**[INSTRUCTIONS: PLEASE READ and remove before submitting the Call.]**

**[This Call shall include a complete description of the process to prepare for and respond to the Call. It shall also provide a detailed description of how each submission will be subsequently evaluated, and how each of the proponents for a winning proposal shall contribute to the selection process. The Call shall furthermore include a description of all steps in the Standard development process including an indication in which of those steps proponents of winning technologies are expected to participate.]**

**[This Call shall be prepared so that persons who are not already members of the WGN can participate. Please refrain from using terminology that would be familiar by only persons who are already in MPEG or who have participated in activities leading to this preparation of this Call.]**

**[All content or sections that are indicated in GREEN highlight are REQUIRED to be included in the Call, unless permission to exclude these items has been obtained from the Convenor to eliminate them from the package of the Call.]**

**[All contents that are marked in YELLOW highlight are examples (and only examples) of items that are helpful to highlight for readers of the Call. These are simply examples to provide some guidance to the editor of this Call of types of items that might be important to highlight in the text of the Call.]**

**[All contents that are highlighted in RED serve as guidance to the editor of this Call. PLEASE REMOVE ALL TEXT THAT IS MARKED IN RED before submitting this Call for publication. Please also remove all GREEN highlight as well.]**

**[PLEASE ENSURE THAT YOU INCLUDE THE CURRENT OUTPUT DOCUMENT COVER SHEET FOR THE WGN AS THE FIRST PAGE OF THE CALL.]**

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 2**

**MPEG TECHNICAL REQUIREMENTS**

**ISO/IEC JTC 1/SC 29/WG 2 N** **0000**

**Online - October 2020**

|  |  |
| --- | --- |
| **Title** | **DRAFT ? MPEG-I Immersive Audio Call for Evidence** |
| **Source** | **WG 2, MPEG Technical requirements** |
| **Status** | **Approved** |
| **Serial Number** | **MPEG number, digits only** |

Table of Contents

[1 Introduction 3](#_Toc61450440)

[2 Who may participate 5](#_Toc61450441)

[3 Code of conduct and rules of engagement 5](#_Toc61450442)

[4 Source code and IPR 5](#_Toc61450443)

[5 Testing Fee 5](#_Toc61450444)

[6 Definitions 5](#_Toc61450445)

[7 Documents of CfE package 6](#_Toc61450446)

[8 Submission Process 6](#_Toc61450447)

[8.1 CfE Time Line table 6](#_Toc61450448)

[8.2 Register your participation 7](#_Toc61450449)

[8.3 Mandatory Equipment, Software and Data Components 7](#_Toc61450450)

[8.4 Get Audio Evaluation Platform 8](#_Toc61450451)

[8.5 Prepare Test Material 8](#_Toc61450452)

[8.6 Get Test Material 8](#_Toc61450453)

[8.7 Measure Submission Latency and Loudness 8](#_Toc61450454)

[8.8 Submit Real-Time Max External, Bitstreams, Latency and Loudness 9](#_Toc61450455)

[8.9 Get Test Code, Metadata and Signals for Subjective Test 10](#_Toc61450456)

[8.10 Conduct Subjective Tests 10](#_Toc61450457)

[8.11 Conduct Objective Evaluations 11](#_Toc61450458)

[8.12 Submit Subjective Test Data 12](#_Toc61450459)

[8.13 Perform Analysis of Subjective Test Data 12](#_Toc61450460)

[8.14 Submit Proponent Documentation 12](#_Toc61450461)

[8.15 Evaluate CfE Submissions 13](#_Toc61450462)

[9 Call Administrator 13](#_Toc61450463)

[10 Email reflector 13](#_Toc61450464)

[11 References 13](#_Toc61450465)

[12 Copyright Header for MPEG-I Immersive Audio Reference Software 14](#_Toc61450466)

[Annex A Proposal description 15](#_Toc61450467)

[Annex B Detailed description of test sequences 17](#_Toc61450468)

[Annex C TEST SEQUENCES, DECODED RESULTS, cross check 20](#_Toc61450469)

[C.1. Distribution formats 20](#_Toc61450470)

[C.2. Decoded results 20](#_Toc61450471)

[C.3. Delivery of bitstreams 20](#_Toc61450472)

[C.4. Cross check 20](#_Toc61450473)

[Annex D Description of Testing Environment and methodology 22](#_Toc61450474)

[D.1. Test procedure 22](#_Toc61450475)

[D.2. Selection of the test method 22](#_Toc61450476)

[D.3. How to express the visual quality opinion with DSIS 23](#_Toc61450477)

[D.4. How to express the visual quality opinion with DSCQS 24](#_Toc61450478)

[D.5. Training and stabilization phase 25](#_Toc61450479)

[D.6. The laboratory set-up 25](#_Toc61450480)

[D.7. Example of a test schedule for a day 26](#_Toc61450481)

[D.8. Overall test effort and subjects’ involvement 27](#_Toc61450482)

[D.9. Statistical analysis and presentation of the results 28](#_Toc61450483)

[D.10. Proponents identification and file names 28](#_Toc61450484)

# Introduction

[Include a thorough description of the purpose for the Call. Do not assume that the reader is already familiar with the MPEG work leading to the issuance of this Call.]

The MPEG-I suite of standards, “Coded Representation of Immersive Media,” is intended to support emerging virtual and augmented reality applications. This Call for Evidence (CfE) is to explore whether or not sufficient technology exists that could be standardized in potential Part 4, “Immersive Audio.”

Along with other parts, Part 3, “Immersive Video” and Part 2, “Systems Support,” the suite of standards will support a Virtual Reality (VR) or an Augmented Reality (AR) presentation in which the user can navigate and interact with the environment using 6 degrees of freedom (6 DoF), that being spatial navigation (x, y, z) and user head orientation (yaw, pitch, roll).

The goal in MPEG-I presentations is to impart the feeling that the user is actually present in the virtual world. Audio in the world (or scene) is perceived as in the real world, with sounds coming from an associated visual figure. That is, perceived with the correct location and distance. Physical movement of the user in the real world is perceived as having matching movement in the virtual world. Furthermore, and importantly, the user can interact with the virtual scene and cause sounds that are perceived as realistic and matching the user’s experience in the real world.

The architecture for Part 4, Immersive Audio is shown in Figure 1 of “MPEG-I Immersive Audio Architecture and Requirements” [1]. A key stipulation of the architecture is that the audio compression engine used in MPEG-I Immersive Audio is MPEG-H 3D Audio (ISO/IEC 23008-3), specifically LC Profile. This is shown in the tan shaded box of Figure 1 in [1]. The other audio decoder, “Low-Delay Decoder”, is out of scope for MPEG-I Immersive Audio, and is shown in the purple shaded box of Figure 1 in [1]. The new technology that is to be standardized in MPEG-I Immersive Audio is shown in the green shaded boxes of Figure 1 in [1], and primarily consists of:

* Technology for rendering the audio presentation while permitting the user to have 6 DoF movement.
* Metadata to support this rendering.
* A bitstream syntax that enables storage and streaming of the MPEG-I Immersive Audio content. Since the coding of all audio signals is done using MPEG-H 3D Audio LC Profile, the MPEG-I bitstream must be able to convey an MPEG-H 3D Audio LC Profile bitstream.

The framework for evaluating the CfE submitted technology is shown in Figure 1 of [4]. This figure shows that the audio media signals are encoded and decoded off-line and then made available to all proponents. This coding is done using MPEG-H 3D Audio LC Profile, and in this way audio signal compression is “out of scope” when responding to this CfE.

Furthermore, an important component of the evaluation of proponent technology is assessing subjective quality of the MPEG-I Immersive Audio presentation, which is done via subjective listening tests in which all proponent MPEG-I Immersive Audio rendering technology must run in real-time on the MPEG-I Audio Evaluation Platform (described in [4]).

Additional components of the evaluation of proponent technology will be done via proponent-supplied description of the functionality of the submitted technology and also via proponent objective measurements of aspects of the technology (e.g. motion-to-sound latency), which must be included in the submitted proponent documentation.

# Who may participate

[update N in WGN and do not change further]

Proponents that respond to this call may include any persons whether they are or are not accredited delegates of ISO/IEC JTC1/SC29/WGN. However, all proponents are expected to attend the meetings at which their respective proposals are evaluated. The meeting during which proposals are evaluated is identified with an \* in Table 1. 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/WGN. Information for how to join National Body committees and to become an accredited delegate for ISO/IEC JTC1/SC29/WGN is available at [How to Get Involved](https://www.iso.org/get-involved.html).

# Code of conduct and rules of engagement

[Do not change]

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.

# Source code and IPR

[Do not change]

There is no obligation of the proponent to release source code, provided experts are still able to conclude that technical evidence to commence standardization work exists.

Furthermore, proponents are advised that this Call 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

[If there is a testing fee, provide exact purpose for the fee, when it is due, and how to pay the fee. Include any information for refunds if relevant. **IDENTIFY IN THE CfE timeline table when the fee is due**.]

# Definitions

[This Call shall include a complete list of terms and definitions used throughout the documents. This may be in a separate file included in the Call package. Do not assume that the reader is already familiar with the terms and definitions used throughout the work in MPEG leading to the issuance of this Call].

The definitions for terms associated with this Call for Proposals are found in [1].

# Documents of CfE package

[The CfE package shall be a zip file including documents for requirements, use cases, test and evaluation procedures, licenses to test content, information for how to access test content, instructions for how to formulate results, forms for providing results including excel spreadsheets if any, information about methodology for core experiments if planned. The CfE package shall be a complete and standalone package.]

The CfE package consists of the following documents:

Nxxxxx MPEG-I Immersive Audio Architecture and Requirements

Nxxxxx MPEG-I Immersive Audio Encoder Input Format

Nxxxxx MPEG-I Immersive Audio Augmented Reality Listener Space Description Format

Nxxxxx MPEG-I Immersive Audio Documentation for the Audio Evaluation Platform

Nxxxxx MPEG-I Immersive Audio Call for Evidence

Nxxxxx MPEG-I Immersive Audio Test and Evaluation Procedures

Nxxxxx MPEG-I Immersive Audio Core Experiment Methodology

Nxxxxx MPEG-I Immersive Audio Licenses for Test Content

At the 3rd WG 6 Meeting, April 2021, the specific logistics for the CfE subjective testing will issue:

Nxxxx MPEG-I Immersive Audio Subjective Test Logistics

# Submission Process

## CfE Time Line table

[The CfE shall include a time line table for processing results of the Call with \* used to indicate which MPEG meetings that the proponents are required to attend. Include other important dates as relevant, e.g., when the testing fee is due.]

Each entry in the table is described in a section below. WG 6 is the SC 29 working group WG 6 MPEG Audio Coding.

|  |  |  |  |
| --- | --- | --- | --- |
| **Meeting** | **Date** | **Who** | **Action** |
| 2 | Jan 2021 | WG 6 | Issue Call for Proposals package |
|  | 2021-04-12 | Proponent | Register |
|  |  | Proponent,  Test labs | Get Audio Evaluation Platform |
|  | 2021-02-12 | WG 6 | Prepare Test Material |
|  | 2021-02-15 | Proponent | Get test material |
|  | 2021-06-01 | Proponent | Submit real-time Max External, bitstream files, and other required documentation. (e.g. latency, loudness) |
|  | 2021-06-04 | Test Lab | Begin subjective tests. |
|  | 2021-06-25 | Test Lab | Submit subjective test data |
|  |  | Test Administrator | Performs analysis of subjective test data and submit result as contribution to 4th WG 6 meeting |
|  |  | Proponent | Submit proponent documentation as contribution to 4th WG 6 meeting |
| 4\* | Jul 2021 | WG 6 | Evaluate Call for Proposals submissions and select technology\* Respondents must be present at the meeting. |
| 5† | Oct 2021 | Proponent | Submits Working Draft (WD) and Reference Model source code (RM) for selected technology. |

Table 1 – CfE Time Line (\*Indicates attendance at the meeting is required.)

## Register your participation

[The CfE shall include exact instructions for how the respondent is required to register a response or otherwise indicate interest.]

Proponent must register on or before the date shown in the CfE Time Line table above, an intention to participate in the CfE. Registering an intent is not binding and registered parties are not required to submit proposals. However, parties that do not register will not be able to submit proposals. Register by sending an email to Schuyler Quackenbush (Convenor, WG 6 MPEG Audio Coding, srq@audioresearchlabs.com). Email should indicate

* Company name
* Contact name and contact email address
* The envisioned scope of the proposal (e.g. will it process all or only some content types, will it be valid for all or only some subjective tests, will it meet all or only some requirements). This envisioned scope is not binding and is not a restriction on proponent response, but is rather for planning purposes only.

After registration, the proponent will receive two “ProponentIDs” (<Proponent\_ID>) for use in submission of Max External and other materials (e.g. bitstream files). This permits each proponent to have two submissions, e.g. each tuned to a different operating point.

## Mandatory Equipment, Software and Data Components

[Identify mandatory equipment, software, and other data components. Ensure that licenses for software are provided to respondents. Include instructions for how to access any required software. Do not assume that all respondents have MPEG credentials for logging into MPEG repositories.]

The use of the following equipment, software and data components is mandatory in the CfE:

* The Beyerdynamic DT-990 Pro headphones in the CfE tests.
* Specific computer hardware to run the Audio Evaluation Platform (AEP) on. See [4] for details .
* Specific software to run the AEP. See [4] for details.
* The use of the HR filter set ‘FABIAN with diffuse field equalization’ without headphone equalization by all renderers. The filter set is available as part of the AEP platform.

## Get Audio Evaluation Platform

[Include instructions for how to acquire the evaluation platform or test model software, and any licenses that may be needed.

Refer to WG2 N00032 for “Guidelines for Test Material for CfP and CfEs.]

The MPEG-I Immersive Audio CfE Audio Evaluation Platform (AEP) is needed to perform the CfE subjective tests, and a description of the AEP can be found in [4]. Proponents and organizations wishing to participate in the CfE subjective tests will need to obtain the AEP.

Proponents are encouraged to obtain the AEP to verify that the Max External for their submission can run in real time within the constraints of the system.

## Prepare Test Material

[If necessary, include a description of how the test material or anchors are prepared. Describe how the material is organized, e.g., bitrates, ANCHORS FOR TEST CLASS A, B, or C]

WG 6 experts will encode and decode all audio signals that are components of all test scenes and create a complete package for each scene that is according to [2]. Signal encoding and decoding will be at a bit rate per audio object, channel set or HOA signal as indicated in the tables in [6], Annex A, using MPEG-H 3D Audio LC profile. The Test Administrator, supported by WG 6 experts, shall put encoded/decoded test items in appropriate folder location of the AEP, as indicated in the AEP documentation [4].

## Get Test Material

[Include exact instructions for how to access test material or anchors. Identify which configuration files are associated with which anchors.]

Proponent should get the test material on the date shown in the CfE Time Line table, above. Test material is available by downloading the AEP master branch, indicated in [4]. The test items used in the evaluation process shall be made available in the format defined in [2]. The list of test material is shown in Annex A. Test material that requires execution of a use agreement is indicated, along with contact person for obtaining the agreement and information on how to access the test material.

## Measure Submission Latency and Loudness

[Include instructions for what the respondents are to measure beyond objective test results, e.g., encoding/decoding times, other measurements, etc… Include references to other specifications or documents that are required to understand how to perform these measurements. Such references shall be accessible to all respondents.]

In order for the test administrator to create appropriate test configuration files, proponent should determine the signal latency of each of its Max External plugins according to the signal latency definition provided in section 4.1 of [6].

Furthermore, loudness data for each of the Max External plugins for all scenes must be provided. Proponent shall use the loudness measurement functionality of the AEP, as described in section 7.9 of [4].

## Submit Real-Time Max External, Bitstreams, Latency and Loudness

[Specify exact and complete instructions for what the respondents are to provide, including references to Excel spreadsheets. Provide detailed instructions including any references to username/passwords that are needed for submitting responses to the Call.]

Proponent must submit proposal materials no later than the date shown in the CfE Time Line table, above. Submission is via FTP with a site URL and username/password communicated by the Test Administrator to all registered proponents.

Two randomly assigned proponent IDs will be delivered to each proponent by the Test Administrator, sufficient to permit two proponent submissions. These IDs must not be disclosed, as they are visible to all Test Labs and Test Labs should not be able to determine the source of the submissions (i.e. name of proponent).

Proponents will submit a bitstream (i.e. MPEG-I metadata) file for each of the test material items in each of the subjective tests for each of the proponent submissions. Bitstream (i.e. MPEG-I metadata) files shall enable random access in time on 1 second time intervals. This is additionally discussed in Table B-1, found in [6], Requirement 8 of [1]. The size of these bitstreams (or metadata files) is not constrained. A lower bitrate (i.e. file size) is preferred according to Requirement 2 of [1] as described in [6].

For each submission, proponents must provide:

* A real-time Max External that performs the proposed technology
* Bitstream (i.e. MPEG-I Immersive Audio metadata) files for each of the test material items in each of the subjective tests
* Signal latency of the Max External (for each configuration used in each of the tests)
* Loudness output of the Max External as specified in AEP [4] for each of the test material items in each of the tests.

The Max External shall be named as:

ME\_<Proponent\_ID>~.mxe64 (windows platforms only)

The Bitstreams shall be named as:

<Test\_ID>\_<Scene\_ID>\_<Proponent\_ID>.bin (or whatever extension the proponent wishes)

Test Administrator, supported by WG 6 experts, shall put proponent Max Externals and bitstreams in appropriate folder location of the AEP, as indicated in the AEP documentation [4].

## Get Test Code, Metadata and Signals for Subjective Test

[If participants can participate in subjective tests, indicate which resources are needed to do so.]

Test code, test items and associated metadata is available by downloading the AEP master branch, indicated in [4].

## Conduct Subjective Tests

[Include this section if relevant to participation in the Call.]

Any WG 6 member organization with the appropriate listening environment, laboratory equipment and expertise can participate in the subjective tests. Subjective testing will be conducted as set forth in [6], which gives details on the requirements for test rooms for HMD/Headphone listening. Logistics for conducting the subjective tests are set forth in [7]. Since submissions are anonymized via Proponent ID, proponents are welcome to participate.

### Test 1 -- Virtual Reality

[Describe each of the subjective tests if relevant. Provide references or access to the specifications that describe the methodology.]

This test will form the “core” technology for RM0.

|  |  |
| --- | --- |
| **Test 1** |  |
| Presentation | HMD (HTC Vive or Vive Pro) with headphones. |
| Material | Objects, Channels and 3DoF HOA (i.e. “interior” HOA, in which user position does not influence rendering)  Material for Test 1 is listed in Table A-1 of Annex A, found in [6]. |
| Methodology | TBD  Subject uses 6DoF full body motion in assessment. |

### Test 2 -- Augmented Reality

Technology selected in this test will be merged into the Test 1 “core” technology.

|  |  |
| --- | --- |
| **Test 2** |  |
| Presentation | HMD (Microsoft Hololens or Hololens 2) with headphones. |
| Material | Objects, Channels  Material for Test 2 is listed in Table A-2 of Annex A, found in [6]. |
| Methodology | TBD  Subject uses 6DoF full body motion in assessment. |

### Test 3 -- Virtual Reality with 6DoF HOA signals

Technology selected in this test will be merged into the Test 1 “core” technology.

|  |  |
| --- | --- |
| Test 3 |  |
| Presentation | HMD (HTC Vive or Vive Pro) with headphones. |
| Material | 6DoF HOA, in which user position influences rendering, including:   * “interior/exterior” HOA, and * “multi-point” HOA, in which there are HOA signals located at more than one position,   and Objects, Channels  Material for Test T3 is listed in Table A-3 of Annex A, found in [6]. |
| Methodology | TBD  Subject uses 6DoF full body motion in assessment. |

Tests 1, 2 and 3 are intended to assess subjective perception of all aspects of the VR/AR experience, including those listed in Table 3, below.

|  |
| --- |
| **Characteristic** |
| Audio timbral quality |
| Audio distortions and artefacts |
| Audio source localization quality |
| Audio source spatial extent quality |
| Audio source directivity quality |
| Audio reverberation, occlusion and diffraction quality |
| Consistency of audio with respect to user motion (i.e. sufficiency low motion to sound latency) |
| Consistency of audio with respect to visual cues (i.e. co-localization of visual and auditory sound objects) |
| Overall auditory immersion elicited by presented scene |
| Plausibility of overall experience |

Table 3 – Non-exhaustive List of Perceptual Characteristics to be Tested

## Conduct Objective Evaluations

[Provide complete instructions for what respondents are to include in their objective test measurements. Reference Excel spreadsheets or other resources that are needed to obtain such test measurements.]

Proponents must provide the objective measurement of the following parameters. Details on how to measure are given in [6], Annex B.

* computational complexity,
* latency (motion to sound latency as well as signal latency),
* bitrate,
* memory usage.

In addition, proponents must provide objective or descriptive characterizations of their proposals as detailed in Table B-1 of Annex B.

* ***Other objective measurements or written descriptions***

A template for the submission of this information is given in Annex B. The proponent must submit documentation that contain a numerical measurement of the objective criterion listed in this section, along with a description of how the measurement was made, including all assumptions made for the measurement setup.

In addition, the proponent must show the extent that their submission fulfills the requirements listed in [1]. A template for the submission of this information is also given in Annex B.

## Submit Subjective Test Data

[If relevant, provide instructions for how to submit subjective test data.]

All labs participating in the subjective tests will submit test data to the Test Administrator in a format according to “MPEG-I Immersive Audio Subjective Test Logistics” [7] (to issue at 3rd WG 6 meeting) no later than the date indicated in the CfE Time Line table, above. This will allow an analysis of test results to be submitted as a WG 6 meeting contribution.

## Perform Analysis of Subjective Test Data

[If relevant, provide instructions for the analysis of subjective test data.]

The Test Administrator will collect subjective test data and perform an analysis. Data analysis and computation of a Figure of Merit will be done as indicated in [6]. It is envisioned that both Excel Pivot Table with Grand Mean and 95% Confidence Interval and also ANOVA will be performed. The ANOVA will be used to assess the suitability of merging data from the several Test Labs.

## Submit Proponent Documentation

[Provide a complete list of what shall be submitted by each proponent according to the CfE timeline table. If necessary, provide one or more response forms in an Annex to this Call.]

Proponents submit as a contribution to the MPEG meeting indicated in the CfE Time Line table, above:

* A written description of the technology having sufficient detail to permit technical discussions.
* Objective test results, as indicated in ???
* Description of how requirements are met (for the methodology “Description”, See ???). Also, Documentation that permit Audio subgroup to check objective test results.
* Evidence of the performance of the technology, as described in [6]

Proponents that are WG 6 members shall register these documents as contributions to the WG 6 meeting and send title and author information to Schuyler Quackenbush, Convenor of WG 6, prior to the close of contribution upload deadline. All proponents are urged to become WG 6 members. However, proponents that are not WG 6 members shall email the documents to Convenor of WG 6 two weeks prior to the WG 6 meeting at which proponent documentation materials are due, so that he can register and upload them as contributions. The documents should be written in Microsoft Word. The Convenor of WG 6 will extend a one-time-only invitation to the WG 6 meeting so that a non-member proponent can present their contributions and participate in the selection process.

The results of the subjective tests also will be available as a contribution to this WG 6 meeting.

## Evaluate CfE Submissions

[Describe the steps that are taken to evaluate submissions, and when these steps are taken. Provide any additional information here for what is expected of respondents to the Call.]

At the WG 6 meeting indicated in the CfE Time Line table, above, submissions will be evaluated by the WG 6 audio experts. It is strongly urged that proponents have experts familiar with the proposed technology attend in order to allow discussions on details of the proposals.

Submissions shall be evaluated, considering all submitted information including subjective listening test results.

# Call Administrator

[Indicate who to contact for additional information.]

For any questions related to this Call for Proposals or associated evaluation procedures please contact the Test Administrator:

Dr. Schuyler Quackenbush, Convenor, WG 6 MPEG Audio Coding

Audio Research Labs

336 Park Ave, Suite 200

Scotch Plains, NJ 07076

Phone: +1 908 490 0700

email: [srq@audioresearchlabs.com](mailto:srq@audioresearchlabs.com)

# Email reflector

[Include information and or recommendations about email reflectors if necessary. Provide instructions for how to subscribe to the reflector.]

# References

[Required. Please do not refer to documents that are in the MPEG Document Management System. All documents referenced shall be available to all respondents.]

1. Nxxxx MPEG-I Immersive Audio Architecture and Requirements
2. Nxxxx MPEG-I Immersive Audio Encoder Input Format
3. Nxxxx, MPEG-I Immersive Audio Augmented Reality Listener Space Description Format
4. Nxxxx MPEG-I Immersive Audio Documentation for the Audio Evaluation Platform
5. Nxxxx MPEG-I Immersive Audio Call for Proposals (*this document*)
6. Nxxxx MPEG-I Immersive Audio Test and Evaluation Procedures
7. Nxxxx MPEG-I Immersive Audio Subjective Test Logistics (*to issue at 4th WG 6 meeting*)
8. Nxxxx MPEG-I Immersive Audio Core Experiment Methodology

# Copyright Header for MPEG-I Immersive Audio Reference Software

[Provide the copyright header that proponents shall include in software contributed to the reference software or test model.]

All MPEG-I Immersive Audio Reference Software files shall contain the following header:

##########################################################################

This software module was originally developed by

<CN>

in the course of development of ISO/IEC 23090-4 for reference purposes and

its performance may not have been optimized. This software module is an

implementation of one or more tools as specified by the ISO/IEC 23090-4

standard. ISO/IEC gives you a royalty-free, worldwide, non-exclusive,

copyright license to copy, distribute, and make derivative works of this

software module or modifications thereof for use in implementations or

products claiming conformance to ISO/IEC standards incorporating ISO/IEC

23090-4 reference software and which satisfy any specified conformance

criteria of that standard.

Those intending to use this software module in products are advised that its

use may infringe existing patents and that the license in the previous

paragraph grants no licenses under such patents. ISO/IEC have no liability

for use of this software module or modifications thereof.

Copyright is not released for products that do not conform to an ISO/IEC

standard.

<CN> retains full right to modify and use the code for its own purpose,

assign or donate the code to a third party and to inhibit third parties

from using the code for products that do not conform to ITU Recommendations

and/or ISO/IEC International Standards.

This copyright notice must be included in all copies or derivative works.

Copyright (c) ISO/IEC 202X.

############################################################################

1. Proposal description

[For example only. Please delete if not relevant.]

Proponents need to complete and submit the spreadsheets provided in the attached Excel file.

The objective metrics listed in this spreadsheet must be derived using the HDRMetrics tool from the HDRTools software package, tag 0.9, accessible at the following location using the most recent MPEG meeting login/password credentials:

<http://wg11.sc29.org/svn/repos/Explorations/XYZ/HDRTools/tags/0.9>

The configuration files located in the directory ‘bin\CfE\_cfgFiles’ of the HDRTools package shall be used.

**Attachments:**

See Section 2 of Annex B

The proponents are encouraged to also submit technical details as follows:

1. An overview of the proposal:

* Single layer or Dual layer concept?
  + If Dual layer, are there inter-layer prediction mechanisms between layers?
* Usage of metadata – for each type of metadata, provide:
  + Purpose & Short description
  + Static (per title) or dynamic – if dynamic, per GOP, per frame
  + Type: signalled in VUI (are new entries required?), signalled in SEI (are new SEIs required?),
  + Typical size, Maximum size
  + Is their generation performed automatically?
* Pre-processing necessary before encoding
  + Purpose and Short description
  + Complexity (processing needs, internal bit-depth, memory needs)
  + Are the corresponding (non-normative, encoder only) parameters derived automatically?
* Post-processing necessary after decoding
  + Purpose and Short description
  + Complexity (processing needs, internal bit-depth, memory needs)
* Recommended parameters
  + HEVC profile, bit-depth, chroma format, colour difference space, coding transfer function
* Changes in core HEVC decoder specification
* Backward compatibility with SDR displays
  + Conform to input SDR?
  + If yes, with or without adaptation mechanisms after HEVC Main10 decoding?
  + Viewable on SDR displays?
  + If yes, with or without adaptation mechanisms after HEVC Main10 decoding?

1. A technical description of the proposal sufficient for the full conceptual understanding and generation of equivalent performance results by experts, and for conveying the degree of optimization required for replicating the performance. This description should include all data processing paths and individual data processing components used to generate the bit streams.
2. A detailed description of the enhancement layer coding structure (for sub-category 2a).

The technical description should contain sufficient information suitable to allow assessment of the complexity of the implementation of the technology.

1. Detailed description of test sequences

[For example only. Please delete if not relevant.]

**Class A:** Size 2560x1600 (pixel resolution as in original 4Kx2K) 30 fps

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sxx | Name | Original size/framerate | Duration | Cropped area position | source/owner/copyright |
| S01 | Traffic | 4096x2048p 30 fps | 5s | Line 80,  Column 1200 | Plannet Inc./ C1 |
| S02 | People on Street | 3840x2160p 30 fps | 5s | Line 480,  Column 540 | Samsung Electronics Co., Ltd. / C2 |

**Class B1:** Size 1920x1080p 24 fps

|  |  |  |  |
| --- | --- | --- | --- |
| Sxx | Name | Duration | source/owner/copyright |
| S03 | Kimono | 10s | Tokyo Institute of Technology, Nakajima Laboratory / C3 |
| S04 | ParkScene | 10s | Tokyo Institute of Technology, Nakajima Laboratory / C3 |

**Class B2:** Size 1920x1080p 50-60 fps

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sxx | Name | fps | Duration | source/owner/copyright |
| S05 | Cactus | 50 | 10s | EBU/RAI / C4 |
| S06 | BasketballDrive | 50 | 10s | NTT DOCOMO Inc. / C5 |
| S07 | BQTerrace | 60 | 10s | NTT DOCOMO Inc. / C5 |

**Class C:** Size 832x480p (WVGA) 30-60 fps

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sxx | Name | fps | Duration | source/owner/copyright |
| S08 | BasketballDrill | 50 | 10s | NTT DOCOMO Inc. / C5 |
| S09 | BQMall | 60 | 10s | NTT DOCOMO Inc. / C5 |
| S10 | PartyScene | 50 | 10s | NTT DOCOMO Inc. / C5 |
| S11 | RaceHorses | 30 | 10s | NTT DOCOMO Inc. / C5 |

**Class D:** Size 416x240p (WQVGA) 30-60 fps

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sxx | Name | fps | Duration | source/owner/copyright |
| S12 | BasketballPass | 50 | 10s | NTT DOCOMO Inc. / C5 |
| S13 | BQSquare | 60 | 10s | NTT DOCOMO Inc. / C5 |
| S14 | BlowingBubbles | 50 | 10s | NTT DOCOMO Inc. / C5 |
| S15 | RaceHorses | 30 | 10s | NTT DOCOMO Inc. / C5 |

**Class E:** Size 1280x720p 60 fps

|  |  |  |  |
| --- | --- | --- | --- |
| Sxx | Name | Duration | source/owner/copyright |
| S16 | Vidyo1 | 10s | Vidyo, Inc. / C6 |
| S17 | Vidyo3 | 10s | Vidyo, Inc. / C6 |
| S18 | Vidyo4 | 10s | Vidyo, Inc. / C6 |

C1:

*Individuals and organizations extracting sequence from this archive agree that the sequence and all intellectual property rights therein remain the property of Plannet inc.. This material may only be used for the purpose of developing, testing and promulgating technology standards. This material cannot be distributed with charge. Plannet inc. makes no warranties with respect to the material and expressly disclaims any warranties regarding its fitness for any purpose.*

C2:

*These sequences are generated by SAMSUNG ELECTRONICS CO., LTD. for developing image processing algorithms. All intellectual property rights for the sequences therein remain the property of SAMSUNG ELECTRONICS CO., LTD. These materials shall be used, copied, modified, or distributed only for developing MPEG and VCEG High-Performance Video coding standards and for testing and promoting such standards. This Copyright and permission notice shall be duplicated whenever the data are copied or distributed. These materials shall not be distributed with charge. For distributing of altered or processed versions of the data, any user must clearly indicate that the data has been altered and thus cannot be relied upon as representing the original material. SAMSUNG ELECTRONICS CO., LTD. makes no warranties with respect to the material and expressly disclaims any warranties regarding its fitness for any purpose. Unless the above conditions are agreed to by the recipient, no permission is granted for any use, copying, modification and distribution of the accompanying data.*

C3:

*Individuals and organizations extracting sequence from this archive agree that the sequences and all intellectual property rights therein remain the property of Nakajima Laboratory of Tokyo Institute of Technology. This material may only be used for the purpose of developing, testing and promulgating technology standards. The material cannot be distributed with charge. Nakajima Laboratory of Tokyo Institute of Technology makes no warranties with respect to the material and expressly disclaims any warranties regarding its fitness for any purpose.*

C4:

*These sequences and all intellectual property rights therein remain the property of the RAI. These sequences may only be used for the purpose of developing, testing and promulgating technology standards. RAI and EBU Technical make no warranties with respect to the sequences and expressly disclaims any warranties regarding their fitness for any purpose.*

C5:

*These sequences and all intellectual property rights therein remain the property of the NTT DOCOMO, INC. These sequences may only be used for the purpose of developing, testing and promulgating technology standards. NTT DOCOMO, INC. makes no warranties with respect to the sequences and expressly disclaims any warranties regarding their fitness for any purpose.*

C6:

*Individuals and organizations using these video sequences agree that the sequences and all intellectual property rights therein remain the property of the owner, Vidyo, Inc.  This material may only be used for the purpose of developing, testing, and promulgating technology standards.  This material cannot be distributed with charge.  The owner makes no warranties with respect to the material and expressly disclaims any warranties regarding its fitness for any purpose.*

1. TEST SEQUENCES, DECODED RESULTS, cross check

[For example only. Please delete if not relevant.]

* 1. Distribution formats

Distribution of original video material files containing test sequences is done in YUV files with extension “.yuv”. Colour depth is 8 bit per component. A description of the YUV file format is available at http://www.fourcc.org/ web site, designated as “planar iyuv”.

AVC Anchor bitstreams are provided with extension “.264”. Bitstream formats of proposals can be proprietary, but must contain all information necessary to decode the sequences at a given data rate (e.g. no additional parameter files). The file size of the bitstream will be used as a proof that the bitrate limitation from Table 2 has been observed. The file extension of a proposal bitstream shall be “.bit”.

* 1. Decoded results

Decoded sequences shall be provided in the same “.yuv” format as originals, and additionally as AVI files (“.avi” extension). A tool that converts YUV into AVI format is available from the test coordinator by request.

All files delivered (bitstreams, decoded sequences and binary decoders) must be accompanied by a checksum file to enable identification of corrupted files. MD5 checksum tools shall be used for that purpose. Such a tool is available typically as part of UNIX/LINUX operating systems where it should be run with option “-b” (binary). For Windows operating systems, a compatible tool can be obtained from http://www.pc-tools.net/win32/md5sums/. This tool should be run with additional option “-u” to generate the same output as under UNIX.

* 1. Delivery of bitstreams

Hard disc should be shipped (for handling in customs) with a declaration “used harddisc for scientific purposes, to be returned to owner” and low value specification (e.g. 20€). The use of a harddisc with substantially larger size than needed is discouraged. The harddisc should be a 3½-inch SATA drive without any additional enclosure (no case, no power supply, no USB interface etc.), NTFS file format shall be used.

* 1. Cross check

Before the evaluation meeting, a one-day cross-check meeting will be held. Proponents shall bring another harddisc, which can be connected via USB 2.0 to a Windows PC, containing original and decoded sequences in YUV and AVI formats, bitstreams, binary decoder executables and all related checksum files. An adequate computer system shall also be brought to this meeting. Proponents shall specify the computing platform (hardware, OS version) on which the binary can be run. Should such a computing platform not be readily available, the proponent shall provide a computer adequate for decoder verification at this meeting. Further information will be exchanged with the proponents after the registration deadline.

1. Description of Testing Environment and methodology

[For example only. Please delete if not relevant.]

* 1. Test procedure

The test procedure foreseen for the formal subjective evaluation will consider two main requirements:

* to be as much as possible reliable and effective in ranking the proposals in terms of subjective quality (and therefore adhering the existing recommendations);
* to take into account the evolution of technology and laboratory set-up oriented to the adoption of FPD (Flat Panel Display) and video server as video recording and playing equipment.

Therefore, two of the test methods described in Recommendation ITU-R BT.500 [5] are planned to be used, applying some modification to them, relating to the kind of display, the video recording and playing equipment.

* 1. Selection of the test method

The anticipated test methods are:

1. DSIS (Double Stimulus Impairment Scale)
2. DSCQS (Double Stimulus Continuous Quality Scale)
   * 1. DSIS

This test method is commonly adopted when the material to be evaluated shows a range of visual quality that well distributes across all quality scales.

This method will be used under the schema of evaluation of the quality (and not of the impairment); for this reason a quality rating scale made of 11 levels will be adopted, ranging from "0" (lowest quality) to "10" (highest quality). The test will be held in three different laboratories located in countries speaking different languages: This implies that it is better not to use categorical adjectives (e.g. excellent good fair etc.) to avoid any bias due to a possible different interpretation by naive subjects speaking different languages.

All the video material used for these tests will consist of video clips of 10 seconds duration.

The structure of the Basic Test Cell (BTC) of DSIS method is made by two consecutive presentations of the video clip under test; at first the original version of the video clip is displayed, immediately afterwards the coded version of the video clip is presented; then a message displays for 5 seconds asking the viewers to vote (see ***Figure C1***)



*Figure C1 - DSIS BTC*

The presentation of the video clips will be preceded by a mid-grey screen displaying for one second.

* + 1. DSCQS

This test method is typically selected when the range of visual quality presented to the viewing subjects is in the upper part of the quality scale.

In this method the original and the coded samples of a video clip are presented two times.

This allows the viewing subjects to evaluate the sequences they are seeing more accurately.

An important aspect of DSCQS is that the viewing subject is asked to evaluate the two original and the coded clips separately. Furthermore the viewer does not know which of the two clips is the original or the coded one.

The position of the original and the coded one is randomly changed for each BTC.

The structure of the Basic Test Cell of the DSCQS method therefore contains two consecutive pairs of presentations.

The first presentation is announced by a mid grey screen with the letter "A" in the middle displaying for one second; the second presentation is announced by another mid grey screen with the letter "B"; these are repeated during the second pair of presentations changing A and B into A\* and B\*; the message "VOTE" displays on the screen for 5 seconds after the second pair of presentations is done (see ***Figure C2***)



*Figure C2 - DSCQS BTC*

* 1. How to express the visual quality opinion with DSIS

The viewers will be asked to express their vote putting a mark on a scoring sheet.

The scoring sheet for a DSIS test is made of a section for each BTC; each section is made of a column of 11 vertically arranged boxes, associated to a number from 0 to 10 (see ***Figure C3***).

The viewers have to put a check mark on one of the 11 boxes; checking the box "10" the subject will express an opinion of "best" quality, while checking the box "0" the subject will express an opinion of the "worst” quality.

The vote has to be written when the message "Vote N" appears on the screen. The number "N" is a numerical progressive indication on the screen aiming to help the viewing subjects to use the appropriate column of the scoring sheet.



*Figure C3 -Example of DSIS test method scoring sheet*

* 1. How to express the visual quality opinion with DSCQS

The viewers will be asked to express their vote putting two marks on a scoring sheet.

The scoring sheet for a DSCQS test is made of a section for each BTC; each section is made of two continuous columns with 100 horizontal marks, (see Figure C4).

The viewers have to put a check mark for each of the two vertical lines; checking the upper side of the bar the subject will express an opinion of "best" quality, while checking the lower side of the bar the subject will express an opinion of the "worst” quality.

The vote has to be written when the message "Vote N" appears on the screen. The number "N" is a numerical progressive indication on the screen aiming to help the viewing subjects to use the appropriate column of the scoring sheet.



*Figure C4 -Example of DSCQS test method scoring sheet*

* 1. Training and stabilization phase

The outcome of a test is highly dependent on a proper training of the test subjects.

For this purpose, each subject has to be trained by means of a short practice (training) session.

The video material used for the training session must be different from those of the test, but the impairments introduced by the coding have to be as much as possible similar to those in the test.

The stabilization phase uses the test material of a test session; three BTCs, containing one sample of best quality, one of the worst quality and one of medium quality, are duplicated at the beginning of the test session. By this way, the test subjects have an immediate impression of the quality range they are expected to evaluate during that session.

The scores of the stabilization phase are discarded. Consistency of the behaviour of the subjects will be checked inserting in the session a BTC in which original is compared to original.

* 1. The laboratory set-up

The laboratory for a subjective assessment is planned to be set up following recommendation ITU-R BT.500 [5], except for the selection of the display and the video play server.

High quality LCD displays with controllable colour rendition will be used (e.g. Eizo LCD Color Graphic CG301W).

High quality Plasma display could also be used, mainly when displays with diagonal size wider or equal than 50” are selected.

When a video clip is shown at a resolution lower than the native resolution of the display itself, the video has to be presented in the centre of the display; the active part of the display (i.e. that is actually showing the video signal) must have a dimension equal in rows and columns to the raster of the video; the remaining part of the screen has to be set to a mid grey level (128 in 0-255 range). This constraint guarantees that no interpolation or distortion artefacts of the video images will be introduced.

The video play server, or the PC used to play video has to be able to support the display of video formats from WVGA to HDTV, at 24, 30, 50 and 60 frames per second, without any limitation, or without introducing any additional temporal or visual artefacts.

* + 1. Viewing distance, seats and monitor size

The viewing distance varies according to the physical dimensions of the active part of the video; this will lead to a viewing distance varying from 2H to 4H, where H is equal to the height of the active part of the screen.

The number of subjects seating in front of the monitor is a function of the monitor size and type; for example, a monitor equal or superior to 30" is expected to allow the seating of two or three subjects at the same time; monitors with 24" would allow two subjects; monitors of 21" would allow just one subject. Monitors with diagonal lower than 21" should not be used. In any case monitors must support a wide screen aspect ratio without any picture adaptation (e.g. letter box etc.); i.e. they must give native support for the 16:9 aspect ratio.

For class D, the use of DLP projectors will be considered. Also in this case, the distance of the viewers from the screen will still be ranging from 2H to 4H according to the laboratory set-up.

* + 1. Viewing environment

The test laboratory has to be carefully protected from any external visual or audio pollution.

Internal general light has to be low (just enough to allow the viewing subjects to fill out the scoring sheets) and a uniform light has to be placed behind the monitor; this light must have an intensity as specified in Recommendation ITU-R BT.500. No light source has to be directed to the screen or create reflections; ceiling, floor and walls of the laboratory have to be made of non-reflecting material (e.g. carpet or velvet) and should have a colour tuned as close as possible to D65.

* 1. Example of a test schedule for a day

This is an example of the schedule of the planned test activity for one day.

A time slot of one hour is dedicated every morning and afternoon to welcome, to screen and to train the viewers.

After the subjects’ screening for visual acuity and colour blindness, the subjects will be grouped in testing groups.

In the following example, four groups of subjects are created, according to the laboratory set-up[[1]](#footnote-1) and to the time constraints.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | DAY 1 | | DAY 2 | |
| 9:00 – 10:00 | Screening / training | | Screening / training | |
| 10:00 -10:40 | G1-S1 |  | G1-S1 |  |
| 10:40 -11:20 |  | G2-S1 |  | G2-S1 |
| 11:20 -12:00 | G1-S2 |  | G1-S2 |  |
| 12:00 -12:40 |  | G2-S2 |  | G2-S2 |
| 13:00 – 14:00 | Screening / training | | Screening / training | |
| 14:00 -14:40 | G3-S3 |  | G3-S3 |  |
| 14:40 -15:20 |  | G4-S3 |  | G4-S3 |
| 15:20 -16:00 | G3-S4 |  | G3-S4 |  |
| 16:00 -16:40 |  | G4-S4 |  | G4-S4 |

* 1. Overall test effort and subjects’ involvement

The duration of the test will depend on the number of submissions.

A tentative computation of the time necessary to complete the subjective test is provided below for the case of 33 submissions.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Class** | **Test sessions** | **EBU** | **EPFL** | **FUB** | **TOTAL** |
| **B** *(HDTV)* | **70** | **38** | **38** | **3** | **79 *(\*)*** |
| **C** *(WVGA)* | **35** | **3** | **3** | **35** | **41 *(\*)*** |
| **D** *(WQVGA)* | **35** | **3** | **35** | **3** | **41 *(\*)*** |
| **E** *(720p)* | **21** | **12** | **12** | **1** | **25 *(\*)*** |
|  |  | **56** | **88** | **42** | **186 *(\*)*** |

***(\*) Approximately 10% of the test sessions are run by all labs to perform a cross-check of the results***

* 1. Statistical analysis and presentation of the results

The data collected from the score sheets, filled out by the viewing subjects, will be stored in an Excel spreadsheet.

Seven spreadsheets will be prepared for Class B/C/D (each for constraint sets 1 and 2) and Class E (constraint set 2 only).

For each coding condition the Mean Opinion Score (MOS) and associated Confidence Interval (CI) values will be given in the spreadsheets.

The MOS and CI values will be used to draw graphs. The Graphs will be drawn grouping the results for each video test sequence. No graph grouping results from different video sequences will be considered.

* 1. Proponents identification and file names

Each Proponent submitting to the CfE will be identified with a two digit code preceded by the letter “P” (e.g. P01 P02 … Pnn).

Each coded video file provided for a submission will be identified by a name formed by the below listed combination of letters and numbers:

PnnSxxRyCz.<filetype>

where:

* Pnn identifies the Proponent,
* Sxx identifies the original video clip used to produce the coded video, as identified in the tables of Annex A;
* Ry identifies the rate y, as identified in Table 2;
* Cz identifies the constraint set z (z=1 or z=2), as identified in section 4.2 (Table 1 and above)
* <filetype> identifies the kind of file:
  + .bit = bitstream
  + .yuv = decoded video clip in YUV format
  + .avi = decoded video clip in AVI format

1. The viewing rooms of the three laboratories could be different according to the test material and/or to the design of the laboratory. Large displays (e.g. monitor equal or wider than 50’’) will allow to seat three (or more) subjects at the same time; a laboratory set-up, in which three wide monitors are available, will allow the creation of wider groups of viewers (three or more). [↑](#footnote-ref-1)