

International Telecommunication Union

ITU-R
Radiocommunication Sector of ITU

Recommendation ITU-R BS.2051-0
(02/2014)

**Advanced sound system for
programme production**

BS Series
Broadcasting service (sound)



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 Recommendations

(Also available online at <http://www.itu.int/publ/R-REC/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
SNG	Satellite news gathering
TF	Time signals and frequency standards emissions
V	Vocabulary and related subjects

Note: This ITU-R Recommendation was approved in English under the procedure detailed in Resolution ITU-R 1.

Electronic Publication
Geneva, 2016

© ITU 2016

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without written permission of ITU.

RECOMMENDATION ITU-R BS.2051-0*

Advanced sound system for programme production

(Question ITU-R 135/6)

(2014)

Scope

This Recommendation specifies an advanced sound system with and without accompanying picture, beyond systems specified in Recommendation ITU-R BS.775. An advanced sound system uses audio data in combination with an appropriate set of metadata to specify a sound scene to be delivered/broadcasted. The specifications include requirements for signalling the properties of advanced sound content and loudspeaker layout to be used in content production for advanced sound systems. The advanced sound system can apply to the sound component of television and expanded large screen digital imagery (LSDI) programmes, as well as for sound only programmes.

Keywords

Advanced sound system, channel-based sound system, object-based sound system, scene-based sound system, multichannel audio, loudspeaker layouts, immersive audio

The ITU Radiocommunication Assembly,

considering

- a) that Recommendation ITU-R BS.775 – Multichannel stereophonic sound system with and without accompanying picture, specifies a multichannel stereophonic sound system with three front channels and two rear/side channels together with an optional low frequency effect (LFE) channel, as the highest level in a hierarchy of multichannel sound systems that range from 1/0 (monophonic) up to 3/2;
- b) that Recommendation ITU-R BT.1769 – Parameter values for an expanded hierarchy of LSDI image formats for production and international programme exchange, specifies an expanded hierarchy of LSDI image formats for production and international programme exchange;
- c) that Recommendation ITU-R BT.709 – Parameter values for the HDTV standards for production and international programme exchange, specifies HDTV system image parameters;
- d) that Recommendation ITU-R BT.2020 – Parameter values for ultra-high definition television systems for production and international programme exchange, specifies UHD TV image system parameters;
- e) that the wide field of view image presentation of expanded-LSDI, HDTV, and ultra-high definition television (UHD TV) programmes benefit from spatially enhanced sound beyond the 5.1 channel sound;
- f) that Recommendation ITU-R BS.1909 – Performance requirements for an advanced multichannel stereophonic sound system for use with or without accompanying picture, specifies the requirements for an advanced sound system with or without accompanying picture;

* Radiocommunication Study Group 6 made editorial amendments to this Recommendation in the year 2016 in accordance with Resolution ITU-R 1.

g) that Report ITU-R BS.2159 – Multichannel sound technology in home and broadcasting applications, includes the results of subjective evaluation experiments on loudspeaker layout to meet the requirements described in Recommendation ITU-R BS.1909,

recommends

1 that static or dynamic metadata/descriptors meeting the requirements described in Annex 1 should be used to signal the properties of all audio signals used in an advanced sound system in order to fully represent the desired audio content;

2 that the loudspeaker positions and configurations described in Annex 1 should be considered for production of advanced sound programmes;

3 that the appropriate number of audio elements¹ and loudspeaker layout configurations in programme production should be a choice by agreement between the producer and the recipient in the programme exchange;

4 that consumer interactions with the received audio should be facilitated in the production and broadcast sound system and that its characteristics should be a choice by agreement between the producer and the recipient in the programme exchange,

and further recommends

that further work should be conducted to provide information about the features of any advance sound system according to this Recommendation to fulfil the quality requirements of Recommendation ITU-R BS.1909.

Annex 1 **(normative)**

Advanced sound system for programme production

1 Introduction

The sound system specified in this Recommendation is defined as an advanced sound system which allows the metadata associated with each audio stream to be static or dynamic for the duration of a programme. This, for example, allows a programme to be represented by elements made by a combination of object signals and channel signals.

Channel signals are defined as audio signals that are mixed to a predefined number of channels and accompanied by metadata being static for the duration of a programme. Each of these channels is associated with a specific loudspeaker position. The production workflows, broadcasting networks and reproduction systems are defined by a set of loudspeaker positions. Examples are systems according to Recommendation ITU-R BS.775.

Object signals are audio signals that, when accompanied by dynamic metadata for the duration of a programme, represent audio elements allowing a renderer to play back the audio objects in a way

¹ An audio element is considered to be a signal with metadata which is either static for the duration of a programme or is dynamic. This enables the delivery of channel-based, object-based and scene-based content.

most appropriate to the playback system and listening environment. An object-based approach also may allow users to fully interact with the audio content.

Object-based elements and channel-based elements can be associated with each other or exist independently. To allow any combinations of channel or object-based elements, all signals should be accompanied by necessary metadata/descriptors, including time-independent (static) and/or time-dependent (dynamic) spatial position of the signal. These signals can be reproduced via a configuration of loudspeakers using a variety of rendering and/or mapping techniques.

Advanced sound programmes consist of the audio signals and the accompanying metadata.

Section 2 specifies requirements for metadata for sound content of the advanced sound system.

Section 3 describes loudspeaker layouts for the systems in production environments. Because a rendering or mapping process is needed for the reproduction of audio signals, the numbers and positions of loudspeakers are required to be well defined. This information enables rendering of the audio signals according to a predefined loudspeaker configuration in a reproduction scenario.

2 Requirements for metadata/descriptors for sound content of advanced sound system content

All audio files and streams used in an advanced sound system will require suitable metadata to accompany them. Unlike simple fixed channel-based systems where channel ordering is often enough to define the channels, the advanced system will need complete descriptions for all the audio elements used to ensure they are handled correctly. Therefore, a metadata model that has been standardized by a standards organization is required to provide consistent definitions for the audio. This model should have the following requirements:

- Contain all information required to reproduce/render a programme in all reproduction scenarios given by Recommendation ITU-R BS.1909 based on a single representation.
- Be able to describe the format of any channel-, object- and scene-based audio element.
- Be flexible enough to describe any combination of elements.
- The metadata items should be fully described such that they can be used by any renderer.
- Be specified in an open XML schema to allow the metadata to be represented in XML (as its primary method, it could of course be translated to other formats such as JSON).
- Be able to be added to an existing audio file format.
- To allow commonly used definitions (in particular existing channel-based configurations) to be open and freely accessible from a reference set of definitions.

3 Loudspeaker configurations for advanced sound system

Channel-based signals (including those in an advanced sound system) require a loudspeaker setup where the number and positions of the loudspeakers are well defined. Object-based signals can be reproduced via loudspeakers configured for channel-based signals or additional loudspeakers for other advanced rendering systems.

To ensure the loudspeaker configuration for the advanced multichannel sound system has a consistent definition, a set of parameters have been defined that specify each loudspeaker label, name, its position as described in Table 1:

- SP Label: an abbreviation of the channel name.
- Azimuth: the azimuth angle expressed in degrees, positive values rotate to the left when facing the front.

- Elevation: the elevation angle expressed in degrees, positive values go up from the horizontal plane.

TABLE 1

List of possible loudspeaker positions for advanced sound system, identification of the loudspeaker layouts in form of “Upper + Middle + Bottom loudspeakers”

SP Label	Azimuth	Elevation	A	B	C	D	E	F	G	H
			0+2+0	0+5+0	2+5+0	4+5+0	4+5+1	3+7+0	4+9+0	9+10+3
M+000	0	0		X	X	X	X	X	X	X
M+022	+22.5	0								
M-022	-22.5	0								
M+SC	Left edge of display	0							X	
M-SC	Right edge of display	0							X	
M+030	+30	0	X	X	X	X	X	X	X	X
M-030	-30	0	X	X	X	X	X	X	X	X
M+045	+45	0								
M-045	-45	0								
M+060	+60	0								X
M-060	-60	0								X
M+090	+90	0						X	X	X
M-090	-90	0						X	X	X
M+110	+110	0		X	X	X	X			
M-110	-110	0		X	X	X	X			
M+135	+135	0						X	X	X
M-135	-135	0						X	X	X
M+180	+180	0								X
U+000	0	+30								X
U+022	+22.5	+30								
U-022	-22.5	+30								
U+030	+30	+30			X	X	X			
U-030	-30	+30			X	X	X			
U+045	+45	+30						X	X	X
U-045	-45	+30						X	X	X
U+060	+60	+30								
U-060	-60	+30								
U+090	+90	+30								X
U-090	-90	+30								X
U+110	+110	+30				X	X		X	
U-110	-110	+30				X	X		X	
U+135	+135	+30								X
U-135	-135	+30								X
U+180	+180	+30								X
UH+180	+180	+45						X		
T+000	–	+90								X
B+000	0	-30					X			X

TABLE 1 (*end*)

SP Label	Azimuth	Elevation	A	B	C	D	E	F	G	H
			0+2+0	0+5+0	2+5+0	4+5+0	4+5+1	3+7+0	4+9+0	9+10+3
B+022	+22.5	-30								
B-022	-22.5	-30								
B+030	+30	-30								
B-030	-30	-30								
B+045	+45	-30								X
B-045	-45	-30								X
B+060	+60	-30								
B-060	-60	-30								
B+090	+90	-30								
B-090	-90	-30								
B+110	+110	-30								
B-110	-110	-30								
B+135	+135	-30								
B-135	-135	-30								
B+180	+180	-30								
LFE1	+45	-30		X	X	X	X	X	X	X
LFE2	-45	-30						X		X

NOTE – The loudspeakers are assumed to be on a sphere. When this is not the case they should preferably be time aligned (at the point of the listener) with an accuracy of 100 μ s, including the acoustic pathway. It is useful to identify, by means of metadata, whether the production environment was time aligned.

The loudspeaker layouts shown in Table 1 are illustrated in Table 2.

TABLE 2

Loudspeaker layouts for advanced sound system

<p>Lower layer Includes LFE loudspeakers</p>	
<p>Middle layer Same level as the listener</p>	
<p>Upper layer Includes overhead loudspeaker (T+000)</p>	

3.1 Loudspeaker positions in production environments

As various conditions may need to be considered for the practical design of loudspeaker set-ups in production studios such as room size, accompanying picture or other constraints, loudspeaker positions may have some azimuth angle and elevation angle variations. To ensure proper adaptation and rendering of channel-based elements, the loudspeaker positions used in the particular production environment should be stored as part of the metadata according to § 2. When the content is transferred to a different reproduction system/location, it should be ensured that the programme is adapted if necessary so that all quality requirements of the advanced sound system are met. The quality requirements are specified in Recommendation ITU-R BS.1909.

However, loudspeakers should be placed within the sectors defined by the azimuth and elevation range as given in Tables 3 to 6 to lessen the sound quality variations by loudspeaker positions differences. The notations “a .. b” in the Tables should be taken to mean the smaller of the two sectors that might be obtained by rotating clockwise or anti-clockwise between angle “a” and angle “b”.

TABLE 3
Loudspeaker configuration for Sound system H (9+10+3)

SP Label	Channel		Azimuth	Elevation
	Label	Name	Range	Range
M+060	FL	Front left	+45 .. +60	0 .. +5
M-060	FR	Front right	-45 .. -60	0 .. +5
M+000	FC	Front centre	0	0 .. +5
LFE1	LFE1	Low frequency effects-1	+30 .. +90	-15 .. -30
M+135	BL	Back left	+110 .. +135	0 .. +15
M-135	BR	Back right	-110 .. -135	0 .. +15
M+030	FLc	Front left centre	+22.5 .. +30	0 .. +5
M-030	FRc	Front right centre	-22.5 .. -30	0 .. +5
M+180	BC	Back centre	+180	0 .. +15
LFE2	LFE2	Low frequency effects-2	-30 .. -90	-15 .. -30
M+090	SiL	Side left	+90	0 .. +15
M-090	SiR	Side right	-90	0 .. +15
U+045	TpFL	Top front left	+45 .. +60	+30 .. +45
U-045	TpFR	Top front right	-45 .. -60	+30 .. +45
U+000	TpFC	Top front centre	0	+30 .. +45
T+000	TpC	Top centre	-	+90
U+135	TpBL	Top back left	+110 .. +135	+30 .. +45
U-135	TpBR	Top back right	-110 .. -135	+30 .. +45
U+090	TpSiL	Top side left	+90	+30 .. +45
U-090	TpSiR	Top side right	-90	+30 .. +45
U+180	TpBC	Top back centre	+180	+30 .. +45
B+000	BtFC	Bottom front centre	0	-15 .. -30
B+045	BtFL	Bottom front left	+45 .. +60	-15 .. -30
B-045	BtFR	Bottom front right	-45 .. -60	-15 .. -30

TABLE 4

Loudspeaker configuration for Sound system G (4+9+0)

SP Label	Channel		Loudspeaker location, Polar		Loudspeaker location, Cartesian		
			Azimuth	Elevation	X	Y	Z
	Label	Name	Range	Range	(left/right)	(rear/front)	(bottom/top)
M+000	C	Centre	0	0	0	+1	0
M+SC	Lsc	Left screen	Left screen edge	0	-1*	+1	0
M-SC	Rsc	Right screen	Right screen edge	0	+1*	+1	0
M+030	L	Left	+30 .. +45	0	-1	+1	0
M-030	R	Right	-30 .. -45	0	+1	+1	0
M+090	Ls	Left surround	+90 .. +110	0	-1	0	0
M-090	Rs	Right surround	-90 .. -110	0	+1	0	0
M+135	Lrs	Left rear surround	+135 .. +150	0	-1	-1	0
M-135	Rrs	Right rear surround	-135 .. -150	0	+1	-1	0
U+045	Ltf	Left top front	+30 .. +45	+30 .. +45	-0.5	+0.5	+1
U-045	Rtf	Right top front	-30 .. -45	+30 .. +45	+0.5	+0.5	+1
U+110	Ltr	Left top rear	+110 .. +135	+30 .. +45	-0.5	-0.5	+1
U-110	Rtr	Right top rear	-110 .. -135	+30 .. +45	+0.5	-0.5	+1
LFE1	LFE	Low frequency effects	-	-	-	-	-

* Referenced to screen, -1 = left edge of screen, +1 = right edge of screen.

TABLE 5

Loudspeaker configuration for Sound system F (3+7+0)

SP Label	Channel		Azimuth	Elevation
	Label	Name	Range	Range
M+000	C	Centre	0	0
M+030	L	Left	+30	0
M-030	R	Right	-30	0
U+045	LH	Left height	+30 .. +45	+30 .. +45
U-045	RH	Right height	-30 .. -45	+30 .. +45
M+090	LS	Left side	+60 .. +150	0
M-090	RS	Right side	-60 .. -150	0
M+135	LB	Left back	+60 .. +150	0
M-135	RB	Right back	-60 .. -150	0
UH+180	CH	Centre height	180	+45 .. +90

TABLE 5 (*end*)

SP Label	Channel		Azimuth	Elevation
	Label	Name	Range	Range
LFE1	LFE1	Left low frequency effects	+30 .. +90	-15 .. -30
LFE2	LFE2	Right low frequency effects	-30 .. -90	-15 .. -30

TABLE 6

Loudspeaker configuration for Sound system B (0+5+0)
(from Recommendation ITU-R BS.775)

SP Label	Channel		Azimuth	Elevation
	Label	Name	Range	Range
M+030	L	Left	+30	0
M-030	R	Right	-30	0
M+000	C	Centre	0	0
LFE1	LFE	Low frequency effects	–	–
M+110	Ls	Left surround	+100 .. +120	0 .. +15
M-110	Rs	Right surround	-100 .. -120	0 .. +15

The loudspeaker layouts shown in Tables 3 to 6 are illustrated in Table 7.

TABLE 7

Loudspeaker layouts for advanced sound system

Sound system	Upper layer	Middle layer	Bottom layer
H (9+10+3)			
Upper layer 3/3/3 (Note)			
Middle layer 5/2/3			
Bottom layer 3/0/0.2			
G (4+9+0)			
Upper layer 2/0/2			
Middle layer 5/2/2			
Bottom layer 0.0.0.1			
F (3+7+0)			
Upper layer 2/0/1			
Middle layer 3/2/2			
Bottom layer 0/0/0.2			
B (0+5+0)			
Upper layer 0/0/0			
Middle layer 3/0/2			
Bottom layer 0/0/0.1			

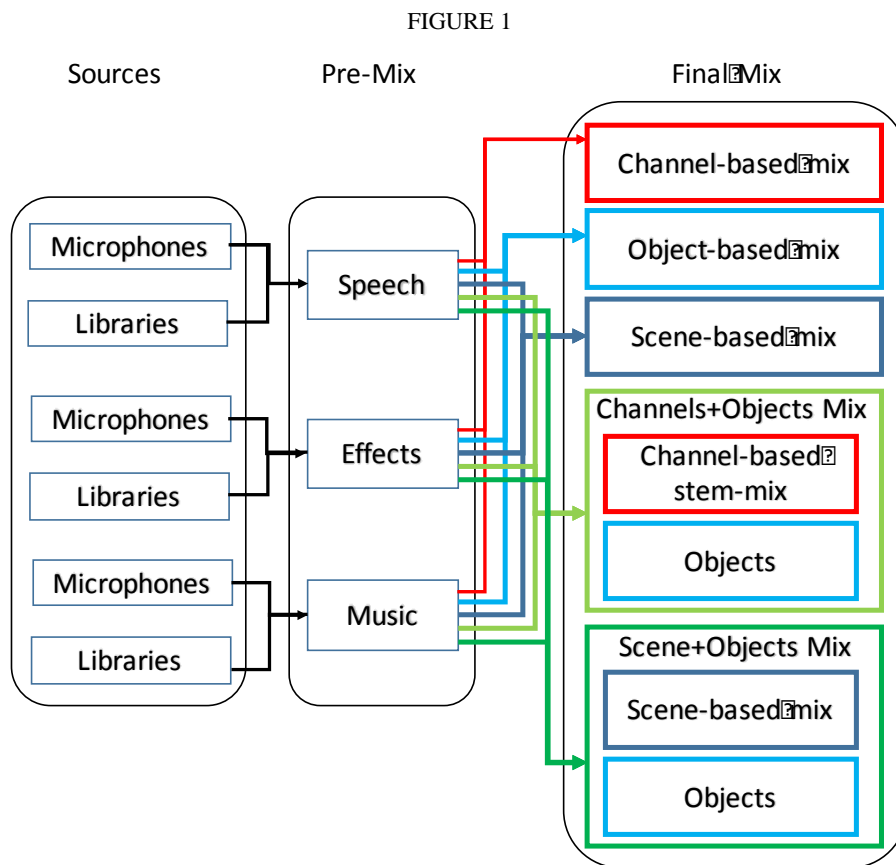
NOTE – X/Y/Z.LFE identifies the number of front/side/rear (plus LFE) loudspeakers in each layer.

Appendix 1

Sound programme production and background of advanced sound system

In the process of mixing, a number of elements that depend on the nature of the production and signal acquisition process, typically determined by the audio engineer, are reduced to a representation of the creator intent, keeping only a reduced number of elements separate. The difference in production between a channel-based, object-based, scene-based and approaches which combine them is the method of mixing for the pre and final mixes. The channel-based approach means mixing all the elements into a predefined set of channels, whereas the hybrid channel-based + object-based approach allows the objects to be either mixed into channels or kept as separate objects. Similarly, in the hybrid scene-based + object-based approach the elements are either stored in a scene-based format (e.g. HOA) or kept as separate objects. In a pure object-based approach, all elements to recreate a certain experience are kept separate.

Existing production, live mixing and the post-production workflow of channel-based systems are the same as the hybrid model of channel-based stem-mix and objects. The delivery format of the final mix is different between the channel-based model and the hybrid model which delivers channel-based stem-mix signals and object signals.



BS.2051-01

Figure 2 shows how an example of a typical broadcast chain could look for an advanced sound system. The production can use any type of audio source and content, but it should be fully annotated with the correct metadata to describe the audio signals, and this is stored in a file format that can support this metadata.

The distribution stage will adapt the representations from production into a more compact representation which will retain or generate new metadata to allow further rendering. The distribution file or stream is then passed to the broadcast stage which will render it towards a particular broadcast format. Higher bandwidth broadcasts will allow to deliver and rendering many object and channels, whereas low bandwidth broadcasts may have to render down to more traditional stereo formats. The broadcast format should retain as much of the metadata as required for the receiver end.

Each receiver device will have its own renderer designed for the loudspeaker layouts possible for that device. For example, a Hi-fi will need something very flexible to allow multiple loudspeakers to have different placements, whereas a TV will have its internal loudspeakers fixed in known positions. Future content delivery will potentially receive a representation, which will maintain full flexibility for user interaction or personalization.

FIGURE 2

