IMF IAB Interoperability

Guidelines

1

About this documentation

The purpose of this documentation is to provide information to assist application developers who wish to implement support for Interoperable Master Format (IMF) immersive audio bitstream (IAB) track files in a manner that enables interoperability with Dolby tools and other industry tools.

- References
- Definition of IAB track file
- Channel abbreviations

1.1 References

Standards provide additional information to assist you in designing your product.

Standards

IAB

- SMPTE ST 2098-2:2019, *Immersive Audio Bitstream Specification*, available from http://www.smpte.org.

IMF IAB

- SMPTE ST 2067-201:2019, *IAB Level 0 Plug-in*, available from http://www.smpte.org.

1.2 Definition of IAB track file

An IAB track file is an IAB essence wrapped in the Material Exchange Format (MXF) format as defined by ST 2067-201 for use in IMF.

1.3 Channel abbreviations

This documentation uses several channel abbreviations.

Abbreviation	Channel
L	Left
R	Right
C	Center
Lc	Left Center
Rc	Right Center
LFE	Low-Frequency Effects
Lfh	Left Front Height
Rfh	Right Front Height
Ls	Left Surround
Rs	Right Surround
Lss	Left Side Surround
Rss	Right Side Surround
Lrs	Left Rear Surround
Rrs	Right Rear Surround
Lw	Left Wide
Rw	Right Wide
Tsl	Left Top Middle
Tsr	Right Top Middle
Ltf	Left Top Front
Rtf	Right Top Front
Ltr	Left Top Rear
Rtr	Right Top Rear

Abbreviation	Channel
Lrh	Left Rear Height
Rrh	Right Rear Height
Lts	Left Top Surround
Rts	Right Top Surround

2

IMF IAB guidelines

Dolby provides specific guidelines for implementing support for IMF IAB files in a manner that enables interopability with Dolby tools and other industry tools.

- Background
- Dolby IAB Implementation
- Concatenation

2.1 Background

The IAB track file carries audio essence and metadata necessary to reproduce a complete audio program, and consists of a sequence of independent IAFrames.

Each IAFrame corresponds to a single video frame and carries audio and metadata. The audio essence may be thought of as being made up of multiple tracks, where each track represents a bed channel or an object. When each track is encoded into an IAFrame, there must be a mechanism to reassemble each track across frames. The *Immersive Audio Bitstream Specification* does not explicitly state how this can be accomplished, but does recommend that the MetaID bitstream field be used for this purpose. The Dolby tools use an alternative method to reassemble each track from multiple IAFrames.

The *Immersive Audio Bitstream Specification* is constrained by the *Level-0 Plug-in* that defines how IAB track files shall be structured. However, the *Level-0 Plug-in* does not constrain the number of ObjectDefinition elements, indicating that the number of objects may vary between frames. This flexibility can lead to problems with encoding and decoding tools. To avoid this, we have made some design choices to simplify the implementation.

This documentation outlines some of the presumptions and implementation choices made by Dolby.

2.2 Dolby IAB Implementation

The Dolby implementation of IAB includes how Dolby tools represent continuous audio and maintain the same number of beds and objects in Dolby-generated IAFrames belonging to the same IAB track file.

2.2.1 Continuous audio

The IAB specification mentions that elements with the same ElementID and MetaID in contiguous IAFrames may represent continuous audio. Dolby tools use a different mechanism to represent continuous audio.

In Dolby implementations:

- Each track persists for the entire program.
- IAFrames generated by Dolby encode subelements in a particular order. That order is used to represent continuous tracks of audio across frames.
- BedDefinition elements are encoded first. Each BedDefinition element with N bed channels is preceded by at most N AudioDataPCM elements that link back to it. Each AudioDataPCM is optional when the channel is silent in the current frame.
- ObjectDefinition elements are encoded next. Each ObjectDefinition element is preceded by an optional AudioDataPCM element that links back to it. The AudioDataPCM element can be absent if the channel is silent in the current frame.
- Each IAFrame has the same number of BedDefinition and ObjectDefinition elements.
- The continuous audio track index is derived from the BedDefinition and ObjectDefinition element position in the list of IAFrame subelements.

2.2.2 Example of Dolby audio track mapping in an IAFrame using subelement position

Logical tracks that appear in the source content are mapped into an IAB frame.

This diagram provides an example of this, and also shows how the essence from this frame might appear in a digital audio workstation (DAW) before and after IAB encoding using Dolby tools
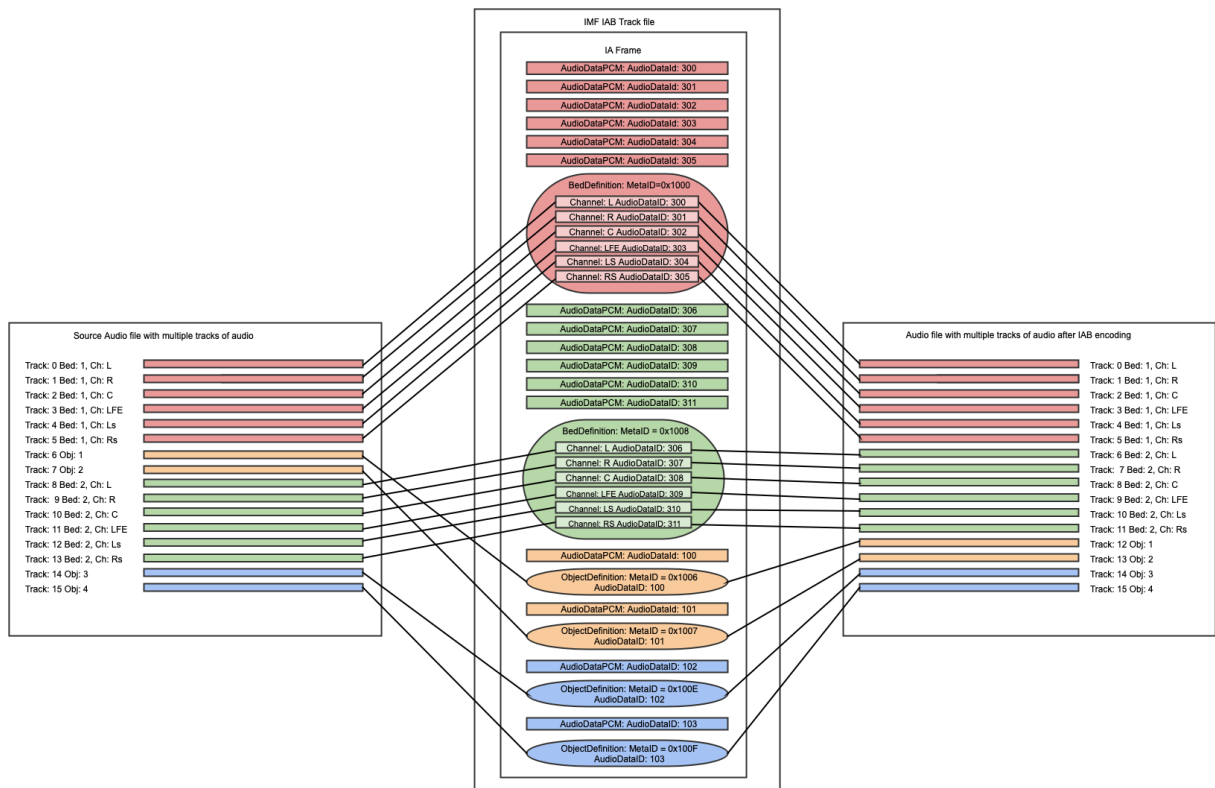
2.2.3 Bed and bed channel counts

The *Level-0 Plug-in* constrains a bitstream in an IAB track file to contain the same number of beds, and each bed must have the same channel configuration. All IAFrames generated by Dolby and belonging to the same IAB track file conform to this constraint.

2.2.4 Object counts

All IAFrames generated by Dolby and belonging to the same IAB track file have the same number of objects.

The IAB track file can have a maximum of 118 objects.

2.2.5 Channel count

The IAB track file can have a maximum of 128 channels.

2.3 Concatenation

It may be necessary to concatenate different IAB track files to create a single, continuous IAB track file.

For instance, it may be necessary to add an ident clip to a feature. In some cases, there may be different bed and bed channel counts in each IAB track file that is to be concatenated. To create a valid output IAB track file, different bed layouts of the input IAB track files need to be reconciled. Dolby uses a method known as flattening as a way to solve this problem.

2.3.1 Flattening

Flattening converts an arbitrary number of beds of different layouts into a single 7.1.2-channel bed. It may also increase the number of objects to accommodate the largest number of objects contained in each IAB track being concatenated.

This pseudocode describes summing and upmixing the beds into one output bed:

```
# up-mix all input bed channels
for each input_bed:
    for each input_bed_channel_id in input_bed.channels:
        for each output_bed_channel_id in output_bed.channels:
            G = get_coefficient(input_bed_channel_id, output_bed_channel_id)
            output_bed[output_bed_channel_id] += G * input_bed[input_bed_channel_id]
# clip output bed values to be within [-1.0, 1.0]
for each channel in output_bed.channels:
    output_bed[channel] = min(1.0, max(output_bed[channel], -1.0))
```

Table 1: Matrix of coefficients for 5.1 to 7.1.2 bed channel upmixing

5.1 channels	7.1.2 channels									
	L	**R**	**C**	**LFE**	**LSS**	**RSS**	**LRS**	**RRS**	**LTS**	**RTS**
L	1	0	0	0	0	0	0	0	0	0
R	0	1	0	0	0	0	0	0	0	0
C	0	0	1	0	0	0	0	0	0	0
LFE	0	0	0	1	0	0	0	0	0	0
Ls	0	0	0	0	**sqrt(0.5)**	0	**sqrt(0.5)**	0	0	0
Rs	0	0	0	0	0	**sqrt(0.5)**	0	**sqrt(0.5)**	0	0

Table 2: Matrix of coefficients for 7.1 to 7.1.2 channel upmixing

7.1 channels	7.1.2 channels									
	L	**R**	**C**	**LFE**	**LSS**	**RSS**	**LRS**	**RRS**	**LTS**	**RTS**
L	1	0	0	0	0	0	0	0	0	0
R	0	1	0	0	0	0	0	0	0	0
C	0	0	1	0	0	0	0	0	0	0
LFE	0	0	0	1	0	0	0	0	0	0
Ls	0	0	0	0	1	0	0	0	0	0
Rss	0	0	0	0	0	1	0	0	0	0
Lrs	0	0	0	0	0	0	1	0	0	0
Rrs	0	0	0	0	0	0	0	1	0	0

This diagram is an example concatenation of two source IAB track files, A and B, whose layout is flattened to that of IAB track file C. Rectangles represent beds, and circles represent objects.
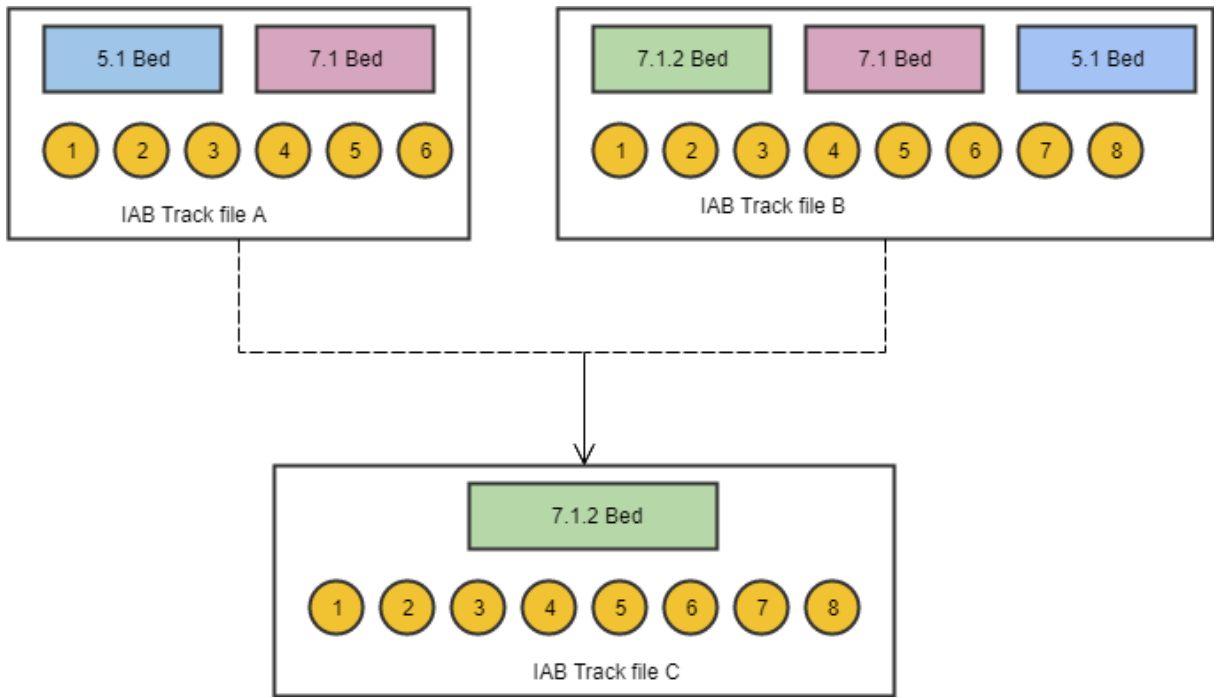
3

Dynamic metadata

IAFrames created by Dolby do not support all element field values defined in the *Immersive Audio Bitstream Specification*. Unsupported field values will not cause parsing to halt, but information present in those fields is ignored.

ObjectDefinition

Table 3: Element fields for ObjectDefinition

Element field	Supported	Notes
ObjectGainPrefix	Yes	Only values `0x00` and `0x01` are supported. A value of `0x01` indicates that an object should be muted.
ObjectGain	No	
ObjectSnap	Yes	A value of 1 indicates that object should be snapped to the loudspeaker; a value of 0 signals otherwise.
ObjectSnapTolExists	No	
ObjectSnapTolerance	No	
ObjectZoneControl	Yes	
ZoneGainPrefix	Yes	A value of `0x01` indicates that the zone is active; a value of `0x00` indicates that the zone is inactive.
ZoneGain	No	
ObjectSpreadMode	Yes	Only mode OBJECT_SPREAD_1D (`0x2`) is supported.
ObjectSpreadX ObjectSpreadY ObjectSpreadZ	No	
ObjectSpread	Yes	
ObjectDecorCoefPrefix	Yes	The values `0x0` and `0x1` are supported.
ObjectDecorCoef	No	
AudioDescription	No	
AudioDescriptionText	No	

BedDefinition

Table 4: Element fields for BedDefinition

Element field	Supported
ChannelGainPrefix	No
ChannelGain	No
ChannelDecorInfoExists	No
ChannelDecorCoefPrefix	No
ChannelDecorCoef	No
AudioDescription	No
AudioDescriptionText	No

Zones

Only these combinations of values in the nine-item ZoneGainPrefix array, as defined in the *Immersive Audio Bitstream Specification*, are meaningful to Dolby tools. All other value combinations are interpreted as Dolby Zone=Floor plus overhead.

Table 5: Zones for Dolby tools

Dolby zone	ZoneGainPrefix array								
	index = 0 All screen speakers left of center	index = 1 Screen center speakers	index = 2 All screen speakers right of center	index = 3 All speakers on left wall	index = 4 All speakers on right wall	index = 5 All speakers on left half of rear wall	index = 6 All speakers on right half of rear wall	index = 7 All overhead speakers left of center	index = 8 All overhead speakers right of center
No back	1	1	1	1	1	0	0	0	0
No back plus overhead	1	1	1	1	1	0	0	1	1
No sides	1	1	1	0	0	1	1	0	0
No sides plus overhead	1	1	1	0	0	1	1	1	1
Center back	0	1	0	0	0	1	1	0	0
Center back plus overhead	0	1	0	0	0	1	1	1	1
Screen only	1	1	1	0	0	0	0	0	0
Screen only plus overhead	1	1	1	0	0	0	0	1	1
Surround only	0	0	0	1	1	1	1	0	0

Table 5: Zones for Dolby tools (continued)

	ZoneGainPrefix array								
Dolby zone	**index = 0** **All screen speakers left of center**	**index = 1** **Screen center speakers**	**index = 2** **All screen speakers right of center**	**index = 3** **All speakers on left wall**	**index = 4** **All speakers on right wall**	**index = 5** **All speakers on left half of rear wall**	**index = 6** **All speakers on right half of rear wall**	**index = 7** **All overhead speakers left of center**	**index = 8** **All overhead speakers right of center**
Surround only plus overhead	0	0	0	1	1	1	1	1	1
Floor	1	1	1	1	1	1	1	0	0
Floor plus overhead	1	1	1	1	1	1	1	1	1

Bed ChannelIDs

Dolby tools use a subset of the bed ChannelIDs that are defined in the *Immersive Audio Bitstream Specification*, as listed in this table.

Table 6: Bed ChannelIDs for Dolby tools

ChannelID code	ChannelID	Notes
0x0	Left	
0x2	Center	
0x4	Right	
0x6	Left Surround	Used for 5.1-channel beds only
0xA	Right Surround	Used for 5.1-channel beds only
0x5	Left Side Surround	Used for 7.1- and 7.1.2-channel beds
0x9	Right Side Surround	Used for 7.1- and 7.1.2-channel beds
0x7	Left Rear Surround	Used for 7.1- and 7.1.2-channel beds
0x8	Right Rear Surround	Used for 7.1- and 7.1.2-channel beds
0xB	Left Top Surround	
0xC	Right Top Surround	
0xD	Low-Frequency Effects (LFE)	

Glossary

channel configuration
A standard for describing a sound system with front, surround, and overhead channels. The first numeral represents the number of front channels, the second represents the number of surround channels, and the third represents the number of height channels. For example, 3/2/2 indicates that there are three front channels, two surround channels, and two height channels.

DAW
Digital audio workstation. An electronic device or computer software application used to record, edit, and produce audio files.

frame
In audio, a series of PCM samples or encoded audio data representing the same time interval for all channels in the configuration. Metadata pertaining to the frame can be carried within the frame or separately, depending on context.

IAB
Immersive audio bitstream. A frame-based audio bitstream that includes audio channels and/or audio objects, plus metadata.

IMF
Interoperable Master Format. A SMPTE standard that defines an interoperable, file-based framework designed to facilitate the management and processing of multiple versions of the same high-quality finished work. See SMPTE ST 2067-2 and related documentation.

LFE
Low-Frequency Effects. A band-limited channel specifically intended for deep, low-pitched sounds.

MXF
Material Exchange Format. A file format used to transfer and store different types of content (for example, audio, video, data, or metadata). MXF currently supports various compression and encoding formats, and its specification can be extended to new essence formats, if needed.

screen center
The axis that runs along the center of the screen.