

SMPTE Education
SMPTE - Enabling Global Education



What is Immersive Audio & Why is it So Cool?

Welcome!

2018-08-09

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)

Your Host

Joel E. Welch

*Director of Education
SMPTE*



SMPTE - Enabling Global Education

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)

2

SMPTE Education Webcast Series
SMPTE - Enabling Global Education



Guest Speaker

Brian Vessa

Sony Pictures Entertainment
Founding chair of SMPTE
TC-25CSS Cinema Sound Systems



© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)

SMPTE Education
SMPTE - Enabling Global Education



What is Immersive Audio & Why is it So Cool?

SMPTE Webcast August 9, 2018

Brian Vessa, Sony Pictures Entertainment
Founding chair of SMPTE TC-25CSS Cinema Sound Systems

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



SMPTE Education
SMPTE - Enabling Global Education



WELCOME!

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



Itinerary



SMPTE - Enabling Global Education

1. Sound and Audio Basics
2. Traditional Cinema Soundfields
3. Immersive Audio Basics and SMPTE ST 2098-5
4. Current Cinema Immersive Sound Systems
5. Immersive Audio Standards To The Rescue!
6. Deep Dive Into SMPTE ST 2098-1 and Immersive Audio Metadata
7. Why Is Immersive Audio So Cool?
8. Immersive Audio For IMF and the Home

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



First-Some Basics



SMPTE - Enabling Global Education

SOUND

- A pressure disturbance in a medium that can be perceived by hearing, and thus is “audible”

AUDIO

- An electrical or digital data representation of sound
- Sound is converted to audio via a transducer such as a microphone, and audio is converted into sound by a transducer such as a loudspeaker or headphone
- Audio in and of itself cannot be perceived by hearing and therefore is not audible
- An Audio File is audio data represented in a defined format such as .wav
- Can be played in real time or be processed by “Digital Signal Processing” (DSP)

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



First-Some Basics



SMPTE - Enabling Global Education

SOUND SYSTEM

- A collection of audio processors and transducers that takes audio input, applies processing and outputs sound

SOUNDFIELD

- The acoustical space created by simultaneously reproducing one or more audio sources
- Traditional Soundfields are two dimensional in the horizontal plane

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



First-Some Basics



SMPTE - Enabling Global Education

SOUNDFIELD CONFIGURATION

- Defined arrangement or configuration of loudspeakers that convey the intended Soundfield

CHANNEL:

- Distinct collection of sequenced audio samples that is intended for delivery to a single loudspeaker or an array of loudspeakers

SOUNDFIELD GROUP:

- Collection of Audio Channels meant to be played out simultaneously through a given Soundfield Configuration

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



Traditional Cinema Soundfields



SMPTE - Enabling Global Education

- Monaural (C)
- Stereo (L, R)
- 3.0 (L, C, R)
- Todd AO (L, Lc, C, Rc, R)
- 70mm (L, Lc, C, Rc, R, S)
- 70mm “baby boom” (L, LFE1/Ls, C, LFE2/Rs, R)
- Surround (L, C, R, S)
- 5.1 (L, C, R, Ls, Rs, LFE)
- 7.1SDS (SDDS) (L, Lc, C, Rc, R, Ls, Rs, LFE)
- 6.1 (L, C, R, Ls, Rs, Cs, LFE)
- 7.1DS (L, C, R, Lss, Rss, Lrs, Rrs, LFE)

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)





SMPTE - Enabling Global Education

Immersive Audio Basics

- All traditional cinema Soundfields are in the horizontal plane
 - The layer of loudspeakers that produce traditional cinema Soundfields is called the “Base Layer”
- Base Layer (from SMPTE ST 2098-5):
- Refers to the nominally horizontal layer of Loudspeakers used in 5.1 and 7.1 Soundfield Configurations

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



Immersive Audio Basics



SMPTE - Enabling Global Education

Name	Symbol	Description
Left	L	Intended to drive the Left loudspeaker (see Annex B).
Right	R	Intended to drive the Right loudspeaker (see Annex B).
Center	C	Intended to drive the Center loudspeaker (see Annex B).
LFE	LFE	Intended to drive the Screen Low Frequency Effects loudspeaker (see Annex B).
Left Surround	Ls	Intended to drive the Left Surround (see Annex B).
Right Surround	Rs	Intended to drive the Right Surround (see Annex B).
Left Side Surround	Lss	Intended to drive the Left Side Surround (see Annex B).
Right Side Surround	Rss	Intended to drive the Right Side Surround (see Annex B).
Left Rear Surround	Lrs	Intended to drive the Left Rear Surround loudspeaker(s) (see Annex B).
Right Rear Surround	Rrs	Intended to drive the Right Rear Surround loudspeaker(s) (see Annex B).
Left Center	Lc	Intended to drive the Left Center loudspeaker (see Annex B).
Right Center	Rc	Intended to drive the Right Center loudspeaker (see Annex B).
Center Surround	Cs	Intended to drive the Center Surround loudspeaker (see Annex B).

Base Layer Audio Channels from ST 428-12

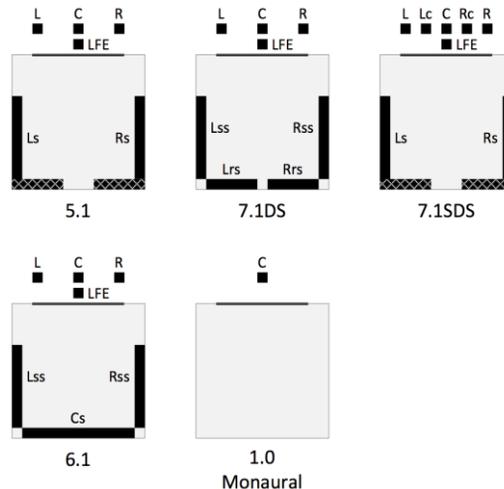
© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



Immersive Audio Basics



SMPTTE - Enabling Global Education



Standard Soundfield Configurations from ST 428-12

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTTE®)



Immersive Audio Basics



SMPTTE - Enabling Global Education

Immersive Sound adds the third dimension of height-sounds above the listener.

Immersive Sound

- Includes sound that emanates from sources at and beyond the Base Layer
- There are two additional loudspeaker layers defined for Immersive Sound in ST 2098-5:
 - Height Layer
 - Refers to the layer of Loudspeakers placed on the walls, above the Base Layer
 - Top Layer
 - Refers to the layer of Loudspeakers placed on the ceiling over the audience

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTTE®)



Immersive Audio Basics



SMPT - Enabling Global Education

Name	Abbreviation	Description
Left Height	Lh	Intended to drive the Left Height Loudspeaker.
Right Height	Rh	Intended to drive the Right Height Loudspeaker.
Center Height	Ch	Intended to drive the Center Height Loudspeaker.
Left Surround Height	Lsh	Intended to drive the Left Surround Height Loudspeaker(s).
Right Surround Height	Rsh	Intended to drive the Right Surround Height Loudspeaker(s).
Left Side Surround Height	Lssh	Intended to drive the Left Side Surround Height Loudspeaker(s).
Right Side Surround Height	Rssh	Intended to drive the Right Side Surround Height Loudspeaker(s).
Left Rear Surround Height	Lrsh	Intended to drive the Left Rear Surround Height Loudspeaker(s).
Right Rear Surround Height	Rrsh	Intended to drive the Right Rear Surround Height Loudspeaker(s).
Left Top Surround	Lts	Intended to drive the Left Top Surround Loudspeaker(s).
Right Top Surround	Rts	Intended to drive the Right Top Surround Loudspeaker(s).
Top Surround	Ts	Intended to drive the Top Surround Loudspeaker(s).

Immersive Audio Channels from ST 2098-5

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPT®)



Immersive Audio Basics



SMPT - Enabling Global Education

SG Name	Audio Channels
9.1OH	L, C, R, Lss, Rss, Lrs, Rrs, LFE, Lts, Rts
9.1HT	L, C, R, Ls, Rs, LFE, Lh, Rh, Lsh, Rsh
11.1HT	L, C, R, Ls, Rs, LFE, Lh, Ch, Rh, Lsh, Rsh, Ts
13.1HT	L, C, R, Lss, Rss, Lrs, Rrs, LFE, Lh, Ch, Rh, Lsh, Rsh, Ts
15.1HT	L, C, R, Lss, Rss, Lrs, Rrs, LFE, Lh, Ch, Rh, Lssh, Rssh, Lrsh, Rrsh, Ts

Immersive Soundfield Groups from ST 2098-5

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPT®)



Immersive Soundfield Configurations



SMPTÉ - Enabling Global Education

Name	Base Layer Loudspeakers	Height/Top Layer Loudspeakers
9.1OH		
9.1HT		
11.1HT		

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTÉ®)



Immersive Soundfield Configurations



SMPTÉ - Enabling Global Education

Name	Base Layer Loudspeakers	Height/Top Layer Loudspeakers
13.1HT		
15.1HT		

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTÉ®)



Immersive Sound Systems

IOSONO



SMPT - Enabling Global Education



© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPT®)

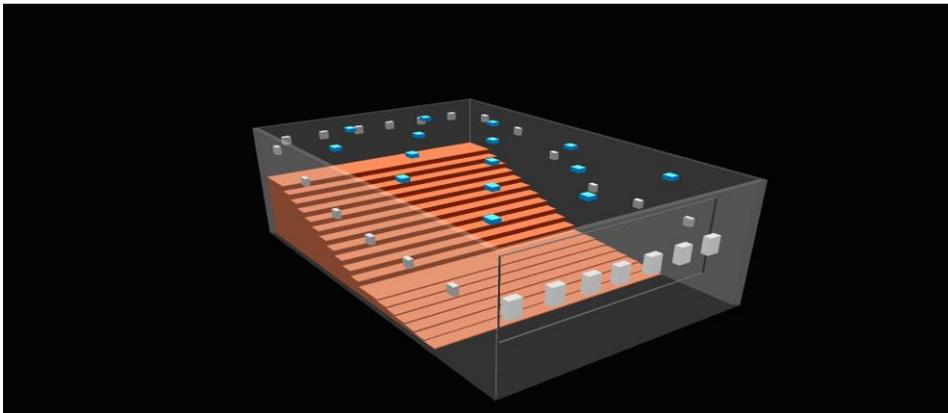


Immersive Sound Systems

IOSONO



SMPT - Enabling Global Education



© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPT®)

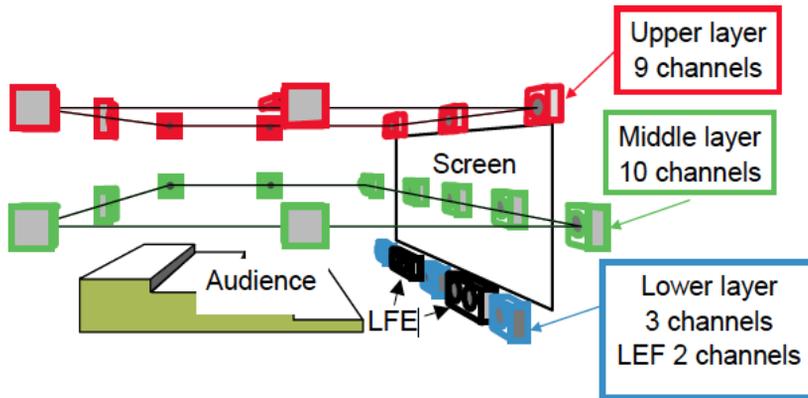


Immersive Sound Systems

NHK 22.2



SMPTe - Enabling Global Education



© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTe®)

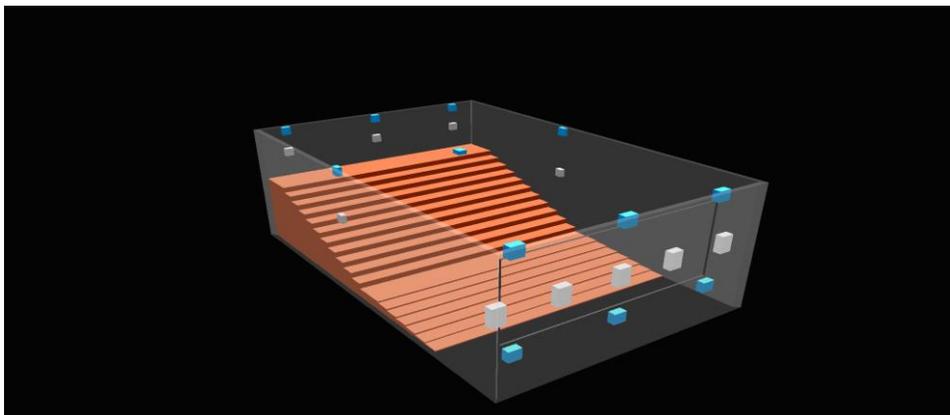


Immersive Sound Systems

NHK 22.2



SMPTe - Enabling Global Education



© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTe®)

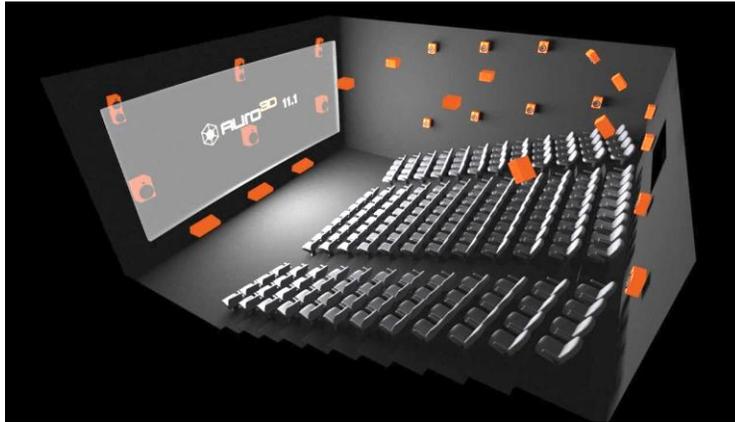


Immersive Sound Systems

Auro-3D



SMPTe - Enabling Global Education



© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTe®)

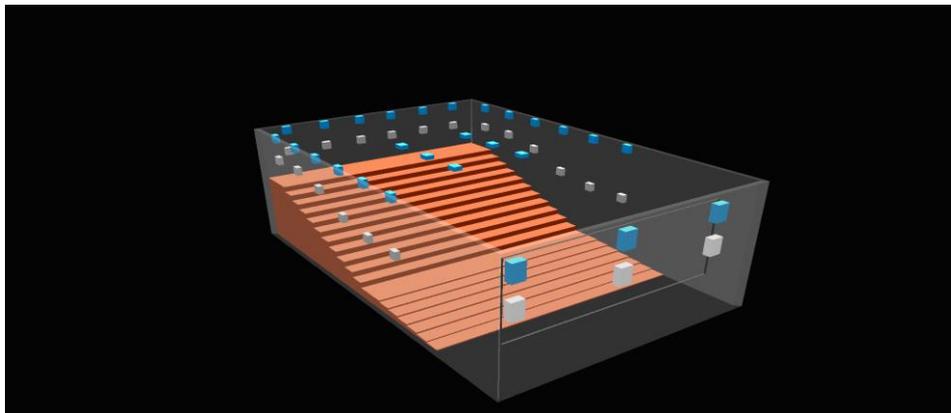


Immersive Sound Systems

Auro-3D



SMPTe - Enabling Global Education



© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTe®)



Immersive Sound Systems



SMPT - Enabling Global Education

Auromax



© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPT®)



Immersive Sound Systems



SMPT - Enabling Global Education

Dolby Atmos



© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPT®)

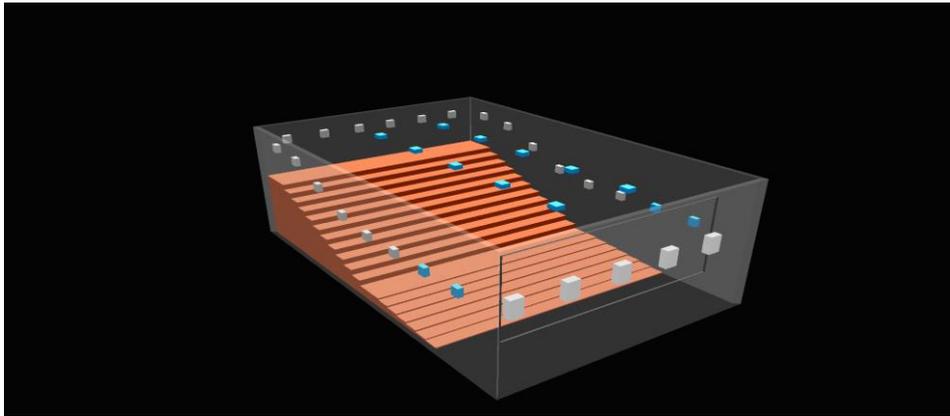


Immersive Sound Systems

Dolby Atmos



SMPTTE - Enabling Global Education



© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTTE®)



Immersive Sound Systems

DTS:X



SMPTTE - Enabling Global Education



© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTTE®)



Immersive Audio Basics



SMPTE - Enabling Global Education

In addition to being able to utilize immersive audio channels, immersive audio can utilize “objects”

Audio Object (From SMPTE ST 2098-1):

- A segment of audio essence with associated metadata describing positional and other properties which may vary with time
- A set of audio samples and associated metadata intended for reproduction according to the position in space and other properties as indicated by the metadata. The position may or may not be associated with a single Loudspeaker

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



Immersive Audio Basics



SMPTE - Enabling Global Education

Audio Object Attributes

- Can be audio of any duration that has associated metadata that describes how it is to be reproduced within a soundfield
- This metadata describes the position, spread, motion characteristics and other rendering information
- Objects can move within the soundfield, be reproduced in a single position, or by a specific loudspeaker
- An object may be defined to behave very much like a channel

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



Immersive Audio Basics



SMPTE - Enabling Global Education

Immersive Audio (from ST 2098-1 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

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



Immersive Audio Basics



SMPTE - Enabling Global Education

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
- In an immersive sound system, rendering is the process of taking immersive audio as input and transforming it to play back as intended into a given immersive soundfield configuration
- A renderer may utilize technologies such as Wave Field Synthesis (WFS), Higher Order Ambisonics (HOA), or Vector Base Amplitude Panning (VBAP), among others

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



Immersive Audio Basics



SMPT - Enabling Global Education

The renderer is the “magic box” that takes the mix intent conveyed by the immersive audio and its metadata, and decides how it should be played back in the immersive sound system to which it is connected, such that the mix intent is conveyed in the best possible way

Current rendering systems only accept one audio distribution format and are designed for one type of sound system

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPT®)



Immersive Audio Standards



SMPT - Enabling Global Education

The big issue for the industry up to now has been that one needs to mix and distribute specifically for the type of sound system that will play the audio. So multiple mixes and DCP's have been required to distribute to theaters with different sound systems. This is not a sustainable model, as there is a lot of work and time required to do this with little or no ROI

Theaters who purchase immersive sound systems can only play content mixed and distributed for that system, so theater owners are hesitant to make the investment

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPT®)



Immersive Audio Standards



SMPTE - Enabling Global Education

SMPTE saw the problem and started working on studying immersive audio in 2013 and began to create standards in early 2014

SMPTE realized that this was a complex issue, as there are many aspects that needed to be addressed to achieve interoperability. A working group was formed to address these, which has a number of drafting groups that are actually creating the standards

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



Immersive Audio Standards



SMPTE - Enabling Global Education

THE GOAL:

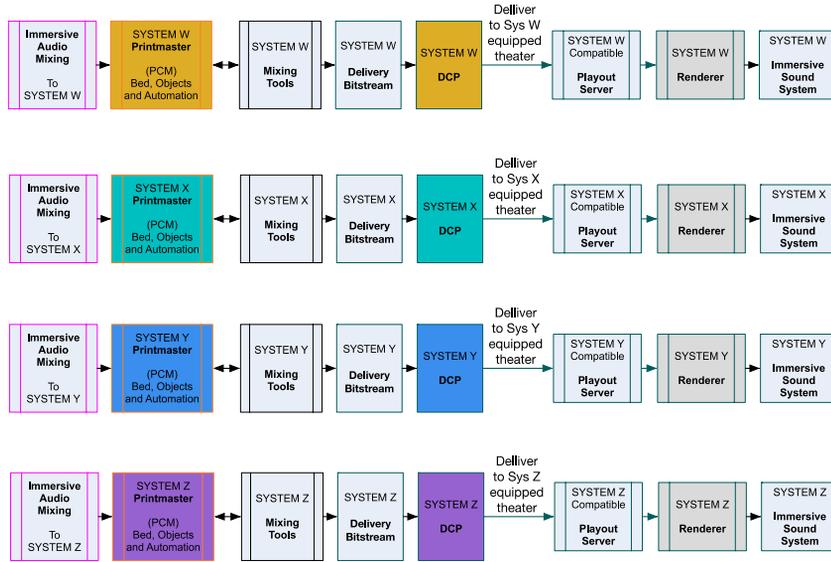
Create one immersive mix using any set of tools that can output the SMPTE standard delivery bitstream, package that bitstream in a standardized Immersive Audio Digital Cinema Package (DCP), and distribute it to any theater with an immersive sound system that conforms to the SMPTE immersive audio standards

➤ Single Mix-Single Inventory-Single Distribution

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



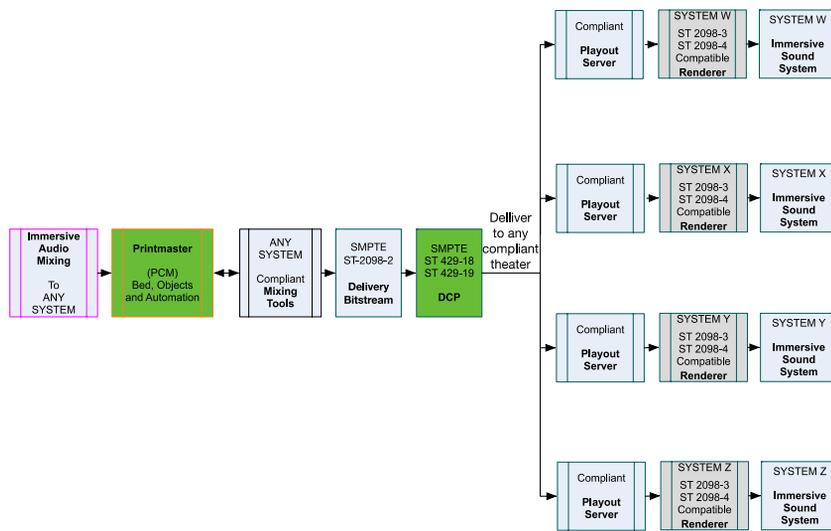
CURRENT IMMERSIVE AUDIO BLOCK DIAGRAM



© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTÉ®)



INTEROPERABLE IMMERSIVE AUDIO BLOCK DIAGRAM



© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTÉ®)



Immersive Audio Standards



SMPTE - Enabling Global Education

- Immersive Audio Study Group report
- ST 2098-1 Immersive Audio Metadata
- ST 2098-2 Immersive Audio Bitstream Specification
- EG 2098-3 Immersive Audio Renderer Expectations
- RP 2098-4 Immersive Audio Renderer Interoperability Testing Procedure
- ST 2098-5 D-Cinema Immersive Audio Channels and Soundfield Groups
- ST 429-14 Aux Data Track File
- ST 429-18 D-Cinema Packaging - Immersive Audio Track File
- ST 429-19 D-Cinema Packaging —DCP Operational Constraints for Immersive Audio
- ST 430-14 D-Cinema Operations —Digital Sync Signal and Aux Data Transfer Protocol
- ST 430-17 SMS-OMB Communications Protocol Specification
- RP 430-18 Immersive Audio OMB Implementation
- Immersive Audio Implementation Study Group Report (internal document)

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



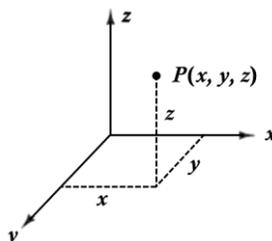
ST 2098-1: Immersive Audio Metadata



SMPTE - Enabling Global Education

Three dimensional space representation and coordinate system

- Audio Object positional metadata shall indicate the placement of an Audio Object using a Cartesian coordinate system. The x coordinate shall represent the left-right dimension, the y coordinate shall represent the front-back dimension, and z coordinate shall represent the down-up dimension as shown below.



© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



ST 2098-1: Immersive Audio Metadata



SMPTE - Enabling Global Education

Three dimensional space representation and coordinate system:

- The Cartesian coordinate values used for Audio Object position shall be normalized relative to reference points of a cube, which represents an idealized cinema model.
- The front plane is the location of the screen; “left” is relative to an observer in the cube, facing the front. The metadata may support Audio Object locations inside, on, and outside the cube. At a minimum, location metadata shall support locations on and inside the cube from the Z axis midpoint to the top of the cube. The reference points are defined as (next slide):

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



ST 2098-1: Immersive Audio Metadata



SMPTE - Enabling Global Education

Three dimensional space representation and coordinate system:

- For the x-axis, the reference points shall be the left and right cube faces. These reference points shall have fixed, defined values.
- For the y-axis, the reference points shall be the front and back cube faces. These reference points shall have fixed, defined values.
- For the z-axis, the reference points shall be the bottom, midpoint, and top of the cube. These reference points shall have fixed, defined values.
- Locations within the cube shall be represented using linear interpolation between the adjacent reference points on each axis. Locations outside the cube, if supported, shall be represented based on a linear extrapolation from the reference points on the cube.

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



ST 2098-1: Immersive Audio Metadata



SMPTe - Enabling Global Education

Three dimensional space representation and coordinate system:

- The actual cube reference values are to be determined as part of the bitstream definition and are outside the scope of this document. The following is one possible definition.
- The origin, $[X,Y,Z] = [0,0,0]$, is set to be the bottom left front corner of the cube. The defined values are as follows:
- X axis: left face value = 0; right face value = 1
- Y axis: front face value = 0; rear face value = 1
- Z axis: bottom value = 0; top value = 1; midpoint value = 0.5
- A coordinate of $[0.5, 0, 0.5]$ will be located on the center of the front face.

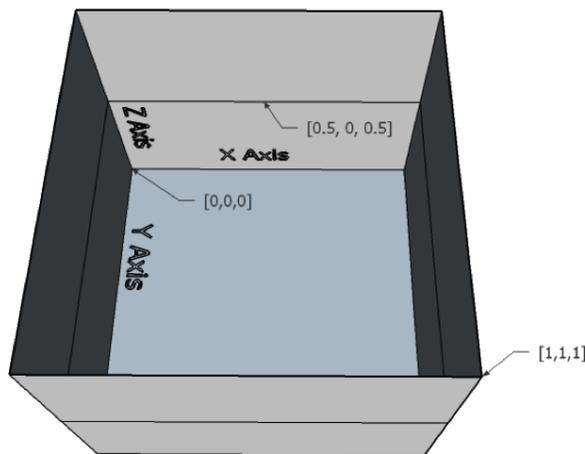
© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTe®)



ST 2098-1: Immersive Audio Metadata



SMPTe - Enabling Global Education



Cube with example coordinates

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTe®)



ST 2098-1: Immersive Audio Metadata



SMPTE - Enabling Global Education

Three dimensional space representation and coordinate system:

- Mapping of object positions within the cube to cinema (or mixing stage) Loudspeakers is the function of the renderer. However, the reference points on the cube shall have defined meaning regardless of the shape of the room into which the locations are mapped.
- The front face of the cube shall map to the nominal front wall of the cinema. Furthermore, the frontmost Loudspeaker-mounting surface shall be considered the nominal front wall of the cinema.
- The left face of the cube shall map to the nominal left wall of the cinema. Furthermore, the leftmost Loudspeaker-mounting surface shall be considered the nominal left wall of the cinema.
- The right face of the cube shall map to the nominal right wall of the cinema. Furthermore, the rightmost Loudspeaker-mounting surface shall be considered the nominal right wall of the cinema.

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



ST 2098-1: Immersive Audio Metadata



SMPTE - Enabling Global Education

Three dimensional space representation and coordinate system:

- The back face of the cube shall map to the nominal rear wall of the cinema. Furthermore, the rearmost Loudspeaker-mounting surface shall be considered the nominal rear wall of the cinema.
- The mid-height plane of the cube shall map to the height of a legacy 2-dimensional Loudspeaker system (e.g. 5.1 or 7.1).
- The top of the cube shall map to the ceiling of the cinema. Furthermore, the overhead Loudspeaker-mounting surface shall be considered the nominal ceiling of the cinema.
- The bottom of the cube shall map to the nominal floor of the cinema.

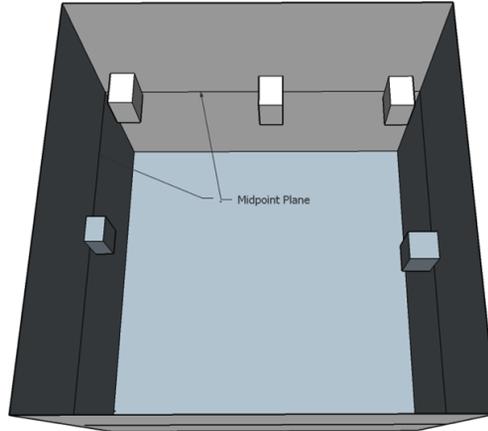
© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



ST 2098-1: Immersive Audio Metadata



SMPTE - Enabling Global Education



Mapping of midpoint plane to base layer Loudspeaker height

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



ST 2098-1: Immersive Audio Metadata



SMPTE - Enabling Global Education

Example - Mapping object position to a cinema Loudspeaker array:

- By using reference points that map to any room, the intent is not to preserve object positions relative to an observer, but relative to the shape and size of the playback environment.
- On the following slide, the diagram on the left shows a possible panner interface, the one on the right shows the simplified overhead view of a cinema. In the panner diagram, the mixer has placed a sound source (green dot) 1/3 of the way from front to back, and 1/4 of the way from left to right. The light blue line represents the room perimeter of a nominal, square room.
- The cinema diagram shows how the object position could be mapped to a cinema that is not a square, but is longer than wide, and has the screen left and right Loudspeakers mounted inside the left and right walls. The dotted-blue line shows the perimeter of the room as now defined by the Loudspeaker mounting. This will vary from cinema to cinema. Note that the position relative to the reference points (walls) remains the same. The sound source is 1/3 of the way from front to back, and 1/4 of the way from left to right.

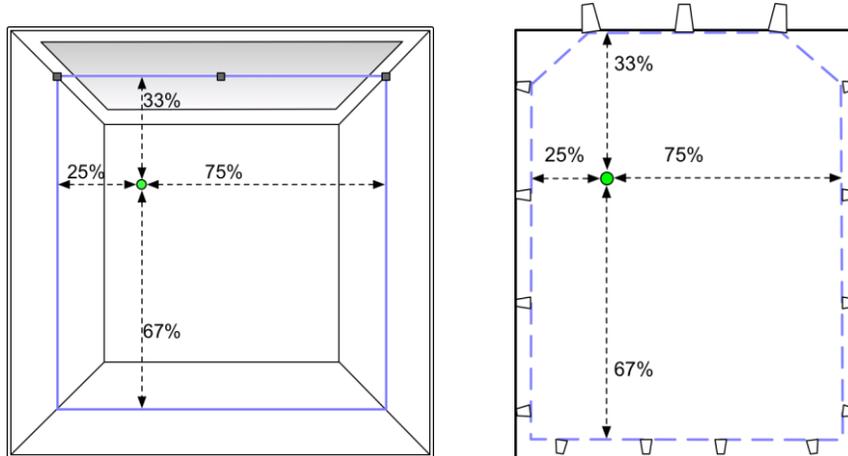
© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



ST 2098-1: Immersive Audio Metadata



SMPTe - Enabling Global Education



Mapping ideal cube to example cinema

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTe®)



ST 2098-1: Immersive Audio Metadata



SMPTe - Enabling Global Education

Some General Terms and Definitions

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

It may be desirable to identify certain Audio Objects or Audio Channels as belonging to a Group for the purposes of common processing or action. Immersive audio metadata may support the function of allowing Audio Objects or Audio Channels to be associated with a Group for this purpose.

- **Sample**-A discrete number representing the amplitude of an audio signal at an instant in time.
- **Target Environment**-A specific set of conditions that is present in the playback environment. (For example, the available soundfield configurations, number of speakers and their placement in the room, the dimensions of the room, etc.)

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTe®)



ST 2098-1: Immersive Audio Metadata



SMPTTE - Enabling Global Education

Channel metadata

- **Channel Identifier**-Uniquely identifies an Audio Channel. No two Audio Channels shall have the same Channel Identifier at any instant in time.
- **Routing Destination**-Identifies the single Loudspeaker or other reproduction device associated with the channel. "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.
- **Waveform Reference**-This metadata item references audio essence associated with the channel. The reference shall allow unambiguous identification of the audio essence.

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTTE®)



ST 2098-1: Immersive Audio Metadata



SMPTTE - Enabling Global Education

Bed metadata

- **Bed** (a.k.a. Immersive Audio 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. (Channels in immersive audio are always within a Bed.)
- **Bed Identifier**-This Identifier uniquely identifies the Bed. No two Beds shall have the same Bed Identifier at any instant in time. This allows referencing multiple Beds if desired.
- **Bed Channel List**-This metadata item lists the Audio Channels in the Bed.

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTTE®)



ST 2098-1: Immersive Audio Metadata



SMPT - Enabling Global Education

Bed metadata

- **Remap Coefficients**-These specify how to map the original channels to a different target configuration. This is a set of values indicating how much gain is to be applied to each Audio Channel of a Bed to generate each output for the target Soundfield Configuration. If remap coefficients are provided, additional metadata shall identify the conditions (Target Environments) under which they should be used.
- **Conditional Bed**-A mixer may want to include one or more alternative Beds for different target Soundfield Configurations. The metadata shall support identifying alternative Beds and the conditions (Target Environments) under which they should be used.
- At a minimum, the following target Soundfield Configurations, defined in SMPTE ST 428-12, shall be supported:
 - 5.1: (L, R, C, LFE, Ls, Rs) 7.1DS: (L, R, C, LFE, Lss, Rss, Lrs, Rrs)

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPT®)



ST 2098-1: Immersive Audio Metadata



SMPT - Enabling Global Education

Object metadata

Note that all object metadata is considered to be dynamic (time-varying) unless explicitly stated differently

- **Object Identifier**-This metadata item provides a unique identity for an object. No two objects shall have the same Object Identifier at any instant in time. An Object Identifier is static for the duration of an object. The cardinality of the identifier space shall be large enough to support the highest number of simultaneous Audio Objects that exist at any point in time during the presentation.
- **Waveform Reference**-This metadata item references audio essence associated with the Audio Object. The reference shall allow unambiguous identification of audio essence. The waveform reference may include a (non-zero) sample offset.

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPT®)



ST 2098-1: Immersive Audio Metadata



SMPTE - Enabling Global Education

Object metadata

- **Object Position**-This metadata item locates the Audio Object in a three-dimensional space. The coordinate system and frame of reference to be used in this standard was defined earlier in this presentation.
- **Object Spread**-This metadata element describes the Audio Object's size and shape in a three-dimensional space. The coordinate system and frame of reference to be used in this standard was defined earlier in this presentation.
- **Object Gain**-This metadata item specifies the gain applied to audio essence associated with the Audio Object and shall allow a gain of 0 dB.

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



ST 2098-1: Immersive Audio Metadata



SMPTE - Enabling Global Education

Object metadata

- **Object Lifetime**-An Audio Object lifetime may be explicitly included. The metadata item specifies the intervals (start and duration for a single interval) in time when an Audio Object is active (or present), i.e. the intervals in time where Audio Object metadata may impact Audio Object rendering.
- **Object Audio Description**-This metadata describes a characteristic of the Audio Object that may be used to target the Audio Object for processing. Supported types shall include: dialog, music, and effects.

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



ST 2098-1: Immersive Audio Metadata



SMPT - Enabling Global Education

Object metadata

- **Decorrelation**-The decorrelation metadata item refers to processing the source signals used to reproduce an auditory event to alter their relationship while maintaining the original sound for each individual signal. The minimum value indicates that no decorrelation effect is intended, and the maximum value indicates that the maximum decorrelation effect is intended.

General Description: The perceived sound image when reproducing an audio object across two or more Loudspeakers in an immersive sound system can be localized or diffuse depending on whether the source signals representing the audio object are correlated or decorrelated, respectively. Reproducing a sound with multiple correlated signals yields an easily locatable sound image. Reproducing a sound with multiple related but uncorrelated sources yields a broader, more diffuse sound image. This can be called a decorrelation effect.

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPT®)



ST 2098-1: Immersive Audio Metadata



SMPT - Enabling Global Education

Object metadata

- **Snap Tolerance**-This metadata item indicates the degree to which preservation of object timbre has priority over preservation of object position. This property has extreme values indicating 'preserving object timbre has highest priority' and 'preserving object position has highest priority', respectively.

General Description: If there is no loudspeaker near the location that an object has been directed to go, it can be advantageous to reproduce it in a single loudspeaker to preserve its timbre rather than spread between loudspeakers in order to attempt to preserve its location. If the value is set to 'preserve object timbre', an object will "snap" to the nearest loudspeaker-otherwise, it will be reproduced in multiple loudspeakers in order to best preserve the positional intent.

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPT®)



ST 2098-1: Immersive Audio Metadata



SMPTe - Enabling Global Education

Object metadata

- **Conditional Object**-A mixer may want to include one or more alternative Objects for different target Soundfield Configurations. The metadata shall support identifying alternative Objects and the conditions (Target Environments) under which they shall be used. Note: A conditional Object can replace another Object.

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTe®)



ST 2098-1: Immersive Audio Metadata



SMPTe - Enabling Global Education

Zone metadata

- **Zone**-A defined region within a generic listening environment.
- **Zone Control**-This metadata item, when present, indicates the degree to which specified set of Loudspeakers (Zones) are to be excluded from rendering.
 - The actual mapping of Zones to Loudspeakers will be defined, for a cinema, at the time of renderer configuration.
 - Zones should be configured as a non-overlapping partition of the set of all available Loudspeakers.
- **Zone Gain**-This metadata item represents the degree to which a Zone is included in sound reproduction. This property shall support a range of values. The range shall include extreme values indicating 'fully enabled' and 'fully disabled'. An object may have a separate gain value for each basic Zone as defined in the following table.

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTe®)



ST 2098-1: Immersive Audio Metadata



SMPTE - Enabling Global Education

Zone metadata

- General Description:* When an object is intended to be located at a particular place, or if an object is being panned around the room in a specific way, it may be advantageous to not allow it to be reproduced by all of the possible loudspeakers that the renderer has available. For example, when panning from the rear, over head and to the front, the pan may sound closer to the intention if the side speakers are excluded. When locating an object very high up, it may be better to exclude the height layer and only use the top layer. Zone metadata allows the user to define this for each object. Most panning GUI's will have panning "modes" that use zone exclusion as part of the panning algorithm, invisible to the user.

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



ST 2098-1: Immersive Audio Metadata



SMPTE - Enabling Global Education

Number	Description
1	Base layer screen Loudspeakers left of center
2	Base layer center screen Loudspeakers
3	Base layer screen Loudspeakers right of center
4	Height layer screen Loudspeakers left of center
5	Height layer center screen Loudspeakers
6	Height layer screen Loudspeakers right of center
7	Base layer rear wall Loudspeakers left of center
8	Base layer center rear wall Loudspeakers
9	Base layer rear wall Loudspeakers right of center
10	Height layer rear wall Loudspeakers left of center
11	Height layer center rear wall Loudspeakers
12	Height layer rear wall Loudspeakers right of center
13	Base layer left wall Loudspeakers
14	Height layer left wall Loudspeakers
15	Base layer right wall Loudspeakers
16	Height layer right wall Loudspeakers
17	Top layer Loudspeakers left of center
18	Top layer center ceiling Loudspeakers
19	Top layer Loudspeakers right of center

Zones

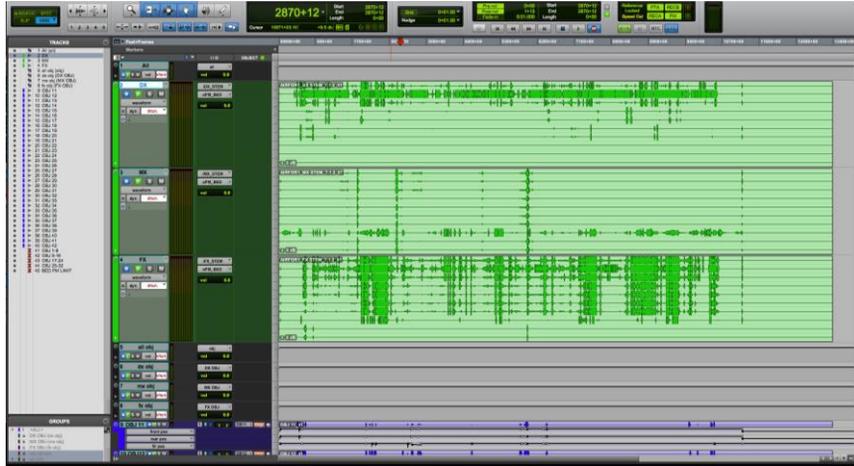
© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



Immersive Audio in Practice



SMPT - Enabling Global Education



Stem Beds

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPT®)



Immersive Audio in Practice



SMPT - Enabling Global Education



Objects

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPT®)



Why is Immersive Audio so cool?



SMPTTE - Enabling Global Education

Immersive Sound: Telling the story

- Sound localization, perspective and imaging is in 3D space, enhancing width and depth as well as adding height
- New possibilities for capturing sounds in nature in 3D and playing them in 3D in the theater
- Gives the sound designers and mixers a larger space to design, place and move sounds-new ways to tell the story through sound
- The moviegoer can be made to feel they are in the same space as the action
- Enveloping the viewer brings more engagement with the story
- Works well for 2D image, can bring additional life to 3D images

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTTE®)



Why is Immersive Audio so cool?



SMPTTE - Enabling Global Education

Immersive Sound: Design

- Backgrounds and ambiances-opens up the upper reflections in rooms and the sky in outdoor scenes. Can now design for 3D-bring the listener into the environment. Change with scene cuts. Design for point of view
- Music-opens up upper reflections, makes the recording room translate more realistically to the theater. Can spread in 3D space like an orchestra or design completely uniquely. Create a virtual soundstage
- Sound effects-flyovers are more convincing, pass-byes wider, whooshes have more space...room treatments on foley and local effects to bring the listener into the space
- Dialog-very convincing location treatment. Offstage dialog can be located to POV, PA/radio calls from above. Great use in SciFi as well as drama.

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTTE®)



Immersive Audio In Movies Today



SMPTE - Enabling Global Education

Over the past five years, hundreds of movies have been mixed in immersive audio

Almost all Hollywood “tentpole” titles are now natively mixed in immersive audio

Some broadcast content providers are mixing all new content in immersive audio and deriving other deliverables from there

Library titles are finding a new audience, being remixed with immersive audio for release in cinema, UHD Blu-ray and streaming.

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



Immersive Audio Mastering: IMF



SMPTE - Enabling Global Education

The Interoperable Master Format (IMF) was standardized by SMPTE in 2013

- IMF is an excellent master format from which to create the many and varied deliverables for home and mobile devices
- Many content providers now master in IMF instead of video tape

Current IMF Immersive Audio Delivery:

- Immersive Audio is created by proprietary tools and delivered as a proprietary file outside the IMF (“sidecar”)
- The IMF CPL and IMP asset list do not know of its presence and the IMF CPL does not reference it in playback

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



Immersive Audio Mastering: IMF



SMPTe - Enabling Global Education

Future IMF Immersive Audio Delivery:

- SMPTE ST 2067-201 (in process) is standardizing an IMF Immersive Audio Bitstream track file, which uses the ST 2098-2 bitstream to create an IMF IAB track file
- This will be referenced and played by the IMF CPL, and can be put into an IMF package for delivery
- A single IAB IMF can be transcoded to create multiple specific deliverables for physical disc and broadcast (e.g. TrueHD, DTS-MA, AC4, MPEG-H, DD+, etc)

More on Immersive Audio for IMF in a future Webinar!

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTe®)



Immersive Audio In The Home



SMPTe - Enabling Global Education

Immersive Audio is already a reality in the home

- Available on UHD Blu-ray, OTT and streaming
- Many home AVR's are now available that can play immersive audio
- Smart TV's are starting to be equipped with immersive sound
- Immersive speaker systems and soundbars are readily available
- Recent headphone technology, processors and apps can now convey immersive sound in headphones
- Can be played on mobile devices through speakers or headphones

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTe®)





SMPTE Education
SMPTE - Enabling Global Education



IMMERSIVE AUDIO IS COOL!

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



SMPTE Education
SMPTE - Enabling Global Education



THANK YOU!

Brian Vessa, Sony Pictures Entertainment
Brian_Vessa@spe.sony.com

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)



SMPT Standards Webcast Series
SMPT Education Webcast Series
SMPT Professional Development Academy - Enabling Global Education
SMPT - Enabling Global Education



Questions

Brian Vessa

Sony Pictures Entertainment

Founding chair of SMPT
TC-25CSS Cinema Sound Systems



© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPT®)