MPEG Visual Quality Assessment: Tasks and Perspectives

Mathias Wien, Convenor of ISO/IEC JTC1 SC19 Advisory Group 5 MPEG VQA
wien@lfb.rwth-aachen.de

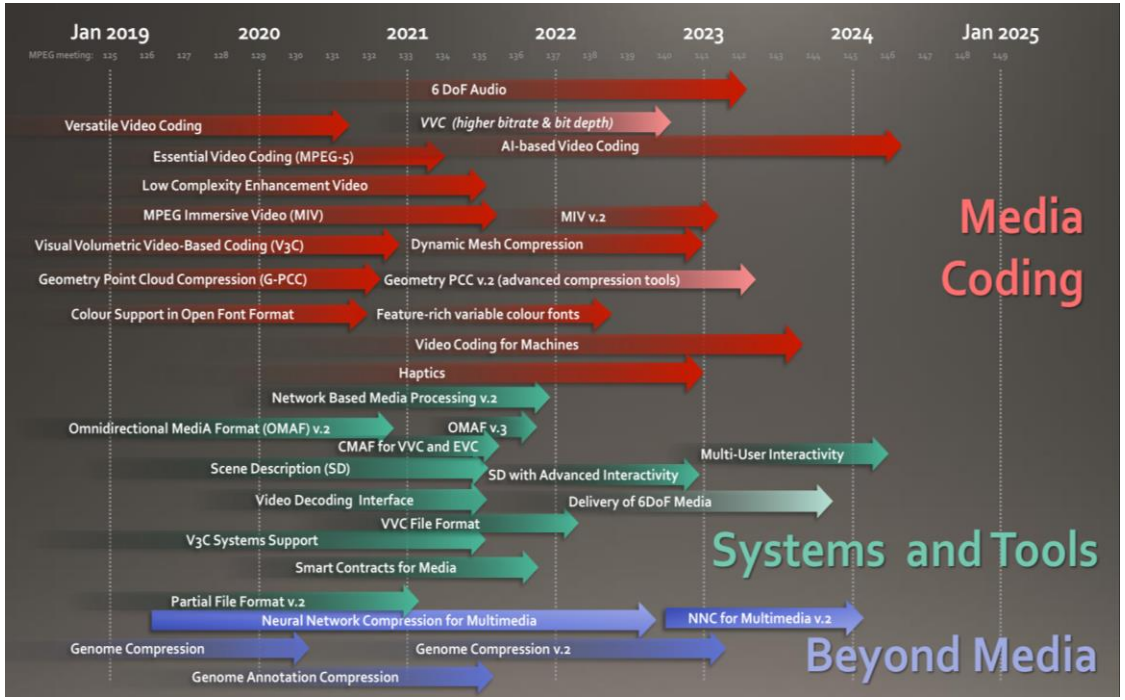
Lehrstuhl für Bildverarbeitung
RWTH Aachen University

Outline

- MPEG
  - New structure in ISO/IEC JTC1 SC29
- Organization of AG5
  - Focus groups
  - Ad-hoc groups
- First achievements and perspectives
  - Verification tests
  - Remote experts viewing
  - Outreach
- Outlook and summary

ISO/IEC SC29 MPEG – The new structure

- Reorganization of SC29 in July 2020
  - Converted from WG11 into 7 WGs and 3 AGs
  - Continuous update of MPEG work plan



SC 29

WG 1	JPEG Coding of digital representations of images
WG 2	MPEG Technical Requirements
WG 3	MPEG Systems
WG 4	MPEG Video Coding
WG 5	MPEG Joint Video Coding Team(s) with ITU-T SG 16
WG 6	MPEG Audio Coding
WG 7	MPEG Coding of 3D graphics
WG 8	MPEG Genomic Coding
AG 1	Chair Support Team and Management (ex. AGM)
AG 2	MPEG Technical Coordination
AG 3	MPEG Liaison and Communication
AG 4	JPEG and MPEG Collaboration
AG 5	MPEG Visual Quality Assessment

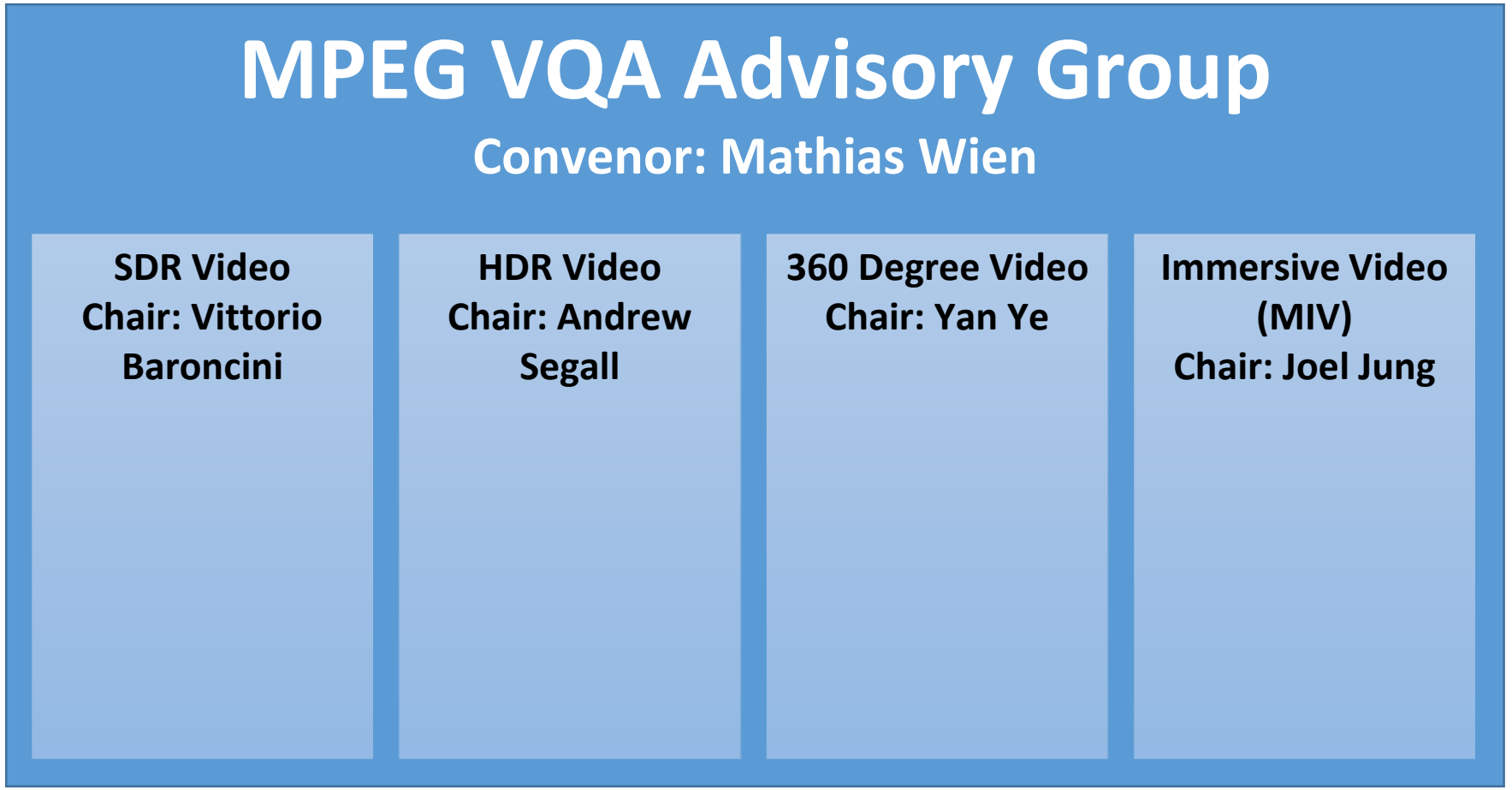
ISO: International Standardization Organization | IEC: International Electrotechnical Commission | JTC1: Joint Technical Committee | MPEG: Moving Pictures Experts Group
SC29: Sub-committee 29 “Coding of Audio, Picture, Multimedia and Hypermedia Information” | AG: Advisory Group | WG: Working Group | CfE: Call for Evidence | CfP: Call for Proposals

MPEG AG5 | Terms of Reference (per SC29 N190201)

- To assess the visual quality of new technologies to be considered to begin a new standardization project;
- To contribute to the definition of Calls for Proposals (CfPs) for new standardization work items;
- To select and design subjective quality evaluation methodologies and objective quality metrics for the assessment of visual coding technologies, e.g. in the context of a Call for Evidence (CfE) and CfP;
- To contribute to the selection of test material and coding conditions for a CfP;
- To define the procedures useful to assess the visual quality of the submissions to a CfP;
- To design and conduct visual quality tests, process and analyse the raw data, and make the report of the evaluation results available conclusively;
- To support in the assessment of the final status of a standard, verifying its performance compared to the existing standard(s);
- To maintain databases of test material;
- To recommend guidelines for selection of testing laboratories (verifying their current capabilities);
- To liaise with ITU and other relevant organizations on the creation of new Quality Assessment standards or the improvement of the existing ones.

MPEG AG5 | Structure

- Focus groups on dedicated areas
- Ad-hoc groups on defined topics



MPEG AG5 | Ad-Hoc Groups and Activities

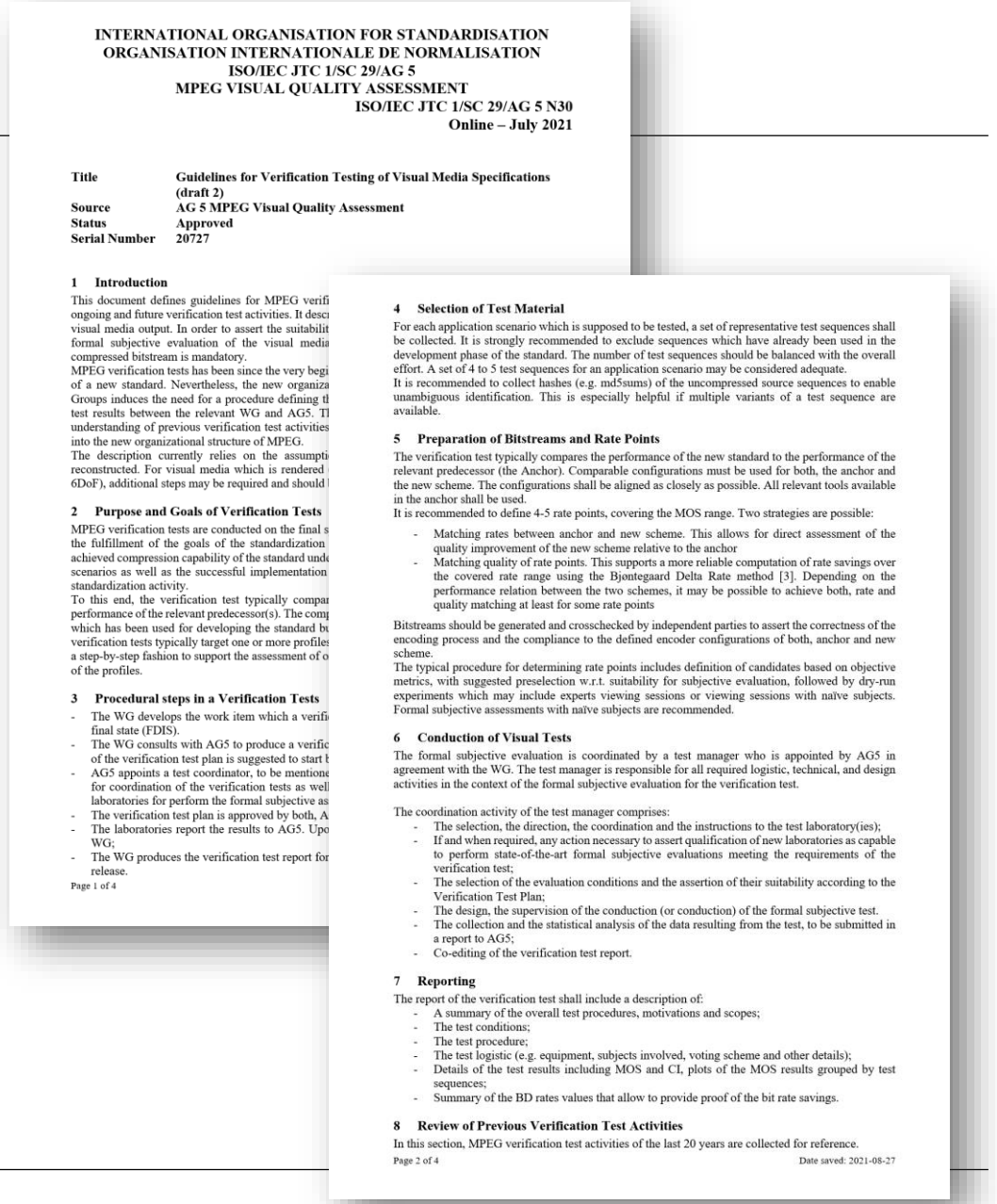
- **AhG on Quality of Immersive Visual Media**
  - Update overview on quality metrics and methodologies for immersive visual media
  - Inputs for subjective evaluation methods and objective metrics
  - Preparation of workshop on 2021-10-05

- **AhG on Learning-based quality metrics for 2D video**
  - Compilation of databases, check for suitability and availability
  - Learning-based quality metrics for 2D video
  - Study correlation between metrics and subjective scores

- **AhG on Guidelines for subjective visual quality evaluation**
  - Guidelines for verification tests
  - Guidelines for remote experts viewing

MPEG AG5 | Guidelines for Verification Tests

- Definition of purpose and goals of verification tests
- Definition of procedures
- Selection of test material
- Preparation of bitstreams and rate points
- Conduction of visual tests
- Reporting



INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/AG 5
MPEG VISUAL QUALITY ASSESSMENT
ISO/IEC JTC 1/SC 29/AG 5 N30
Online – July 2021

Title Guidelines for Verification Testing of Visual Media Specifications (draft 2)
Source AG 5 MPEG Visual Quality Assessment
Status Approved
Serial Number 20727

1 Introduction

This document defines guidelines for MPEG verifi
ongoing and future verification test activities. It descr
visual media output. In order to assert the suitabilit
formal subjective evaluation of the visual media
compressed bitstream is mandatory.
MPEG verification tests has been since the very begi
of a new standard. Nevertheless, the new organiza
Groups induces the need for a procedure defining th
test results between the relevant WG and AG5. Th
understanding of previous verification test activities
into the new organizational structure of MPEG.
The description currently relies on the assumpti
reconstructed. For visual media which is rendered
6DoF), additional steps may be required and should

2 Purpose and Goals of Verification Tests

MPEG verification tests are conducted on the final s
the fulfillment of the goals of the standardization
achieved compression capability of the standard unde
scenarios as well as the successful implementation
standardization activity.
To this end, the verification test typically compar
performance of the relevant predecessor(s). The comp
which has been used for developing the standard bu
verification tests typically target one or more profiles
a step-by-step fashion to support the assessment of o
of the profiles.

3 Procedural steps in a Verification Tests

- The WG develops the work item which a verifi
final state (FDIS).
- The WG consults with AG5 to produce a verific
of the verification test plan is suggested to start b
- AG5 appoints a test coordinator, to be mentione
for coordination of the verification tests as well
laboratories for perform the formal subjective as
- The verification test plan is approved by both, A
- The laboratories report the results to AG5. Upo
WG;
- The WG produces the verification test report for
release.

Page 1 of 4

4 Selection of Test Material

For each application scenario which is supposed to be tested, a set of representative test sequences shall be collected. It is strongly recommended to exclude sequences which have already been used in the development phase of the standard. The number of test sequences should be balanced with the overall effort. A set of 4 to 5 test sequences for an application scenario may be considered adequate.
It is recommended to collect hashes (e.g. md5sums) of the uncompressed source sequences to enable unambiguous identification. This is especially helpful if multiple variants of a test sequence are available.

5 Preparation of Bitstreams and Rate Points

The verification test typically compares the performance of the new standard to the performance of the relevant predecessor (the Anchor). Comparable configurations must be used for both, the anchor and the new scheme. The configurations shall be aligned as closely as possible. All relevant tools available in the anchor shall be used.
It is recommended to define 4-5 rate points, covering the MOS range. Two strategies are possible:

- Matching rates between anchor and new scheme. This allows for direct assessment of the quality improvement of the new scheme relative to the anchor
- Matching quality of rate points. This supports a more reliable computation of rate savings over the covered rate range using the Bjøntegaard Delta Rate method [3]. Depending on the performance relation between the two schemes, it may be possible to achieve both, rate and quality matching at least for some rate points

Bitstreams should be generated and crosschecked by independent parties to assert the correctness of the encoding process and the compliance to the defined encoder configurations of both, anchor and new scheme.
The typical procedure for determining rate points includes definition of candidates based on objective metrics, with suggested preselection w.r.t. suitability for subjective evaluation, followed by dry-run experiments which may include experts viewing sessions or viewing sessions with naïve subjects. Formal subjective assessments with naïve subjects are recommended.

6 Conduction of Visual Tests

The formal subjective evaluation is coordinated by a test manager who is appointed by AG5 in agreement with the WG. The test manager is responsible for all required logistic, technical, and design activities in the context of the formal subjective evaluation for the verification test.

The coordination activity of the test manager comprises:

- The selection, the direction, the coordination and the instructions to the test laboratory(ies);
- If and when required, any action necessary to assert qualification of new laboratories as capable to perform state-of-the-art formal subjective evaluations meeting the requirements of the verification test;
- The selection of the evaluation conditions and the assertion of their suitability according to the Verification Test Plan;
- The design, the supervision of the conduction (or conduction) of the formal subjective test.
- The collection and the statistical analysis of the data resulting from the test, to be submitted in a report to AG5;
- Co-editing of the verification test report.

7 Reporting

The report of the verification test shall include a description of:

- A summary of the overall test procedures, motivations and scopes;
- The test conditions;
- The test procedure;
- The test logistic (e.g. equipment, subjects involved, voting scheme and other details);
- Details of the test results including MOS and CI, plots of the MOS results grouped by test sequences;
- Summary of the BD rates values that allow to provide proof of the bit rate savings.

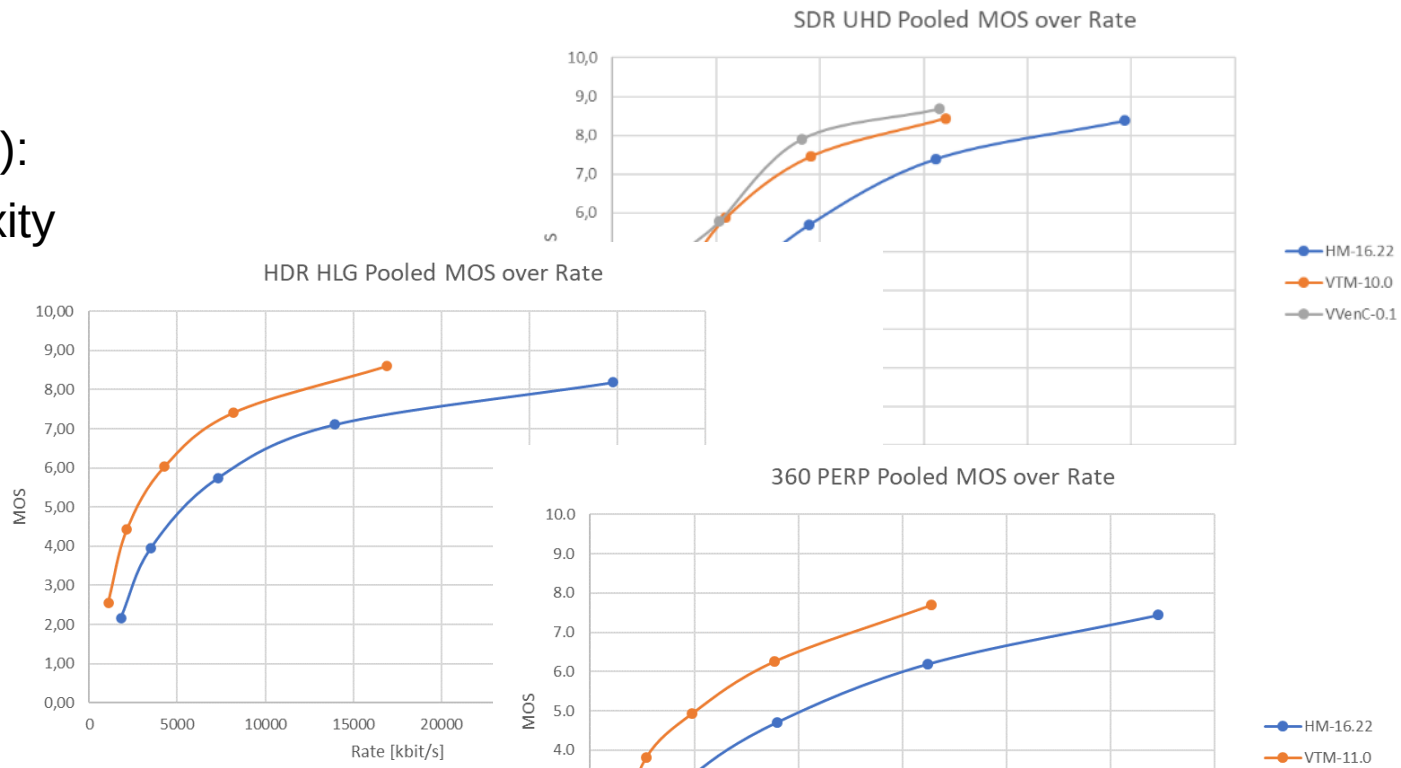
8 Review of Previous Verification Test Activities

In this section, MPEG verification test activities of the last 20 years are collected for reference.

Page 2 of 4 Date saved: 2021-08-27

MPEG AG5 | Verification Tests

- Completed verification test activities
  - *MPEG-5 Part 1*: Essential Video Coding (EVC): SDR and HDR. *MPEG-5 Part 2*: Low-Complexity Enhancement Video Coding (LCEVC): AVC, EVC, HEVC, VVC + enhancement layer
  - *MPEG-I Part 3 / ITU-T H.266* Versatile Video Coding (VVC): SDR, HDR, 360° video

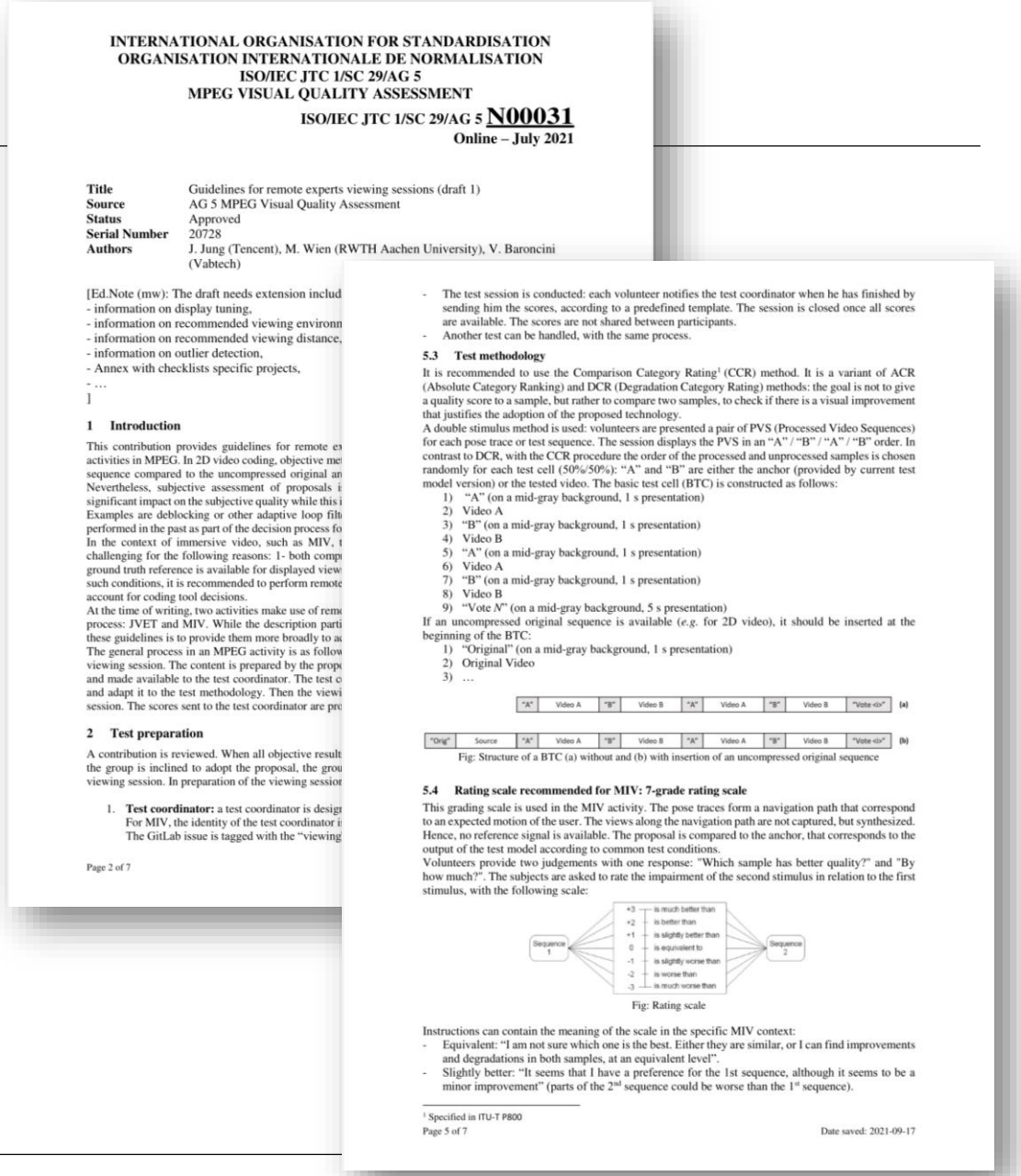


- Emerging verification test activities
  - MPEG-I Part 4: Video-based Point Cloud Compression (V-PCC)
  - MPEG Immersive Video (MIV)
  - VVC: Scalability, Screen Content, ...

[JVET-T2020] Mathias Wien and Vittorio Baroncini, “VVC Verification Test Report for Ultra High Definition (UHD) Standard Dynamic Range (SDR) Video Content”, JVET-T2020, ITU-T/ISO/IEC JVET, 20th meeting: Oct.`20.
[JVET-V2020] Mathias Wien and Vittorio Baroncini, “VVC Verification Test Report for High Definition (HD) and 360 Standard Dynamic Range (SDR) Video Content”, JVET-V2020, ITU-T/ISO/IEC JVET, 22nd meeting: Apr.`21.
[JVET-W2020] Mathias Wien and Vittorio Baroncini, “VVC Verification Test Report For High Dynamic Range Video Content”, JVET-W2020 of ITU-T/ISO/IEC JVET, 23rd meeting: Jul.`21.

MPEG AG5 | Remote Experts Viewing

- **Guidelines for remote experts viewing (REV)**
  - Remote testing in focus due to pandemic situation
  - Adaptation of experts viewing methods commonly used at meetings
    - Currently A-B comparisons, e.g. test model against variant with proposed change
      - 4-grade scale
      - 7-grade scale
    - Needed for decisions e.g. when objective metrics are difficult or not available
- Remote method applied in context of Joint Video Experts Team (JVET), MPEG Immersive Video (MIV), Video-based Point Cloud Coding (V-PCC)



INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/AG 5
MPEG VISUAL QUALITY ASSESSMENT

ISO/IEC JTC 1/SC 29/AG 5 N00031
Online – July 2021

Title Guidelines for remote experts viewing sessions (draft 1)
Source AG 5 MPEG Visual Quality Assessment
Status Approved
Serial Number 20728
Authors J. Jung (Tencent), M. Wien (RWTH Aachen University), V. Baroncini (Vabtech)

[Ed.Note (mw): The draft needs extension includ
- information on display tuning,
- information on recommended viewing environn
- information on recommended viewing distance,
- information on outlier detection,
- Annex with checklists specific projects,
- …
]

1 Introduction

This contribution provides guidelines for remote e activities in MPEG. In 2D video coding, objective me sequence compared to the uncompressed original an Nevertheless, subjective assessment of proposals i significant impact on the subjective quality while this i Examples are deblocking or other adaptive loop filt performed in the past as part of the decision process fo In the context of immersive video, such as MIV, t challenging for the following reasons: 1- both compu ground truth reference is available for displayed view such conditions, it is recommended to perform remote account for coding tool decisions.
At the time of writing, two activities make use of remo process: JVET and MIV. While the description parti these guidelines is to provide them more broadly to a The general process in an MPEG activity is as follow viewing session. The content is prepared by the propo and made available to the test coordinator. The test c and adapt it to the test methodology. Then the viewi session. The scores sent to the test coordinator are pr

2 Test preparation

A contribution is reviewed. When all objective result the group is inclined to adopt the proposal, the grou viewing session. In preparation of the viewing sessio

1. **Test coordinator:** a test coordinator is desig For MIV, the identity of the test coordinator i The GitLab issue is tagged with the "viewing

Page 2 of 7

- The test session is conducted: each volunteer notifies the test coordinator when he has finished by sending him the scores, according to a predefined template. The session is closed once all scores are available. The scores are not shared between participants.
- Another test can be handled, with the same process.

5.3 Test methodology

It is recommended to use the Comparison Category Rating[1] (CCR) method. It is a variant of ACR (Absolute Category Ranking) and DCR (Degradation Category Rating) methods: the goal is not to give a quality score to a sample, but rather to compare two samples, to check if there is a visual improvement that justifies the adoption of the proposed technology.
A double stimulus method is used: volunteers are presented a pair of PVS (Processed Video Sequences) for each pose trace or test sequence. The session displays the PVS in an "A" / "B" / "A" / "B" order. In contrast to DCR, with the CCR procedure the order of the processed and unprocessed samples is chosen randomly for each test cell (50%/50%): "A" and "B" are either the anchor (provided by current test model version) or the tested video. The basic test cell (BTC) is constructed as follows:

1) "A" (on a mid-gray background, 1 s presentation)
2) Video A
3) "B" (on a mid-gray background, 1 s presentation)
4) Video B
5) "A" (on a mid-gray background, 1 s presentation)
6) Video A
7) "B" (on a mid-gray background, 1 s presentation)
8) Video B
9) "Vote N" (on a mid-gray background, 5 s presentation)

If an uncompressed original sequence is available (*e.g.* for 2D video), it should be inserted at the beginning of the BTC:

1) "Original" (on a mid-gray background, 1 s presentation)
2) Original Video
3) …

Fig: Structure of a BTC (a) without and (b) with insertion of an uncompressed original sequence

5.4 Rating scale recommended for MIV: 7-grade rating scale

This grading scale is used in the MIV activity. The pose traces form a navigation path that correspond to an expected motion of the user. The views along the navigation path are not captured, but synthesized. Hence, no reference signal is available. The proposal is compared to the anchor, that corresponds to the output of the test model according to common test conditions.
Volunteers provide two judgements with one response: "Which sample has better quality?" and "By how much?". The subjects are asked to rate the impairment of the second stimulus in relation to the first stimulus, with the following scale:

Fig: Rating scale

Instructions can contain the meaning of the scale in the specific MIV context:
- Equivalent: "I am not sure which one is the best. Either they are similar, or I can find improvements and degradations in both samples, at an equivalent level".
- Slightly better: "It seems that I have a preference for the 1st sequence, although it seems to be a minor improvement" (parts of the 2nd sequence could be worse than the 1st sequence).

[1] Specified in ITU-T P800

Page 5 of 7 Date saved: 2021-09-17

MPEG AG5 | Perspectives

- **Communication**
  - Liaison and collaboration with other organizations
  - Organization of workshops
  - Output documents public by default
- **Next steps**
  - Finalization of guidelines for verification tests
    - Under review in AG5
    - Iteration towards release at October MPEG meeting
  - Further development and study of guidelines for remote experts viewing
    - Refinement and extension of procedures and methods
    - Study REV results
  - Drafting of new guidelines for (immersive) visual media
    - Study of available and emerging metrics
    - Study of congruence of objective and subjective assessment

Summary and Conclusions

- **ISO/IEC JTC1 SC29/AG5 MPEG Visual Quality Assessment**
  - New structure established
    - Focus groups
    - Ad-hoc groups
  - Verification test activities and core / exploration experiment contributions with MPEG Video, JVET, and 3D Graphics
  - Guideline documents under development
    - Verification tests
    - Remote experts viewing
  - Study towards further guidelines
    - Quality metrics for immersive visual media
    - Learned quality metrics
  - Outreach to other organizations and to public