International Telecommunication Union

ITU-R

Radiocommunication Sector of ITU

Report ITU-R BS.2266-1

(11/2013)

Framework of future audio representation systems

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, 2014

© ITU 2014

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

REPORT ITU-R BS.2266-1

Framework of future audio representation systems

(05/2013-11/2013)

Summary

This Report is intended to present a framework of future audio representation systems and establish a speaker layout superset naming convention. The need for a production exchange file format is depicted.

TABLE OF CONTENTS

Page

1 Introduction 2

2 Framework of future audio representation systems 2

3 Convention for identifying the position of a loudspeaker in a superset of loudspeaker layouts 3

3.1 Definition of a superset of loudspeaker layouts 3

3.2 Labels and layouts within the superset 5

4 Exchange file format 7

1 Introduction

In order to incorporate the current developments in the field of advanced sound systems, it is proposed to separate the elements of such a system with clear interfaces and organize them as a framework for future audio representation systems as proposed here.

This Report reflects the current development in the broadcast industry and enables a more universal/flexible understanding of the current state-of-the art and future audio representation systems. It recognizes several representations defined thus:

Channel-based

Microphone signals are mixed to a predefined number of channels. Each of these channels is associated with a specific speaker position. The production workflows, broadcasting networks and reproduction systems are defined by a set of speaker positions. Examples are stereo, plus formats 5.1, 7.1 or 22.2.

Object-based

The scene is represented by audio signals that (either separately or combined) represent audio objects. They are accompanied by dynamic metadata that allows a renderer to play back the audio objects in a way most appropriate to the playback system and listening environment. Audio objects can be rendered using different algorithms (amplitude panning, Wave Field Synthesis, etc.) for different loudspeaker setups. An object-based approach also allows users to fully interact with the sound scene.

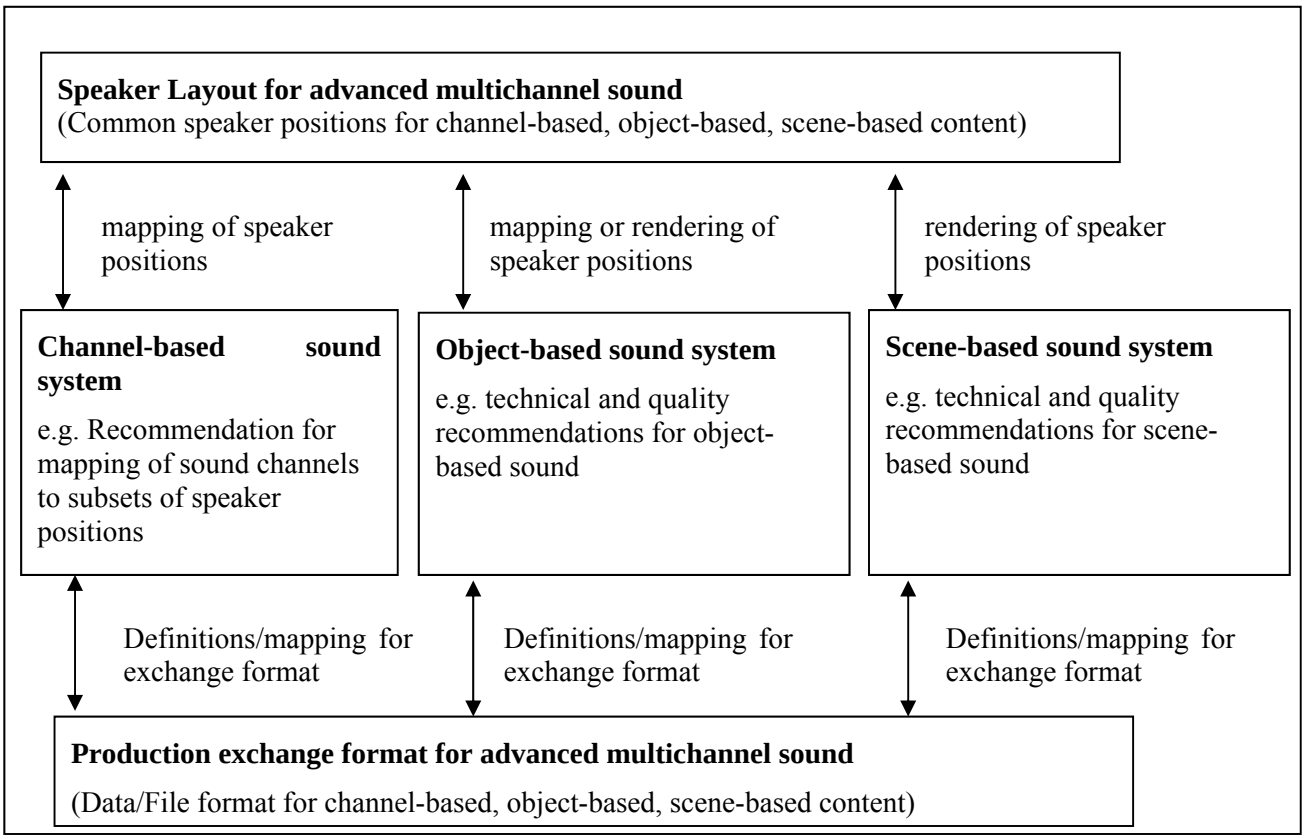
Scene-based

The sound scene is represented by a set of signals that can be decoded to provide a set of speaker feeds. Although there may be some constraints to the number and placement of speakers, the signals representing the sound scene are independent of the speaker positions. Examples are B-Format and Higher-Order Ambisonics.

2 Framework of future audio representation systems

Given the different approaches currently under development and the need for a common understanding of these different systems in an optimal way, a framework is proposed that addresses several working areas. The framework can act as a basis to operate with different possible audio content representations, as there are channel-based audio, object-based audio and scene-based audio.

Framework of future audio representation systems



One important part of the framework specifies common speaker positions for advanced sound systems. This information can be independent of the approach of content representation and rendering. It is intended as a reference for describing advanced loudspeaker setups (e.g. in perceptual evaluations). For the definition of a possible speaker layout, it is necessary to gain a common base of knowledge about naming, positions and labelling of speakers in multichannel sound systems.

The second part is a production exchange format incorporating channel-based, object-based and scene-based representations to allow flexibility between audio content representations, as well as speaker numbers and layouts. This may be based on Recommendation ITU-R BR.1352. (ITU-R BR.1352 is closely related to the EBU formats known as “BWF”, “MBWF” and “RF64”.)

3 Convention for identifying the position of a loudspeaker in a superset of loudspeaker layouts

In this section, the Report collects common speaker positions for advanced sound systems and aims for a naming and labelling convention to establish a possibility for common understanding and discussion.

3.1 Definition of a superset of loudspeaker layouts

Based on the already existing Recommendations and Reports:

a) 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) 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) Recommendation ITU-R BT.2020 – Parameter values for UHDTV systems for production and international programme exchange, specifies UHDTV image system parameters;

d) 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;

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

this Report identifies:

- that there are three approaches to represent spatially enhanced sounds: a channel-based approach, an object-based approach and a scene-based approach;
- that common speaker positions for channel-based, object-based, scene-based content are desired for advanced sound programme production;
- and that the wide field of view image presentation of expanded-LSDI and UHDTV programme requires spatially enhanced sound beyond the 5.1 channel sound.

Therefore, it is necessary to define a way to describe a superset of speaker layouts used in these systems, as well as in future systems for programme production. This Report only specifies a unified way to describe existing and future speaker layouts for common and advanced sound systems, not the method of signal representation.

The signals for the speakers may be derived from any channel-based, object-based or scene-based representation (or any combination of those), and this is not defined in this Report.

There are previous studies on subjective evaluation on advanced sound systems beyond 5.1ch sound that investigate speaker layouts to meet the sound quality requirements. These studies are included in Report ITU-R BS.2159 – Multichannel sound technology in home and broadcasting applications (see Table 1) which also summarizes the results of subjective evaluation experiments on speaker layout to meet the requirements described in Recommendation ITU-R BS.1909:

- elevation perception of phantom sound images in the frontal hemisphere, which is desired for UHDTV applications;
- sensation of “listener’s envelopment (LEV)”, which is one of the primary features of a three-dimensional spatial impression;
- localization and localization uncertainty of phantom sound images in the elevation direction generated by two speakers located above the listener;
- influence of listening position on directional perception of frontal sound images.

The results of these studies motivate for a definition of a superset of speaker layouts and a descriptive basis as given in this Report.

TABLE 1

Speaker layouts to meet the requirements

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

NOTE 1 – Listener envelopment (LEV) is generally defined as "The extent to which the sound source envelops/surrounds/exists around you. The feeling of being surrounded by the sound source" (definition from J. Berg and F. Rumsey. "Validity of selected spatial attributes in the evaluation of 5-channel microphone techniques". Presented at AES 112th Convention. Audio Eng. Soc. (2002)). In the subjective listening tests, listeners were instructed to consider LEV as being homogeneously-enveloped in spatially arranged sound reproduced by loudspeakers placed around the listener.

NOTE 2 – 7 680 × 4 320 Image system: horizontal viewing angle 96°, vertical viewing angle 54°.

NOTE 3 – Listen at out-of-centre listening position.

3.2 Labels and layouts within the superset

Speaker positions for advanced sound systems are composed of three vertical layers: M-middle, U-upper[1] and B-bottom layer.

To ensure the compatibility of speaker layout definitions, a consistent definition of a set of parameters has been defined that specifies the included speakers and their positions:

- SP-Label: A short label that abbreviates the position.
- Azimuth: The azimuth angle expressed in degrees, positive values rotate to the left when facing the front.

[1] Upper layer sometimes is called Top layer.

- Elevation: The elevation angle expressed in degrees, positive values go up from the horizontal plane.
- Distance: A normalized distance value that should be used as a ratio to the actual distance of speakers from the origin.

A list of possible speaker positions has been defined as a superset of established loudspeaker layouts. New speaker layouts can then use a subset of these positions. The layouts can then be used for reproduction of channel-based, object-based or scene-based content using appropriate signal processing.

Table 2 lists the speakers with the parameters explained above. However, as distance of speakers from the origin is assumed to equal, the column of distance is not included in Table 2.

TABLE 2

List of possible speaker positions for advanced sound system, identification of the speaker layouts in form of "Upper speakers + Middle speakers + Bottom speakers"

SP-Label	Azimuth	Elevation (Note 1)	A	B	C (Note 3)	D	E (Note 3)	F	G	H (Note 2, Note 3)
			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

TABLE 2 (*end*)

SP-Label	Azimuth	Elevation (Note 1)	A	B	C (Note 3)	D	E (Note 3)	F	G	H (Note 2, Note 3)
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		X
T+000	–	+90								X
B+000	0	–30					X			X
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 1 – Note that MPEG-H 3D-Audio uses an elevation angle of –15° for the bottom loudspeaker layer and an elevation angle of 35° for the upper loudspeaker layer.

NOTE 2 – Note that Japan uses the 22.2ch speaker layout with the azimuth angle of 45° to 60° for front left and front right channels and the azimuth angle of 22.5° to 30° for front centre left and front centre right channels.

NOTE 3 – As used in MPEG-H 3D-AUDIO.

4 Exchange file format

For exchanging audio content, a definition is required regarding how to store a sound scene so that it includes all necessary information to be delivered, rendered, coded and/or stored. In such a file format, the carriage of content with respect to the different speaker layouts derived from the superset in Table 2 should be possible, as well as the carriage of channel-based, object-based and scene-based content or a mixture of that. The files of that exchange format should be suitable or useable for all (targeted) playback systems.

A detailed requirement list of such a file format is to be defined and the definition of such a production exchange format for advanced multichannel sound could lead to a Recommendation in future. A generic list of requirements is given here:

– the format should enable the delivery of an audio experience beyond the state-of-the-art;

– it should deliver the artistic intent as good as possible and support and promote new creative and technological opportunities;
– the format shall provide a reasonable path for upgrading;
– it should be able to carry existing channel-based formats as well as future channel-based, object-based and scene-based content;
– the specification should be open and royalty-free;
– it should minimize the changes needed on the production and delivery process while transporting the information needed for advanced rendering;
– it should allow for advances in all related areas while ensuring compatibility;
– it should enable content, equipment and software to remain future-proof;
– it should be based on international standards bodies and industry standards.

The definition of an exchange format is an important component to ensure that content creators can confidently create programme material that will be interoperable and able to maximize the potential of next generation audio rendering technologies beyond the state-of-the-art. It would minimize the risk of market fragmentation and confusion for consumers. It could simplify interoperability between different stages of the audio content production, delivery chain and playback.