 ISO/IEC JTC 1/SC 29/WG 6 N0396

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

**MPEG Audio Coding Convenorship: DE**

**Document type:** Output Document

**Title:** Summary ofMPEG-I immersive audio verification test report

**Status:** Approved

**Date of document:** 2026-01-23

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

**Expected action:** None

**Action due date:** None

**No. of pages:** 6 (without cover page)

**Email of Convenor:** Thomas.Sporer@idmt.fraunhofer.de

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

**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

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

**MPEG AUDIO CODING**

**ISO/IEC JTC 1/SC 29/WG 6 N0396**

**January 2026, Online**

|  |  |
| --- | --- |
| **Title** | **Summary of MPEG-I immersive audio verification test report** |
| **Source** | **WG 6, MPEG Audio Coding** |
| **Editors** | **Andreas Silzle, Pablo M. Delgado, Sascha Disch, Alejandro Restrepo Garcia, Leon Terentiv, Erlendur Karlsson, Jussi Leppanen, Hiroyuki Ehara, Sam Jelfs** |
| **Serial Number** | **25970** |

# Abstract

The verification test evaluated the perceptual quality of the finalized international standard (IS) of the MPEG‑I immersive audio renderer [3][4] with expert listeners across six labs using an ITU‑R BS.2132 [1] multi‑stimulus method on a 100‑point scale for three conditions (IS, mid anchor, low anchor) in 10 virtual reality (VR) scenes plus two repetitions. Results showed a clear separation of the conditions (IS > mid > low); the low anchor was stable (~16 points) while the mid anchor varied by scene (around 47 points). The IS was rated with a median of 84 points among all labs, which is the “excellent” region of the scale. The ratings showed some scene dependency. The rating variance for some scenes reached ~20 points. Repetition analysis indicated good listener consistency. Across all six labs, rankings were consistent, but absolute IS median scores differed slightly.

# Overview

A verification test plan was agreed upon for a subjective evaluation of the performance of the MPEG-I immersive audio renderer. This document reports and discusses the results; the individual results obtained by each test site (Dolby, Ericsson, Fraunhofer, Nokia, Panasonic, and Philips), and their respective statistical analysis.

# Background

This section presents a summary of the important aspects agreed upon.

## Test Method

The following test method was applied:

* Multi-stimulus test ITU-R Rec. BS.2132-0 [1] and BS.1284-2 [2], without external reference
* Listener instructions: “*Your task is to rate the* ***Overall Quality and Plausibility*** *of the scene with a single score.*”
* As verbal anchors for the 100-point scale, the following terms are defined for the five 20-point regions of the scale: *excellent*, *good*, *fair*, *poor*, and *bad*.
* Listeners were required to be experts in audio evaluation. There is also a distinction between general audio evaluation experts and MPEG-I listening experts (developers of the standard and well experienced with 6DoF interactive rendering).

**Conditions:**

(1) MPEG-I immersive audio international standard (IS)

(2) Mid anchor

(3) Low anchor

Conditions (3) and (2) were designed with the intent to resemble a stable lower anchor satisfying minimum requirements for audio reproduction (3.5kHz audio bandwidth, mono mixdown), and a suitable mid anchor simulating a basic 6DoF audio rendering with 7kHz audio bandwidth and stereo amplitude panning that, in turn for its simplicity, was found to lack proper externalization.

## VR Test Scenes

10 different in- and outdoor VR scenes were selected for the test, covering all aspects of rich acoustical environments. 2 scenes were repeated for consistency checks.

# Verification Test Results

## Test Data

59 expert listeners from the different labs participated in the test. With the consistency check the results of 4 listeners were removed. One listener repeated the test successfully. So, the results of 56 listeners were further analyzed. 30 of them were MPEG-I experts and 26 were other listening test experts. The average test duration per lab was 37 minutes with a standard deviation of 14 minutes. Which leads to a total test time of 37 hours.

## Statistical Analysis and Combined Results of all Labs

An ANOVA test across all data for multiple factors was calculated. All factors are significant, with the Condition being the factor that explains most of the variance in the data. A Tukey HSD (Honestly Significant Difference) pairwise comparison over conditions shows that all conditions are significantly different from each other. The Tukey HSD comparisons for the labs reveals that only Fraunhofer IIS and Nokia show a slight significant difference over all pairwise-comparisons. Given these results, it would appear fair to average across scenes and labs for the final analysis.

Figure 1 and Figure 2 present the pooled results of all labs, with 56 listeners, after post-screening.

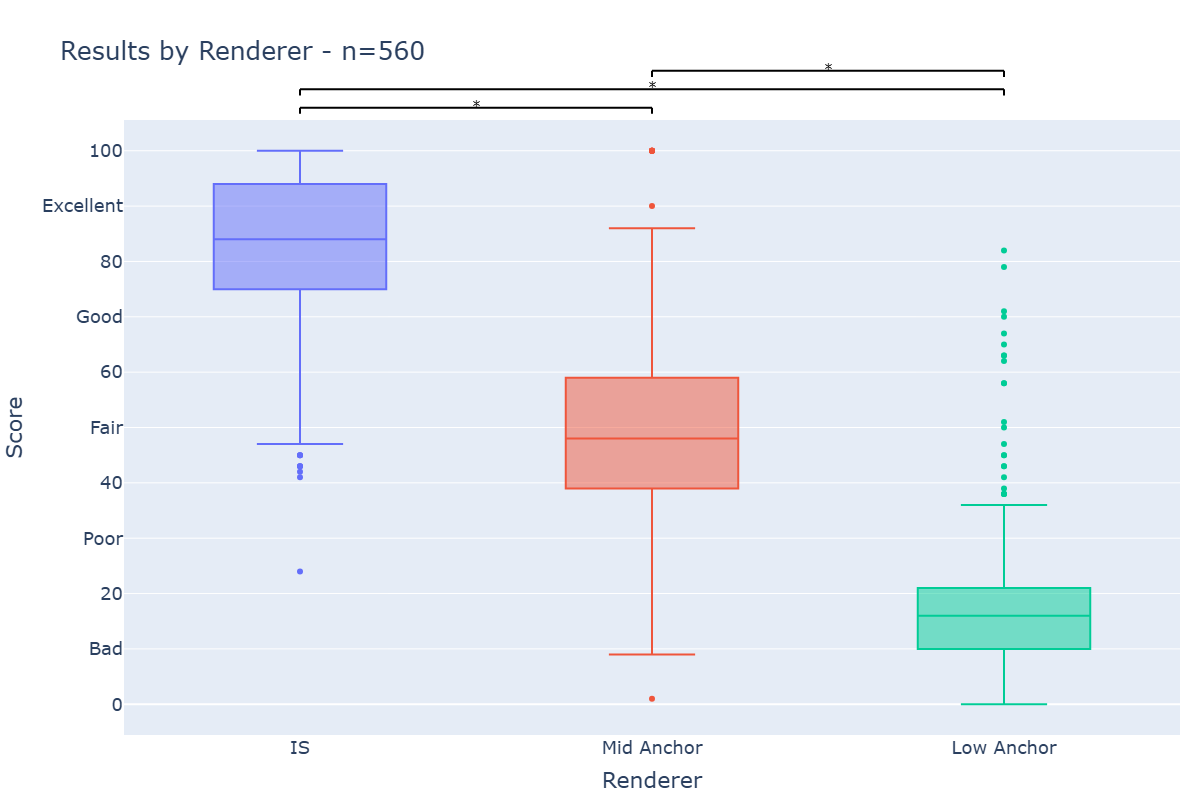


Figure 1: Overall results by render condition as box plots, N = 560 (56 listeners, 10 trials per listener).

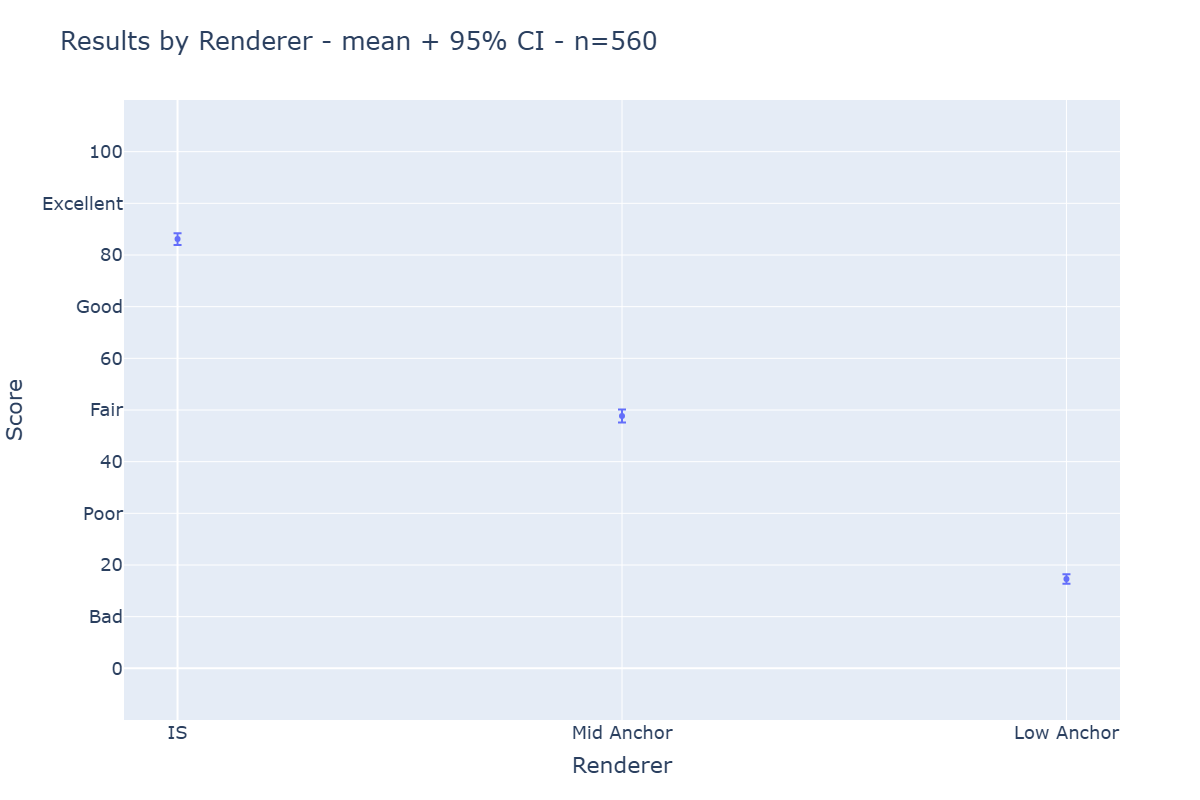


Figure 2: Overall results by render condition as mean and 95% confidence interval, N = 560.

## Detailed Test Results

The results are presented as box plots, with median and quartile values. The non-parametric presentation does not require normal distribution of the results and is preferred for results from a small number of listeners. In Figure 3, the scene results of the pooled labs are presented. The listening test results show:

1. A clear separation of the conditions (IS > mid > low)
2. The low anchor was stable (around 16 points median value)
3. The mid anchor varied by scene (around 47 points).
4. The IS is rated on average around 84 points among all labs, which is the “excellent” region of the scale.
5. The individual scenes are rated differently.
6. The quartile range for some scenes can reach 20 points.

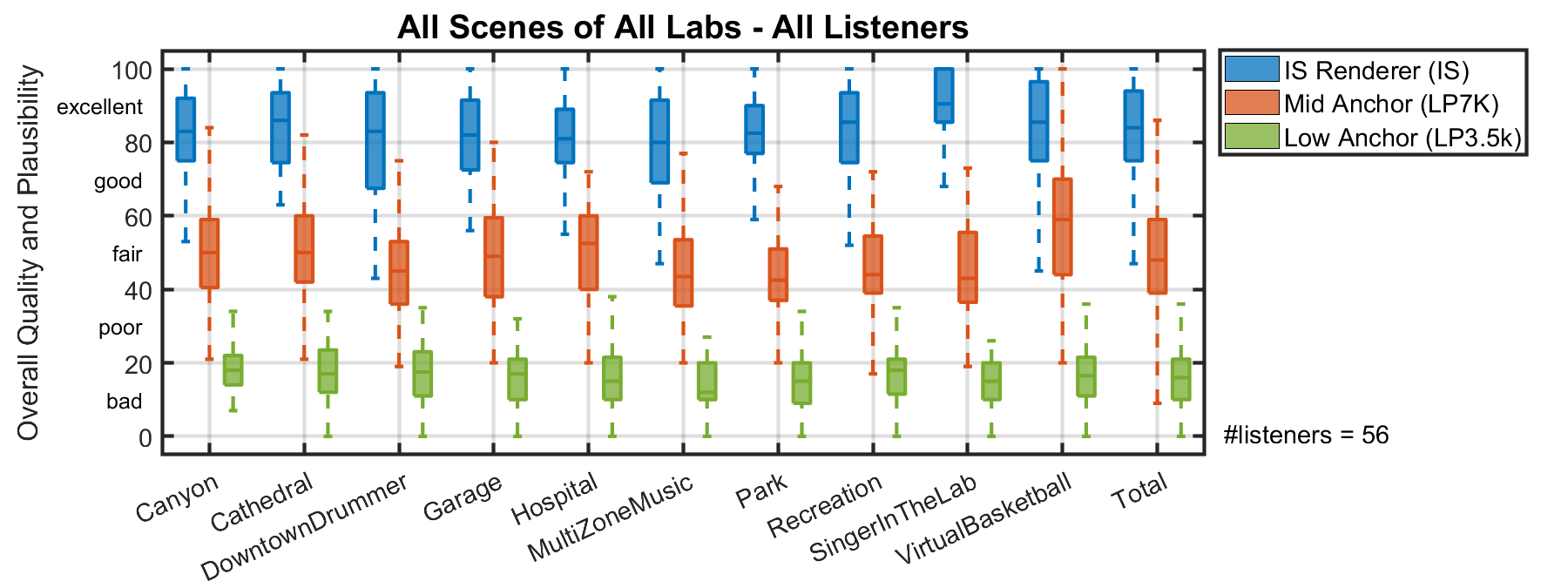


Figure 3: Scene results of all labs, N = 56.

Two different expert groups evaluated the MPEG-I immersive audio renderer:

1. The group of MPEG-I developers, who are very familiar with the algorithm and all the test scenes, see Figure 4.
2. Audio expert listeners from other fields like e.g. sound engineering, automotive audio tuning, game development and others, see Figure 5.

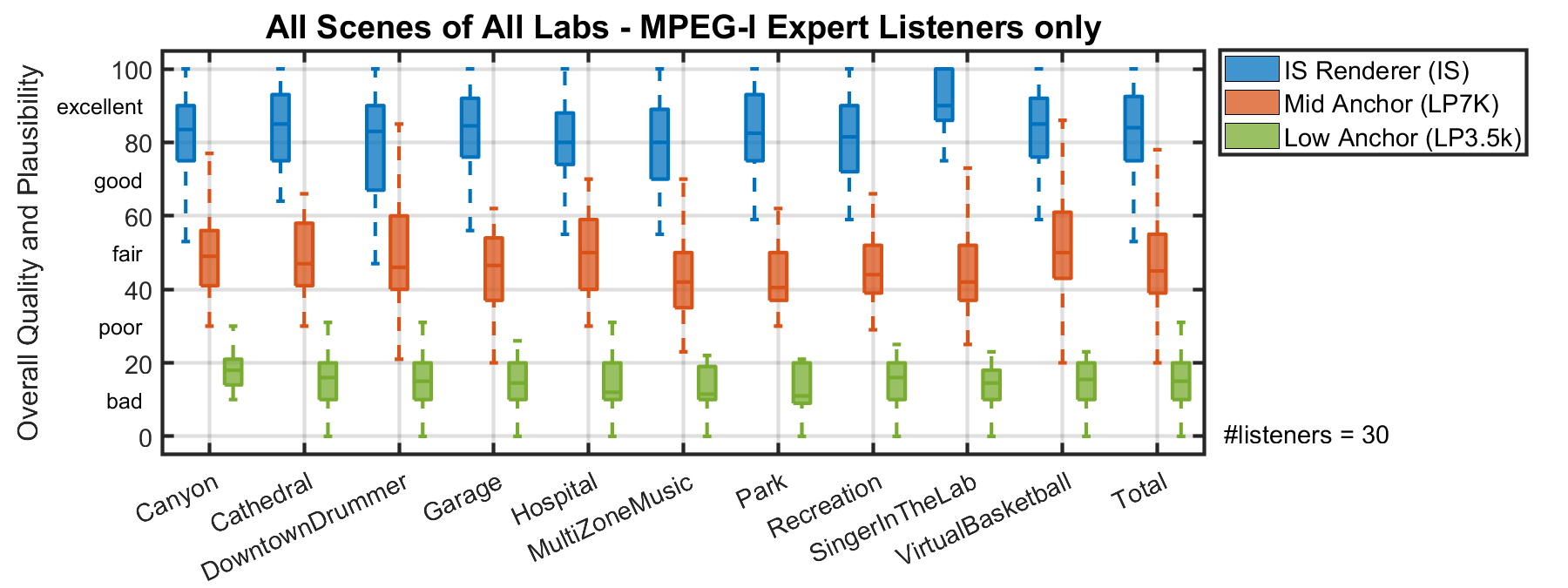


Figure 4: Scene results of all labs, MPEG-I experts only, N = 30.

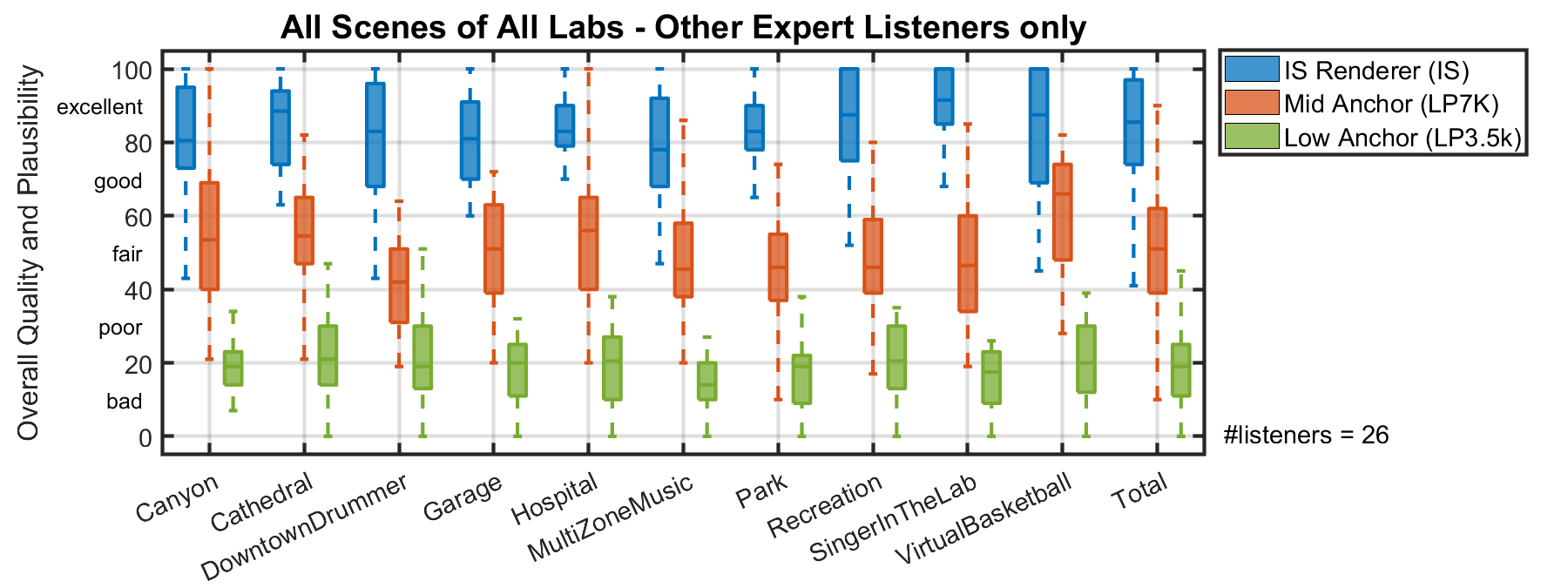


Figure 5: Scene results of all labs, other experts only, N = 26.

## Summary of Results

The listening test data show that:

* All selected scenes exhibit a clear and significant quality separation between the IS renderer, the mid- and the low-quality anchor.
* The listener’s repeatability shows quite good consistency for the two repeated scenes.
* The IS is rated with a median of 84 points among all labs, which is the “excellent” region of the scale.
* The group sizes of the MPEG-I experts and the other listening test experts were similar, 26 to 30. But the quartile range for all three conditions, per lab or per scene, and the repeatability is a bit larger for the non-MPEG-I experts. This repeated the experience of many former MPEG audio listening tests that the developers of the system are the more critical listeners.

# Conclusion

In the MPEG-I immersive audio verification test, the overall perceptual quality and plausibility of the standardized renderer have been assessed by expert listeners. Grounded on two anchor conditions, a lower anchor satisfying minimum requirements for reasonable audio reproduction, and a mid anchor simulating a basic 6DoF audio rendering using well-known stereo amplitude panning, listeners rated the MPEG-I immersive audio condition overall as “excellent”, confirming that the initial development goal of achieving high-quality interactive audio rendering has been accomplished.

# Literature

1. ITU-R, Rec. BS.2132-0, Method for the Subjective Quality Assessment of Audible Differences of Sound Systems Using Multiple Stimuli Without a Given Reference, 2019, Intern. Telecom Union, Geneva, Switzerland. Available from: <https://www.itu.int/rec/R-REC-BS.2132/en>.
2. ITU-R, Rec BS.1284-2, General Methods for the Subjective Assessment of Sound Quality, 2019. Intern. Telecom Union, Geneva, Switzerland. <https://www.itu.int/rec/R-REC-BS.1284/en>
3. Herre, J. and S. Disch, MPEG-I Immersive Audio – Reference Model for the New Virtual / Augmented Reality Audio Standard. J. Audio Eng. Soc., 2023. 71(5): p. 229..240 DOI: <https://doi.org/10.17743/jaes.2022.0074>
4. S. Disch, L. Terentiv, J. Koppens, T. Falk, J. Leppänen, and I. Munoz, “MPEG-I Immersive Audio – The Technology of the New Standard for Virtual / Augmented Reality Audio,” AES 159th AES Convention, Long Beach, USA. (2025). <https://aes2.org/publications/elibrary-page/?id=23078>