**SMPTE EG 2098-3:2020**

# SMPTE ENGINEERING GUIDELINE

## Immersive Audio Renderer Expectations and Testing Recommendations



**Table of Contents**





**Approved July 6, 2020**

## 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 its Standards Operations Manual. This SMPTE Engineering Document was prepared by Technology Committee 25CSS.

## 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

This section is entirely informative and does not form an integral part of this Engineering Document.

This document provides eight audio samples via eight non-prose elements, SMPTE EG 2098-3a through -3h, as detailed in section 8.2.

Rendering is the process whereby channel-based and Object-based audio content is converted to audio signals suitable for reproduction in a cinematic sound system, e.g., a dubbing stage or a cinema. Object-based audio includes audio essence plus associated metadata describing position within the soundfield or other aspects of how the audio is intended to be reproduced. The Renderer uses this metadata along with Loudspeaker configuration information to determine how best to produce signals for each available output channel.

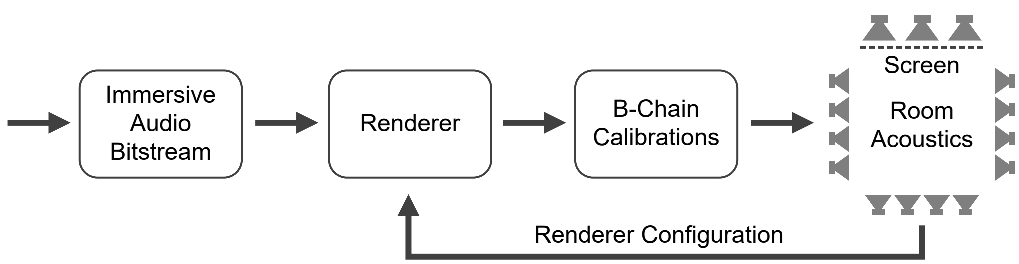



**Figure 1. Typical immersive sound system playback structure.**

Traditionally, movie soundtracks are prepared in dubbing stages built to enable mixers to predict how the mix will translate to commercial cinemas. In part this is achieved by maintaining a degree of physical, acoustic, and calibration consistency between dubbing stages and cinemas, even though cinemas are usually much larger than dubbing stages. Another aspect of consistency regards the worldwide adoption of the 5.1 soundtrack format itself.

Immersive audio brings a greater diversity in the way Loudspeaker systems are defined, both in the numbers and locations of Loudspeakers, and in the ability to provide individual signals for every Loudspeaker. This greatly expands the range of expression available to movie mixers, but brings with it greater challenges in maintaining the subjective aspects of a soundtrack as it is presented across a variety of immersive cinemas. This concept is known as interoperability. It is the job of the Renderer to do its best to achieve subjective interoperability by adapting itself to the presentation circumstances at hand.

For purposes of evaluating Renderer performance, subjective criteria are evaluated with respect to an allocentric presentation model. This means that sounds recorded in the production room's “box” are mapped to the Room Coordinate Model, and conversely upon playback the recorded sound coordinates are mapped to fit within the “box” of the playback room. This ensures that sounds maintain the wall-by-wall allocations heard in production, regardless of the playback room's L/W/H proportions.

It can be said that subjective interoperability has been achieved when a single bitstream, presented in a variety of playback systems, achieves results that are consistent with the experience heard during production. The interoperability of a Renderer is judged within the context of a given playback environment.

The key provisions of this document fall into two areas:

- Expected Renderer responses for ideal and non-ideal conditions are described in sections 6 and 7.
- Renderer test procedures and example tests are detailed in sections 9 and 10.

# 1 Scope

This Engineering Guideline specifies the baseline expected behavior of a generic Renderer in response to Abstract Bitstream Expressions and playback environment parameters, with sufficient detail to define the expected behavior in varied playback circumstances. Nothing in this document is intended to define or limit the technology used in a Renderer.

Methods and examples for evaluating Renderer behavior relative to the baseline expectations are provided.

The Renderer expectations and tests in this document are intended for cinema applications.

# 2 Conformance Notation

This Engineering Guideline is purely informative and meant to provide tutorial information to the industry. It does not impose Conformance Requirements and avoids the use of Conformance Notation.

Engineering Guidelines frequently provide tutorial information about a Standard or Recommended Practice and when this is the case, the user should rely on the Standards and Recommended Practices referenced for interoperability information.

# 3 Terms and Definitions

For the purposes of this document, the following terms and definitions apply. Table 1 lists additional terms whose definitions appear in the cited SMPTE documents.

**Table 1. Terms defined in other SMPTE documents.**

| Term | SMPTE document |
|---|---|
| Soundfield Configuration | ST 377-4 |
| Bed<br>Object, Audio Object<br>Renderer<br>Soundfield<br>Snap Tolerance<br>Target Environment | ST 2098-1 |
| Base Layer<br>Height Layer<br>Top Layer | ST 2098-5 |
| Loudspeaker | ST 428-12 |

## 3.1 Abstract Bitstream Expressions

Abstract bitstream expressions express metadata intent without reference to specific bitstream formats or values.

### 3.2 Active Metadata

Active Metadata is metadata that is used to render a result. The ST 2098-2 bitstream allows carriage of Conditional Metadata, and provides rules for when they are to be used. The Renderer is expected to determine when metadata is active and is expected to use all valid active metadata to render a result.

### 3.3 Conditional Metadata

Conditional Metadata include Conditional Beds, Conditional Objects, and Bed Remap Coefficients. These metadata are used by the Renderer when the associated condition (Target Environment) matches the Soundfield Configuration supported by the renderer.

### 3.4 Immersive Audio Renderer (Renderer)

An Immersive Audio Renderer translates an immersive audio distribution format (e.g., ST 2098-2 Immersive Audio Bitstream) to the playback environment for reproduction. A Renderer has knowledge of all the playback environment's Loudspeakers defined by the Renderer Configuration.

### 3.5 Renderer Configuration

Renderer Configuration is a term used to express the set of Loudspeakers that a Renderer has been configured to support. This includes every addressable Loudspeaker, plus any Loudspeakers grouped into arrays (e.g., Lss array). Note that an environment can include Loudspeakers that are not used.

### 3.6 Ideal Environment

In order to establish the intended effects of the various renderer functions without any potential limitations imposed by real-world cinemas, the concept of the Ideal Environment was developed.

An Ideal Environment for immersive sound is a reproduction environment with the following characteristics:

a) Supports a Base Layer that includes at least a 7.1DS Soundfield Configuration with the addition of a Top Layer of Loudspeakers, and can include a Height Layer of Loudspeakers.

b) Supports the ability to locate sound sources anywhere within the room including all walls (at or above the Base Layer) and ceiling.

c) Supports the ability to reproduce Objects whose timbre remains consistent regardless of its location in the room.

d) Has been calibrated at a Reference Listening Position so that Objects of constant level do not change in loudness as they are panned around the room.

e) Has been calibrated to produce a specific sound pressure level and frequency response in response to a reference audio signal.

f) Is "shoe box" shape, i.e., rectangular plan view and longer on the viewing axis than it is wide.

g) The room acoustics support good sound localization. These characteristics normally exist in modern cinemas and the characteristics are well known in acoustic design literature.

### 3.7 Test Vector

An audio waveform plus associated metadata.

## 3.8 Test

A Test Vector combined with a specific Environment Circumstance.

## 3.9 Environment Circumstance

Aspects of the environment that bear on Renderer expectations, e.g., room shape and Loudspeaker configuration.

## 3.10 Expectation

The desired outcome of the test.

## 3.11 Reference Listening Position (RLP)

The listening position located 2/3 the distance from the screen to the rear wall, centered on the screen.

## 3.12 Room Coordinate Model

All measurements of position relating to Objects or Loudspeakers use Cartesian coordinates referred to the origin, [X,Y,Z] = [0,0,0], at the lower left front corner of a nominal cube. For purposes of creating input vectors for testing, 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
- The horizontal plane at Z = 0 represents the Base Layer
- The screen is located on the front face (Y = 0)

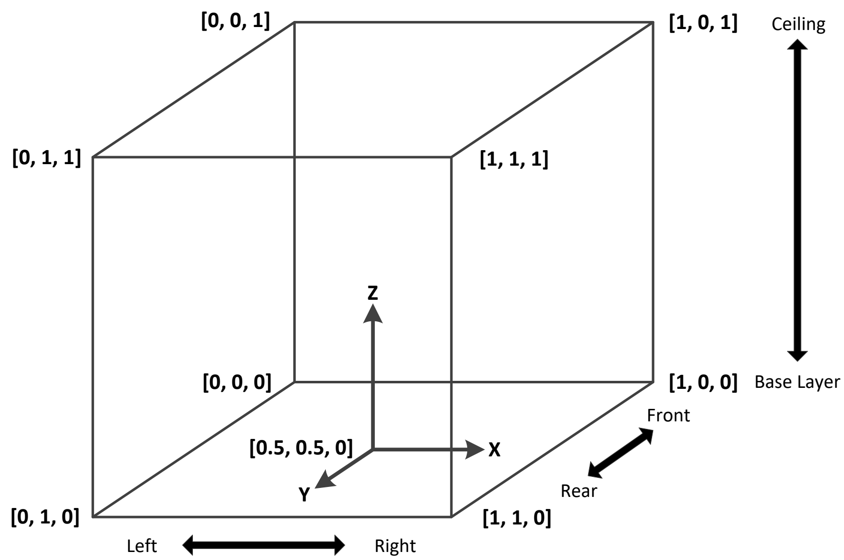



**Figure 2. Nominal cubic room definition.**

# 4 Process

Figure 3 illustrates the processing model for defining Renderer expectations.

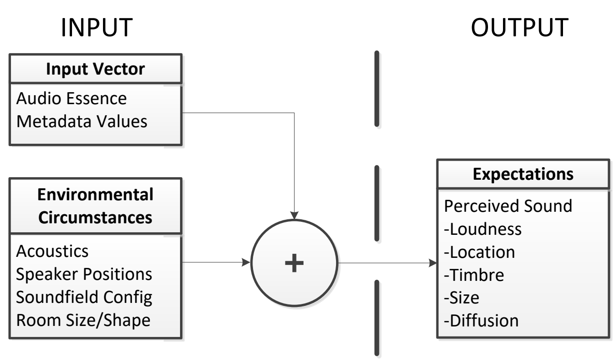



**Figure 3. Simplified processing model for setting Renderer expectations.**

On the input side, the input vector represents content delivered in a bitstream; audio essence and associated metadata. The Input Vector conveys inputs from the mixing engineer which are expected to control how the audio is presented in some perceptible way, e.g., position or movement. Also on the input side are Environmental Circumstances, which affect expectations.

On the output side, expectations are described in terms of perceived sound attributes, e.g., loudness, location, timbre, and size. Other methods, such as defining specific Loudspeaker output levels, could be used to define expectations, but this method constricts Object panning to a single algorithm, so it will not be used.

The task of defining Renderer expectations is divided into two phases. The first phase (section 6) defines high-level expectations assuming an Ideal Environment, and therefore, sets aside the inputs related to non-ideal acoustics, Loudspeaker location, and room size and shape. A second phase (section 7) takes such Environmental Circumstances into account.

## 5 Metadata items that affect rendering

The numbered items and descriptions in Table 2 (and Table 3) are derived from ST 2098-1, and are presented here to establish their association with the symbols in ST 2098-2. Refer to ST 2098-1 for complete metadata definitions.

**Table 2. List of metadata that can affect rendering.**

| SMPTE ST2098-1 Reference | Metadata Name | Description | SMPTE ST 2098-2 Symbol |
|---|---|---|---|
| 5.3 | Routing Destination | Identifies the single Loudspeaker or other reproduction device associated with the channel. | ChannelID |
| 6.4 | Remap Coefficients<br>Target Environment | Specifies how to map the original channels to a different Soundfield Configuration, and the Target Environment under which the Remap Coefficients are to be used. | RemapGain<br>RemapUseCase |
| 6.5 | Conditional Bed<br>Target Environment | An alternative bed and the Target Environment under which it is to be used. | ConditionalBed,<br>BedUseCase |
| | Channel Gain | Specifies the gain applied to audio essence of a channel. | ChannelGain |
| | Channel Decorrelation | Indicates whether the source signal is correlated or decorrelated when reproducing an audio channel across two or more Loudspeakers. | ChannelDecorCoef |
| 7.4 | Object Position (X,Y,Z) | Locates the Object in a three-dimensional space. | ObjectPos[X,Y,Z] |
| 7.5 | Object Spread | The Object’s size and shape in a three-dimensional space. | ObjectSpreadMode,<br>ObjectSpread,<br>ObjectSpread[X,Y,Z]<br><br>These three variables combine to describe the Object’s size and shape in 3D space. The spread mode determines whether the spread is equal in all dimensions or spread can be controlled in each dimension. |
| 7.6 | Object Gain | Specifies the gain applied to the audio essence of an Object. This gain can be unity, often depicted as 0 dB. | ObjectGain |

| | | | |
|---|---|---|---|
| 7.9 | Object Decorrelation | Indicates whether the source signal is correlated or decorrelated when reproducing an audio Object across two or more Loudspeakers | ObjectDecorCoef ObjectDecorCoefPrefix |
| 7.10 | Snap Tolerance | Indicates the degree to which preservation of Object timbre has priority over preservation of Object position. | ObjectSnap, ObjectSnapTolExists, ObjectSnapTolerance |
| 7.11 | Conditional Object Target Environment | An alternative Object and the Target Environment under which it is to be used. | ConditionalObject, ObjectUseCase |
| 7.12 | Zone Gain | The degree to which a zone is included in sound reproduction. | ZoneGain |

### 5.1 Systemwide constants

Systemwide constants are conditions to be supported by the system that are not explicitly expressed by bitstream metadata.

**Table 3. Systemwide constants that relate to rendering.**

| SMPTE ST2098-1 Reference | Common Name | Description |
|---|---|---|
| 7.12 | Zone definitions | Identifies which areas of the room comprise each zone. |
| 9.1 | Coordinate System | Audio Object positional metadata indicates the placement of an Object using a Cartesian coordinate system. |
| 9.2 | Frame of reference and location coding | The Cartesian coordinate values used for Object position are normalized relative to reference points of a cube, which represents an idealized cinema model. |

## 6 Expected Renderer responses

Each channel-based and Object-based audio source can include various kinds of metadata (Table 2) to express how the audio is to be presented. This section describes the subjective results to be expected from the rendering process.

Only one Renderer Configuration can be in effect at a time. Support for a Soundfield Configuration can be derived from the Renderer Configuration. Only one Target Environment can test true for a given Renderer Configuration.

Channel-based and Object-based audio rendering can be evaluated independently as each has unique metadata controls and expectations.

Tables of metadata descriptions will suffice for representing Abstract Bitstream expressions.

Unless otherwise stated, subjective evaluations are performed at the Reference Listening Position.

### 6.1 Subjective Attributes

To aid in the description of sound perception, the following Subjective Attributes are used.

Loudness. The perceived intensity of the sound. Loudness is different than sound pressure level (SPL), which can be measured objectively.

Timbre. The characteristic of a complex sound that reflects its tonal balance. For example, a complex sound can be composed of a fundamental tone and its harmonics. If the perceived level of the fundamental changes in relation to the harmonics, the timbre will change.

Location. The perceived direction and distance of a sound source.

Size. The apparent dimension (height or width) of the sound source. The minimum size is a point source, which is the dimension of the sound from a single Loudspeaker. The maximum size is determined by the listening environment. Movement of the listener (or gathering responses from multiple listening positions) can help with the determination of size. As size increases, the sound will appear to originate over a larger area, and the location could become less precise.

Diffusion. The apparent size of a sound source combined with the ability to locate the sound. A point source is an example of a non-diffuse sound. Its location is perceived as being easy to determine and its apparent size is small. A diffuse sound will be difficult to locate precisely and will appear to originate over a larger area.

### 6.2 Expectations relevant to channel-based audio

Channels are conveyed in the bitstream as a Bed. Multiple Beds can exist in the ST 2098-2 bitstream with the expectation that they be presented simultaneously (e.g.,, unconditional, or condition met). It is expected that the Renderer plays the sum of such beds based on any embedded gain settings and on their channel assignments.

Metadata that can affect a channel’s presentation fall into one of these general control categories:

- Routing destination
- Remap Coefficients
- Conditional use
- Gain
- Decorrelation. For more on Decorrelation, refer to 6.3.5.

**Table 4. Metadata expressions and expectations for channel-based audio.**

| SMPTE ST 2098-1 Reference | Description | Other Conditions | Expectation |
|---|---|---|---|
| 5.3 | Routing Destination (No Conditional Beds) | Renderer Configuration supports the Routing Destination | Sound emanates per the Routing Destination. |
| 5.3 | Routing Destination (No Conditional Beds) | Renderer Configuration does not support the Routing Destination | Channels for which there are no matching Loudspeakers are rendered such that the sound is perceived from the intended position, and not discarded. (See Section 7.2) |
| 6.4 | Remap Coefficients are present | The Remap Coefficients are active. | Sound is routed into the Renderer Configuration in a manner consistent with the remap parameters. |
| 6.5 | Conditional Bed | Conditional Bed is active. | Conditional Bed is presented. If Conditional Bed replaces another Bed, that Bed is not presented. |
| | Gain | Static gain and dynamic gain changes.<br><br>Permissible to measure level with SPL meter. | Change in relative Loudness (or SPL) correlates with Gain value change.<br>Notes 1, 2 |
| 7.9 | Decorrelation | Vary Decorrelation from minimum to maximum. | Decorrelation of the specified bed channel(s) are affected accordingly.<br><br>Timbre, Size, and Diffusion can vary. Loudness will not vary.<br>Note 2 |

Note:

1. Sounds maintain consistent Timbre and Size.
2. Sounds are free of artifacts (e.g., “zipper” noise, discontinuities, clicks)

## 6.3 Expectations relevant to Object-based audio

Metadata that can affect an Object’s presentation fall into one of these general control categories:

- Conditional
- Position
- Spread
- Gain
- Zone Gains
- Decorrelation

While any given Object can be subject to several of these controls, it will be useful in the understanding of Renderer expectations to describe how each of these controls individually affect the resulting sound.

Multiple Objects can exist and be presented simultaneously. Conditional Objects are presented when active. If a Conditional Object replaces another Object, that Object is not presented.

### 6.3.1 Position (static or dynamic)

The apparent location of the sound corresponds with the position per the room coordinate model. Positional coordinates are updated continuously by the bitstream. The values can remain unchanged for a period of time (static position), or they can vary regularly with sufficient granularity to give the illusion that the sound source is moving (dynamic position).

A positional value can also represent a substantial change in Location from those preceding, such that the sound appears to jump to the new Location. While the jump might be startling to the listener, the sound will ideally be otherwise free of unintended artifacts such as clicks, pops or zipper effect.

**Table 5. Metadata expressions and expectations for Object audio position.**

| SMPTE ST 2098-1 Reference | Description | Other Conditions | Expectation |
|---|---|---|---|
| 7.4 | Point source, static | Check several locations including typical L, W, H room extents | Sound Location correlates with positional coordinates, with the size of a point source. Note 1 |
| 7.4 | Point source, moving | Check several paths including exercising X, Y and Z axes independently | Sound Location tracks positional coordinate trajectory, with the size of a point source. Notes 1, 2 |
| 7.11 | Conditional Object, static point source | Conditional Object is active, with different positional coordinates than the Object it replaces. | Conditional Object is presented with sound located per position coordinates of the Conditional Object, with the size of a point source. If Conditional Object replaces another Object, that Object is not presented. |

Notes:

1. Sounds maintain consistent Timbre, Loudness, and Size.
2. Sounds are free of artifacts (e.g., “zipper” noise, discontinuities, clicks)

### 6.3.2 Snap Tolerance

Snap Tolerance indicates the degree to which preservation of object timbre has priority over preservation of object position. For purposes of this document, timbre preservation is not a variable, only the point where timbre preservation takes effect can vary. The Snap Tolerance value indicates the maximum acceptable positional displacement between the encoded object position and the timbre preserved position.

The minimum value means no Snap action occurs.

A common method of timbre preservation is to move the object position to the nearest Loudspeaker. In this case, low values indicate that the object's intended position needs to be relatively close to a Loudspeaker for Snap to take effect, High values indicate that the object's intended position can be further from a Loudspeaker for Snap to take effect.

How a given renderer achieves timbre preservation is outside the scope of this document.

**Table 6. Metadata expressions and expectations for Snap Tolerance.**

| SMPTE ST 2098-1 Reference | Description | Other Conditions | Expectation |
|---|---|---|---|
| 7.4<br>7.10 | Point source, static, with Snap Tolerance enabled at a moderate value | The displacement between the encoded object position and the Snapped Object position is within the Snap Tolerance | Object timbre is preserved<br>Note 1 |
| 7.4<br>7.10 | Point source, static, with Snap Tolerance enabled at a moderate value | The displacement between the encoded object position and the Snapped Object position is not within the Snap Tolerance | Object position is preserved |
| 7.4<br>7.10 | Point source, moving, with Snap Tolerance enabled at a moderate value | Intermediate tolerance value | Sound can transition abruptly to a timbre preserved condition. Snap action occurs at intended distance from closest Loudspeaker based on Snap Tolerance value.<br>Note 2 |
| 7.4<br>7.10 | Point source, with Snap Tolerance enabled at maximum value | Object position can be static or dynamic | Sound exhibits maximum preservation of timbre. |

Notes:

1. Sounds maintain consistent Timbre, Loudness, and Size.
2. Sounds are free of artifacts (e.g., “zipper” noise, discontinuities, clicks)

### 6.3.3 Spread

The perception of a Spread Object's (an Audio Object with a non-zero Spread value) location can vary depending on the position of the listener. In Figure 4, all three listeners identify the point source Object at the same location. With Spread Object A, the source Location appears to be closer to two of the listeners, but is mostly unchanged for the third. However, the sound can appear more Diffused for the listeners than it was for the point source. With Spread Object B, all three listeners hear the sound toward their sides, again with some potential Diffusion.

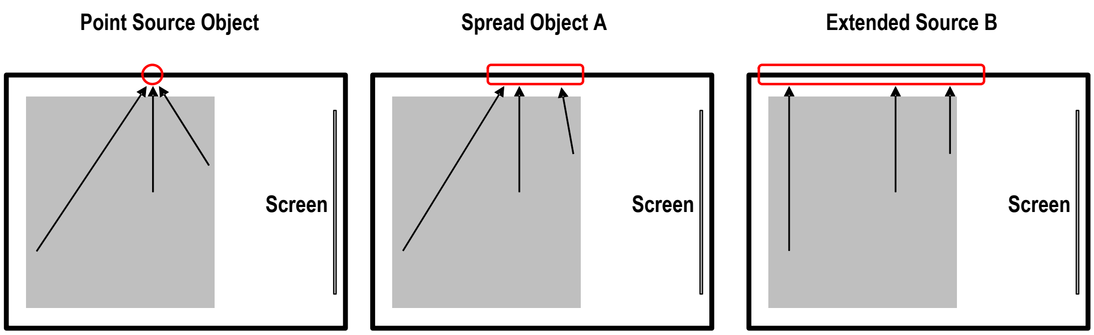



**Figure 4. Perceived location of extended sounds vs listener position.**

While no single listening position can determine the exact size and shape of a sound source, it is possible to improve that determination with data from multiple listening positions.

**Table 7. Metadata expressions and expectations for Object audio spread.**

| SMPTE ST 2098-1 Reference | Use case description | Other Conditions | Expectation |
|---|---|---|---|
| 7.5 | Spread Object, static | Check various values of spread for X, Y and Z dimensions<br><br>Include equal X/Y/Z Spread | Sound size correlates with Spread value in each dimension, as determined by sampling multiple seat locations.<br>Notes 1, 2, 3 |
| 7.5 | Spread Object, moving | Check various values of spread for X, Y and Z dimensions<br><br>Include equal X/Y/Z Spread | Sound size correlates with Spread value and tracks positional coordinate trajectory as determined by sampling multiple seat locations.<br>Notes 1, 2, 3, 4 |

Notes:
1. Sounds maintain consistent Loudness.
2. Sounds are free of artifacts (e.g., "zipper" noise, discontinuities, clicks)
3. Timbre change relative to a single Loudspeaker reference might be observed.
4. Spread and size remain consistent relative to Object position as Object moves.

#### 6.3.4 Gain

**Table 8. Metadata expressions and expectations for Object audio gain.**

| SMPTE ST 2098-1 Reference | Use case description | Other Conditions | Expectation |
|---|---|---|---|
| 7.6 | Object Gain | Static gain and dynamic gain changes.<br><br>Permissible to measure level with SPL meter. | Change in relative Loudness (or SPL) correlates with Gain value change.<br>Notes 1, 2 |

Notes:

1. Sounds maintain consistent Timbre, and Size.
2. Sounds are free of artifacts (e.g., “zipper” noise, discontinuities, clicks)

#### 6.3.5 Decorrelation

Decorrelation can affect sound perception in various ways such as described below. Any of these can be the intended effect, depending on what the mixer had in mind, so all are considered acceptable within Renderer expectations. Note that a change in Decorrelation is only audible when a sound is spread across multiple Loudspeakers, as can be the case with an extended Object-based source, or a channel feeding an array of Loudspeakers.

- Timbre

  When multiple Loudspeakers present the same waveform, an interference pattern is created. Under certain circumstances the resulting timbre of the sound can be audibly altered relative to the sound from a single Loudspeaker, or it can be perceived as unnatural when stronger comb filtering exists. One way to mitigate this result is to apply Decorrelation, thus reducing coherence and changing the interference pattern in the sound field.

- Size

  Decorrelation has the effect of increasing the perceived size of an extended sound source. In Figure 5, Case A shows a Spread Object with low Decorrelation (dotted line), which sounds like a narrow source directly to the side of the listener (circle). Case B shows substantial Decorrelation applied to the same Spread Object, with the result being perceived as a broader sound source covering the region occupied by the Spread Object.

- Diffusion

  When Decorrelation is low, Diffusion is minimum. As Decorrelation is increased, Diffusion increases, making it more difficult to locate the source precisely due to the increase in apparent size.

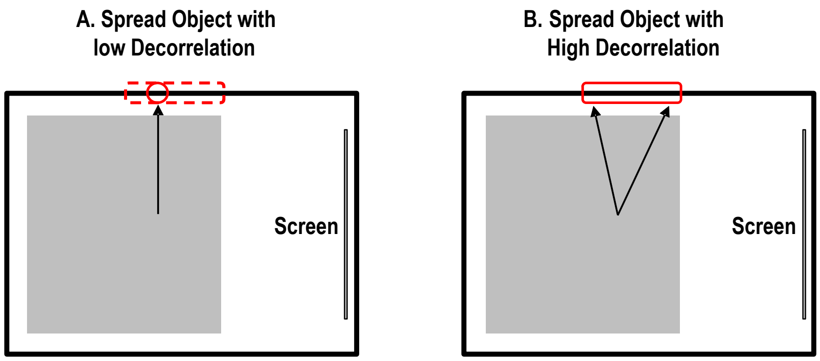



**Figure 5. Perceived source size vs. Decorrelation.**

**Table 9. Metadata expressions and expectations for Object audio Decorrelation.**

| SMPTE ST 2098-1 Reference | Use case description | Other Conditions | Expectation |
|---|---|---|---|
| 7.9 | Decorrelation | Spread is >0.<br>Maximum Decorrelation | Sound is heard as spread over the region described by the Object Spread; Maximum Diffusion is observed.<br>Notes 1, 3 |
| 7.9 | Decorrelation | Spread is >0.<br>Vary Decorrelation from minimum to maximum. | Timbre, Size, and Diffusion can vary.<br>Note 2 |

Notes:
1. Sounds maintain consistent Loudness.
2. Sounds are free of artifacts (e.g., “zipper” noise, discontinuities, clicks)
3. Timbre change relative to a single Loudspeaker reference might be observed.

### 6.3.6 Zone exclusions

A zone is a defined region within a generic listening environment. The actual mapping of zones to Loudspeakers will be defined for a cinema at the time of Renderer configuration. Zones are to be configured as a non-overlapping partition of the set of all available Loudspeakers.

Zone exclusion means the Loudspeakers within the specified zone(s) have limited or no availability for the rendering of that Object depending on the value of the Zone Gain parameter. The intention is that this will not affect the loudness or location of that sound at the Reference Listening Position. However, for sounds directed to excluded or attenuated zones, some reduction in localization accuracy can occur. In addition, listeners away from the reference position could potentially perceive changes in any of the following:

- Location (or trajectory)
- Loudness
- Timbre
- Size
- Diffusion

To the extent any of the above subjective attributes are affected by the use of zone exclusions, they can be considered acceptable.

**Table 10. Metadata expressions and expectations for Object audio zone control.**

| SMPTE ST 2098-1 Reference | Use case description | Other Conditions | Expectation |
|---|---|---|---|
| 7.12 | Point source, static, in a zone with minimum Zone Gain in that zone | Loudspeakers outside the affected zone are muted. | Zone is silent |
| 7.12 | Point source, static, in a single zone with fractional Zone Gain in that zone | Object is positioned within affected zone. Loudspeakers outside the affected zone are muted. | Loudness (or SPL) of Object reflects fractional Zone Gain value |
| 7.12 | Point source, static, in a zone with monotonically varying Zone Gain in that zone from minimum to maximum | Object is positioned within affected zone. Loudspeakers outside the affected zones are muted. | For a monotonically increasing Zone Gain, the loudness is expected to change monotonically from silence to maximum<br>Notes 2, 3 |
| 7.12 | Point source, static, in a zone with monotonically varying Zone Gain in that zone from maximum to minimum | Object is positioned within affected zone. | For a monotonically decreasing Zone Gain, the loudness and location are expected to remain constant, while the contribution from Loudspeakers within the zone decreases.<br>Notes 1, 2 |
| 7.5<br>7.12 | Spread Object, static, that utilizes a zone with minimum Zone Gain in that zone | Object extent includes affected zone. Loudspeakers outside the affected zones are muted. | Zone is silent |
| 7.12 | Point source, moving, in a zone with minimum Zone Gain in that zone | Sound trajectory spans multiple zones, including passing through affected zone. | Point source sound moves per coordinate trajectory while not utilizing Loudspeakers in excluded zones. Zones with a gain of zero generate no sound.<br>Notes 1, 2 |
| 7.5<br>7.12 | Spread Object, static, spanning multiple zones with monotonically varying Zone Gain in one zone from maximum to minimum | Object center is positioned within the gain affected zone. | For a monotonically decreasing Zone Gain, the loudness and location are expected to remain constant, while the contribution from Loudspeakers within the zone decreases.<br>Notes 1, 2 |

Notes:

1. Sounds maintain consistent timbre, loudness, and size.
2. Sounds are free of artifacts (e.g., “zipper” noise, discontinuities, clicks)
3. Sounds maintain consistent timbre and size.

# 7 Expectations in non-ideal environments

Section 6 describes high-level expectations for the rendering of channel-based and Object-based audio assuming an Ideal Environment. This section describes how expectations can be modified when a Renderer is used in a non-ideal (a.k.a. real) environment.

## 7.1 Limitations due to Loudspeaker density

Even in well-equipped immersive cinemas, audio Objects will regularly be directed to positions where no Loudspeakers exist (between Loudspeakers). In that case, a Renderer is expected to use existing Loudspeakers to reproduce the sound to ensure it remains audible. This can affect the perception of Location, Loudness, Timbre, Size, or Diffusion to varying degrees. These side-effects need to be minimized as much as possible.

In the case of a point source Object panned from a single Loudspeaker to a position between loudspeakers, the Renderer cannot guarantee to maintain consistent timbre. No action is expected from a Renderer to compensate for this.

## 7.2 Limitations due to a missing Loudspeaker associated with a channel

Renderers are expected to support the routing of channel-based audio to its respective Loudspeaker. However, if the designated Loudspeaker is not present, the Renderer can, for example, present the channel based on the implied Loudspeaker position as the intended destination. This can affect the perception of Location, Loudness, Timbre, Size, or Diffusion to varying degrees. These side-effects need to be minimized as much as possible.

## 7.3 Limitations due to a missing region of Loudspeakers

In this section, the term ‘region’ is defined more generally to mean an area of the cinema to which bitstream metadata can direct sound (e.g., the ceiling). Regions with no Loudspeakers cannot be expected to reproduce sound. In the case where bitstream metadata directs an Object to a region with no Loudspeakers, a Renderer is expected to use other Loudspeakers in its Renderer Configuration to reproduce the sound to ensure it remains audible. This can affect the perception of Location, Loudness, Timbre, Size, or Diffusion to varying degrees. These side-effects need to be minimized as much as possible.

## 7.4 Limitations due to Loudspeaker signal sharing

Some environments use Loudspeaker arrays (a group comprising two or more Loudspeakers that receive the same signal) in certain locations as the smallest addressable reproduction element, wherein all the Loudspeakers share the same signal. In such a case, the perception of a point source Object could include alterations in Location, Loudness, Timbre, Size, or Diffusion to varying degrees depending on the size of the array. No action is expected from a Renderer to compensate for this.

## 7.5 Limitations due to Loudspeaker characteristics

Some environments have Loudspeakers that are different in power handling, frequency response and dispersion pattern. Such an environment cannot be expected to maintain consistent timbre for a point source Object as it is panned to Loudspeakers of different characteristics. No action is expected from a Renderer to compensate for this. Note: B-Chain Loudspeaker equalization is not considered part of a Renderer, but can certainly be used to compensate for some Loudspeaker differences.

# 8 Renderer testing process overview

## 8.1 Renderer testing system

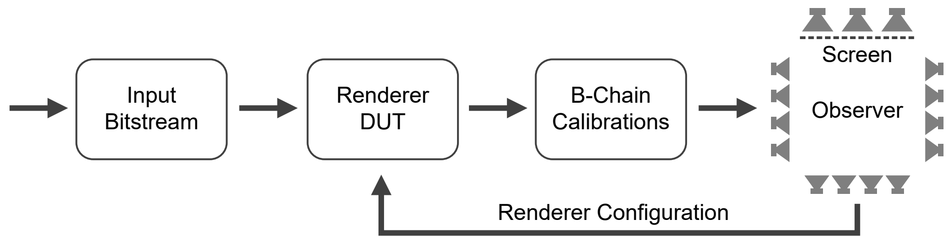



**Figure 6. Test system**

In general, the tests in this document have a direct correspondence with Renderer expectations as defined in section 6. Each Renderer expectation is associated with certain metadata states. These metadata states are combined with one or more test waveforms to produce a test vector. Each test vector is then converted to an input bitstream per ST 2098-2.

B-chain calibrations and room acoustics can impact the perception of immersive audio rendering. These factors are out of scope except to say that rooms are expected to comply with industry standards and practices, such as ST 202 and RP 2096-1.

## 8.2 Audio Signal Considerations

Different kinds of audio content can be used to evaluate particular aspects of Renderer performance. Certain audio files might be better suited to assessing localization while others might be better for perceiving timbre. Still others might be better for use with measurement instrumentation. Table 11 is a list of representative audio signals applicable to Renderer testing. The list is not intended to limit the use of any other signals as might be deemed appropriate.

This document provides eight monophonic audio samples via eight non-prose elements as listed in Table 11.

**Table 11. List of audio waveforms.**

| | File name | Description |
|---|---|---|
| 1 | eg2098-3a-2020.wav | Large audience applause from a moderate distance |
| 2 | eg2098-3b-2020.wav | Male talking, close mic'd |
| 3 | eg2098-3c-2020.wav | Unfiltered pink noise 20 Hz -20 kHz |
| 4 | eg2098-3d-2020.wav | Male voice enumerates 7.1 channels |
| 5 | eg2098-3e-2020.wav | Male voice enumerates 9.1 channels |
| 6 | eg2098-3f-2020.wav | Sine wave generator at 1 kHz, -30 dBFS |
| 7 | eg2098-3g-2020.wav | Cabasa played by hand, 80 BPM, close mic'd |
| 8 | eg2098-3h-2020.wav | Male voice enumerates 23.1 channels |

### 8.3 Renderer Configuration

A Renderer can be tested with any applicable Renderer Configuration. If it is intended to qualify a Renderer for general use, it can be evaluated across a range of configurations representative of relevant immersive cinemas.

### 8.4 Expected Results

The expectations for each metadata parameter that affect the presentation of either channel-based or Object-based audio is described in section 6 assuming ideal conditions. Section 7 provides information on how expectations can be modified in non-ideal (real world) environments. Those modified expectations can apply to the observed testing results. For example, since all testing will be performed in "non-ideal" conditions, the results will depend on the Renderer Configuration in use.

## 9 Test Procedure

For purposes of illustrating the concepts of Renderer testing, the following procedure is offered.

### 9.1 Document test conditions

Record the following information for future reference.

- Loudspeaker configuration: X-Y-Z position (feet or meters) of each Loudspeaker.
- Auditorium: stadium or flat seating; balconies; number of seats; listener position(s); screen size/position.
- Renderer under test: model; number of outputs; output Loudspeaker mapping.

### 9.2 Preparing for tests

A well-equipped auditorium can potentially support a variety of Renderer Configurations for testing purposes. The term Renderer Configuration as used herein refers to the specific Loudspeaker configuration currently in use by the Renderer.

- Set up the Renderer under test to the manufacturer's specifications.

The process of Renderer installation involves mapping Renderer outputs to the room's Loudspeakers in order to establish:

- Which of the available Loudspeakers to use (some can remain unused depending on test design)
- The physical positions of the Loudspeakers
- Which Loudspeakers are driven individually, which are driven as arrays, and which are members of zones

Repeat the Renderer setup for each Renderer Configuration to be tested.

- Listener training.
  - If the listener is expected to express observations relative to ideal expectations, the listener needs an understanding of such expectations. If the listener is expected to express

observations relative to an alternative reference presentation, the listener needs an understanding of that alternative. The latter can be especially challenging if the alternative requires use of a different venue, due to the limitations of auditory memory.

- Provide acoustic examples to set baseline expectations for sound characteristics: Location, Loudness, Timbre, Size and Diffusion, e.g.,, use a single Loudspeaker as a reference for Timbre and point source Size.
- Provide examples of unwanted artifacts (e.g.,, zipper noise, clicks, distortions, level shifts, Timbre changes, motion discontinuities). These can be prerecorded examples.

### 9.3 Test Procedure Overview

- Define the test plan—which tests are needed to cover the desired aspects of performance. Define how to quantify and record results.
- Create or obtain test streams. If the monophonic audio samples from Table 11 are used, they will need to be mixed to utilize the channels or positional coordinates described in the respective tests.
- Position listener. Most tests are done at the RLP. Some tests use multiple listening locations.
- Inform the listener of the nature of the test at hand. e.g.,, sound position, Size, level. (Note: Generally, do not inform the listener of the intended Location of sounds for tests related to position. Randomizing a sequence of position tests will also reduce the ability of the listener to infer a pattern.)
- Listeners are asked to record any unexpected changes in apparent Loudness, Timbre, or Size, and to note any artefacts observed.

## 10 Example Tests

Tests are designed to confirm expected Renderer behavior. The following tests illustrate the testing of several rendering expectations that are important to immersive audio presentation, but they are not intended to be comprehensive. For example, only cinema applications are considered. Additionally, each test typically focuses on a single metadata parameter. Bitstreams are able to structure multiple metadata commands to occur simultaneously, and it is expected that Renderers will perform each expectation accordingly. It is left to the testing administrator to decide whether to evaluate multiple conditions simultaneously.

Only the audio parameters necessary for each test are shown in the following tables. All other audio metadata parameters are assumed to be zero or inactive.

Channel-based and Object-based audio rendering can be evaluated independently.

For purposes of describing the state or the degree of effect for a metadata parameter, a scale factor between 0 and 1 is used. A metadata value of 0 means the metadata is either at its minimum value, inactive, or not present in the bitstream. A metadata value of 1 means the metadata is in its maximally effective state. Fractional values, if present, denote intermediate degrees of effect for the metadata parameter. If the metadata parameter is binary by nature, fractional values would not apply.

## 10.1 Channel-based tests

### 10.1.1 Routing Destination

Bitstream has one Bed comprising every possible channel. This test confirms Renderer behavior for every bitstream channel. Results will vary depending on whether the Renderer Configuration supports the channel's intended Loudspeaker.

**Table 12. Conditions and results for Routing Destination test.**

| Test item | Waveform (Table 11) | Metadata | SMPTE EG 2098-3 Expectations (Table 4) |
|---|---|---|---|
| | | Routing Destination | |
| **1** | 8 | L, Lc, C, Rc, R, Ls, Rs, Lss, Rss, Lrs, Rrs, Lsh, Rsh, Lssh, Rssh, Lrsh, Rrsh, Lts, Rts, Lh, Ch, Rh, Ts, LFE | Channels for which there are matching Loudspeakers, sound is rendered per the Routing Destination.<br><br>Channels for which there are no matching Loudspeakers, sound is rendered such that it is perceived from the intended position, and not discarded. |

### 10.1.2 Remap Coefficients

The bitstream contains an unconditional Bed and Remap Coefficients (always conditional). This test confirms Renderer behavior for Remap Coefficients. Results will vary depending on whether the Remap Coefficients are active.

**Table 13. Conditions and results for Remap Coefficients test.**

| Test item | Waveform (Table 11) | Metadata | | Expectations (Table 4) |
|---|---|---|---|---|
| | | Routing Destination | Remap Coefficients | |
| **2** | 5 | L, C, R, Lss, Rss, Lrs, Rrs, LFE, Tsl, Tsr | Present (provides obviously audible remapping) | For Renderers that have determined the Remap Coefficients are active, sound is routed into the Renderer Configuration in a manner consistent with the remap parameters.<br><br>For Renderers that have not determined the Remap Coefficients are active, Remap Coefficients are not applied. |

This test can be repeated for any Bed with a variety of Remap Coefficients. It is recommended to test both conditions: when the Remap Coefficients are active and when they are not active.

### 10.1.3 Conditional Bed

The bitstream contains a Bed and a Conditional Bed. This test confirms Renderer behavior for Conditional Bed when the Conditional Bed is active.

**Table 14. Conditions and results for Conditional Bed test.**

| Test item | Waveform (Table 11) | Metadata | Expectations (Table 4) |
|---|---|---|---|
| | | **Bed Channels** | |
| **3** | Bed<br>4 | L, C, R, Lss, Rss, Lrs, Rrs, LFE | For Renderers that have determined the Conditional Bed is active, the Conditional Bed is presented. |
| | Conditional Bed<br>1<br>(sounds obviously different from Bed) | L, C, R, Lss, Rss, Lrs, Rrs, LFE | |

This test can be repeated for each possible Bed configuration.

### 10.1.4 Multiple Beds

The bitstream contains multiple Beds (none are Conditional Beds). This test confirms Renderer behavior for multiple Beds.

**Table 15. Conditions and results for Multiple Beds test.**

| Test item | Waveform (Table 11) | Metadata | Expectations (Table 4) |
|---|---|---|---|
| | | **Bed Channels** | |
| **4** | Bed<br>4 | L, C, R, Lss, Rss, Lrs, Rrs, LFE | For supported channels, sound emanates per the routing destination.<br><br>Channels for which there are no matching Loudspeakers are rendered appropriately, and not discarded. |
| | Bed<br>7<br>(provides obviously audible difference) | L, C, R, Lss, Rss, Lrs, Rrs, LFE | |

This test can be repeated for additional Beds or with different Bed configurations.

### 10.1.5 Gain

The bitstream contains a channel whose Gain value changes from 1.0 (0 dB reference) to one or more other values. Gain values can be either static or vary dynamically over time.

**Table 16. Channel gain tests and expectations.**

| Test item | Waveform (Table 11) | Metadata | Expectations (Table 4) |
|---|---|---|---|
| | | Gain | |
| **5** | 3, 6 | 1.0, 0.5 | Change in relative Loudness (or SPL) correlates with Gain value change. Notes 1, 2 |

Notes:
1. Sounds maintain consistent Timbre and Size.
2. Sounds are free of artifacts (e.g., “zipper” noise, discontinuities, clicks)

This test can use waveform 3 for acoustic measurement and waveform 6 for electrical measurement.

### 10.1.6 Channel Decorrelation

The bitstream contains a channel intended for a Loudspeaker array, with different values of Decorrelation. Subjective terms Timbre, Size, and Diffusion, are discussed in section 6.3.5.

**Table 17. Conditions and results for channel Decorrelation test.**

| Test item | Waveform (Table 11) | Metadata | | Expectations (Table 4) |
|---|---|---|---|---|
| | | Routing Destination | Decorrelation | |
| **6** | 1 | Lss | 0 | Minimum Size and Diffusion are observed. Note 1 |
| **7** | 1 | Lss | 1 | Maximum Size and Diffusion are observed Note 1 |
| **8** | 1 | Lss | Varies smoothly over range | Size and Diffusion vary smoothly. Loudness will not vary. Note 2 |

Note:
1. Timbre change relative to a single Loudspeaker reference might be observed.
2. Sounds are free of artifacts (e.g., “zipper” noise, discontinuities, clicks)

This test can be repeated for any channel that plays from a Loudspeaker array.

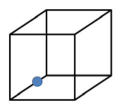
### 10.2 Object-based tests

The following tests describe Object position in [X, Y, Z] coordinates, with a graphical depiction of same. Refer to Figure 1 for details of the nominal Room Coordinate Model.

#### 10.2.1 Point Source, Static

The bitstream contains a point source Object at fixed position.

**Table 18. Static point source test and expectations.**

| Test item | Waveform (Table 11) | Metadata | | | Expectations (Table 5) |
|---|---|---|---|---|---|
| | | Coordinates [X,Y,Z] | | Object Gain | |
| **9** | 7 | [0, 0.25, 0] |  | 1 | Sound Location correlates with positional coordinates, with the size of a point source. Notes 1 |

Note:

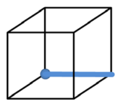
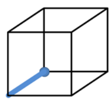
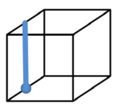
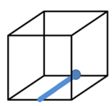
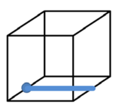
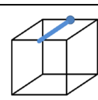
1. Sounds maintain consistent Timbre, Loudness and Size regardless of position

This test can be repeated for a variety of Object positions to confirm range in X, Y and Z axes.

### 10.2.2 Point Source, Moving

The bitstream contains a point source Object moving along the room boundaries. This test confirms Renderer behavior for panning in X, Y and Z axes.

**Table** 19**. Moving point source tests and expectations.**

| Test item | Waveform (Table 11) | Metadata | | | Expectations (Table 5) |
|---|---|---|---|---|---|
| | | Coordinates [X,Y,Z] | | Object gain | |
| **10** | 7 | [0, 0, 0]<br>to<br>[1, 0, 0] |  | 1 | Sound Location tracks trajectory along X-axis.<br>Notes 1, 2 |
| **11** | 7 | [0, 0, 0]<br>to<br>[0, 1, 0] |  | 1 | Sound Location tracks trajectory along Y-axis.<br>Notes 1, 2 |
| **12** | 7 | [0, 0.5, 0]<br>to<br>[0, 0.5, 1] |  | 1 | Sound Location tracks trajectory along Z-axis.<br>Notes 1, 2 |
| **13** | 7 | [0.5, 0, 0]<br>to<br>[0.5, 1, 0] |  | 1 | Sound Location tracks trajectory along Y-axis.<br>Notes 1, 2 |
| **14** | 7 | [0, 0.5, 0]<br>to<br>[1, 0.5, 0] |  | 1 | Sound Location tracks trajectory along X-axis.<br>Notes 1, 2 |
| **15** | 7 | [0.5, 0, 1]<br>to<br>[0.5, 1, 1] |  | 1 | Sound Location tracks trajectory along Y-axis.<br>Notes 1, 2 |

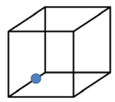
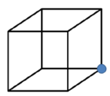
Notes:
1. Sounds maintain consistent Timbre, Loudness and Size.
2. Sounds are free of artifacts (e.g., “zipper” noise, discontinuities, clicks)

This test can be repeated for a variety of trajectories and speeds.

### 10.2.3 Conditional Object is Present

The bitstream contains an Object and a Conditional Object that, when active, replaces the first Object. This test confirms Renderer behavior for a Conditional Object when the Conditional Object is active.

**Table 20. Conditional Object tests and expectations.**

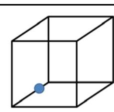
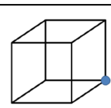
| Test item | Waveform (Table 11) | Metadata | | Expectations (Table 5) |
|---|---|---|---|---|
| | | Coordinates [X,Y,Z] | | |
| **16** | Object 2 | [0, 0.25, 0] |  | For Renderers that have determined the Conditional Object is active, Conditional Object is presented. If Conditional Object replaces another Object, that Object is not presented. |
| | Conditional Object 7 | [1, 0, 0] |  | |

This test will only confirm an active Conditional Object replaces another Object. It is recommended to also test the condition when the Conditional Object is not active, in which case the Conditional Object is not presented.

### 10.2.4 Conditional Object is Present

The bitstream contains an Object and a Conditional Object that, when active, is presented along with the first Object. This test confirms Renderer behavior for a Conditional Object when the Conditional Object is active.

**Table 21. Conditional Object tests and expectations.**

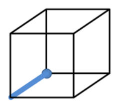
| Test item | Waveform (Table 11) | Metadata | | Expectations (Table 5) |
|---|---|---|---|---|
| | | Coordinates [X,Y,Z] | | |
| **17** | Object 2 | [0, 0.25, 0] |  | For Renderers that have determined the Conditional Object is active, Conditional Object is presented. If Conditional Object replaces another Object, that Object is not presented. |
| | Conditional Object 7 | [1, 0, 0] |  | |

This test will only confirm that an active Conditional Object plays in addition to another Object. It is recommended to also test the condition when the Conditional Object is not active, in which case the Conditional Object is not presented.

### 10.2.5 Point Source, Moving, with Snap Tolerance (intermediate)

The bitstream contains a point source Object moving along the room boundaries, with partial Snap Tolerance enabled.

**Table 22. Moving point source with Snap Tolerance tests and expectations.**

| Test item | Waveform (Table 11) | Metadata | | | Expectations (Table 6) |
|---|---|---|---|---|---|
| | | Coordinates [X,Y,Z] | | Snap Tolerance | |
| **18** | 7 | [0, 0, 0] to [0, 1, 0] |  | 0.7? | Sound can transition abruptly to a timbre preserved condition. Snap action occurs at intended distance from closest Loudspeaker based on Snap Tolerance value.<br>.<br>Notes 1, 2 |

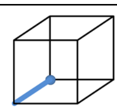
Notes:
1. Sounds maintain consistent Timbre, Loudness and Size.
2. Sounds are free of artifacts (e.g., “zipper” noise, discontinuities, clicks)

This test can be repeated for a variety of trajectories and intermediate Snap Tolerance values.

### 10.2.6 Point Source, Moving, with Snap Tolerance (max)

The bitstream contains a moving point source Object with Snap Tolerance enabled to preserve Timbre.

**Table 23. Moving point source with Snap Tolerance tests and expectations.**

| Test item | Waveform (Table 11) | Metadata | | | Expectations (Table 6) |
|---|---|---|---|---|---|
| | | Coordinates [X,Y,Z] | | Snap Tolerance | |
| **19** | 7 | [0, 0, 0] to [0, 1, 0] |  | 1 | Sound exhibits maximum preservation of timbre. Notes 1, 2 |

Notes:
1. Sounds maintain consistent Timbre, Loudness, and Size.
2. Sounds are free of unwanted artifacts (e.g., “zipper” noise, clicks)

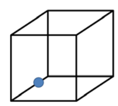
This test can be repeated for a variety of trajectories.

### 10.2.7 Spread Object, static

The bitstream contains a Spread Object at a fixed position. Spread can be applied in one, two or three axes.

For testing purposes, it is beneficial to collect data from multiple listening positions to better determine the Size of the sound.

**Table 24. Extended static source tests and expectations.**

| Test item | Waveform (Table 11) | Metadata | | | Expectations (Table 7) |
|---|---|---|---|---|---|
| | | Coordinates [X,Y,Z] | | Object Spread | |
| **20** | 1 | [0, 0.25, 0] |  | 00 X<br>0.5 Y<br>0.0 Z | Sound Size correlates with Spread value, as determined by sampling multiple seat locations.<br>Notes 1, 2, 3 |

Notes:
1. Sounds maintain consistent Loudness.
2. Sounds are free of artifacts (e.g., “zipper” noise, discontinuities, clicks)
3. Timbre change relative to a single Loudspeaker reference might be observed.

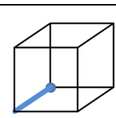
This test can be repeated for Objects at various positions or with various Spread in X, Y, or Z dimensions.

### 10.2.8 Spread Object, moving

The bitstream contains an Spread Object moving along the room boundaries. Spread can be applied in one, two or three axes.

For testing purposes, it is beneficial to collect data from multiple listening positions to better determine the Size of the sound.

**Table 25. Fractionally extended moving source tests and expectations.**

| Test item | Waveform (Table 11) | Metadata | | | Expectations (Table 7) |
|---|---|---|---|---|---|
| | | Coordinates [X,Y,Z] | | Object Spread | |
| **21** | 1 | [0, 0, 0] to<br>[0, 1, 0] |  | 0 - 0.5 X<br>0 - 0.5 Y<br>0 - 0.5 Z | Sound Size correlates with Spread value and tracks positional coordinate trajectory as determined by sampling multiple seat locations.<br>Notes 1, 2, 3, 4 |

Notes:
1. Sounds maintain consistent Loudness.
2. Sounds are free of artifacts (e.g., “zipper” noise, discontinuities, clicks)
3. Timbre change relative to a single Loudspeaker reference might be observed.
4. Spread and size remain consistent relative to Object position as Object moves.

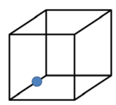
This test can be repeated for various extents in X, Y, or Z dimensions, over various trajectories.

### 10.2.9 Spread Object, Maximum

The bitstream contains an Object with maximum Spread.

For testing purposes, it is beneficial to collect data from multiple listening positions to better determine the Size of the sound.

**Table 26. Maximally extended source tests and expectations.**

| Test item | Waveform (Table 11) | Metadata | | | Expectations (Table 7) |
|---|---|---|---|---|---|
| | | Coordinates [X,Y,Z] | | Object Spread | |
| **22** | 1 | [0, 0.25, 0] |  | 1 | Sound is spread uniformly across all walls and ceiling, as determined by sampling multiple seat locations.<br>Notes 1, 2, 3 |

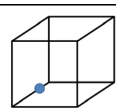
Notes:

1. Sounds maintain consistent Loudness.
2. Sounds are free of artifacts (e.g., “zipper” noise, discontinuities, clicks)
3. Timbre change relative to a single Loudspeaker reference might be observed.

### 10.2.10 Object Gain

The bitstream contains an Object whose Gain value changes from 1.0 (0 dB reference) to one or more other values.

**Table 27. Object gain tests and expectations.**

| Test item | Waveform (Table 11) | Metadata | | | Expectations (Table 8) |
|---|---|---|---|---|---|
| | | Coordinates [X,Y,Z] | | Object gain | |
| **23** | 3, 6 | [0, 0.25, 0] |  | 1.0, 0.5 | Change in relative Loudness (or SPL) correlates with Gain value change.<br>Notes 1, 2 |

Notes:

1. Sounds maintain consistent Timbre and Size.
2. Sounds are free of artifacts (e.g., “zipper” noise, discontinuities, clicks)

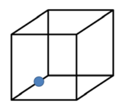
This test can use waveform 3 for acoustic measurement and waveform 6 for electrical measurement.

### 10.2.11 Spread Object with Maximum Decorrelation

The bitstream contains a Spread Object with maximum Decorrelation applied.

For testing purposes, it is beneficial to collect data from multiple listening positions to better determine the Size of the sound.

**Table 28. tests and expectations for decorrelation.**

| Test item | Waveform (Table 11) | Metadata | | | | Expectations (Table 9) |
|---|---|---|---|---|---|---|
| | | Coordinates [X,Y,Z] | | Object extent | Decorrelation | |
| **24** | 1 | [0, 0.25, 0] |  | 0.5 | 1 | Sound is heard as spread over the region described by the Object extents; Maximum Diffusion is observed.<br>Notes 1, 2 |

Note:

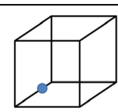
1. Sounds maintain consistent Loudness.
2. Timbre change relative to a single Loudspeaker reference might be observed.

This test can be repeated for Objects with different degrees of Decorrelation applied.

### 10.2.12 Spread Object with Variable Decorrelation

The bitstream contains a Spread Object with a varying degree of Decorrelation applied.

**Table 29. Variable decorrelation tests and expectations.**

| Test item | Waveform (Table 11) | Metadata | | | | Expectations (Table 9) |
|---|---|---|---|---|---|---|
| | | Coordinates [X,Y,Z] | | Object Spread | Decorrelation | |
| **25** | 1 | [0, 0.25, 0] |  | 0.7 | 0 - 1 | Timbre, Size, and Diffusion can vary.<br>Note 1 |

Notes:

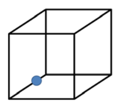
1. Sounds are free of artifacts (e.g., “zipper” noise, discontinuities, clicks)

### 10.2.13 Point Source, Static, with Minimum Zone Gain in One Zone

The bitstream contains a point source Object positioned within a zone having minimum gain. This test confirms Renderer behavior for Zone Gain.

With all Loudspeakers unmuted, the audio plays normally. For testing purposes, in order to confirm audio is presented as expected in the attenuated zone, Loudspeakers outside the attenuated zone can be muted in the B-chain.

**Table 30. Minimum Zone Gain tests and expectations.**

| Test item | Waveform (Table 11) | Metadata | | | Expectations (Table 10) |
|---|---|---|---|---|---|
| | | Coordinates [X,Y,Z] | | Zone Gains | |
| **26** | 7 | [0, 0.25, 0] |  | All Loudspeakers on the left wall, 0 | Zones with a gain of zero generate no sound. |

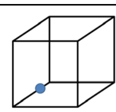
This test can be repeated for any zone.

### 10.2.14 Point Source, Static, with Fractional Zone Gain in One Zone

The bitstream contains a point source Object positioned within a zone having fractional gain.

For testing purposes, Object is positioned within affected zone, and Loudspeakers outside the affected zone are muted in the B-chain. Fractional gain can be determined by comparing the fractional output to the maximum output of that zone.

**Table 31. Fractional Zone Gain tests and expectations.**

| Test item | Waveform (Table 11) | Metadata | | | Expectations (Table 10) |
|---|---|---|---|---|---|
| | | Coordinates [X,Y,Z] | | Zone Gains | |
| **27** | 7 | [0, 0.25, 0] |  | 0.5 | Loudness (or SPL) of Object reflects fractional Zone Gain value. |

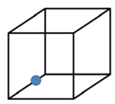
This test can be repeated for any zone.

### 10.2.15 Point Source, Static, with Monotonically Varying Zone Gain in One Zone

The bitstream contains a point source Object positioned within a zone having monotonically varying gain.

For testing purposes, Object is positioned within affected zone, and Loudspeakers outside the affected zone are muted in the B-chain.

**Table 32. Monotonically varying Zone Gain tests and expectations.**

| Test item | Waveform (Table 11) | Metadata | | | Expectations (Table 10) |
|---|---|---|---|---|---|
| | | Coordinates [X,Y,Z] | | Zone Gain | |
| **28** | 3 | [0, 0.25, 0] |  | 0-1 | For a monotonically increasing Zone Gain, the Loudness is expected to change monotonically from silence to maximum.<br>Notes 1, 2 |

Notes:
1. Sounds maintain consistent Timbre and Size.
2. Sounds are free of artifacts (e.g., “zipper” noise, discontinuities, clicks)

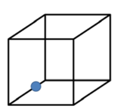
This test can be repeated for any zone.

### 10.2.16 Point Source, Static, with Monotonically Varying Zone Gain in One Zone

The bitstream contains a point source Object positioned within a zone having monotonically varying gain.

For testing purposes, Object is positioned within affected zone.

**Table 33. Monotonically varying Zone Gain tests and expectations.**

| Test item | Waveform (Table 11) | Metadata | | | Expectations (Table 10) |
|---|---|---|---|---|---|
| | | Coordinates [X,Y,Z] | | Zone Gain | |
| **29** | 3 | [0, 0.25, 0] |  | 1-0 | For a monotonically decreasing Zone Gain, the loudness and location are expected to remain constant, while the contribution from Loudspeakers within the zone decreases.<br>Notes 1, 2 |

Notes:
1. Sounds maintain consistent Timbre and Size.
2. Sounds are free of artifacts (e.g., “zipper” noise, discontinuities, clicks)

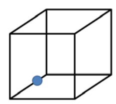
This test can be repeated for any zone.

### 10.2.17 Spread Object, Static, with Minimum Zone Gain in One Zone

The bitstream contains a Spread Object and a zone having minimum gain. The Object can be positioned within the zone, or the Object Spread extent can overlap the affected zone.

With all Loudspeakers unmuted, the audio plays normally. For testing purposes, in order to confirm audio is presented as expected in the attenuated zone, Loudspeakers outside the attenuated zone can be muted in the B-chain.

**Table 34. Spread Object with minimum Zone Gain tests and expectations.**

| Test item | Waveform (Table 11) | Metadata | | | | Expectations (Table 10) |
|---|---|---|---|---|---|---|
| | | Coordinates [X,Y,Z] | | Object extent | Zone Gains | |
| **30** | 3 | [0, 0.25, 0] |  | 0.5 | 0 | Zones with a gain of zero generate no sound. |

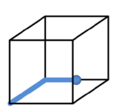
This test can be repeated for various Object positions and Spread extent parameters.

### 10.2.18 Point Source, Moving, with Minimum Zone Gain in One Zone

The bitstream contains a point source Object moving along the room boundaries, and a zone having minimum gain. Sound trajectory spans multiple zones, passing through the attenuated zone.

With all Loudspeakers unmuted, the audio plays normally. For testing purposes, in order to confirm audio is presented as expected in the attenuated zone, Loudspeakers outside the attenuated zone can be muted in the B-chain.

**Table 35. Moving point source with minimum Zone Gain tests and expectations.**

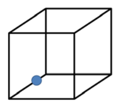
| Test item | Waveform (Table 11) | Metadata | | | Expectations (Table 10) |
|---|---|---|---|---|---|
| | | Coordinates [X,Y,Z] | | Zone Gains | |
| **31** | 7 | [0.5, 0, 0] to [0, 1, 0] |  | Main screen Loudspeakers left of center, 0 | Point source sound moves per coordinate trajectory while not utilizing Loudspeakers in excluded zones. Zones with a gain of zero generate no sound. |

This test can be repeated for various zones.

### 10.2.19 Spread Object, Static, with Varying Zone Gain in One Zone

The bitstream contains a Spread Object spanning multiple zones. Zone Gain will be adjusted for the zone intersecting the Object's coordinates.

**Table 36. Spread Object with variable Zone Gain tests and expectations.**

| Test item | Waveform (Table 11) | Metadata | | | | Expectations (Table 10) |
|---|---|---|---|---|---|---|
| | | Coordinates [X,Y,Z] | | Object extent | Zone Gains | |
| **32** | 3 | [0, 0.25, 0] |  | 0.5 | Varies from 0 to 1 | For a monotonically decreasing Zone Gain, the loudness and location are expected to remain constant, while the contribution from Loudspeakers within the zone decreases. Notes 1, 2 |

Notes:

1. Sounds maintain consistent Timbre, Loudness, and Size.
2. Sounds are free of artifacts (e.g., "zipper" noise, discontinuities, clicks)

This test can be repeated for various Object positions and Spread extent parameters.

## Bibliography (informative)

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

SMPTE, ST 2098-1:2018, “Immersive Audio Metadata”

SMPTE, ST 2098-2:2019, “Immersive Audio Bitstream Specification”

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

SMPTE, ST 202:2010, “Dubbing Stages (Mixing Rooms), Screening Rooms and Indoor Theaters — B-Chain Electroacoustic Response”

SMPTE, RP 2096-1:2017, “Cinema Sound System Baseline Setup and Calibration”