

SMPTE STANDARD

Immersive Audio Metadata



Table of Contents

Page

1	Scope	4
2	Conformance Notation	4
3	Normative References	4
4	Terms and Definitions	5
4.1	Audio Object	5
4.2	Bed	5
4.3	Group	5
4.4	Renderer	5
4.5	Sample	5
4.6	Target Environment	5
4.7	Zone	5
5	Channel Metadata	5
5.1	General	5
5.2	Channel Identifier	5
5.3	Routing Destination	5
5.4	Waveform Reference	6
6	Bed Metadata	6
6.1	General	6
6.2	Bed Identifier	6
6.3	Bed Channel List	6
6.4	Remap coefficients.	6
6.5	Conditional Bed	6
7	Object Metadata	6
7.1	Introduction	6
7.2	Object Identifier	7
7.3	Waveform Reference	7
7.4	Object Position	7
7.5	Object Spread	7
7.6	Object Gain	7
7.7	Object Lifetime	7

7.8	Object Audio Description	7
7.9	Decorrelation	7
7.10	Snap Tolerance	8
7.11	Conditional Object	8
7.12	Zone control	8
8	Structural Metadata	9
8.1	Group	9
9	Coordinate System and Frame of Reference	9
9.1	Coordinate System	9
9.2	Frame of reference and location coding	9
9.3	Mapping the cube to a cinema	11
9.4	Example - Mapping object position to a cinema Loudspeaker array (informative)	12
	Bibliography (Informative)	14

Foreword

SMPTE (the Society of Motion Picture and Television Engineers) is an internationally recognized standards developing organization. Headquartered and incorporated in the United States of America, SMPTE has members in over 80 countries on six continents. SMPTE's Engineering Documents, including Standards, Recommended Practices, and Engineering Guidelines, are prepared by SMPTE's Technology Committees. Participation in these Committees is open to all with a bona fide interest in their work. SMPTE cooperates closely with other standards-developing organizations, including ISO, IEC and ITU.

SMPTE Engineering Documents are drafted in accordance with the rules given in Part XIII of its Administrative Practices. The Technology Committee 25CSS Working Group on Interoperability of Immersive Sound in Digital Cinema prepared this SMPTE Engineering Document.

Intellectual Property

At the time of publication no notice had been received by SMPTE claiming patent rights essential to the implementation of this Engineering Document. However, attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. SMPTE shall not be held responsible for identifying any or all such patent rights.

Introduction

Immersive audio consists of Audio Channels and/or Audio Objects, which can be utilized by the content creator to design a sound track with sounds above and around the listener. Immersive audio combines metadata with audio essence, which allows the Audio Objects and Audio Channels in the sound track to be rendered successfully into multiple Loudspeaker configurations. This document identifies metadata required for immersive audio for cinema.

1 Scope

This standard defines metadata for use in creating immersive audio content for cinema. The standard defines the metadata items that shall be supported in immersive audio content and if appropriate, the range of values, value precisions, and cardinal values. The format of the metadata items when carried in a bitstream or file is outside the scope of this specification and is provided in a separate document.

2 Conformance Notation

Normative text is text that describes elements of the design that are indispensable or contains the conformance language keywords: "shall", "should", or "may". Informative text is text that is potentially helpful to the user, but not indispensable, and can be removed, changed, or added editorially without affecting interoperability. Informative text does not contain any conformance keywords.

All text in this document is, by default, normative, except: the Introduction, any section explicitly labeled as "Informative" or individual paragraphs that start with "Note:"

The keywords "shall" and "shall not" indicate requirements strictly to be followed in order to conform to the document and from which no deviation is permitted.

The keywords, "should" and "should not" indicate that, among several possibilities, one is recommended as particularly suitable, without mentioning or excluding others; or that a certain course of action is preferred but not necessarily required; or that (in the negative form) a certain possibility or course of action is deprecated but not prohibited.

The keywords "may" and "need not" indicate courses of action permissible within the limits of the document.

The keyword "reserved" indicates a provision that is not defined at this time, shall not be used, and may be defined in the future. The keyword "forbidden" indicates "reserved" and in addition indicates that the provision will never be defined in the future.

A conformant implementation according to this document is one that includes all mandatory provisions ("shall") and, if implemented, all recommended provisions ("should") as described. A conformant implementation need not implement optional provisions ("may") and need not implement them as described.

Unless otherwise specified, the order of precedence of the types of normative information in this document shall be as follows: Normative prose shall be the authoritative definition; Tables shall be next; followed by formal languages; then figures; and then any other language forms.

3 Normative References

The following standards contain provisions that, through reference in this text, constitute provisions of this Standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this Standard are encouraged to investigate the possibility of applying the most recent edition of the standards indicated below.

SMPTE ST 377-4:2012 MXF Multichannel Audio Labeling Framework

SMPTE ST 428-12:2013 D-Cinema Distribution Master – Common Audio Channels and Soundfield Groups

4 Terms and Definitions

4.1 Audio Object

A segment of audio essence with associated metadata describing positional and other properties which may vary with time.

4.2 Bed

A Soundfield Group, such as a 5.1, 7.1 or 9.1, that is typically present for the duration of the program and serves as the foundation of the immersive soundtrack mix.

4.3 Group

A set of Audio Channels, Audio Objects or components that belong together (e.g. stereo, 5.1).

4.4 Renderer

A device or algorithm that reads in audio tracks and associated metadata and converts them to another set of audio signals destined for individual reproduction devices.

4.5 Sample

A discrete number representing the amplitude of an audio signal at an instant in time.

4.6 Target Environment

A specific set of conditions that is present in the playback environment.

4.7 Zone

A defined region within a generic listening environment.

5 Channel Metadata

5.1 General

The following subsections specify metadata requirements associated with Audio Channels.

5.2 Channel Identifier

This Identifier uniquely identifies an Audio Channel. No two Audio Channels shall have the same Channel Identifier at any instant in time.

5.3 Routing Destination

This metadata item identifies the single Loudspeaker or other reproduction device associated with the channel. For purposes of this document, "other reproduction device" may include an array of Loudspeakers driven in unison, or it may include a process to modify the audio prior to presentation. The metadata shall be coded such that the desired routing destination is unambiguously identified.