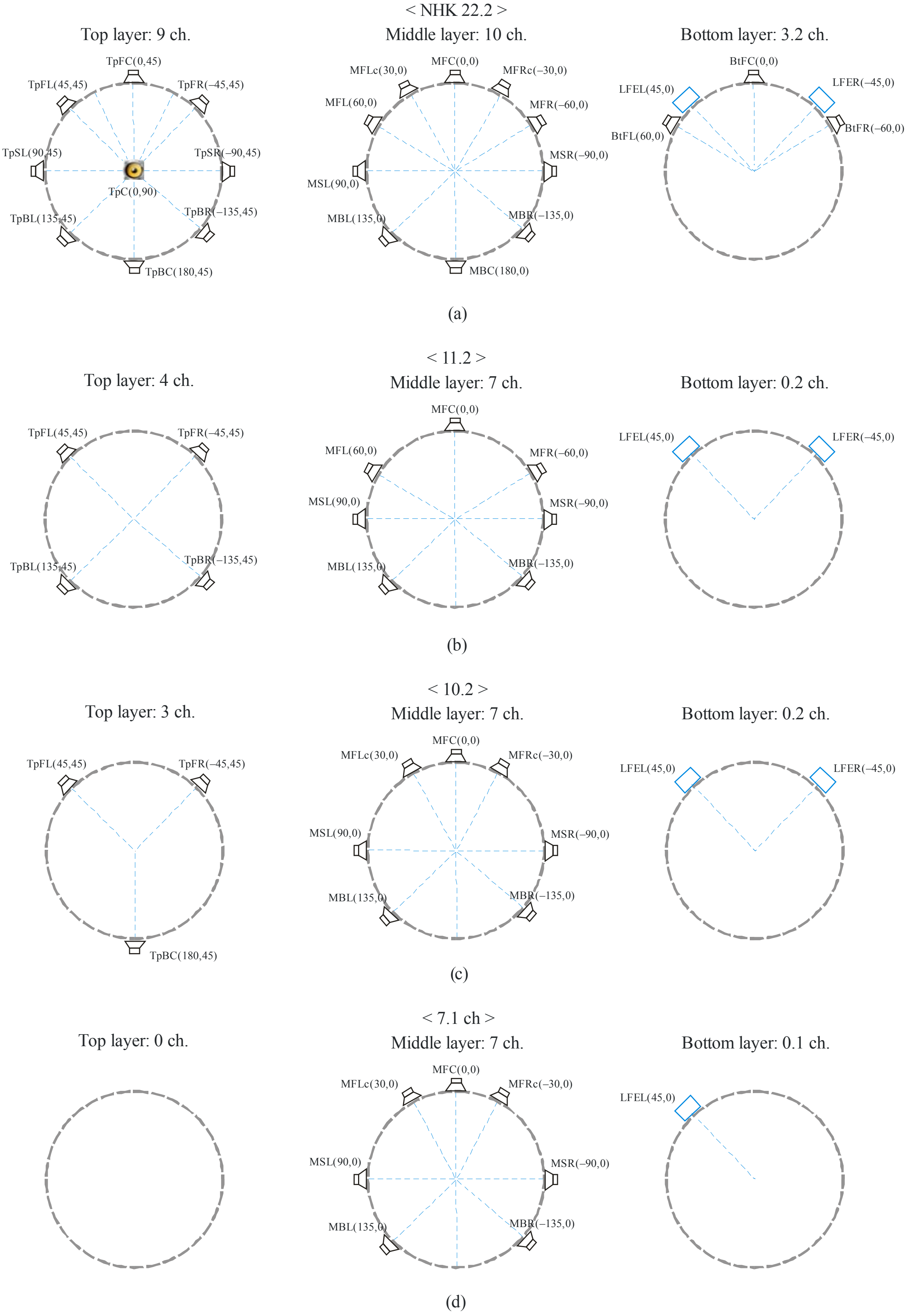
Slightly annoying (60-40).

Annoying (40-20).

Very annoying (20-0).

FIGURE 29

The layout of overall audio quality evaluation



Report BS.2159-29

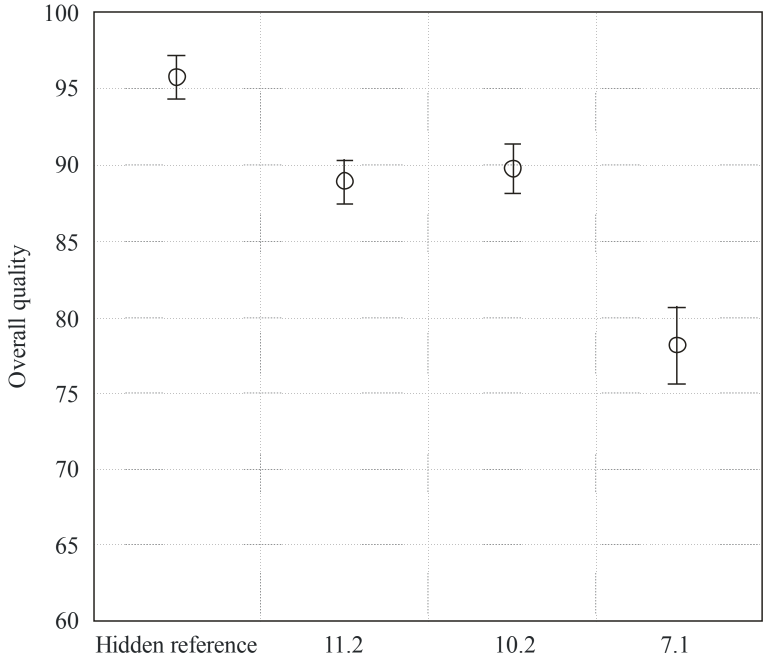
– Result of experiments:

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

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

FIGURE 30

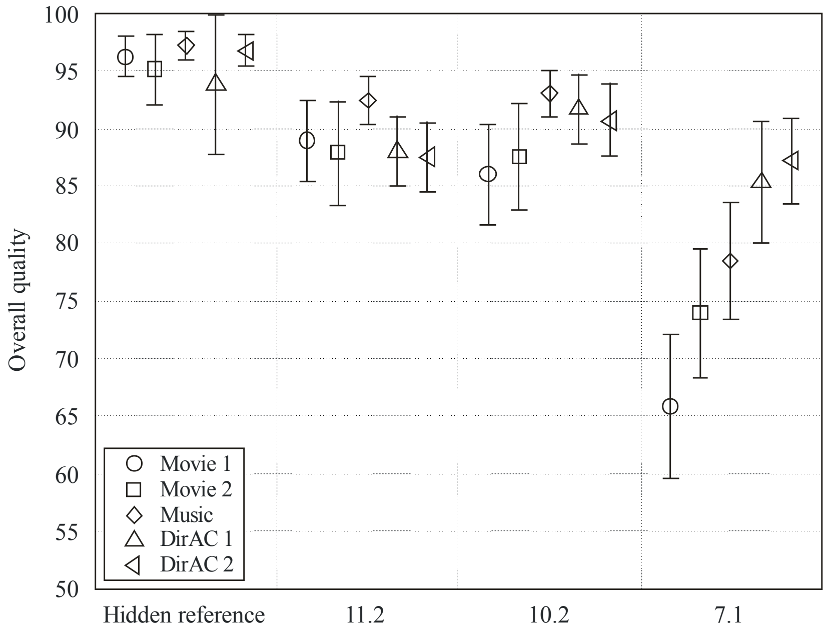
The result of overall audio quality evaluation



Report BS.2159-30

FIGURE 31

The result of overall audio quality evaluation by contents



Report BS.2159-31

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"

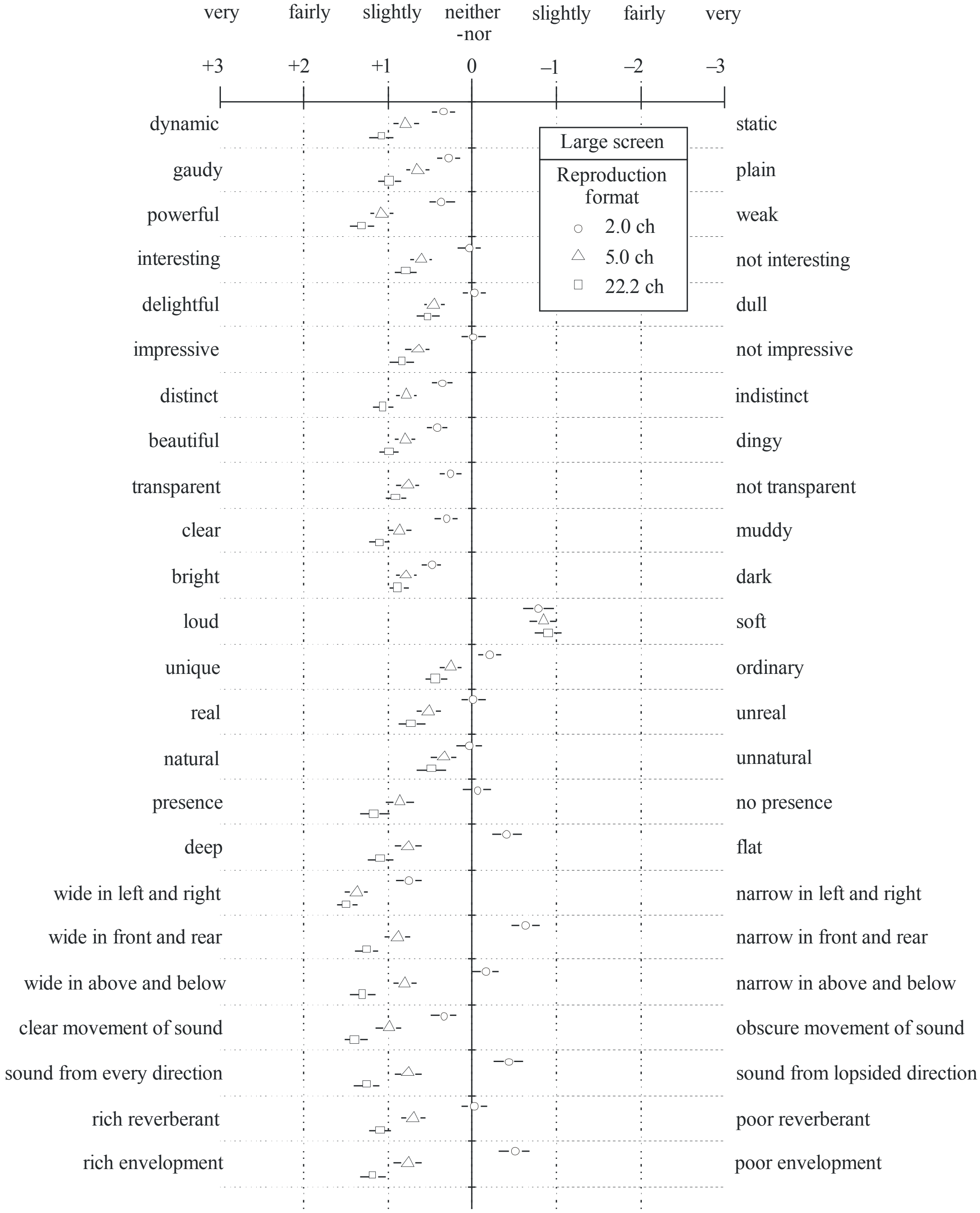
SMPTE 2036 Ultra High Definition Television (UHDTV) suite of documents is in multiple parts:

– Part 1: Image parameter values for programme production.

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

FIGURE 32

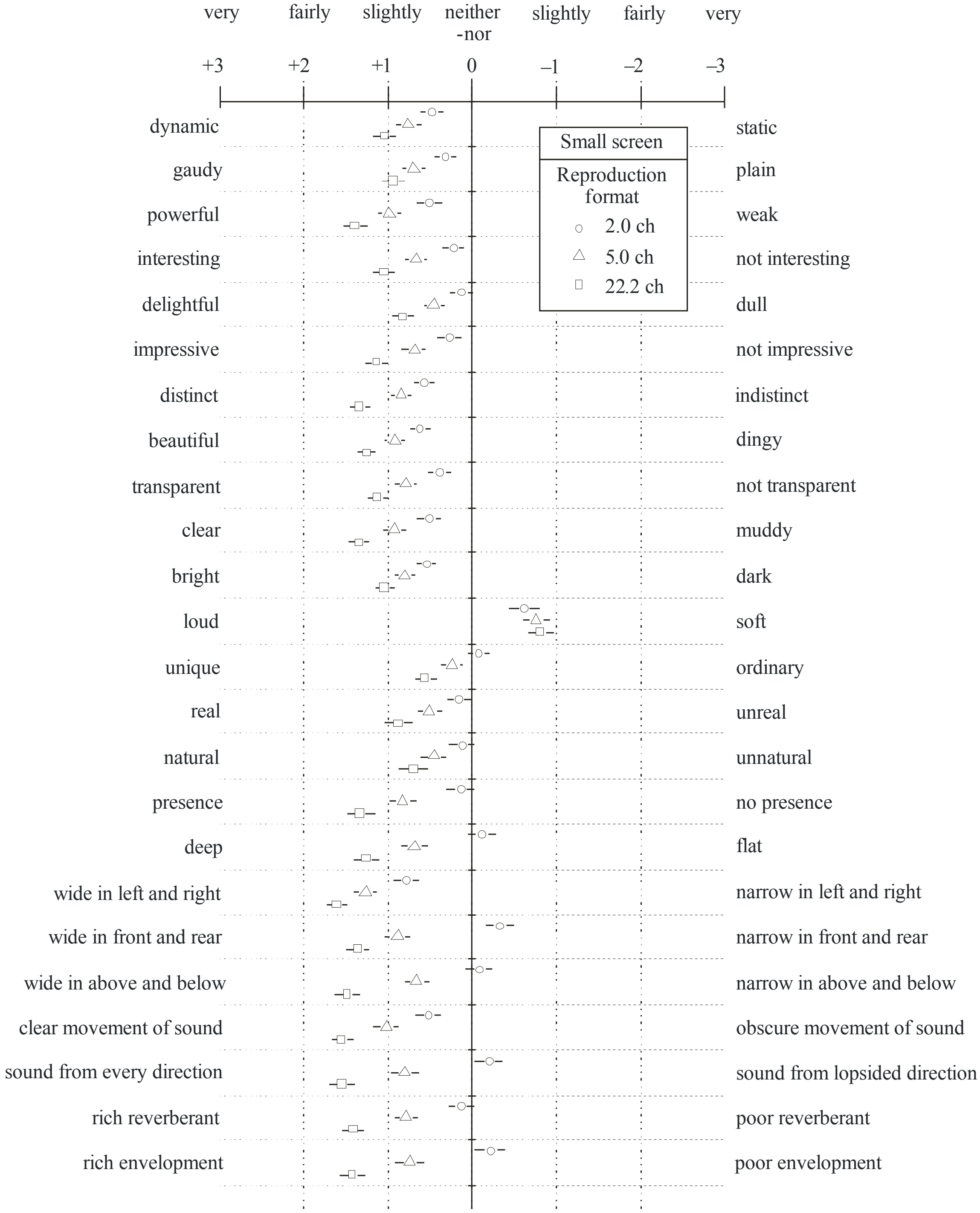
Results of subjective evaluation comparing three different sound systems with a large screen in a large theatre



Report BS.2159-32

FIGURE 33

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



Report BS.2159-33

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 5

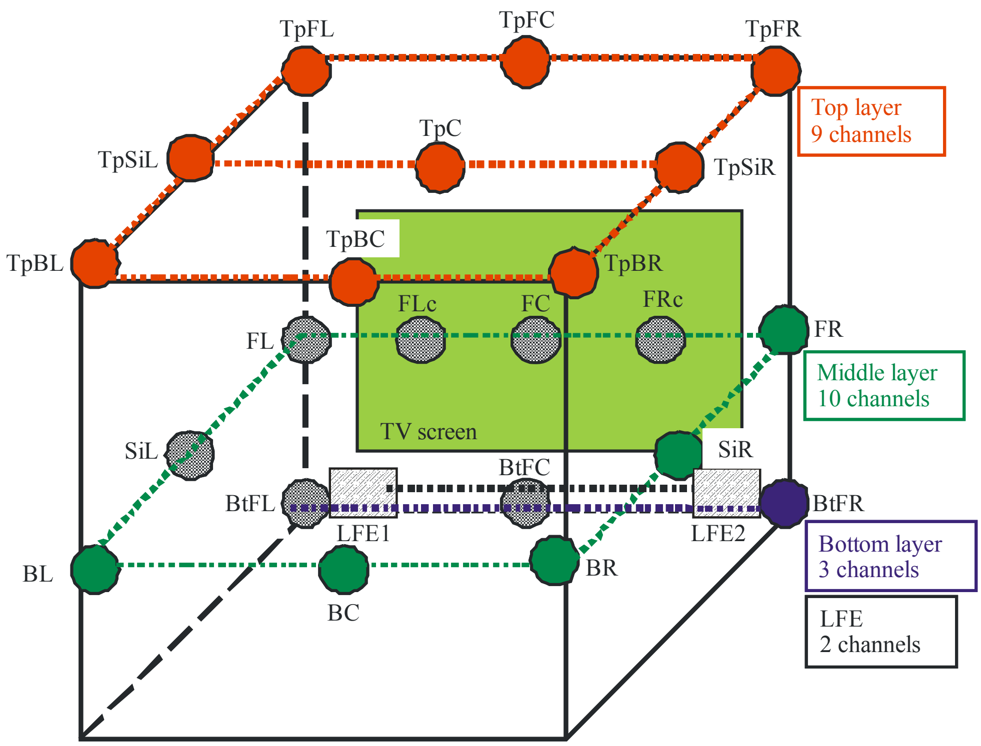
Channel maps and labels of 22.2 multichannel audio

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 34 illustrates the loudspeaker layout of a 22.2 multichannel sound system.

FIGURE 34



Report BS.2159-34

8.2 IEC

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

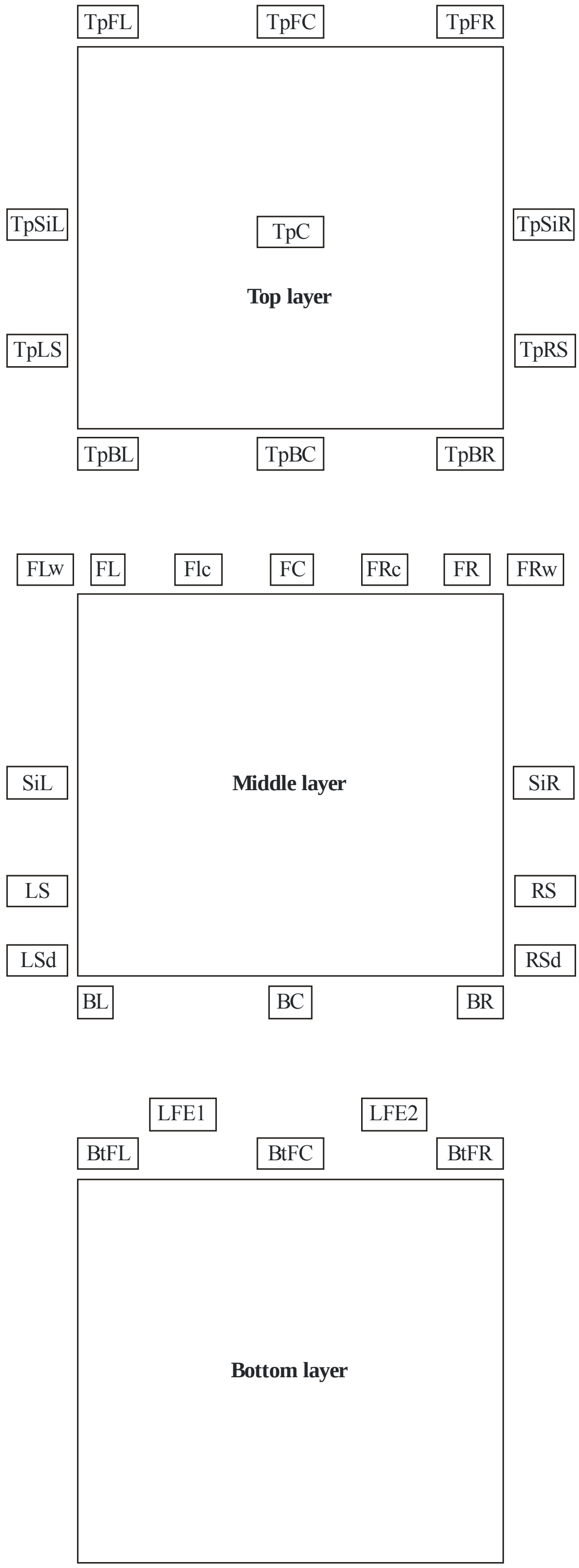
IEC 62574 "Audio, video and multimedia systems – General channel assignment of multichannel audio" specifies one general multichannel assignment. The general channel assignment as a channel mapping and labelling provides the unified usage of channel assignments for source devices, digital audio interfaces 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. 35. These are assigned channels and their labels, but they do not define exact loudspeaker positions.

FIGURE 35

Layers and channel assignment



Report BS.2159-35

Table 6 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 6

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)

MPEG-2 Advanced Audio Coding (AAC), which has been standardized by ISO and IEC, specifies low bit rate audio coding scheme which able to include up to 48 audio channels in one stream. The technical corrigendum ISO/IEC 13818-7:2006/COR 1 on Advance Audio Coding (AAC) has been approved to add the channel mapping for application specific channel configurations. The audio coding signal of 3D sound system including 22.2 multichannel audio can be transmitted by applying the application specific channel configuration.

Table 7 shows the example of an application specific channel alignment for a 22.2 channel configuration.

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 7

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_CENTER            0x00000004
#define SPEAKER_LOW_FREQUENCY           0x00000008
#define SPEAKER_BACK_LEFT               0x00000010
#define SPEAKER_BACK_RIGHT              0x00000020
#define SPEAKER_FRONT_LEFT_OF_CENTER    0x00000040
#define SPEAKER_FRONT_RIGHT_OF_CENTER   0x00000080
#define SPEAKER_BACK_CENTER             0x00000100
#define SPEAKER_SIDE_LEFT               0x00000200
#define SPEAKER_SIDE_RIGHT              0x00000400
```

```
#define SPEAKER_TOP_CENTER            0x00000800
#define SPEAKER_TOP_FRONT_LEFT        0x00001000
#define SPEAKER_TOP_FRONT_CENTER      0x00002000
#define SPEAKER_TOP_FRONT_RIGHT       0x00004000
#define SPEAKER_TOP_BACK_LEFT         0x00008000
#define SPEAKER_TOP_BACK_CENTER       0x00010000
#define SPEAKER_TOP_BACK_RIGHT        0x00020000
```

8.5 Japan

8.5.1 Advancement of satellite digital broadcasting

The advancement of satellite digital broadcasting has been studied in Japan. Due to its utilization of state-of-the-art technologies, the advanced system is expected to enable the effective use of spectrum and the introduction of new services. The Ministry of Internal Affairs and Communications of Japan finally standardized the advancement of satellite digital broadcasting system in February 2009. The technical specifications were also standardized by the Association of Radio Industries and Businesses (ARIB) of Japan in July 2009.

The advanced system has a number of new features, including a transmission capacity increased by 30% or more due to the LDPC code and less roll-off factor, APSK modulations for a much higher capacity, higher resolution imagery and H.264|MPEG-4 AVC, three-dimensional surround sound, and the transport of IP packets targeted for multimedia storage.

The audio coding specifications are as follows.

TABLE 8

Audio coding for advancement of satellite digital broadcasting

Audio input	Maximum channels	22.2 channels per stream
	Sampling frequency	48 kHz
	Bit resolution	16, 20, or 24 bits
	Audio mode	Mono, stereo, multi-channel stereo (3.0, 4.0, 5.0, 5.1, 6.1, 7.1, 10.2, 22.2), dual-mono
Audio coding	Standard	MPEG-2 AAC, SBR (spectral band replication)
	Profile	LC

Attachment 1

Elevation perception of phantom sound images in the frontal hemisphere

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.

2 Subjective evaluation experiments

2.1 Preliminary subjective evaluation

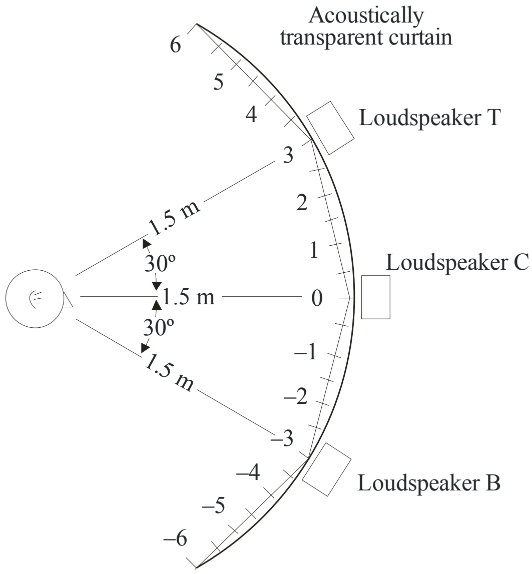
In a preliminary subjective evaluation, while a part of listeners reported that phantom sound images were unperceivable for a two-channel loudspeaker configuration with 60° vertical angle intervals, all listeners reported that phantom sound images were perceivable for a three-channel loudspeaker configuration.

2.2 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. 36). 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 36

Loudspeaker configuration

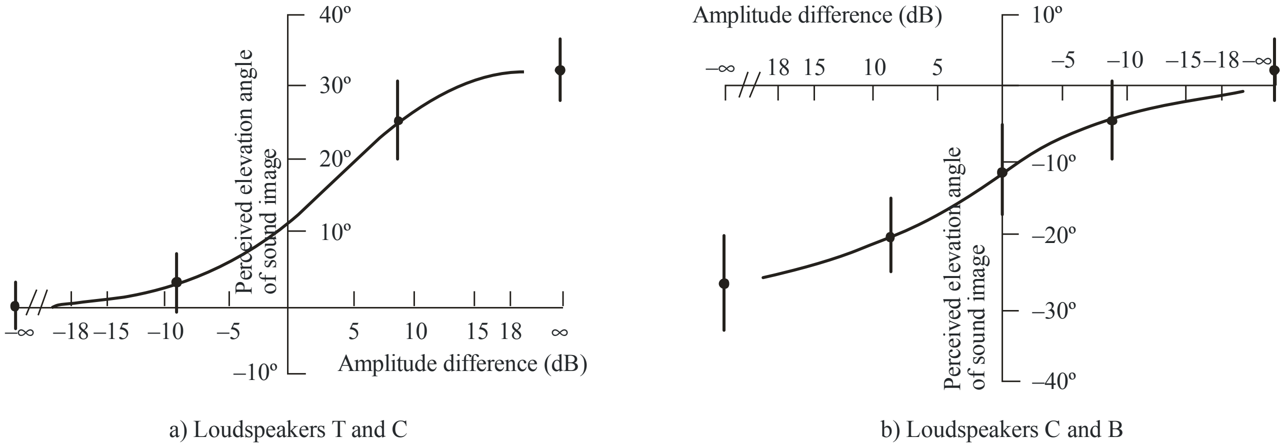


The stimuli comprised white noise with 1 s duration and 50 ms rise and fall time. The stimuli were reproduced by two loudspeakers (T and C or C and B) with amplitude difference. Five amplitude difference conditions were evaluated. The sound pressure level at the listening position was adjusted to 60 dBSPL for all conditions. Listeners were asked to respond to 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 37 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° vertical viewing angle.

FIGURE 37

Results for elevation angles of perceived sound images



Report BS.2159-37

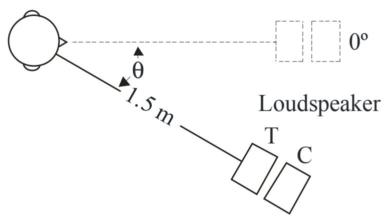
2.3 Elevation perception of phantom images in the frontal side direction

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 § 2.2 with the exception of the azimuth angle of the loudspeakers configuration (Fig. 38).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°.

Figure 39 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°. A vertical angle of 60° corresponds to the optimum vertical viewing angle for UHDTV.

FIGURE 38

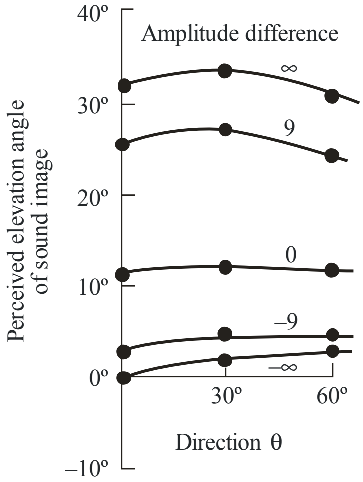
Loudspeaker configuration



Report BS.2159-38

FIGURE 39

Results for perceived elevation angles of sound images



Report BS.2159-39

3 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.

Attachment 2

Sensation of listener's envelopment[3] in an upper hemispherical sound field

1 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.

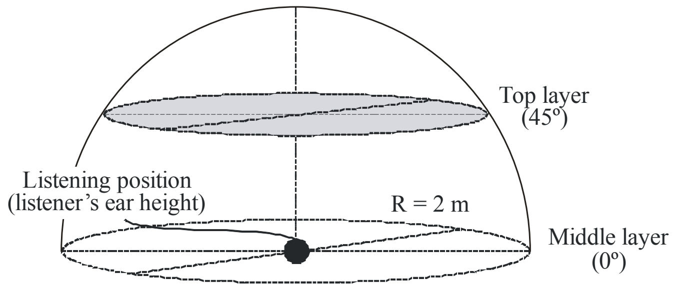
[3] As used in this Report, "listeners envelopment" has the meaning of the listener's sensation when the surrounding space is filled with sound. In the subjective listening tests, listeners were instructed to consider LEV as being indicative of spatial homogeneity of sound reproduced by loudspeakers placed around the listener.

2 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. 40. 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° elevation angle. The distance between the listener and each loudspeaker was 2 m.

FIGURE 40

Loudspeaker layout in upper hemispherical plane

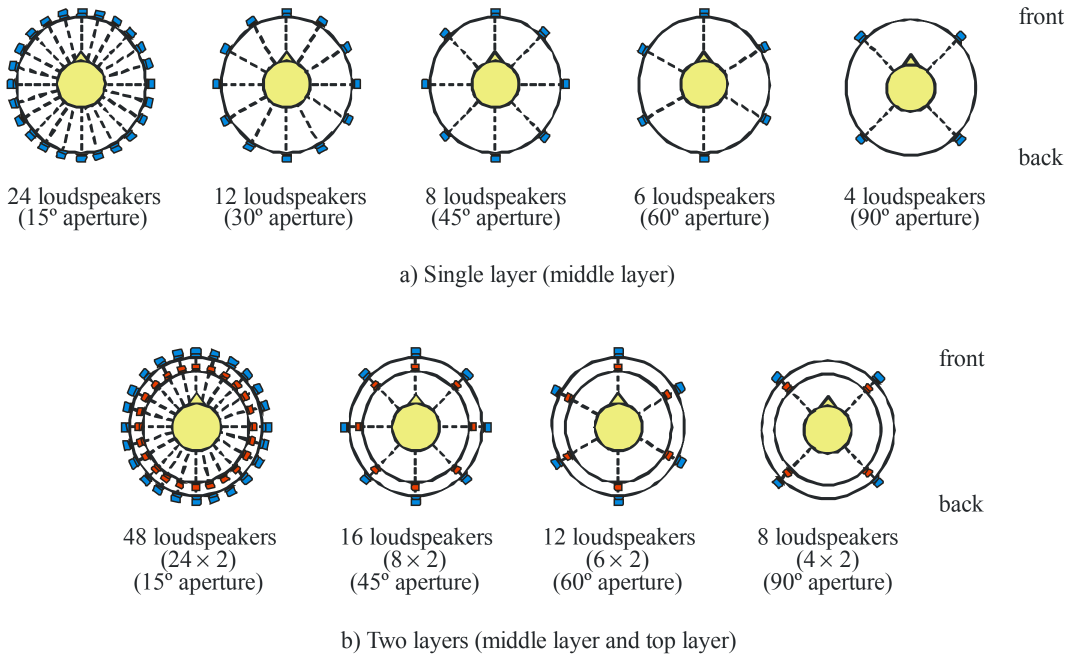


Report BS.2159-40

We evaluated nine test conditions for loudspeaker configurations, five using a single (middle) layer and four using both the middle and top layers. These conditions are shown in detail in Fig. 41.

FIGURE 41

Loudspeaker configurations for subjective test



Report BS.2159-41

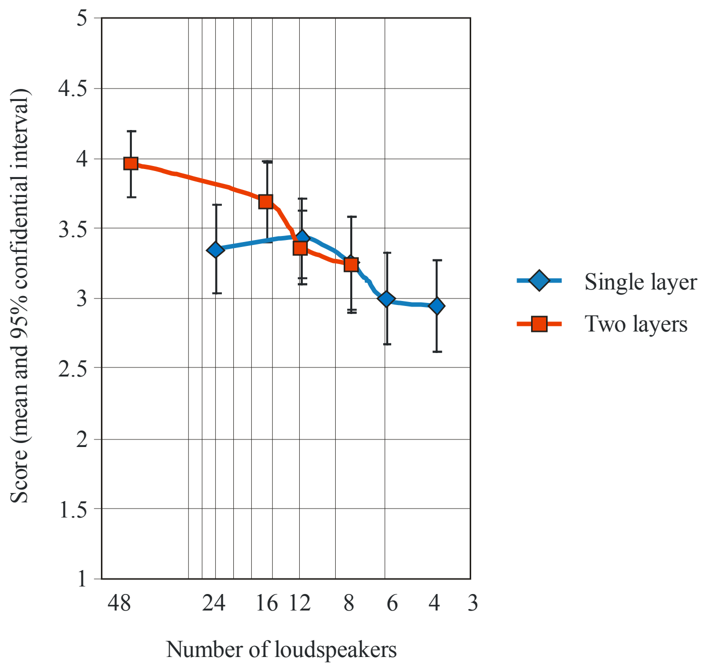
The stimuli comprised uncorrelated white noise with 6 s duration. The sound pressure level for each loudspeaker configuration was adjusted to 71.6 dBA ± 0.5 dBA at the listening point. For the conditions when both the middle and top layers were used, the sound pressure level at the listening point for 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 20s to 40s, 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 42 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 42

Results for sensation of LEV



Report BS.2159-42

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).

Attachment 3

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

1 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.

2 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°) as well as on other apertures and lateral displacements using a so-called quadraphonic loudspeaker configuration (with a 90° aperture for each loudspeaker) [1,2,3], a 6-channel "all round effect" loudspeaker configuration (with a 60° 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° for a loudspeaker pair is allowed, the 6-loudspeaker configuration shown in Fig. 43 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.

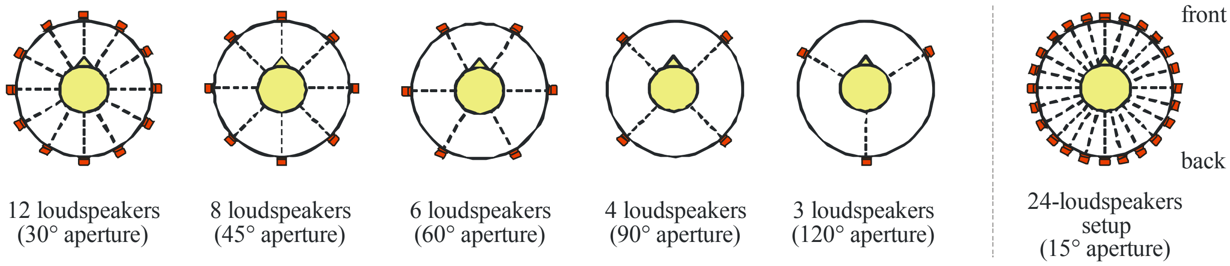
Following up on these studies, we carried out a preliminary subjective test on the localization and localization uncertainty of phantom sound images in the horizontal plane (middle layer) prior to conducting a subjective test on phantom sound images in the elevation direction (top layer). The latter test was conducted using the subjective listening test method described below in § 3. The results we obtained for the directional localization uncertainty of phantom sound images were fairly consistent with those obtained in previous studies.

3 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. 42). Twenty-four loudspeakers were placed at a 15° aperture above ear height with a 45° 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. 43.

FIGURE 43

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 is half the angle between the two loudspeakers

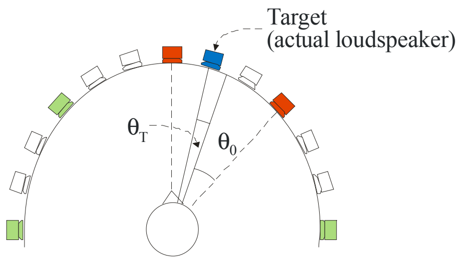
θ_T is the desired angle of the phantom source

g_1, g_2 are 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° (front) to 180° (back) with 15° intervals. The sound pressure level at the listening position was adjusted to 71.5 dBA ± 0.5 dBA at the listening point.

FIGURE 44

Pair-wise amplitude panning method



Report BS.2159-44

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 45

Presentation of subjective test

Report BS.2159-45

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

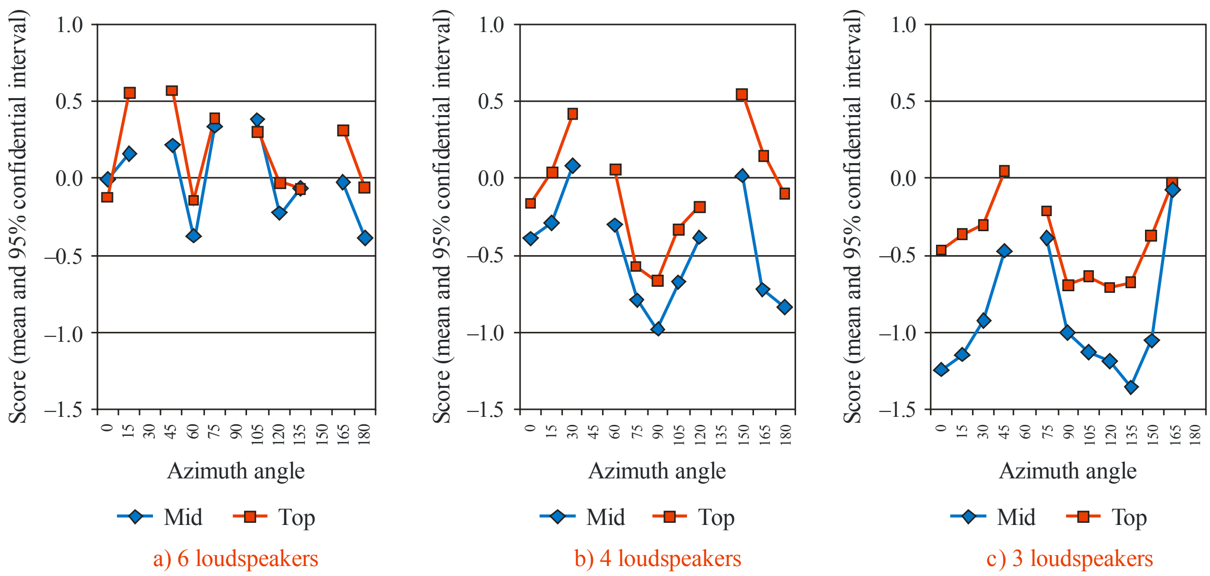
A subjective listening test was carried out in order to investigate the localization and localization uncertainty of phantom sound images above the listener using loudspeakers in the top layer. The test method was described in § 3. As the distribution of the listeners’ responses for perceived sound image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

Figure 46 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 2-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 46

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

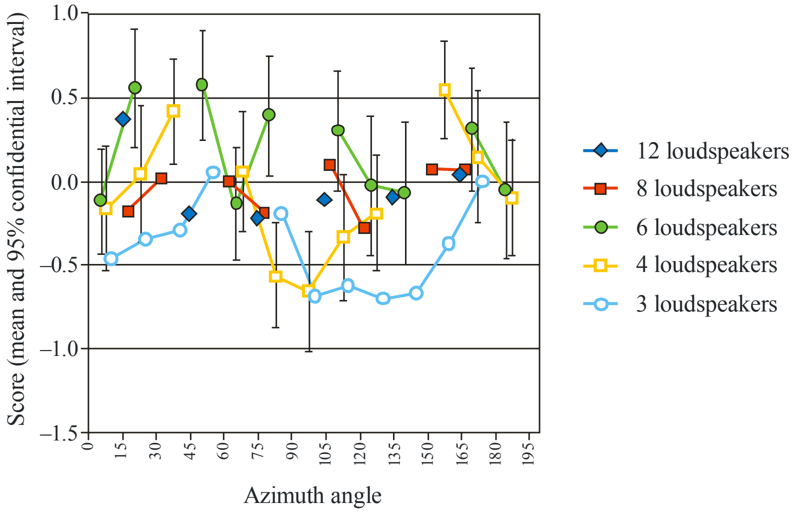


Report BS.2159-46

Figure 47 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 2-channel stereo configuration. The scores for 3 loudspeakers were around –0.5 for some directions except those close to the actual loudspeaker directions. Those for 4 loudspeakers were less than –0.5 for the right and left lateral (90°) directions.

FIGURE 47

Results for localization uncertainty of phantom sound images in top layer



Report BS.2159-47

5 Outline

According to the results we obtained in previous studies and in our 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).

____