Report of the Study Group on Immersive Audio Systems: Cinema B-Chain and Distribution

31 March 2014

Society of Motion Picture and Television Engineers
3 Barker Avenue
White Plains, NY 10601 USA
www.smpte.org

Report of the Study Group on Immersive Audio Systems: Cinema B-Chain and Distribution

March 31, 2014

Society of Motion Picture and Television Engineers
3 Barker Avenue
White Plains, NY 10601 USA
www.smpte.org

Table of Contents

Page

1. Introduction 3
2. Scope 3
3. Glossary 4
4. Immersive Systems Reviewed 5
5. Immersive Sound System A-Chain Characteristics by System 6
   5.1. Distribution Payload 6
   5.2. Sound/Picture Synchronization 7
6. Immersive Sound System B-Chain Characteristics By System 7
7. Immersive Sound System Considerations 7
   7.1. Legacy Compatibility 7
   7.2. Speaker Performance and Calibration Criteria 8
   7.3. Immersive Soundtrack Loudness Considerations 8
8. Considerations and Proposals For Common Track File and B-Chain Description 8
   8.1. Common Distribution Track File 8
   8.2. Shared B-Chain Description 9
   8.3. Translation Considerations In Playing Through Multiple Sound Systems 10
9. Issues not covered 10
10. Recommendations of this Study Group for Further Work 10
11. Bibliography 12
Annex A Comparison of Reviewed Immersive Sound Systems 13
Annex B Reviewed Immersive Sound System Speaker Diagrams 14
Annex C Reviewed Immersive Sound System Descriptions 15

Foreword

The Oscar® and Emmy® Award-winning Society of Motion Picture and Television Engineers® (SMPTE®), a professional membership association, is the worldwide leader in developing and providing motion-imaging standards and education for the communications, technology, media, and entertainment industries. An internationally recognized and accredited organization, SMPTE advances moving-imagery education and engineering across the broadband, broadcast, cinema, and IT disciplines. Since its founding in 1916, SMPTE has published the SMPTE *Motion Imaging Journal* and developed more than 650 standards, recommended practices, and engineering guidelines. More than 6,000 members — motion-imaging executives, engineers, creative and technology professionals, researchers, scientists, educators, and students — who meet in Sections throughout the world, sustain the Society. Participation the SMPTE Standards Community 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.

Society of Motion Picture and Television Engineers
3 Barker Avenue
White Plains, NY 10601 USA
www.smpte.org

This SMPTE Engineering Document was prepared by Immersive Audio Study Group 25CSS.

Trademark Acknowledgements

1. Introduction

Monophonic sound was transformative for the art of motion pictures. Every sound format that followed has attempted to improve the sense of reality by adding more channels feeding ever more speakers. With rare exception, those speakers were located behind the screen and around the audience, effectively on the same horizontal plane. Even though the ensuing decades transitioned from optical film to digital media, the planar sound concept lives on with 5.1 and 7.1 soundfield configurations. (In this document, any references to "5.1" also include 7.1 unless noted.)

The movie industry is now evolving toward an expanding range of audio formats with additional audio signals driving height and overhead speakers to achieve a more immersive experience. Broadly stated, the Immersive Audio Study Group (IASG) was formed to explore ways to deliver these capabilities using a common distribution method and possible shared B-chain attributes.

2. Scope

With the development in the cinema industry of the expanded capability audio formats being offered by various companies, common methods are needed for distribution to multiple types of sound systems. Special consideration needs to be given to the B-chain to maximize shared components.

1. Evaluate the distribution requirements of the cinema sound systems submitted to the Study Group, including a) object-based, b) channel-based, and c) linear-matrixed systems (e.g. Higher Order Ambisonics), and to explore methods for a common distribution track file.
2. Evaluate the B-chain requirements of the systems submitted to the Study Group, and explore a shared B-chain description.

To complete the project the group has considered these specific tasks:

1. Assemble known system requirements
2. Find common approaches
3. Define differences
4. Research possible solutions
5. Create a Report

It should be noted that at no time did this group perform any objective or subjective testing of any of the immersive systems reviewed.

3. Glossary

For the purposes of this document, the following terminology and definitions apply:

- **Sound**. A pressure disturbance in air that can be perceived by hearing.
- **Audio**. An electrical or digital 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. Audio is distinguished from sound, as it is solely a representation of sound and cannot be perceived by hearing.
- **Sound system**. A collection of audio processors and transducers that takes audio as input, applies processing and outputs sound
- **Immersive sound**. The term used to describe sound that emanates from sources beyond the horizontal plane by means of enhanced spatial properties such as additional height and overhead speakers and localized apparent sound sources within the auditorium.
- **Immersive sound system**. A sound system capable of producing immersive sound.
- **Audio essence**. The base audio that is created and delivered to a sound system. In D-Cinema, audio essence is wrapped in MXF to create a D-Cinema track file.
- **Immersive audio**. Audio that is created with the intent of being reproduced as Immersive Sound via an Immersive Sound System. Immersive audio can consist of channel and/or object based audio essence, with associated metadata, including temporal and spatial metadata.
- **Immersive audio track file**. A D-Cinema track file that contains Immersive audio. -
- **Channel**. SMPTE ST 377-4 defines *channel* as a "distinct collection of sequenced audio samples that is intended for delivery to a single loudspeaker or other reproduction device." For purposes of this document, "other reproduction devices" may include an array of speakers driven in unison, or it may include a process to modify the audio prior to presentation. Channels may contain linear matrix encoding that will allow enhanced reproduction by a number of additional loudspeakers in a designated manner.
- **Matrix encoding**. A method whereby multiple audio signals are combined in a prescribed manner. This process is usually done before distribution in order to allow additional processing to achieve the desired reproduction of the source at playback. Examples include MS stereo, or surround systems such as Sansui QS and Ambisonics.
- **Object**. A segment of audio of any duration, with associated metadata describing how it is intended to be reproduced, such as spatial position, spread, motion characteristics and perhaps other attributes. The number of concurrent objects may vary over time. Objects may be defined as channels.
- **Soundfield**. SMPTE ST 377-4 defines *soundfield* as the "acoustical space created by simultaneously reproducing one or more audio channels." For purposes of this document, the definition is expanded to the "acoustical space created by simultaneously reproducing one or more audio channels and/*or objects*." Traditional cinema soundfields have been solely on the horizontal plane.
- **Immersive Soundfield**. A soundfield where sound can be reproduced from all three dimensions.
- **Rendering**. A process to transform a given audio distribution format to the particular playback speaker configuration in use. An object-based renderer may utilize technologies such as Wave Field Synthesis (WFS), Higher Order Ambisonics (HOA), or Vector Base Amplitude Panning (VBAP), among others. (See Appendix C for reference documents covering these rendering techniques.)
- **DCP**. Digital Cinema Package. Encapsulates picture, sound, data, and text D-Cinema track files and one or more Composition Play Lists (CPL) for delivery to a theater (see SMPTE ST 429-2 for DCP operational constraints).

- **D-Cinema player**. The real-time playout device for the DCP track files for transfer to the image projection and sound systems.
- **HI/VI-N**. Hearing Impaired, Visually Impaired-Narrative accessibility tracks.
- **Media block**. A media block decrypts, decodes and forensically marks some or all of the DCP contents for playback.
- **A-chain, B-chain**. Traditionally, the A-chain is that part of a motion-picture audio system that plays the audio out and delivers it to the playback system in the theater, which in a general sense may be considered to be the input to the cinema sound processor. The B-chain is that part of a motion-picture sound reproduction system commencing at the input to the cinema sound processor and extending into the listening area. Note that the screen and room acoustics are part of the B-chain as well as the cinema processor, amplifiers and loudspeakers. An immersive sound system adds a speaker renderer, which may be integrated with the media block or be part of the cinema processor. Figure 1 illustrates the latter case.

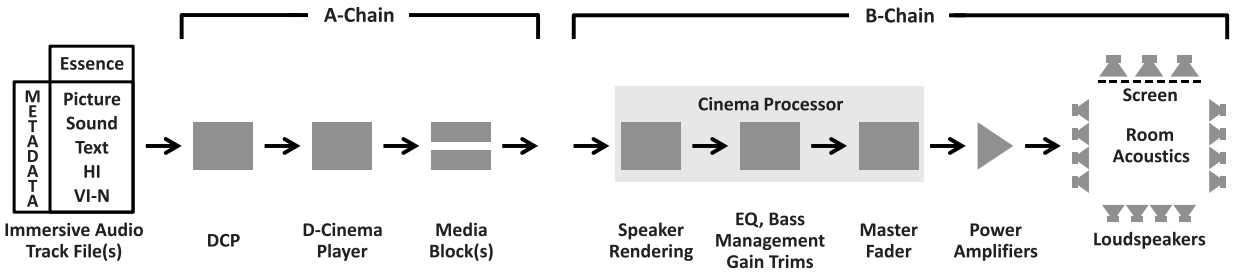

Figure 1. Typical immersive sound system A-chain and B-chain elements.

4. *Immersive Systems Reviewed*

The IASG considered the immersive sound systems in Table 1. All systems support channel-based audio, and four of them also support object-based audio. Two of the systems are in commercial cinema use, with current movies released in their formats.

System Proponent	Immersive Sound System (Name in report)	Channel-based audio	Object-based audio	In commercial cinema use
Auro Technologies	Auro 11.1, Auro 13.1 (Auro-3D®)	✓		✓
	Auro-Max™ (Auro-Max)	✓	✓	
Dolby® Laboratories	Dolby Atmos® (Dolby Atmos)	✓	✓	✓
DTS®	DTS Multi-Dimensional Audio (MDA)	✓	✓	
IOSONO	IOSONO (IOSONO)	✓	✓	
NHK	8K Super Hi-Vision, consistent with 22.2 (8K SHV)	✓		
Technicolor	HOA linear matrixed signals (HOA)	✓		
USL	USL Signal Matrix (USL)	✓		

Table 1. Immersive systems reviewed.

The Appendices have detailed technical information provided by each proponent.

- Appendix A tabulates structural characteristics of each system.
- Appendix B has speaker configuration diagrams for each system.
- Appendix C has descriptive papers about each systems.

Conventional cinema sound formats were also reviewed, as these benchmarks will continue to exist in parallel with immersive systems. In addition, reference channel assignments from SMPTE (ST 428-12:2013, ST 2036-2:2008) and IEC (62574:2011) were noted.

5. Immersive Sound System A-Chain Characteristics by System

The A-chain manages the immersive audio track file assets and streams them to the cinema playback system. This section provides more detail about these assets for each system.

5.1. Distribution Payload

Table 2 shows the nature of the content the track file would carry for each immersive system, as currently implemented or proposed.

The number of objects used in object-based audio essence can vary significantly due to the artistic requirements of the soundtrack. It is also possible to use objects and channels together.

In the HOA system, either channels or objects may be used to carry HOA signals. The number of transport channels needed for HOA varies with the order, *n*, per the relationship:

$$(n+1)^2 = \text{channels}$$

It is recommended by the HOA proponents that the HOA order should be ≥4 for diffuse audio objects and order 6 to 8 for high and very high spatial resolution content.

Payload	Dolby Atmos	MDA	IOSONO	Auro-3D	Auro-Max	HOA	8K SHV	USL
Essence	ch + objects	ch + objects	64 objects or ch	12-14 ch	ch + objects	16, 81 ch	24 ch	14 ch
Audio Coding	lossless	PCM	PCM or AAC-HD	Auro Codec®	PCM + Auro-Codec	PCM or MPEG-H Audio	not specified	PCM
Metadata	yes	yes	yes	yes	yes	yes	yes	no
HI/VI-N	2 ch	2 ch	2 ch	2 ch	2 ch	2 ch	2 ch	2 ch

Table 2. Track file payload for each immersive format.

Table 2 shows that the immersive formats use either PCM or one of multiple forms of audio data-rate compression. Audio encoded with the Auro-Codec represents itself as 6-8 channels of linear PCM and can be handled as such by systems without a decoder.

Representative examples of metadata were collected for four immersive systems, Dolby Atmos, HOA, IOSONO, and MDA. See Appendix A.

5.2. Sound/Picture Synchronization

Sound and picture need to be synchronized, so this must be considered when defining a common immersive audio track file. Some systems carry a sync signal as an audio asset. In other systems, a sync signal is generated in the media block at playout. Some systems, whose essence is transmitted using means that are already synchronized with picture, e.g. via the Main Sound interface, do not require a separate sync signal. See Table 3.

	Dolby Atmos	MDA	IOSONO	Auro-3D	Auro-Max	HOA	8K SHV	USL
Requires Sync Signal	yes	yes	yes	no	yes	no	no	no
Sync Method	ST 430-12 (audio asset)	Media Block Generated	LTC (Linear Time Code)	Main Sound	Media Block Generated	Main Sound	Main Sound	Main Sound

Table 3. Sync support for each system.

6. Immersive Sound System B-Chain Characteristics By System

The B-chain is responsible for converting the soundtrack audio into speaker feed signals. The IASG was tasked with evaluating the B-chain requirements of the various systems, identifying common approaches, and exploring a shared B-chain description.

One of the key areas of focus involved the preferred speaker configurations for each system, as listed in Appendix A and shown in the diagrams of Appendix B. Generally speaking, there is great commonality in the number and placement of surround speakers on the side and rear walls; good commonality for the screen speakers; and rather little commonality for the height and overhead speakers.

Just as with 5.1 cinemas today, the actual number of speakers present in a given immersive cinema sound system may vary based on the size of the room in order to maintain a certain uniformity of coverage across the audience.

7. Immersive Sound System Considerations

7.1. Legacy Compatibility

Legacy compatibility may be considered from two perspectives:

- Backward compatibility: Playing 5.1 content in immersive cinemas.
- Forward compatibility: Playing immersive content in 5.1 cinemas.

Even though every system reviewed utilizes additional speakers to create immersive sound, all the systems incorporate the existing speaker positions, thus fostering compatibility with legacy content presentations with minimal impact.

If the immersive sound system is presented with a legacy 5.1 track file, its primary responsibility is to reproduce it as it was intended. This means the immersive sound system needs to have a “5.1 mode” that replicates the configuration of screen speakers, surround arrays, subwoofer(s), and system calibration as found in a conventional 5.1 cinema sound system, and provides the same speaker drive signals. It is imperative that the system does this

faithfully and does not automatically “upmix” or “enhance” the sound, thus preserving the artistic intent of the content provider.

Regarding forward compatibility — playing immersive programs in legacy 5.1 cinemas — the IASG recognizes that while it is technically possible to automatically render immersive audio into a 5.1 sound system, that should only be permitted if the results reflect the artistic intent of the content provider. The solution today is to carry a separate 5.1 track file and CPL in the DCP that is keyed to play in cinemas with legacy sound systems.

While the immediate intent of defining a common immersive audio file format is to convey immersive audio mixes, it may be possible in the future to expand the definition to include metadata created by the content provider that would allow talent-approved rendering into legacy sound standards.

7.2. Speaker Performance and Calibration Criteria

As the utilization of surround speakers shifts from arrays to individual speakers, the performance requirements of the individual speakers must be reconsidered. Power handling, frequency response, and directivity are just some of the potentially relevant characteristics. Revised practices for the aiming of the speakers relative to the audience and their interaction with each other and the acoustical space are indicated. Bass management may also be effective in allowing smaller speakers to safely handle higher sound pressure levels, and is currently employed in several of the reviewed systems.

Having individual access to smaller arrays or individual speakers also opens the door to different calibration strategies for EQ, delay, and gain, all of which must take into account large audience areas with long path lengths affecting sound arrivals. Each system may have its own preferred optimization strategy.

7.3. Immersive Soundtrack Loudness Considerations

The rendering of immersive soundtracks can lead to a significantly higher number of output signals than conventional soundtrack formats, which in turn drive a significantly higher number of amplifiers and speakers. While the goal is to achieve enhanced spatial capabilities, an additional consequence is the ability to deliver higher overall sound pressure levels to cinemas, as well as high levels from individual speakers located close to the audience. It is imperative that immersive soundtrack loudness be controlled in the creation process.

8. Considerations and Proposals For Common Track File and B-Chain Description

The IASG has researched the considerations necessary for a common distribution track file and related B-chain description, and presents the findings and proposals below.

8.1. Common Distribution Track File

Table 4 lists all the essence types identified in Tables 2 and 3. In order to avoid redundancy, it will be necessary to define a single set of characteristics and metadata for each of the audio essence types, including audio data rate compression or matrixing, so that every playback system’s renderer can operate from the same file assets.

Creation Environment	→	Payload	→	Renderer Support	→	Cinema Configurations
Auro-3D Auro-Max Dolby Atmos IOSONO HOA MDA 8K SHV USL		Channels Objects Sync signal(s) HI/VI-N		PCM or Compressed: • Channels • Objects • Matrixed signals		5.1-22.2 Auro-3D Auro-Max Dolby Atmos IOSONO HOA USL

Table 4. Key stages in proposed immersive audio process.

As can be seen in Table 4, in order to design a single common distribution format, there are a number of requirements that must be reconciled, some of which may conflict. A payload containing channels and/or objects would need to be renderable into multiple sound systems that support various rendering techniques. If the content is matrix encoded, that adds another requirement to support. In a complex soundtrack, there would be a great deal of information contained in the file, which could make it quite demanding when multiple assets occur simultaneously. Real–time playout of such a file could impact data transmission capabilities. Ideally, discrete PCM audio can be utilized, but if compression is needed, shall it be lossy or lossless? Input will be needed from content creators who are most directly impacted, artistically, as well as the system proponents and product makers, since all would likely be required to redefine parts of their systems in order to accommodate the common file format.

In order for the immersive audio track file to be interoperable on a fundamental level, the relationship of digital signal levels to SPL as mixed must be well defined and must carry through to the auditorium consistently. In order to achieve this in the most general sense, the renderer would have to know this relationship for the immersive sound system monitored during the mix and the relationship for the immersive sound system to which it is rendering. While it is certainly possible to create and distribute this metadata, a simpler approach would be to standardize the digital signal level versus SPL relationship in the distribution track file, and apply gain offsets as needed when rendering to a given playback system of known calibration characteristics. Based on this, it is recommended that the relationship of digital signal levels to SPL of audio assets in the distribution track file be standardized. One possible approach would be to specify that all audio assets in the distribution track file have the relationship of -20 dB FS corresponding to 85 dB SPL as is consistent with industry practice. Playback system renderers can easily offset gains as may be necessary to correspond with the existing B-chain requirements.

Due to the dynamic nature of the immersive audio payload, the potential variety of audio coding options (PCM, lossy or lossless compression), the range of rendering techniques (e.g. HOA, VBAP, WFS), and the possibility that many of these options may appear in the same track file, the precise impact on renderer resources is not easily determined. The challenge is compounded if several computationally intensive features are used at the same time in multiple instantiations. This implies a need to ensure any track file will be playable in any cinema without exceeding the capability of the renderer.

8.2. Shared B-Chain Description

Thus far this report has considered interoperability of the common distribution file in terms of functionality and basic sound level calibrations. Another aspect involves the presentation of the soundfield. The varying characteristics among the immersive systems reviewed, primarily regarding height and overhead speakers, as well as how the sounds are presented, implies the range of performance variation can be significantly greater than among 5.1 systems, even if all are presenting the same common distribution track file.

While there are many common elements between the various immersive sound systems' B-chains, there may be many B-chain aspects that remain unique to each immersive audio system in practice, such as the number, placement and use of speakers, and possibly the methods for applying EQ and time delays.

Given a sufficient number of speakers in the right places, a single B-chain model could support every immersive system, and any unique array mappings and calibration requirements. Limited examples of this already exist, with several dubbing stages and some commercial cinemas configured to support both Dolby Atmos and Auro-3D, sharing as many B-chain resources as possible. Other solutions may evolve as the immersive concept gains maturity.

8.3. Translation Considerations In Playing Through Multiple Sound Systems

If a content maker were to take the time to create mixes in two immersive sound systems, A and B, they would be expected to sound different due to system differences, even if the artistic intent never wavered. By that definition, a difference in presentation does not necessarily mean the soundtrack is being poorly represented.

However, if the mixers create the mix on system A, and the common file is played in system B, what is considered a correct translation? Should it sound like system B or system A? If the answer is A, how would this be achieved? The mixers will need to be satisfied with system B's (and other systems') presentation, and therefore will want to ensure this in some fashion. Knowing that as a result of a common distribution file, immersive content will encounter immersive cinemas that do not match the original production system's characteristics, the better the original intent can be captured and conveyed in the content file, the better the chances that the playback system will be able to interpret it. This could start by identifying the characteristics of the system used for the production. If this or other metadata relating to the audio processing and immersive sound system characteristics used at the creation phase could be captured, put into the common file, and used advantageously at playback, the mix could potentially translate better into a wider variety of systems. This possibility should be considered in designing the common distribution track file.

9. Issues not covered

During the course of these discussions, topics were noted that were considered out of scope for this project. These include:

- Methods for interconnecting the A and B chains
- Content security
- Audio forensic marking
- Defining the aesthetics of immersive sound
- Determining the immersive capabilities of the reviewed systems
- Determine the efficacy of audio data-rate compression for immersive audio systems

10. Recommendations of this Study Group for Further Work

1) That SMPTE undertake work to standardize an essence format for immersive audio, capable of supporting channels, objects and additional metadata, to be carried in a D-Cinema Track File that is capable of conveying all detail necessary to reproduce an immersive soundtrack on any immersive sound system.
2) That SMPTE undertake work to define constraints to ensure interoperability of distribution track files and playback systems.
3) That SMPTE undertake work to create immersive sound system calibration standards, recommended practices and/or engineering guidelines.
4) That SMPTE undertake work to define the calibration of audio assets in the distribution track file.
5) That SMPTE undertake work to define the measurement of soundtrack loudness in general, with specific guidelines for immersive soundtracks.

11. Bibliography

SMPTE RP 200:2012, Relative and Absolute Sound Pressure Levels for Motion-Picture Multichannel Sound Systems — Applicable for Analog Photographic Film Audio, Digital Photographic Film Audio and D-Cinema

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

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

SMPTE ST 429-2:2013, D-Cinema Packaging - DCP Operational Constraints

SMPTE ST 2036-2:2008, Ultra High Definition Television — Audio Characteristics and Audio Channel Mapping for Program Production

Annex A Comparison of Reviewed Immersive Sound Systems

System data collected by the IASG

IASG Report Appendix A.xls

Society of Motion Picture and Television Engineers
3 Barker Avenue
White Plains, NY 10601 USA
www.smpte.org

Annex B Reviewed Immersive Sound System Speaker Diagrams

In order to gain a sense of the specifics and differences in the speaker layouts for the reviewed systems, each proponent was asked to utilize a template.

Auro-3D
25CSS Auro Speakers.pdf
25CSS Auro-Max Speakers.pdf

Dolby Atmos
25CSS Dolby Atmos Speakers.pdf

HOA
25CSS HOA Speakers.pdf

IOSONO
25CSS IOSONO Speakers.pdf

MDA
25CSS MDA Speakers.pdf

8K SHV
25CSS NHK 8K SHV Speakers.pdf

USL
25CSS USL Speakers.pdf

Annex C Reviewed Immersive Sound System Descriptions

System Proponent	Documents Submitted
Auro Technologies	Auro-3D One-pager 20131125b.pdf
Dolby Laboratories	Dolby_Atmos_Specifications.pdf
DTS	DTS Object Audio_SMPTE 2013.pdf 25CSS submission_DTS.pdf Pulkki, Ville, Virtual Sound Source Positioning Using Vector Base Amplitude Panning, JAES Vol 45 Issue 6 pp. 456-466; June 1997 (*Please see AES for this document*)
IOSONO	WFS and Object-Based Mixing for Motion Picture Sound_IOSONO.pdf AES128_Tutorial_Spatial_Audio_Reproduction.pdf (*Please see AES for this document*)
NHK	Advanced Multichannel Sound Systems_NHK_2013-01-28.pdf 22.2MultichannelSound_NHK_2013-01-27.pdf SMPTE ST 2036-2:2008, Ultra High Definition Television — Audio Characteristics and Audio Channel Mapping for Program Production (*Please see SMPTE for this document*)
Technicolor	25CSS_HOA_Description_Technicolor.pdf SoundFieldFormat_Technicolor.pdf Sound Field Transmission using Higher Order Ambisonics_Technicolor.pdf
USL	USL13_1v131028.pdf

Table 5. Documents supplied by respective proponents.

