ITUPublications

International Telecommunication Union
Radiocommunication Sector

Report ITU-R BS.2266-3

(03/2023)

BS Series: Broadcasting service (sound)

Framework of future audio representation systems

Foreword

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

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

Policy on Intellectual Property Right (IPR)

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

Series of ITU-R Reports

(Also available online at https://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
TF	Time signals and frequency standards emissions

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, 2023

REPORT ITU-R BS.2266-3

Framework of future audio representation systems

(05/2013-11/2013-2014-2023)

Summary

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

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 format 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.

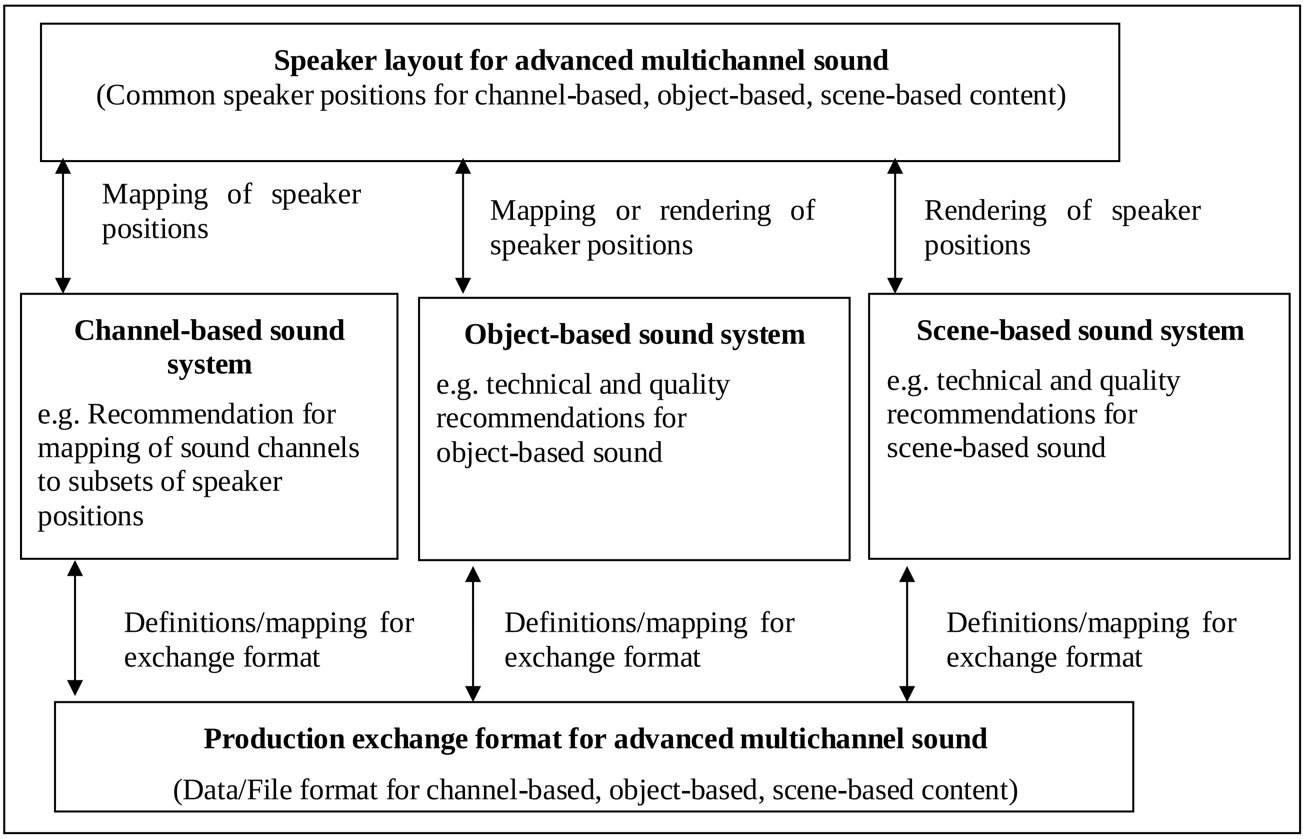
Scene-based

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


Speaker layout for advanced multichannel sound
(Common speaker positions for channel-based, object-based, scene-based content)
Mapping of speaker positions
Mapping or rendering of speaker positions
Rendering of speaker positions
Channel-based sound system
e.g. Recommendation for mapping of sound channels to subsets of speaker positions
Object-based sound system
e.g. technical and quality recommendations for object-based sound
Scene-based sound system
e.g. technical and quality recommendations for scene-based sound
Definitions/mapping for exchange format
Definitions/mapping for exchange format
Definitions/mapping for exchange format
Production exchange format for advanced multichannel sound
(Data/File format for channel-based, object-based, scene-based content)

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 BS.1352. (Recommendation ITU-R BS.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

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.

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)[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[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[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

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 fixed speaker layouts, as well as the carriage of channel-based, object-based and scene-based content or a mixture of that should be possible. The files of that exchange format should be suitable or useable for all (targeted) playback systems.

[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.

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

[3] Listen at out-of-centre listening position.

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 faithfully 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 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.