

8K Subjective Evaluation Experiments



What we are doing to assess 8K compressed videos

Yasuko Sugito

NHK

sugitou.y-gy@nhk.or.jp

Introduction

- NHK (Japan Broadcasting Corporation)
 - Japan's only Public broadcaster
- STRL has developed 8K technologies
 - Camera, display, codec, transmission and etc.
- 8K broadcasting starts on this December 1st

7,680 pixels(~8K)

4,320 pixels



8K Broadcasting Video Parameters³⁻¹

Ultra high definition	7,680×4,320 ³⁻²
High bit depth	10 bit ³⁻²
High frame rate	60p, 120p, and those divided by 1.001 ³⁻²
Wide color gamut	Rec. 2020 ³⁻²
High dynamic range	Hybrid Log-Gamma (HLG) ³⁻³
Viewing distance	0.75 H ³⁻⁴

Upcoming
8K broadcasting
is 59.94p

- Sense of being there
- Pixel structure is invisible

3-1. ARIB STD-B32 Ver. 3.9-E1, 2016.

3-2. Rec. ITU-R BT.2020-2, 2015.

3-3. Rec. ITU-R BT.2100-2, 2018.

3-4. Rec. ITU-R BT.2022, 2012.

1st 8K Video Coding Evaluation

- 8K/60p HEVC/H.265 real-time encoder⁴⁻¹ in 2013
 - 10 sequences at 4 bit-rates: 60, 85, 120, 170 Mbps
 - Considered broadcasting transmission capacity
- Used 85-inch 8K LCD monitor

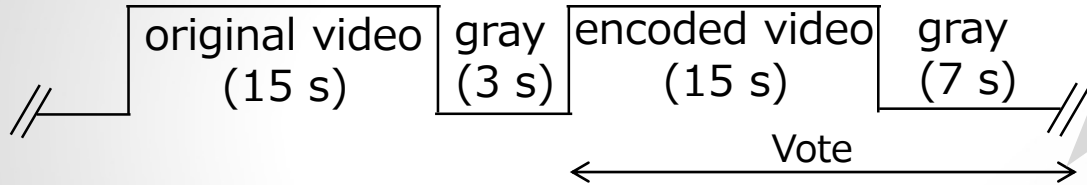
Size	Approx. 1.8 m wide × 1.05 m high
Video format	7,680 × 4,320/59.94p/10 bit
Brightness	300 cd/m ²
Pixel pitch	0.245 mm



4-1. Y. Sugito et al., "Development of the Super Hi-Vision HEVC/H.265 Real-Time Encoder," SMPTE 2013 Annual Technical Conference & Exhibition, Hollywood, CA, USA, pp. 1-16, Oct. 2013.

Evaluation Method

- Based on double-stimulus impairment scale (DSIS) method, Variant I⁵⁻¹



- Five-grade scale:
5. Imperceptible
 4. Perceptible, but not annoying
 3. Slightly annoying
 2. Annoying
 1. Very annoying

- 12 video coding experts participated

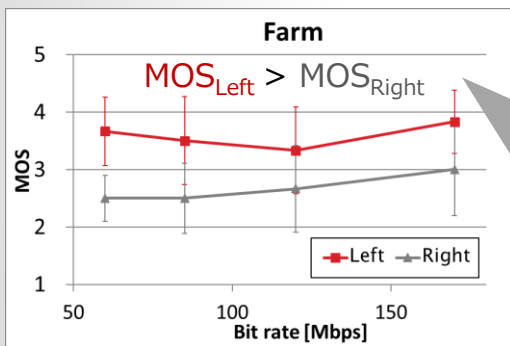
Distance	0.75 H (approx. 0.8 m)
Position	2 viewing points: left and right in front of the monitor (each 6 people).



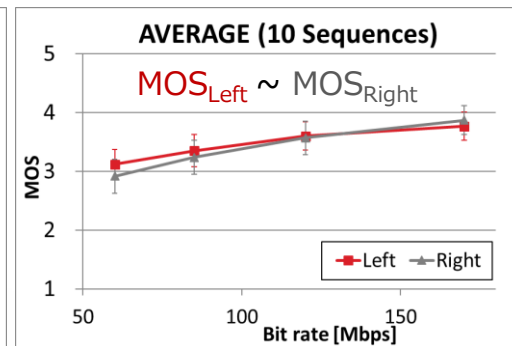
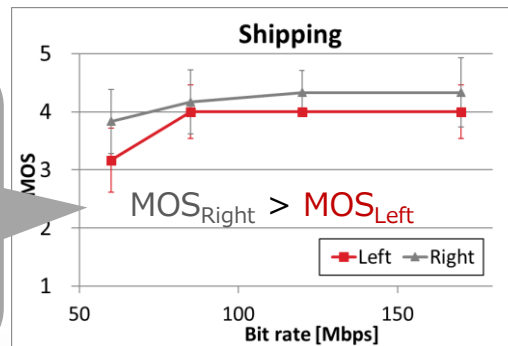
5-1. Rec. ITU-R BT.500-13, 2012.

Results Relevant to Evaluation Method

- 5 in 10 seqs. showed mean opinion score (MOS) of left/right \geq MOS of right/left for all the 4 bit-rates



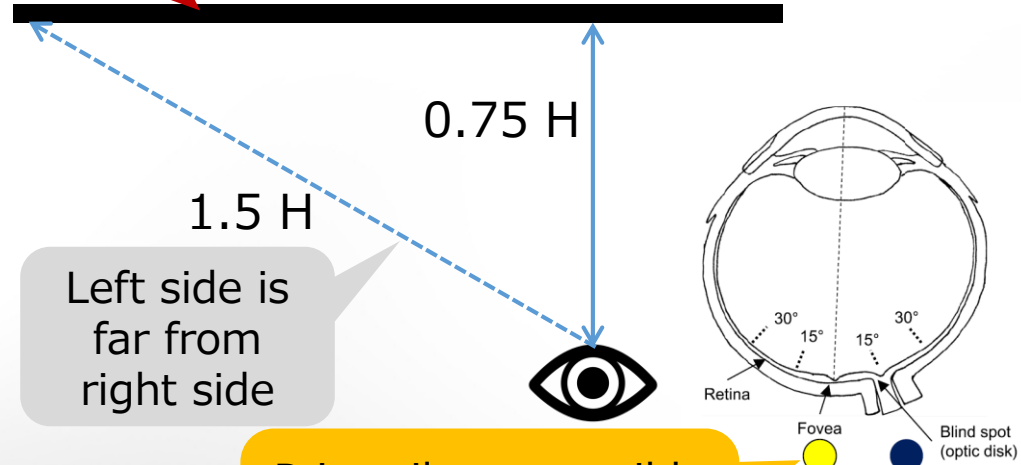
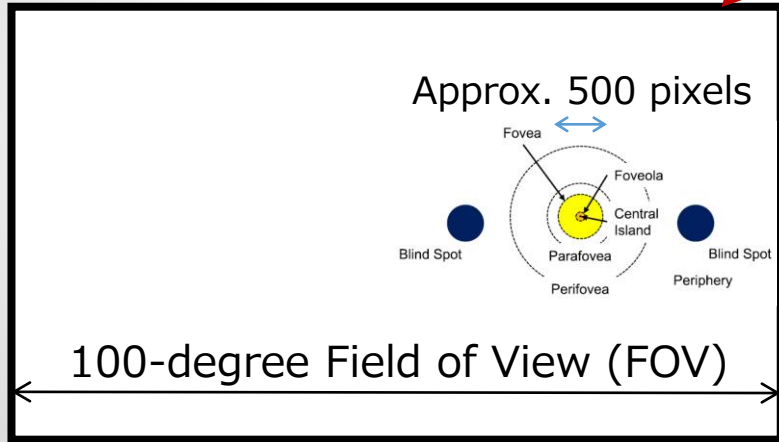
Part where degradation easily can be seen was on right and left, respectively



- All evaluators said, "I saw just in front of sitting position (while moving eyes and head)."
 - Seeing deterioration of another side or entire display was impossible

Considerations

- Results are reasonable considering human visual system (HVS)⁷⁻¹ **8K monitor**

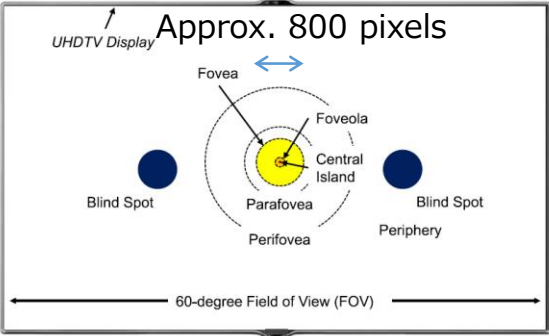


- Blind spots are inside of monitor
 - To look around monitor, head motion as well as eyeball rotation is required

Primarily responsible for high-acuity photopic viewing, 6°

Common 8K Evaluation Method

- Based on DSIS method, Variant I⁵⁻¹
 - Use at least 50-inch monitors to notice deterioration
 - Viewing conditions

Distance	0.75 H	1.5 H (optional)
Position	<ul style="list-style-type: none">■ 2 or more viewing points■ Frequently 3 (left, center, and right) points for 85-inch monitors■ Number of evaluators in each point is equal	<ul style="list-style-type: none">■ Center of monitor to see entire display  <p data-bbox="1251 972 1611 1019">Figure 3 of 7-1</p>

Detailed Subjective Evaluation

- 8K 120p HEVC/H.265 temporal video encoder in 2017
 - Compared slightly different 3 conditions
 - 120p total bit-rate was 85 Mbps, but 60p part bit-rate differed
 - A: 80 Mbps, B: 70 Mbps, and C: 60 Mbps
 - Evaluated both 120 and 60p videos
- Used repeatable pair comparison (RPC) method^{9-1, 9-2} based on the pair comparison (PC) method^{5-1, 9-3}
 - 😊 Pair of videos can be repeated as many times as desired
 - 😊 Simple system (recorder and monitor) is sufficient



9-1. Y. Sugito et al., "Validation of a Repeatable Pair Comparison Method," 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX), Cagliari, 2018, pp. 1-3.

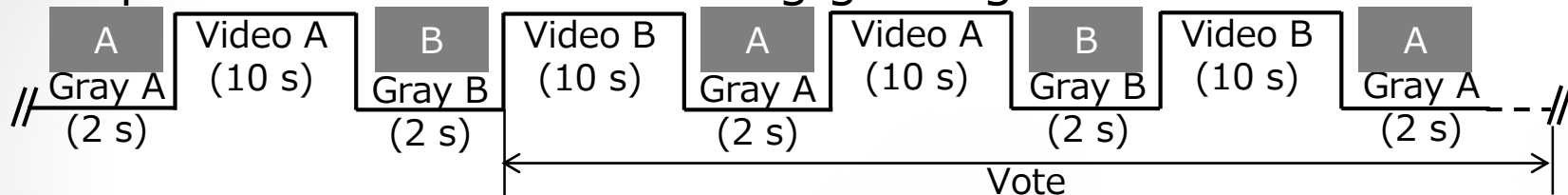
9-2. Y. Sugito et al., "A Study on the Required Video Bit-rate for 8K 120-Hz HEVC/H.265 Temporal Scalable Coding," 2018 Picture Coding Symposium (PCS), San Francisco, CA, 2018, pp. 106-110.

9-3. Rec. ITU-T P.910

Evaluation method

- Based on the Pair Comparison method^{5-1, 9-3}

- Repeated videos until finishing grading score



- Seven-grade scale⁵⁻¹

-3	-2	-1	0	1	2	3
Much worse	Worse	Slightly worse	The same	Slightly better	Better	Much better

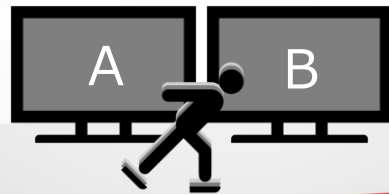
- 9 video coding experts participated
 - Evaluated all the 24 pairs ($= {}_3C_2$ -cond. \times 4-seq. \times 2-FR.)
 - Both conds. B after A (A-B) and B-A were included
 - One person at a time

Viewing Condition

- Used “one” 85-inch 8K SDR LCD monitor
 - Video format: 7,680 × 4,320/119.88p, 59.94p/10 bit

Distance	0.75 H (approx. 0.8 m)
Position	Free, a chair on casters was equipped

- Why not 2 monitors side-by-side?
 - Evaluator can see just in front of them
 - To compare the same part with 2 monitors, gliding from monitor to monitor is required
 - Distance \approx display width: at least 1.25 m

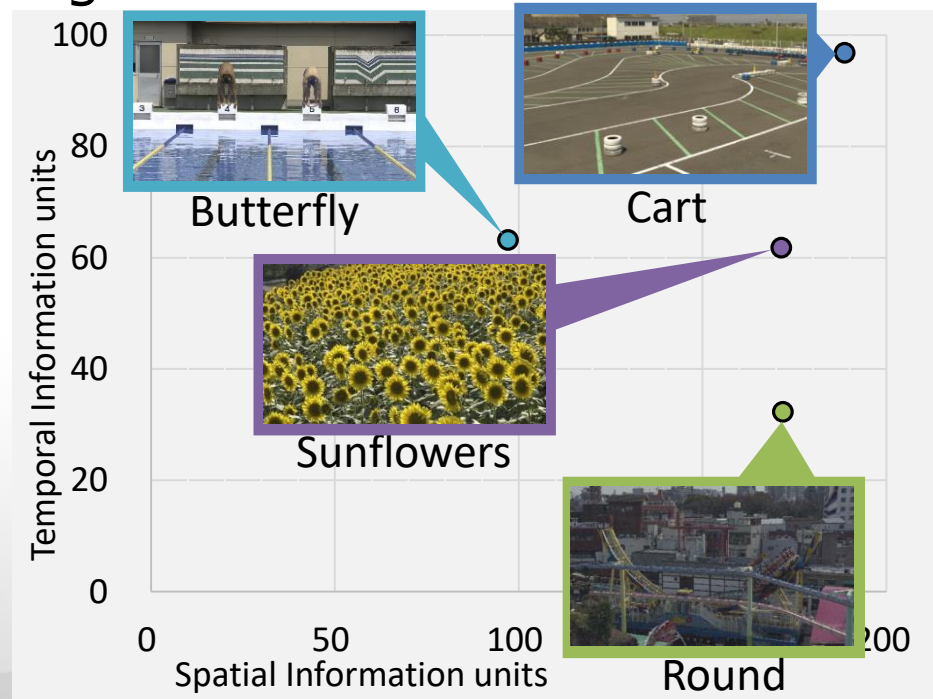


Evaluation Results

- Slightly different conditions were detectable
 - Some conditions showed significant difference^{9-1, 9-2}

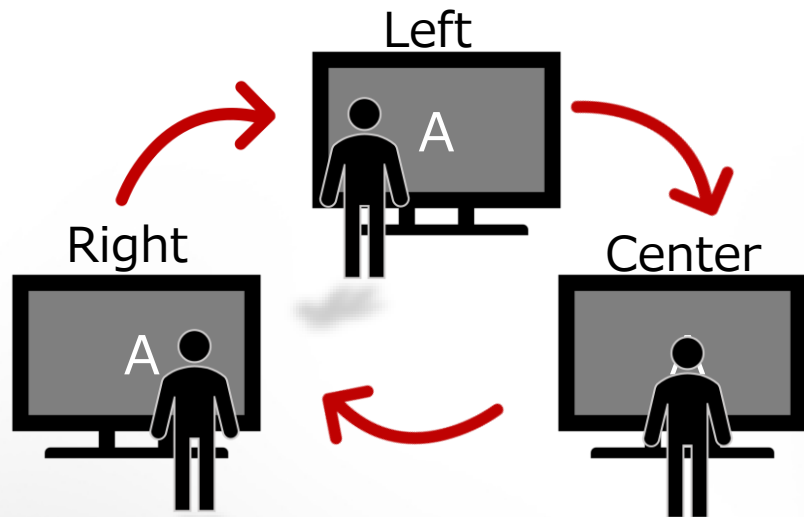
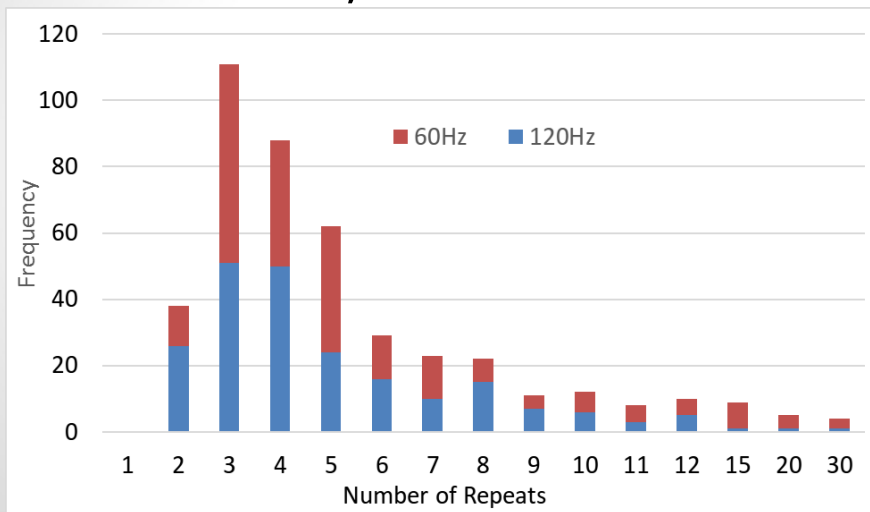
	120 Hz	60 Hz
Butterfly	-	A > B, C
Cart	-	-
Round	-	-
Sunflowers	A, B > C	A, B > C
Average	A, B > C	A > B, C

- Analyzed variation of MOS at significant level of 0.05
 - >: Significant difference
 - -: No significant difference



Number of Repeats

- Average = 4.9 ± 0.34
- Min. = 1.3, and Max. = 28.4



- Changing viewing position in front of monitor while repeating videos might be effective for 8K evaluation

Other Considerations

- Order effect caused by presentation order (e.g. A-B or B-A) was not negligible⁹⁻¹
 - All pairs of conds. must be evaluated in RPC method
- Appropriate num. of repeats: between 4 and 9⁹⁻¹
- Bradley-Terry score¹⁴⁻¹ showed sig. diff. when seven-grade scale was treated as win, lose, and draw
 - three-grade scale (-1, 0, and 1) was sufficient?
 - Perhaps, evaluators could score ± 1 for slight differences since max. score was ± 3

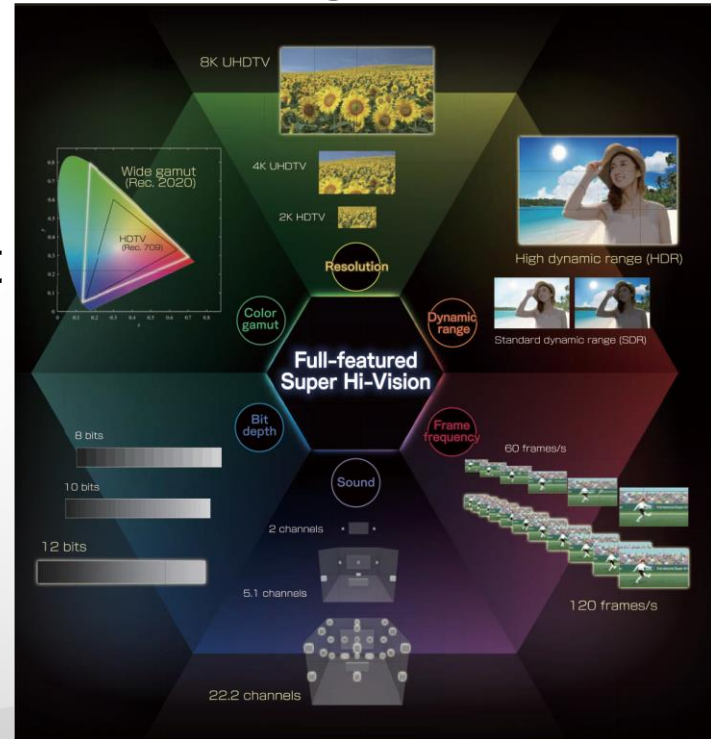


Better?
Not so
much
confidence

14-1. J. Li, M. Barkowsky, and P. Le Callet, "Boosting paired comparison methodology in measuring visual discomfort of 3DTV: performances of three different designs," Proc. SPIE 8648, Stereoscopic Displays and Applications XXIV, 86481V, 2013.

Future Work

- Conduct experiment with RPC method again
 - 15 or more evaluators
- Evaluate 8K HDR HFR videos
 - Flicker is more visible in bright videos⁷⁻¹



Conclusions

- Introduced two subjective evaluation methods for 8K compressed videos
 1. Based on DSIS method
 2. Based on PC method for slightly different videos
- 8K viewing distance, $0.75H$, provides immersive experience; however, it makes evaluations more difficult
- Hope this knowledge will be useful to consider subjective evaluation methods for new immersive media