Comparison of Remote Subjective Assessment Strategies in the Context of the JPEG Pleno Point Cloud Activity

Stuart Perry
Faculty of Engineering and IT
University of Technology Sydney
Sydney, Australia
Stuart.Perry@uts.edu.au

Luis A. Da Silva Cruz
Dep. of Elect. and Computer Engineering and
Instituto de Telecomunicações
University of Coimbra
Coimbra, Portugal
lcruz@deec.uc.pt

Emil Dumic
Department of Electrical Engineering
University North
Varaždin, Croatia
edumic@unin.hr

Nhung Hong Thi Nguyen
Faculty of Engineering and IT
University of Technology Sydney
Sydney, Australia
NhungHongThi.Nguyen@student.uts.edu.au

António Pinheiro
Instituto de Telecomunicações and
Universidade da Beira Interior
Covilhã, Portugal
pinheiro@ubi.pt

Evangelos Alexiou
Multimedia Signal Processing Group
École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland
evangelos.alexiou@epfl.ch

***Abstract*—In this work we compare two different options to perform on-line subjective quality assessment experiments in the context of the Call for Evidence on JPEG Pleno Point Cloud Coding. A deep-learning based point cloud codec submitted to the Call was tested against current MPEG point cloud compression methods. The first option is based on participants downloading the entire set of stimuli and running a set of scripts in MATLAB to perform the experiment. The second option involves the participants accessing a server on the web and viewing and judging the stimuli using a web browser. Quality scores compiled using both methods were compared showing strong correlation. A second analysis compared the quality scores with those obtained in a prior laboratory-based study using higher resolution screens. The entire study also brought to light each option's unique advantages and disadvantages that make each one better suited to specific types of subjective evaluation contexts and situations.**

***Index Terms*—JPEG, point cloud, subjective assessment, crowd sourcing, coding**

I. Introduction

The JPEG Committee has been pursuing standards for the representation and compression of plenoptic data. The JPEG Pleno activity seeks a seamless standards framework to integrate lightfield, holographic and 3D point cloud data. With the lightfield aspect of the activity well-advanced, attention has been turned to standards for the representation of holographic and 3D point cloud data. The Ad hoc Group on JPEG Pleno Point Cloud has identified a need for standardisation in the area of point cloud coding with particular focus on solutions supporting random access and scalable representations [1]. In order to support this activity, the JPEG Committee released the Call for Evidence for JPEG Pleno Point Cloud Coding [2]. This Call for Evidence solicited innovative new point cloud coding solutions with support for scalability and random access and outlined a procedure for comparing submissions with existing point cloud coding solutions in terms of both objective metrics and subjective quality as described in the JPEG Pleno Point Cloud Common Test Conditions Document [3].

Crucial to the Call for Evidence was the subjective evaluation of submissions' coding performance against "anchor codecs" consisting of existing point cloud coding solutions. Although subjective experiments are normally conducted under highly controlled laboratory conditions by a small group of test laboratories, the spread in 2020 of the COVID-19 pandemic forced this activity to be be done with the support of online systems. In this work we describe and compare the use and results of two online subjective testing frameworks used for the Call for Evidence on JPEG Pleno Point Cloud Coding. One system involved the participants downloading an entire dataset of processed and original point clouds represented as 2D videos, together with MATLAB code for viewing the videos and recording the subjective judgments. The second system involved participants accessing a website and viewing and grading the stimuli videos via their web browser. Subjective experiments have always been difficult and expensive to conduct, in terms of the care that needs to be taken in setting up the experiment, the difficulty in finding participants, and the time needed by the participants to complete the evaluation tasks of the experiment. The COVID-19 pandemic added an additional hurdle: the danger of viral transmission between participants and the experiment conductors. An al-

We acknowledge the support of "PLIVE" through FCT project UIDB/EEA/50008/2020.

ternative solution is to perform the evaluations using many geographically distributed observers following crowd-based paradigms. However, crowd-based video quality assessment faces several challenges, as pointed out in [4] where it is argued that despite conceptual, technical, motivational and reliability challenges, crowd-based video quality assessment is a promising alternative to the standard laboratory-based experiments. However, for uncompressed or losslessly compressed video sequences, dedicated applications have to be designed to provide videos without experiencing stalling events [5]. In [6] authors present several improvements for the existing crowd-based test solutions, enhancing the reliability of the results and overall completion time. Related ITU-T Technical Report PSTR-CROWDS [7] provides a general guidance on testing methods for media quality assessments using a crowd-sourcing approach. In [8] these principles were followed to design an online web-based solution using the Chrome browser to evaluate the quality of 3D video. Inspired by this system, an online web based framework was designed to be used in the study described in this article, to evaluate the quality of point clouds, presented as movies showing the point cloud objects rotating around a predefined axis. To avoid stalling during playback caused by network bandwidth fluctuations or insufficiency, the content animations were preloaded into the Chrome browser cache. The two main drawbacks of this approach are the content size limit imposed by the Chrome cache size limit of about 1.5-2 GB per session and the preload waiting time. Unlike the crowdsourced experiments in [8], in this study the qualification task was skipped because the participants were volunteers recruited from amongst the image and video compression research community and so dishonest behavior was not expected. Another solution adopted in this study involved downloading the complete test set to the client's computer and afterwards running the whole experiment with all stimuli videos at once, with the entire process controlled through MATLAB scripts. The use of this direct download approach implies that all participants have access to MATLAB and communication and local storage resources compatible with the data volumes to be downloaded. In this work we will be contrasting these two methodologies for on-line subjective testing, *web browser* and *direct download*, examining their advantages and disadvantages, and assessing the degree of agreement between the subjective judgments obtained through their use.

This paper is divided into the following sections. Section I gave an introduction to this work and the prior art. Section II describes the two on-line methodologies for subjective testing and Section III compares the subjective results obtained by the methodologies. Section IV discusses the results and advantages and disadvantages of the two methodologies, while Section V draws some conclusions.

II. Subjective Quality Assessment

The subjective experiment involved having participants observe videos showing side-by-side pairs of rendered point clouds, with an unprocessed point cloud on one side and a processed one on the other side, and asking them to subjectively grade the difference as detailed in the JPEG Pleno Point Cloud Common Test Conditions Document [3] (CTC).

A. Stimuli

Eight points clouds from the CTC, shown in Fig. 1, were used for the subjective experiments.

(a) *bumbameuboi* (b) *guanyin* (c) *longdress*

(d) *rhetorician* (e) *ricardo* (f) *romanoillamp*

(g) *soldier* (h) *phil*

Fig. 1. Frontal views of point clouds used in this work. Source: [9], [10]

B. Tested Codecs

Four codecs were tested in this experiment, three anchor codecs and one codec submitted to the Call for Evidence. The anchor codecs were MPEG G-PCC with Octree module, MPEG G-PCC with TriSoup module and MPEG V-PCC configured for intra-coding, while the submitted codec (herein denoted "RS-DLPCC+Color") was a deep learning-based point cloud coding solution.

1) G-PCC (Octree and Trisoup): The MPEG G-PCC codec [11] has two alternative encoding modules that may be used to compress geometry information. These modules are denoted Octree and Triangle Soup (TriSoup) [12]. The first method is based on an Octree decomposition of the point cloud wherein the degree of compression can be controlled by the *positionQuantizationScale* parameter. This parameter selects the number of divisions of the octree structure from the root to each leaf node. The second method used the octree structure to a certain extent and then encodes the points within a leaf node with a triangular surface approximation. The size of the block on which the approximation is applied is determined by the *trisoup_node_size_log2* parameter. The color information is encoded by the "Prediction-plus-Lifting" scheme, involving the prediction of a color value from its neighbours [13]. The above parameters were adjusted to encode the content at five rate points for both the Octree and TriSoup encodings. These rate points were designed to range from 0.1 to 10 bits per point covering quality levels from poor to excellent for each of the point clouds.

2) V-PCC: THE MPEG V-PCC codec [12] involves the projection of the point cloud onto a set of 2D planes followed by the encoding of the projected planes using a 2D video codec. The number of projection planes is usually 6 and in the current test model TMC2 [14], the HEVC codec is used to encode the projections. Three types of information are encoded in the projections; texture, depth information and occupancy maps. The texture information is broken into a set of patches and arranged in a compact 2D mosaic image. The distances from the projection plane to each point in the cloud corresponding to each pixel in each texture patch is determined and encoded in the depth image. The occupancy map indicates which pixels in the 2D depth and texture maps contain meaningful information. V-PCC parameters were set according to the MPEG Common Test Conditions document [15] in the C2, *Lossy Geometry – Lossy Attributes* encoding condition with All Intra coding mode used, as we are encoding static point clouds as opposed to dynamic ones. Five rate points were chosen covering rates from 0.1 to 3 bits per point and quality levels from poor to excellent.

3) RS-DLPCC+Color: The Resolution Scalable Deep Learning-based Point Cloud Geometry Coding (RS-DLPCC) codec [16], was a submission to the Call for Evidence on JPEG Pleno Point Cloud Coding. This codec is a deep learning-based point cloud geometry coding solution using the latent representation of point cloud computed by an autoencoder framework. Scalability is obtained by the use of interlacing sampling of blocks of the point cloud each of which is independently encoded. Random-access is enabled by the division of the point cloud into a number of super-blocks. The RS-DLPCC codec encodes geometry only, so the submitters adapted the approach to include color information to allow for the subjective testing. The geometry of each point cloud was encoded with RS-DLPCC and the color was then transferred from the nearest neighbour of the original point cloud. The color for the recolored points is encoded with G-PCC, using the lossless geometry Octree coding mode, and the Predlift color encoder. The lossless Octree coding mode was chosen so that the (decoded) geometry is not changed, minimising the geometry coding effects on the color information from the G-PCC codec. This color information is then textured over the RS-DLPCC lossy decoded geometry. Four rate points were chosen for each point cloud between 0.3 to 20 bits per point.

C. Evaluation Methodology

In accordance with the JPEG Pleno Point Cloud Common Test Conditions [3], the point clouds to be evaluated were rendered as 12 second videos to be displayed at 30 fps. The resolution of the videos was set to be 1920x1080 and they were losslessly compressed using HEVC. In previous experiments with the same anchors [17], videos were rendered at 4K resolution. However, it was thought that Full HD (FHD) 2K resolution was more likely to be compatible with the wider range of monitors expected to be used by on-line participants while also reducing the amount of data to be downloaded. This choice becomes especially crucial in light of the requirement for the *web browser* methodology to preload the stimulus videos. Only a couple of 4K resolution videos would fit in the browser cache at one sitting, rendering the resultant experiment unfeasible. Consequently, we used the same videos with the same, 2K resolution for both *web browser* and *direct download* methodologies. For most point clouds, the videos were created by a complete rotation of the point clouds about the vertical axis. For the *ricardo* and *phil* point clouds, the videos were created by a partial rotation about the vertical axis keeping the front of the point cloud in view. The Double Stimulus Impairment Scale (DSIS) test method [18] was used wherein both the reference and decoded point clouds were shown to the participant with the reference point cloud identified by text underneath it on the video. A 5-level rating scale for comparing the decoded to reference point cloud *(1 - very annoying, 2 - annoying, 3 - slightly annoying, 4 - perceptible, but not annoying, 5 - imperceptible)* was used and hidden reference-reference pairs were included to check for unusual behaviour by participants. At the beginning of each evaluation session the observers were shown the *phil* point cloud at various quality levels in order to familiarise the participants with the types of artefacts present. For every point cloud/codec/rate combination, Mean Opinion Scores and 95% Confidence Intervals were computed. Outlier detection was performed on each of the subjective testing methodologies below according to ITU-R Recommendation BT.500-14 [18]. No outliers were detected.

D. Direct Download Option

In the Direct Download Option, participants were asked to download a 32GB compressed data file containing all 280 stimuli videos together with MATLAB scripts to run the experiment. This approach was used in a previous study [17] with 4K videos and the participants limited to a small set of laboratories with controlled viewing conditions. In this study, the stimuli videos were 2K and potential participants were merely asked to register with the JPEG Committee and confirm that they had access to a 2K monitor and at least MATLAB 2019b. No instructions were given about viewing conditions. The participants were under instruction to run the MATLAB scripts for the training session first, followed by the scripts for the main experiment, but there was no way to enforce this protocol on the participants. Upon the start of an evaluation, the MATLAB scripts randomised the entire set of videos such that each evaluation presented a different order of the stimuli, however the order was such that the participants never saw the same point cloud consecutively. The experiment took around 40 minutes to complete and subjects were instructed to take a break after 20 minutes. At the end of the experiment, participants were asked to email the log files produced by the MATLAB scripts to the JPEG Committee.

E. Web Browser Option

For the Web Browser option, following registration, participants were given a set of instructions to configure their web browsers and monitors. Participants were requested to use the Google Chrome web browser and given instructions to configure the browser for the maximum cache size possible, (1.5GB). The Google Chrome browser was chosen as it was found to have the largest available cache size of commonly used browsers and supported seamless playback of HEVC encoded video. Once the participant had configured their browser correctly and been instructed to clear the cache, they were directed to one of two servers; one in Seoul, Korea and the other in Frankfurt, Germany. After initial questions about the monitor, location, age and gender of the participant, a preload stage loaded a 1.5GB set of stimuli videos into the participant's browser cache. The preload stage could last up to 30 minutes depending on the proximity of the observer to the web server they selected. Following the preload stage, the loaded stimuli videos were displayed to the participant and subjective judgments recorded. Due to the 1.5GB maximum cache size available, only 20 videos out of the complete set of 280 videos could be viewed in a single sitting. Assuming that the participant had followed the instructions correctly, a single evaluation session would take approximately 10 minutes.

The 20 videos loaded in a single session needed to include the training videos using the *phil* point cloud. This limitation on the number of videos that could be viewed in a single evaluation session required adjustments to the experimental design. The training video point clouds were coded with lossy compression at a level that displayed no obvious visual artefacts. A subset of the point cloud and codecs were chosen for the Web Browser Option to ensure that a sufficient sample was collected for each point cloud. Hence the Web Browser Option was restricted to the *longdress*, *ricardo* and *romanoillamp* point clouds encoded with the G-PCC (Octree), G-PCC (TriSoup) (*longdress* only), V-PCC and RS-DLPCC+Color codecs, for a total of 94 testing videos. These videos were randomly split into 6 playlists of approximately 1.5GB each. Three of the playlists contained the stimuli with the reference on the right and other three contained the stimuli with the reference on the left. When an evaluation session was started, the server would randomly choose one of the 6 playlists and load the videos in the playlist to the participant's browser cache in the preload stage. Following the preload stage, the training videos would be shown first, followed by the experiment videos in random order. Due to the few point clouds in each playlist, it was impossible to avoid the participants seeing two consecutive videos with the same point cloud.

III. Results

A. Participation

Participants were asked to register for the experiment and sent instructions for both the Direct Download and Web Browser options. All participants confirmed that they had access to a 2K display. The Direct Download option was used by 27 participants from 7 different academic institutions, in Korea, Portugal, Belgium, Brazil, Vietnam, Italy and Australia. The age and gender of participants were not recorded and returned log files contained no information identifying participants. The Web Browser option was used by 40 participants from 10 different academic institutions in Vietnam, Portugal, France, Croatia, Switzerland, Greece, Brazil and India. For the Web Browser option, information about participant's age and gender were recorded. Due to the manner in which these experiments were performed it was not possible to be completely certain that information entered was accurate. In total there were 67 participants to the experiment and no participants or participating institutions were given any financial compensation.

B. Correlation between Web and Direct Download Scores

Fig. 2 shows a comparison of the MOS values collected by the Direct Download and Web Browser options for the *longdress*, *ricardo*, and *romanoillamp* point clouds for all four of the codecs tested. Table I gives the Pearson and Spearman correlations between the two sets of MOS scores for each of the point clouds and the three combined. For the *ricardo* and *romanoillamp* point clouds, the Spearman correlation is significantly lower than the Pearson Correlation. In practice this is often a function of uneven distribution of the data and by examination of Fig. 2, we can observe clumping of the *ricardo* (blue) and *romanoillamp* (green) data points toward higher MOS values. This is likely an artefact of the rate-distortion tradeoff choices for the codecs used in the experiment. To check whether any of the point clouds have a biased relationship between the MOS values from the two options, lines of best fit were computed for the MOS values of each point cloud and the combined results and the slope and intercept values are shown in Table II. Although the intercept values all appear to be consistent with no offset in MOS values for specific point clouds between the two different testing approaches, the slope of the best fit line for the *longdress* point cloud does appear to have a different slope compared to the *ricardo* and *romanoillamp* point clouds. This effect can also be seen in Fig. 2, with the red *longdress* data points being consistently below the combined line of best fit. This can be interpreted as participants judging the quality of the *longdress* point clouds consistently lower on the web browser method as compared to the direct download method.

C. Correlation between Different Video Resolutions

In [17] the anchor codecs G-PCC (Octree and TriSoup) and V-PCC were subject to a subjective test using the Direct Download Option. In that experiment, controlled laboratory conditions were used and the stimuli were rendered for, and

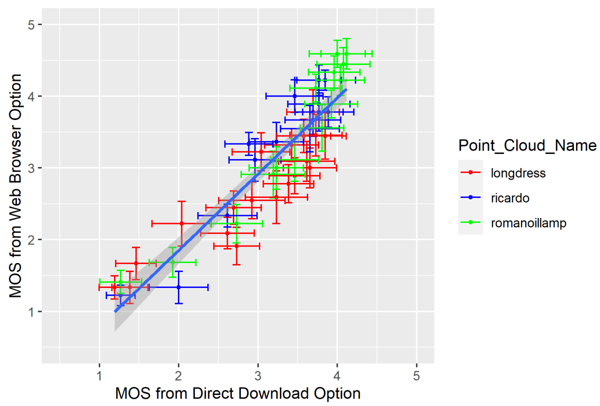

Fig. 2. Comparison of MOS values from Direct Download and Web Browser Options. Error bars on data points indicate 95% confidence intervals. Line of best fit shown in blue with gray area signifying 95% confidence of fit.

TABLE I
CORRELATION BETWEEN MOS VALUES FOR THE DIRECT DOWNLOAD AND WEB BROWSER OPTIONS

Point Cloud	Pearson Correlation	Spearman Correlation
longdress	0.921	0.920
ricardo	0.954	0.872
romanoillamp	0.954	0.872
Combined	0.926	0.959

displayed on, 4K monitors, instead of the 2K resolution used in this work. In this section, we compare MOS values collected with the previous experiment (denoted here as the “4K Direct Download Experiment”) and the current test. In the previous experiment, a reduced set of point clouds were tested and only the *bumbameuboi*, *guanyin*, *rhetorician* and *romanoillamp* point clouds were in common to both experiments. Fig. 3 shows the comparison of the MOS values from the 2K and 4K Direct Download Experiments grouped by point clouds across all the codecs tested. The correlations between the 2K and 4K data for each of the point clouds and the combined results are shown in Table III. In general both the Pearson and Spearman correlations are above 0.94 except for the *rhetorician* point cloud where the correlation is lower than the other point clouds. In Fig. 3, the *rhetorician* point cloud (green) can be seen to have considerable variation about the best fit line and a notable outlier. From Fig. 1 we note that this point cloud has a lack of color, which may result in instability of the subject judgments compared to the other point clouds tested. Table IV shows the intercepts and slopes of best fit lines for each of the point clouds and the combined results. We can notice that the combined results have an intercept that is significantly greater and a slope that is significantly different from unity. This indicates a bias in the results wherein the MOS values (slightly) favour the videos rendered in 2K for poor quality point clouds, but as the quality increases, the judgments tend to be more favourable at 4K, compared to 2K. This effect seems to be absent in the *bumbameuboi* point cloud and strongest in the *romanoillamp* point cloud, although the *guannyin* and *rhetorician* point clouds also displayed this tendency.

TABLE II
SLOPES AND INTERCEPTS OF BEST FIT LINES FOR RELATIONSHIP BETWEEN MOS VALUES FOR THE DIRECT DOWNLOAD AND THE WEB BROWSER OPTIONS (ERRORS ARE 95% CONFIDENCE INTERVALS)

Point Cloud	Intercept	Slope
longdress	0.29±0.49	0.81±0.16
ricardo	-0.51±0.69	1.18±0.21
romanoillamp	-0.56±0.72	1.18±0.21
Combined	-0.27±0.41	1.06±0.12

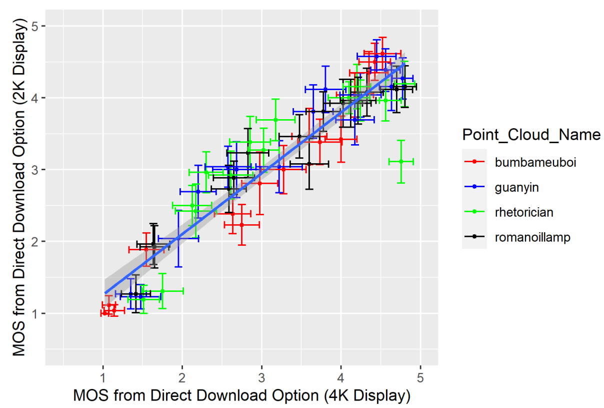

Fig. 3. Comparison of MOS values from the Direct Download Option for 2K versus 4K resolution stimuli videos. Error bars on data points indicate 95% confidence intervals. Line of best fit shown in blue with gray area signifying 95% confidence of fit.

TABLE III
CORRELATION BETWEEN MOS VALUES FROM THE DIRECT DOWNLOAD OPTION FOR 2K VERSUS 4K RESOLUTION STIMULI VIDEOS.

Point Cloud	Pearson Correlation	Spearman Correlation
bumbameuboi	0.982	0.982
guanyin	0.961	0.940
rhetorician	0.849	0.860
romanoillamp	0.968	0.985
Combined	0.942	0.935

TABLE IV
SLOPES AND INTERCEPTS OF BEST FIT LINES FOR THE RELATIONSHIP BETWEEN MOS VALUES FOR THE 2K AND 4K DIRECT DOWNLOAD OPTIONS (ERRORS ARE IN TERMS OF 95% CONFIDENCE INTERVALS)

Point Cloud	Intercept	Slope
bumbameuboi	-0.004±0.322	0.957±0.1
guanyin	0.377±0.484	0.878±0.138
rhetorician	0.853±0.817	0.711±0.240
romanoillamp	0.621±0.385	0.787±0.057
Combined	0.413±0.26	0.846±0.08

IV. DISCUSSION

Although the Direct Download option allowed for the entire experiment to be completed by participants, it had a number of disadvantages. In addition to the large size of the download and the need for the MATLAB environment, the stimuli videos are not encrypted which allows participants to examine and copy them if they wished. This might be undesirable in some use cases. In contrast, the Web Browser option made it difficult for the participant to examine the videos outside of the experiment. The requirement that the participant could only view a total of 1.5GB of stimuli videos in a single evaluation session was a restriction on the size of the experiment that could be conducted using this option. This restriction could be mitigated by supplying dozens of playlists to the system and asking participants to sit the experiment dozens of times with a new preload stage for each attempt which is clearly not feasible. In addition, proximity to the web servers dramatically affected the time for the preload stage. Participants in Europe generally reported short preload stage times for the Frankfurt server, but this server was not usable for participants in Asia and Australia. Attempts to access this server in Australia caused preload times of many hours, followed by videos displaying lag effects. This may be due to the long time period between cache load and viewing for some videos leading them to be removed from the cache before the preload stage was complete. In general the MOS values recorded in the Web Browser and Direct Download options were consistent with a correlation of greater than 0.92 for all point clouds. However there was some evidence of a bias in MOS values between the two experimental options for one of the point clouds. Investigating this effect might be worthwhile future work. Comparing the results of the Direct Download experiment conducted with 2K resolution stimuli with the same mode of experiment conducted with 4K resolution stimuli by a limited set of institutions in controlled laboratory circumstances, we see a bias in the results that is difficult to explain. Compared to the 4K resolution stimuli, it appears that participants tended to slightly over-estimate quality of the 2K resolution stimuli at lower quality levels and underestimate the quality at higher quality levels. One could hypothesize that at low quality levels, the lower resolution display masks perception of the defects, while at higher quality levels, the improved resolution of the 4K stimuli allows better appreciation of the finer detail in the point cloud which tends to lift the quality judgment.

V. Conclusion

In this work the authors contrasted two options for conducting on-line subjective experiments in the context of the Call for Evidence on JPEG Pleno Point Cloud Coding. The first option involved participants downloading the entire experiment together with MATLAB scripts to run the experiment. The second option involved participants accessing a web page to run the experiment. The Direct Download option was found to be effective for large experiments, while the web browser option attracted more participants due to its convenience. The Mean Opinion Scores for both options were compared and found to be in general consistent, however one of the point clouds appeared to display a systematic bias that needs further investigation. The results of the Direct Download experiment were contrasted with an earlier experiment with the same software with a higher resolution display of the stimuli. A bias in results was detected potentially caused by defect masking by the lower resolution display.

Acknowledgment

The authors thank João Prazeres, André Guarda, Fernando Pereira, Nuno Rodrigues, and Peter Schelkens for assistance with the Call for Evidence. We also thank Sejong University (KR), University of Beira Interior (PT), University of Technology Sydney (AU), Vrije Universiteit Brussel (BE), Vietnam National University Hanoi (VN), Universidade Federal do Rio de Janeiro (BR), Roma Tre University (IT), Le Quy Don Technical University (VN), University of Coimbra (PT), Institut National des Sciences Appliquées de Lyon (FR), University North (HR), École Polytechnique Fédérale de Lausanne (CH), University of Patras (GR) and Aligarh Muslim University (IN) for participating in the subjective experiment.

References

[1] ISO/IEC JTC1/SC29/WG1, "JPEG Pleno Point Cloud – Use Cases and Requirements v1.3," Doc.WG1N86012, Jan 2020.

[2] ISO/IEC JTC1/SC29/WG1, "Final Call for Evidence on JPEG Pleno Point Cloud Coding," Doc.WG1N88014, July 2020.

[3] ISO/IEC JTC1/SC29/WG1, "JPEG Pleno Point Cloud Coding Common Test Conditions v3.5," Doc.WG1N90053, Jan 2021.

[4] C. Keimel, J. Habigt and K. Diepold, "Challenges in crowd-based video quality assessment," 2012 Fourth International Workshop on Quality of Multimedia Experience, Melbourne, VIC, Australia, 2012, pp. 13-18.

[5] T. Hoßfeld et al., "Best Practices for QoE Crowdtesting: QoE Assessment With Crowdsourcing," IEEE Transactions on Multimedia, vol. 16, no. 2, Feb. 2014, pp. 541-558.

[6] B. Gardlo, S. Egger, M. Seufert and R. Schatz, "Crowdsourcing 2.0: Enhancing execution speed and reliability of web-based QoE testing," 2014 IEEE International Conference on Communications (ICC), 2014, pp. 1070-1075.

[7] ITU-T Technical Report, "PSTR-CROWDS - Subjective evaluation of media quality using a crowdsourcing approach," *International Telecommunications Union*, May 2018.

[8] E. Dumic, K. Sakic, and L.A. da Silva Cruz, "Crowdsourced subjective 3D video quality assessment," Multimedia Systems, vol. 25, no. 6, 2019, pp. 673–694.

[9] JPEG Pleno Database. [Online]. Available: https://jpeg.org/plenodb/

[10] E. Alexiou, N. Yang and T. Ebrahimi, "PointXR: A Toolbox for Visualization and Subjective Evaluation of Point Clouds in Virtual Reality," 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), 2020, pp. 1-6.

[11] ISO/IEC JTC1/SC29/WG11, "G-PCC Codec Description v5," Doc.N18891, Geneva, Switzerland, Nov 2019.

[12] S. Schwarz et al., "Emerging MPEG standards for point cloud compression," IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 9, no. 1, 2019, pp. 133–148.

[13] E. Alexiou et al., "A comprehensive study of the rate-distortion performance in MPEG point cloud compression," APSIPA Transactions on Signal and Information Processing, vol. 8, 2019, pp. e27.

[14] ISO/IEC JTC1/SC29/WG11, "V-PCC Codec Description," Doc.N18884, Geneva, Switzerland, Nov 2019.

[15] ISO/IEC JTC1/SC29/WG11, "Common test conditions for point cloud compression," Doc.N18474, Geneva, Switzerland, Mar 2019.

[16] A. Guarda et al., "Deep Learning-based Point Cloud Geometry Coding with Resolution Scalability," IEEE International Workshop on Multimedia Signal Processing (MMSP'2020), Tampere, Finland, Sep 2020.

[17] S. Perry et al., "Quality evaluation of static point clouds encoded using MPEG codecs," IEEE International Conference on Image Processing (ICIP2020), Abu Dhabi, United Arab Emirates, Oct 2020, pp. 25-28.

[18] ITU-R BT.500-14, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunications Union, Oct 2019.