Sequence and rate-point selection for a CfE on video compression with capability beyond VVC

Mathias Wien

Convenor of ISO/IEC JTC 1/SC 29/AG 5
MPEG Visual Quality Assessment

VQEG ETG 116 | May 2025

Development towards a Call for Evidence (CfE) in JVET

- **ITU and ISO/IEC JTC 1 SC 29 Joint Workshop** on "Future video coding – advanced signal processing, AI and standards", Geneva, Switzerland, 2025-01-17
  - Review of current status of development w.r.t. conventional tools and learning-based tools
  - Industry view on new standardization project

- **Joint discussion of JVET parent bodies on requirements**
  - Use cases, technical requirements, performance expectations
  - Encoder / decoder complexity as important topic
  - Definition of requirements as prerequisite of new standardization project

- **Preparations towards draft Call for Evidence**
  - Reflecting requirements defined by parent bodies (ISO/IEC JTC 1/SC 29 and ITU-T SG21)
  - Selection of test sequences

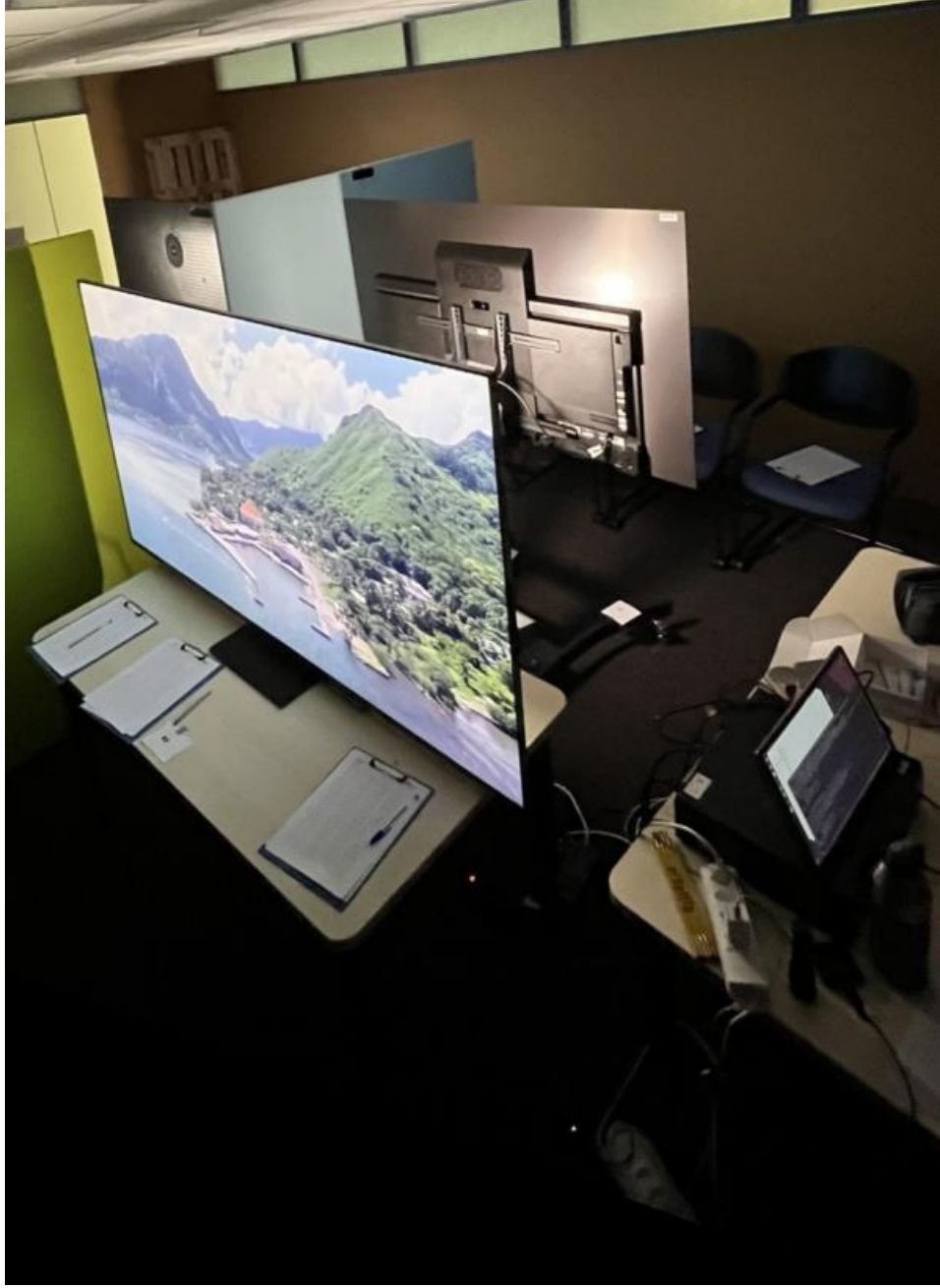
Content categories

- Basic dimensions
  - SDR and HDR content
  - HD and UHD resolutions
  - Streaming and low-delay applications (random access, RA and low-delay, LD)
  - Consideration of portrait / landscape orientation
- New categories
  - User Generated Content
  - Gaming content

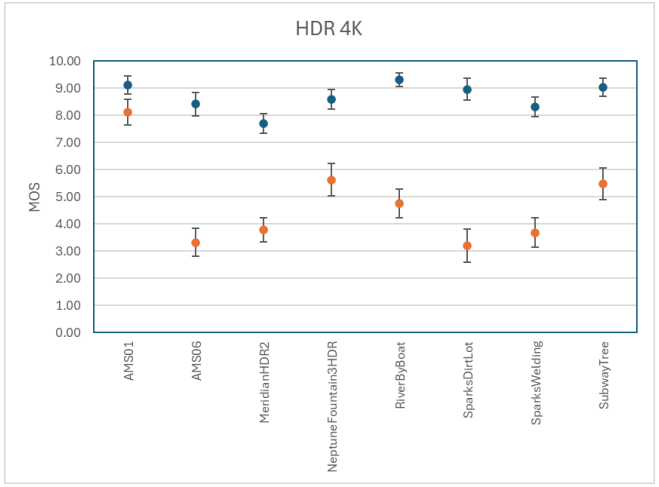
- CfE categories under consideration
  - SDR_RA_UHD
  - SDR_RA_HD
  - SDR_LD_HD
  - HDR_RA_UHD
  - HDR_RA_8Kcrop
  - Gaming_LD_HD
  - UGC_RA_HD

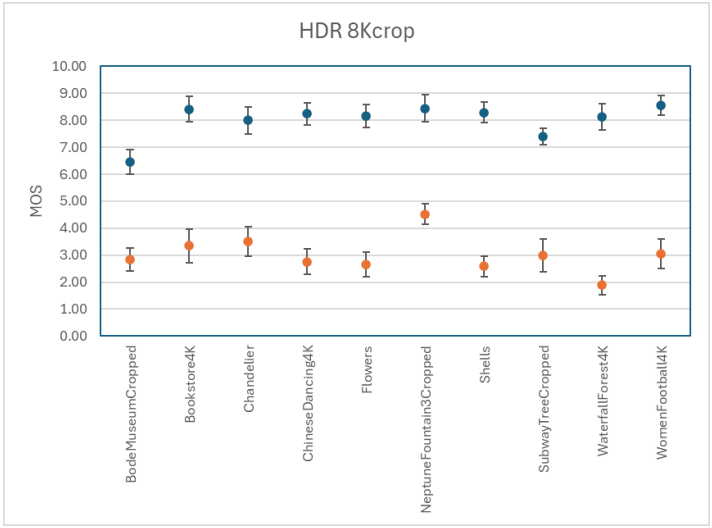
Step 1: On-site expert viewing at JVET January meeting

- **49 test sequences**, evaluated at low and high rate points
- ACR test with 11-grade scale
- Three test sessions
  1. UHD HDR: 42 scores
  2. UHD SDR: 24 scores
     (shown right after UHD HDR)
  3. HD SDR: 51 scores
- 24 experts participating in viewing



Outcome of on-site effort





- Example results: UHD HDR
- Review of available candidate sequence
- Ranking of sequences in candidate set in 3 groups
  - Discarding of some sequences
  - Identification of 35 candidates to be further studied

Class	Sequence	Group
GAMING	DOTA2s360	2
GAMING	GTAVs090	2
GAMING	Minecraft	2
GAMING	Darktree	1
GAMING	EuroTruckSimulator2	1
GAMING	JianlingTemple	1
HD RA	Metro	2
HD RA	DrivingPOV4	1
~~HD RA~~	~~BarScene~~	~~0~~
~~HD RA~~	~~Meridian2~~	~~0~~
LD	Beatriz	2
LD	OfficeWalkAtWall	1
~~LD~~	~~OfficeWalkCeiling~~	~~0~~
UGC	Animal-6v9a-pheasant	2
UGC	Sports-76a2-hockey	2
UGC	VerticalVideo-3709-snow	2
UGC	VerticalVideo-4b92-park	2
UGC	CoverSong-3499-outdoor	1
UGC	Hobby-w5xz-backpack	1
UGC	VerticalVideo-3d96-walk	1
~~UGC~~	~~Cooking-h6qi~~	~~0~~
~~UGC~~	~~Hobby-1fc7-park~~	~~0~~

Class	Sequence	Group
8K-CROP	WaterfallForest4K	2
8K-CROP	WomenFootball4K	2
8K-CROP	Chandelier	1
8K-CROP	ChineseDancing4K	1
8K-CROP	Shells	1
8K-CROP	SubwayTreeCropped	1
~~8K-CROP~~	~~BodeMuseumCropped~~	~~0~~
~~8K-CROP~~	~~Bookstore4K~~	~~0~~
~~8K-CROP~~	~~Flowers~~	~~0~~
~~8K-CROP~~	~~NeptuneFountain3Cropped~~	~~0~~
UHD-HDR	AMS06	2
UHD-HDR	SparksWelding	2
UHD-HDR	AMS01	1
UHD-HDR	MeridianHDR2	1
UHD-HDR	NeptuneFountain3HDR	1
UHD-HDR	SparksDirtLot	1
UHD-HDR	SubwayTree	1
~~UHD-HDR~~	~~RiverByBoat~~	~~0~~
UHD-SDR	CrowdRun	2
UHD-SDR	DrivingPOV3	2
UHD-SDR	NeptuneFountain3	2
UHD-SDR	Marathon2	1
UHD-SDR	MountainBay2	1
UHD-SDR	TallBuildings2	1
~~UHD-SDR~~	~~BaoleiYard~~	~~0~~
~~UHD-SDR~~	~~ToddlerFountain2~~	~~0~~
~~UHD-SDR~~	~~TrafficFlow~~	~~0~~

Step 2: On-site expert viewing at AHG meeting in March

- Input of 28 additional test sequences, partially new content
  - Repetition of ACR test at low and high rate points for these sequences

- Determination of suitable rate points
  - Goal: meaningful coverage of quality range, 4 rate points for evaluation
  - Leave headroom for proposals at highest rate point
  - Anchor may perform quite bad at lowest rate point
  - Quality at the four rate points should be clearly distinguishable
  - Evaluation with VTM (will be the anchor) and ECM (as an example for enhanced compression scheme)

Step 2: Evaluation tasks

- “Succession tests” for proposed VTM and ECM bitstreams
  - “Relative DCR” approach, assess visible delta between neighboring rate points
    - Original → QP1 → QP2 → QP3 → QP4 (→ QP5) [*additional highest point for VTM*]
    - Scoring by expert viewers using a 3-value scale
      - +1: too close to the previous rate point
      - 0 : ok
      - -1 : too large distance from previous rate point
    - Evaluation:
      - *MOS*: Mean value and *Rdiv*: relative number of cases different from 0
      - CM = (*Rdiv* - |*MOS*|)/*Rdiv*: Indicator for viewer confidence (small value indicates higher confident)
  - Assessment of results and suggestion of updates for QP selection

Step 2: Succession test results

- Examples from the tests

sequence	QP	MOS	Rdiv	aggreg.MOS	sum(abs(MOS))	sum(Rdiv)
Seeking	QP1	0.12	0.50	0.12	0.88	1.15
Seeking	QP2	-0.24	0.00	-0.12		
Seeking	QP3	-0.18	0.40	-0.29		
Seeking	QP4	-0.35	0.25	-0.65		
Seeking	QP5	0.00	0.00	-0.65		

sequence	QP	MOS	Rdiv	aggreg.MOS	sum(abs(MOS))	sum(Rdiv)
Camellia	QP1	0.59	0.17	0.59	1.35	0.17
Camellia	QP2	0.47	0.00	1.06		
Camellia	QP3	0.18	0.00	1.24		
Camellia	QP4	-0.12	0.00	1.12		
Camellia	QP5	0.00	0.00	1.12		

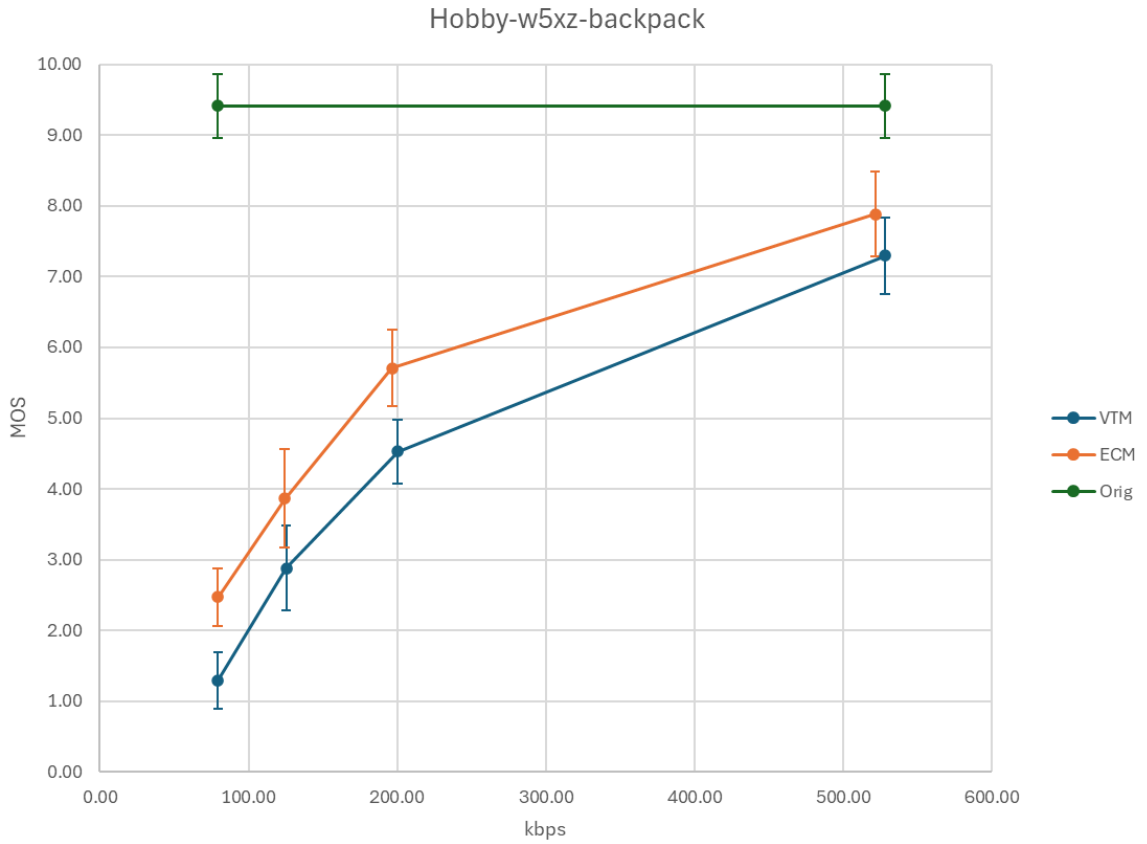
Step 2: Consolidation of test set

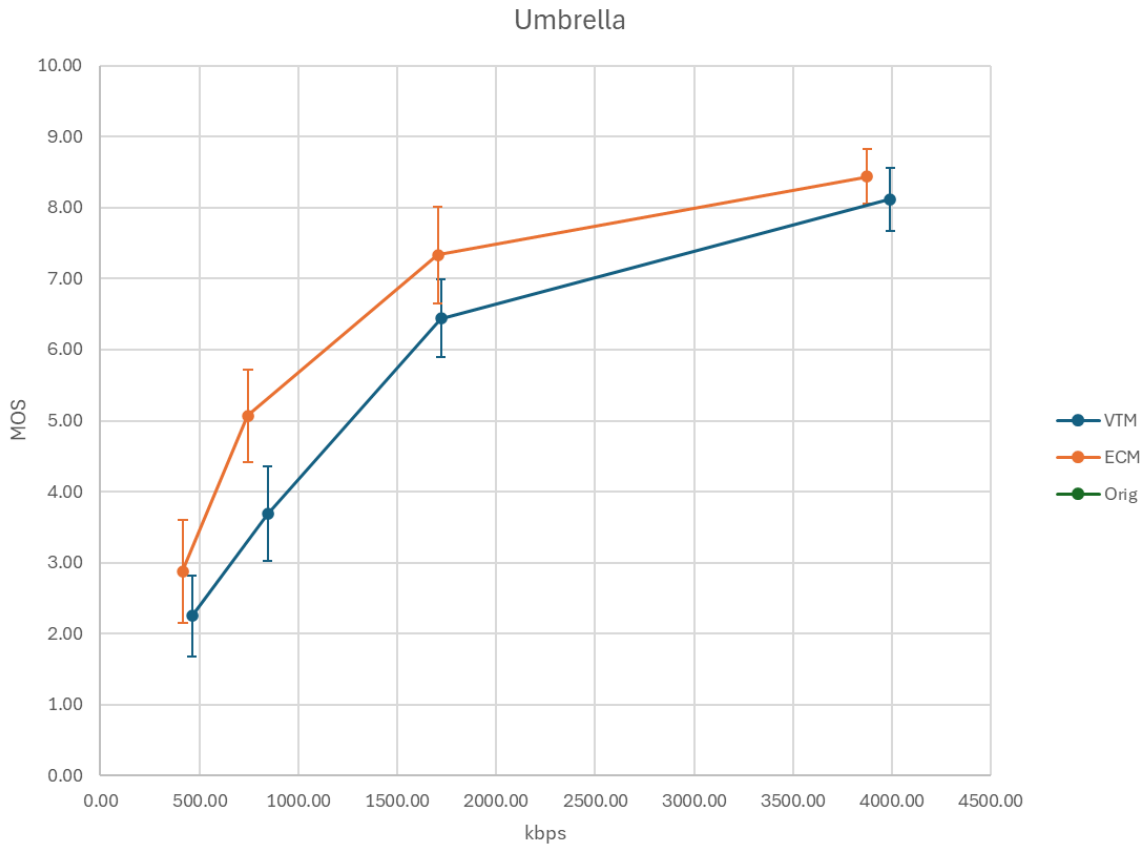
- Selection of candidate sequences based on feedback from viewing sessions
  - Participants in experiment have gained good knowledge of assessed test sequences
  - Five or more test sequences in each category available
- Checking of validity of approach: DCR test (on subset using HD-resolution sequences)
  - New candidates + UGC sequences
  - UGC is compressed ‘original’ compressed again: impact on DCR test?

- Gaming sequences
  - Level1
  - DOTA2s360
  - GTAVs090
  - Minecraft
  - Darktree
- SDR LD HD
  - GregoryCactus2
  - GregoryScarf2
  - Beatriz
  - OfficeWalkAtWall
- SDR HD RA
  - DucksTakeOff
  - Seeking
  - Umbrella
  - DrivingPOV4
  - Metro
- UGC
  - Camellia
  - NightLandscapeRunning
  - Sports-76a2-iceball
  - VerticalVideo-3709-snow
  - VerticalVideo-4b92-park
  - CoverSong-3499-outdoor
  - Hobby-w5xz-backpack
  - VerticalVideo-3d96-walk
- SDR UHD
  - FireDance
  - HallwayScene
  - TaikoDrums
  - CrowdRun
  - DrivingPOV3
  - MountainBay2
- HDR 8Kcrop
  - FashionLadyCrop1
  - MeltMetalCropBC
  - WaterfallForest
  - WomenFootball
  - ChandelierCropBR
- HDR UHD
  - AMS06
  - SparksWelding
  - MeridianHDR2
  - NeptuneFountain3R1
  - SubwayTreeR1

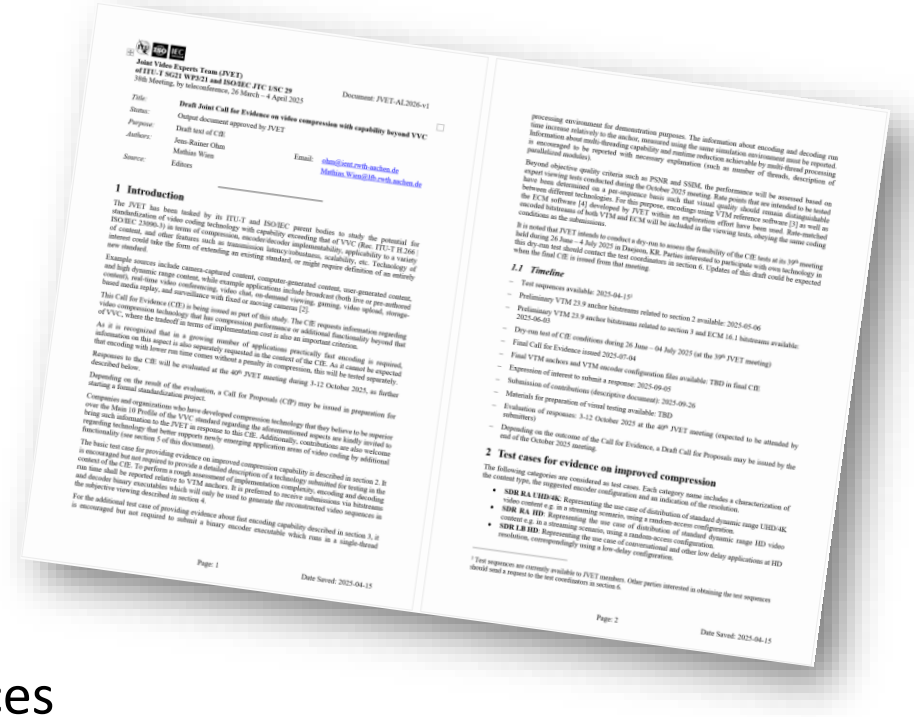
Step 2: Results of DCR assessment

Examples from DCR test





Step 3: Draft CfE description



- Decision on proposed final set of test sequences per category at JVET meeting in April
- Draft CfE document released: JVET-AL2026
  - 6 test sequences for the UGC category
    4 test sequences for the other 6 categories
    → Total of 30 test sequences
  - Two tracks
    - **Improved compression performance**
      - Compression gain on top of VVC
    - **Encoder complexity assessment**
      - Keep (or improve) compression performance at lower computational load
  - Subjective and objective evaluation
- Dry-run experiment with expert at JVET meeting in Daejeon in July

Conclusions

- Draft CfE designed
  - 7 content categories
  - 30 test sequences in total
  - Track for improved compression efficiency
  - Track for reduced encoding complexity
- AG 5 document with report and analysis under preparation (AG5N158)
- CfE issued in July, responses in October (Geneva)
  - Evaluation by on-site expert viewing
- Potential development towards a CfP
  - Could have test set of similar size / shape
  - Will require laboratory tests instead of on-site expert viewing