 ISO/IEC JTC 1/SC 29/AG 05 N0130

**ISO/IEC JTC 1/SC 29/AG 05  
MPEG Visual Quality Assessment   
Convenorship: DE**

**Document type:** Output Document

**Title:** Report on interactive subjective quality assessment test investigation for volumetric video

**Status:** Approved

**Date of document:** 2024-10-20

**Source:** ISO/IEC JTC 1/SC 29/AG 05

**Expected action:** Information

**Action due date:** None

**No. of pages:** 4 (without cover page)

**Email of Convenor:** wien@lfb.rwth-aachen.de

**Committee URL:** <https://isotc.iso.org/livelink/livelink/open/jtc1sc29ag5>

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/AG 05**

**MPEG VISUAL QUALITY ASSESSMENT**

**ISO/IEC JTC 1/SC 29/AG 05 N** **0130**

**July 2024, Sapporo, JP**

|  |  |
| --- | --- |
| **Title** | **Report on interactive subjective quality assessment test investigation for volumetric video** |
| **Source** | AG 05 MPEG Visual Quality Assessment |
| **Status** | Approved |
| **Serial Number** | 24263 |
| **Editors** | Siyuan Zhu, Jiaqi Zhang, Lu Yu |

# Introduction

This document presents the investigation conducted in AG 5 on interactive subjective visual quality assessment test of volumetric video. The subjective test results demonstrate the influence of viewing methods, interactive or passive.

# Test Setup

## Test Contents

Five sequences provided by the common test conditions of MPEG immersive video [1] were selected as the source contents, one of which was used for subject training (W02-*Dancing*) and the others for testing. The characteristics of the contents are listed in Table 1. In particular, all the videos were played at a frame rate of 15fps (slower than its original frame rate), which allows users to have more time to explore the scenes when viewing interactively.

**Table. 1 The characteristics of source contents**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sequence ID** | **Name** | **Length** | **frame rate** | **Type** | **FOV** |
| B01 | Museum | 300 frames | 30 fps | Computer generated | Omnidirectional |
| C01 | Hijack | 300 frames | 30 fps | Computer generated | Semi-omnidirectional |
| D01 | Painter | 300 frames | 30 fps | Natural  content | Semi-omnidirectional |
| E01 | Frog | 300 frames | 30 fps | Natural  content | Semi-omnidirectional |
| W02 | Dancing | 300 frames | 30 fps | Computer generated | Semi-omnidirectional |

The reference software TMIV v11.0 with MIV main anchor was employed to encode four distorted versions with different QPs for each source content. The QPs settings follow MIV Common test conditions. In the configuration files, some parameters were changed. *“attributeOffsetEnabledFlag”* is *true*, *“blockSizeDepthQualityDependent”* is *[16, 16]*, *“maxAtlases”* is *1* and suitable *“maxLumaPictureSize”* were chosen for different source contents for the purpose of real-time rendering. A total of 16 compressed contents were obtained.

## Test Device

A professional 2D monitor Eizo ColorEdge CG303W with a size of 29.8 inches and a resolution of 2560×1600 was employed to display the immersive video. The experiment was conducted on a computer equipped with an Intel i9-14900KF CPU, a NVIDA GeForce RTX 4080 GPU and 64GB of memory. The contents were played through a MIV real-time player developed for interactive and passive subjective quality assessment, which is capable achieving a real-time rendering frame rate of 30+ fps with GPU.

## Test Method

### Passive test method

The subjects passively watch the synthesized video in the passive test. It is just like a normal subjective test on 2D video quality, in which the same subjective quality assessment method as that of the verification test of MIV is used [2]. Synthesized videos of predefined pose traces were generated before the test, so that each subject watch the same video for each test point. The pose traces used in this experiment are predefined in the common test conditions of MIV [1].

### Interactive test method

The subjects interactively control the pose trace, i. e. the position and orientation, of video watching by using a keyboard in the interactive test. The viewport shown to the subject is synthesized in real time according to the controlled pose trace. The controlled pose trace was always restricted in viewing space defined in advance per content to avoid significant quality degradation of synthesized viewport. When the position and orientation of a desired viewport is moved across the boundary of viewing space, the pose will be frozen and an alarm sound will be trigged in the meantime. One viewing space illustration video would be shown to the subjects before the interactive test to provide a rough idea about the viewing space of the tested scene.

### Test protocol and procedure

The Absolute Category Rating with Hidden Reference (ACR-HR) method with a 5-grade quality scale was applied for the subjective assessment [3]:

1. ─ excellent
2. ─ good
3. ─ fair
4. ─ poor
5. ─ bad

At the beginning of training phase, the experimental objectives, experimental procedure, as well as assessment protocol were introduced to the subjects. They were then allowed to watch the training contents until they could skillfully manipulate their movement through the keyboard.

The testing phase was divided into two independent groups, including passive test group and interactive test group. In each group, the subjects rated all the test contents in random order.

In passive test, each basic test cell (BTC) is comprised of:

**“A” (1sec) – [PVS] (20sec) – “Vote” (5sec),**

where A represents for the mid-grey video, and PVS denotes the tested video content under assessment.

In interactive test, each BTC is comprised of:

**“[viewing space illustration video]” (5sec) – “A” (1sec) – [PVS] (20sec) – “Vote” (5sec).**

A total of 17 subjects participated in the experiment, including 14 males and 3 females, aged between 22 and 30.

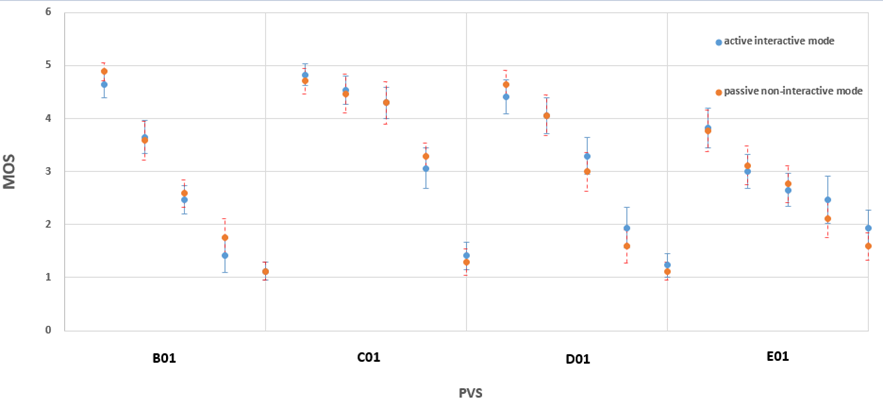
# Experimental Results

The screening method with Pearson analysis recommended by ITU-T P.910 [3] is used. Finally, none of the subjects were excluded.

The MOS values and 95% confidence intervals (CI) of each PVSs is shown in Fig. 1. The test results show that the MOS values and 95% confidence intervals of each PVS of passive test and interactive test are quite close with each other. We believe the major reasons are:

1. QP settings of all source contents are well configured by MIV CTC, so that subjective qualities of a same source content with one QP could almost be isolated with the others. The dominant factor of video quality in this experiment is the QP settings.
2. There is no spatially nonuniform coding conducted on test contents.
3. The viewing spaces of all the selected source contents are relatively small. Impact of the viewing pose trace on visual quality is minor, at least for the tested contents of this experiment.

These reasons can explain the similarity of the MOS values achieved with passive test and interactive test.



**Fig. 1 The MOS values and 95% confidence intervals**

# Conclusion

Passive as well as interactive tests have been conducted in this investigation on 4 coded immersive video contents. The experiment shows no significant difference in visual quality for the same coded content as watched passively or interactively on a 2D screen. So passive test could be sufficient for evaluation of subjective quality of immersive video content coding, especially when there is no spatially nonuniform coding applied.

We can still imagine that the interactive viewing may play a role in assessment of quality of experience, in which it is not limited to visual only experiences.

# Reference

1. “Common test conditions for MPEG immersive video”, ISO/IEC JTC 1/SC 29/WG 04, N0342, April 2023, Antalya.
2. “Verification test report of MPEG immersive video”, ISO/IEC JTC 1/SC 29/WG 04, N0341, Apr. 2023, Antalya.
3. Recommendation ITU-T P.910 (2023), Subjective video quality assessment methods for multimedia applications.