 ISO/IEC JTC 1/SC 29/WG 6 N0083

**ISO/IEC JTC 1/SC 29/WG 6**

**MPEG Audio Coding Convenorship: US**

**Document type:** General

**Title: MPEG-I Immersive Audio CfP Supplemental Information, Recommendations and Clarifications, Version 1**

**Status:** Approved

**Date of document:** 2021-07-16

**Source:** ISO/IEC JTC 1/SC 29/WG 6

**Expected action:** None

**Action due date:** None

**No. of pages:** 7 (including cover page)

**Email of Convenor:** srq @ audioresearchlabs . com

**Committee URL:** https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg6

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC1/SC29/WG6, MPEG AUDIO CODING**

**ISO/IEC JTC1/SC29/WG6 N0083**

**July 2021, Virtual**

|  |  |
| --- | --- |
| **Title** | **MPEG-I Immersive Audio CfP Supplemental Information, Recommendations and Clarifications, Version 1** |
| **Source** | **WG 6, MPEG Audio Coding** |
| **Status** | **Approved** |
| **Serial Number** | **20662** |

# Introduction

This document provides clarifications and recommendations to MPEG-I Immersive Audio CfP proponents that they may find useful in developing the best possible CfP technology.

The document is a strictly informative companion to the other MPEG-I Immersive Audio CfP documents, in particular the MPEG-I Immersive Audio Encoder Input Format (EIF) document [1]. CfP proponents are free to decide whether they want to take anything that is clarified, suggested, or recommended in this document into account in developing their CfP technology.

# Additional information on CfP procedures

[the following is taken from m56709 [MPEG Req General] A proposed template for CfP and CfE]

This section provides additional information on the CfP procedures.

## Who may participate

Proponents that respond to this call may include any persons whether they are or are not accredited delegates of ISO/IEC JTC1/SC29/WG6. However, all proponents are expected to attend the meeting at which submissions are evaluated. A one-time invitation may be extended to proponents to participate in the evaluation process if the proponent is not an accredited delegate of ISO/IEC JTC1/SC29/WG6. If the proponent’s technology is accepted into the Working Draft of the Standard, then the proponents are expected to participate in subsequent WG6 meetings in which the standard is developed. In such a case where the technology is accepted from a proponent who is not an accredited delegate of ISO/IEC JTC1/SC29/WGN, the proponent is expected to initiate the process to join their National Body committees in order to become accredited to participate in subsequent meetings of WG6. Information for how to join National Body committees and to become an accredited delegate for ISO/IEC JTC1/SC29/WG6 is available at [How to Get Involved](https://www.iso.org/get-involved.html).

## Code of conduct

All participants shall be required to familiarize themselves with relevant [ISO Policies and Procedures](https://www.iso.org/resources.html), including in particular [ISO Code of Conduct](https://www.iso.org/publication/PUB100397.html), [ISO Declaration for Participants in ISO Activities](https://www.iso.org/declaration-for-participants-in-iso-activities.html), [ISO Privacy and Copyright](https://www.iso.org/privacy-and-copyright.html) policy, and [ISO Policy on Communication of Committee Work](https://www.iso.org/publication/PUB100382.html), and to consent to be bound by these policies.

## IPR

Proponents are advised that this CfP is being made subject to the common patent policy of ITU-T/ITU-R/ISO/IEC (refer to [www.itu.int/ITU-T/dbase/patent/patent-policy.html](http://www.itu.int/ITU-T/dbase/patent/patent-policy.html) or Appendix I of [ISO/IEC Directives Part 1](http://isotc.iso.org/livelink/livelink?func=ll&objId=4230455&objAction=browse&sort=subtype)).

## Testing Fee

There is no fee for submission to this CfP.

# Clarifications and recommendations on the Encoder Input Format (EIF) document

This section contains clarifications and recommendations related to various aspects of the “MPEG-I Immersive Audio Encoder Input Format” (EIF) document [1].

## Clarification of “permanence of EIF”

The EIF was created to support responses to the CfP. While it may have a purpose after the CfP, for example as the basis for the audio components of a possible MPEG scene graph standard, it is not envisioned that the EIF would be used for this purpose “as is.” Rather, it is envisioned that CfP proponents will propose changes to the EIF as part of their CfP submission, or propose changes as part of the MPEG-I Immersive Audio CE process or as part of a possible future development of a standard for an audio scene graph.

## Clarification of reference distance for object sources with extent

According to the definition in the EIF, the reference distance (refDistance) field of the various source types allows to specify a distance at which – independently from the distance attenuation law used – the computed attenuation of the audio element is 0 dB.

The EIF does however not specify how user-source distance is measured for object sources with an extent, e.g. relative to the origin of the extent (i.e. its “position” attribute), or relative to the extent itself.

To avoid any confusions in this respect (which may lead to unintentional level differences between renderers and/or mismatch between the level intended by a scene provider and the level delivered by a renderer), scene providers of CfP test scenes that contain object sources with extent are encouraged to specify in this clarifications document what was the interpretation of user-source distance (e.g. relative to extent, or relative to origin) that was used when setting the value of the refDistance attribute of each extended source in their scene. If nothing is specified for a source, the default interpretation is “relative to extent”.

## Clarification of reverberation attributes

This section contains clarifications regarding the RT60, predelay, and DDR attributes of EIF acoustic environments.

### **Source of acoustic environment data**

Scene providers are encouraged to describe how they obtained the acoustic parameters of an acoustic environment, e.g. from:

* a (measured or simulated) RIR,
* data from literature or databases,
* authoring by artistic intent.

This may include aspects of technology that proponents submit to the CfP.

### **RT60**

The calculation of the values being provided in the EIF attribute called RT60 is performed by calculating twice the T30 decay time. The T30 is derived from a “non-curved” 30 dB attenuation region of the energy decay curve of the RIR. This region is selected by the scene provider. This does not have to be a “script-driven” process.

### **Predelay**

Predelay indicates the lag in the room response from which it is diffuse. I.e. the same incident level from all directions and a similar level over all positions in the room.

The equation estimating an appropriate predelay value is based on the time of flight (ToF) from a sound wave between the two most distant boundaries of the room:

with m/s, the speed of sound, and dlongest is the largest axis of the bounding volume (in m).

The reasoning is that it takes at least two back-and-forth bounces between the most distant walls before a diffuse sound field has been achieved in the room. In order to capture only the diffuse response in the RIR, predelay must be high enough to avoid the influence of early reflections.

Note that the late reverberation rendering may be starting before this predelay value, depending on the choices of the renderer or proponent tuning.

### **DDR**

This section provides guidance to scene providers on how the value of the DDR attribute can be derived in various authoring scenarios, and to CfP proponents how a renderer’s correct handling of a specified DDR value can be verified.

**Content creation**

The value of the DDR attribute defined in the EIF can be obtained by measuring the conditioned reverberant-to-direct ratio, defined as:

where is a RIR measured with an omnidirectional point source at a distance of m from an omnidirectional mic.

The values of the time constants in the equation above are as follows:

* = obtained through the following procedure:
  + Find the time of the maximum amplitude of the first peak,
  + Then go back in time and find the time at the -20dB of the raising slope
* = + 3 msec [2]
* t2 = predelay, as defined in the previous subsection.
* TRIR = length of the RIR, or the time at which the noise floor is reached.

An RIR of a source at any distance from the mic can be measured and converted to the conditioned RDR defined above (provided it is not too close, not when occluders are interfering and not so far that the integration ranges given by and overlap) from the following modified equation:

where is an RIR measured with an omnidirectional point source at an arbitrary distance rmeas from the mic. This allows using measurements that have not been performed at r0 (1 m).

Likewise, an RIR measured with a non-omnidirectional source with a known directivity pattern can be converted to the conditioned RDR defined above by applying a correction factor to the measured reverberant component (numerator) of the RDR that compensates for the relative source power of the directional source with respect to an omnidirectional source, and a correction factor that compensates the measured direct sound component (denominator) for the directivity of the source in the direction of the microphone.

The correction factor for the direct sound component is , with the directivity factor of the source in the direction of the microphone.

The correction factor for the reverberant component is , with the average of the squared directivity factor of the source over the unit sphere.

Combining the corrections for arbitrary measurement distance and non-omnidirectional source results in:

**Relation between RDR and EIF DDR**

If we add a fixed scale-factor that accounts for the mapping from the direct sound level of the omnidirectional point source to its total emitted energy, we arrive at the EIF definition of the DDR:

where

with m2.

As a result:

or

The latter formula shows that the DDR attribute defined in the EIF can be obtained by measuring the conditioned RDR.

Clarification of the 41 dB relationship between DDR and RDR in the equation above:

The EIF DDR measure is defined as a ratio of diffuse energy and total emitted source energy, whereas the RDR measure is defined as a ratio of diffuse energy and direct sound energy.

This means that to obtain the EIF DDR from a measurement of RDR, we need a mapping from direct sound energy to total emitted source energy.

Generally speaking, this mapping is given by the fraction of the total emitted source energy that reaches the listener’s ear directly, i.e. without any reflections.

For an omnidirectional source, this energy fraction is easily found: it is simply the ratio between the surface of the sphere with radius *r*, with *r* the measurement distance, and the surface of the human ear:

With A = 0.001 m2 the surface of the human ear.

Since we define the RDR measure to be for an omnidirectional source measured at 1 m, this simplifies to:

The DDR can be calculated for different frequencies by feeding the RIR through a filterbank where the bands correspond with the desired frequencies for the DDR values.

In authoring scenarios where no RIR for the acoustic environment is available and the desired balance between direct sound and reverberation is tuned directly (e.g. by tuning the output level of a reverberation module), essentially the same procedure as described above can be used, but instead of deriving the energies of direct sound and reverberation from a RIR they are now measured directly, e.g. by rendering the direct sound and reverberation in isolation and measuring the resulting level at the mic position, or by measuring the output levels of the direct sound and reverberation modules going into the binauralization engine.

**Verification of reverberation rendering**

In order to verify the correctness of the reverberation renderer, the following procedure can be used.

When the DDR value of an acoustic environment in a scene is provided e.g. as the result of an authoring process, the following procedure verifies the correct implementation of the reverberant-to-direct scaling in the renderer:

* Place an object source in the virtual space.
  + Reference distance = 1 m
  + Omnidirectional (directivity gain = 1 for all directions)
* Place the listener at a suitable position[[1]](#footnote-1) in the virtual space, 1 m from the source.
* Render a test signal using a DIRAC HRTF, outputting only the reverberation signal.
  + Reverberation rendering should start from lag .
  + Measure the energy of the reverb signal
* Render a test signal using a DIRAC HRTF, outputting only the direct signal.
  + Measure the energy of the direct signal

The relation between the two should be:

The verification procedure can be done in a frequency dependent manner, by feeding the input and output signals and HRTF IRs through a filterbank where the bands correspond with the center frequencies of the DDR information.

* 1. **Extracting auxiliary metadata from the EIF**

A semi-automatic procedure to extract auxiliary metadata from the EIF is allowed if it fulfills the following criteria:

1. It is described in pseudocode as a sequence of ordered steps, where a step could involve a loop over a sequence of ordered steps
2. Each step carefully describes:
   1. The input data used in the step
   2. The processing performed on the input data
   3. The output data from the step
3. The processing performed on the input data can involve:
   1. Any type of program configured and executed through any type of interface, including both graphical and programmable interfaces.
   2. Human interaction which must be very clearly described together with all assumptions being made, so that it is fully repeatable by other proponents. This interaction cannot be based on any audio output produced by program. Examples of allowed interactions are setting values through a GUI, interactions with displayed graphical representations and reading output parameter values.

Neither the operation of the program nor the human interactions must include any scene specific parameters/steps.

1. The output from each step must be completely reproducible by any proponent executing the step from the provided description of the step. If this is not the case, the proponent using the semi-automatic extraction procedure in their candidate solution will be disqualified from the CfP. To this end, the name, platform and version of the used program/software have to be supplied to enable reproduction of the submitted results.

## User Reachable Region

For VR scenes (Test1 and Test3):

In each test scene / scene task, the region (or better: the volume) that can be reached by the listener (subject) is defined and restricted by the scene’s Unity part. In order to allow optimized encoding/rendering (e.g. acoustic pre-baking), it is useful to provide this information about the user reachable region also to the audio encoder.

In the spirit of simplicity, the user reachable region (URR) is specified as one or several EIF primitives (box, sphere, cylinder) and it should be noted that this is – by definition – an approximation of the reachable area in Unity rather than an exact representation. However, this does not affect its usefulness as a hint for optimized encoding/rendering.

Scene providers shall ensure that teleportation outside the specified regions are not permissible.

* Scene providers shall provide user reachable volume boxes (URR) for all scenes in the EIF file for the scene. Optionally, a visualization of the boxes may be provided in the clarifications document. This URR shall be at least the area accessible via teleportation *plus* the maximum dimension (as specified in AEP document) of the play area.
  + Use specific box naming, which shall be “userregion1 …n”.
* Proponent encoder may use this extra information.
  + Proponents are advised that renders should support rendering anywhere within the URR.

# Clarifications and recommendations on individual MPEG-I Immersive Audio CfP test scenes

This section contains clarifications and recommendations from CfP test scene providers to CfP proponents related to individual CfP test scenes.

# References

[1] ISO/IEC JTC1/SC29/WG6 N0054 “MPEG-I Immersive Audio Encoder Input Format”, MPEG134, Virtual, April 2021.

[2] Larsen, E., N. Iyer and a.A.S.F. Charissa R. Lansing, “On the Minimum Audible Difference in Direct-to-reverberant Energy Ratio”, J. Acoust. Soc. Am., 2008. 124: p. 450. Available from: <https://asa.scitation.org/doi/10.1121/1.2936368>.

1. Not with any occluder between the source and receiver, or directly next to a reflecting surface. [↑](#footnote-ref-1)