IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Subjective Video QoE with a Hybrid Cognitive-motor Task

Jan Holub[1], Member, IEEE, **Hakob Avetisyan**[1], Member, IEEE, and **Paulo Victor Lima**[2]

[1]Department of Measurement, Faculty of Electrical Engineering, Czech Technical University, Prague, 160 00, Czech Republic

[2]Department Electronics and Computing Engineering, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 21941-901, Brazil

Corresponding author: Hakob Avetisyan (e-mail: avetihak@fel.cvut.cz).

ABSTRACT This paper reports on the subjective video quality analysis with and without a parallel task. It aims to understand how the quality perceived by the human observer brain is affected when the subject is not entirely focused to the judgement task. For this purpose, two subjective video quality tests were performed according to the ITU-T P.910 recommendation - Subjective video quality assessment methods for multimedia applications. They consisted of two parts involving a total of 24 subjects. Unlike the first test, the second test included an additional (parallel) psychomotor task where subjects had to sort different colors of pasta into three bowls. Various analyses were provided, including preliminary analysis, group analysis, and RMSE analysis. Surprisingly, the results show that not all the spectrum of quality is affected equally by parallel task introduction.

INDEX TERMS Video signal processing, Human Factors, Quality of Service

I. INTRODUCTION

Video quality is a feature of the video transmitted or processed by a system. In some cases, undesirable distortions and artifacts can affect the signal, which negatively impacts the subjective video perception of the viewer. Ensuring video quality is an essential task for many stakeholders, such as content or service providers [1]–[3].

The objective way to evaluate a video is by using various algorithms and mathematical models to estimate results from the subjective evaluation. According to [4], the term 'objective' is commonly a reference to instrumental measurement techniques in the literature. The primary purpose of these techniques is to reduce most of the issues encountered in subjective assessment [5]. Moreover, its methods are full-reference (FR) where full access of the reference image is required, no-reference (NR) that assume no access to the reference and reduced-reference (RR) where full access is not required. However, only partial information is provided in the RR features form [6].

The subjective video quality assessment is the most reliable technique known since the ultimate users are humans [7]. According to [5], the traditional method, and still the mainstream approach, is to conduct the tests in a controlled laboratory environment where it is needed to follow the standard recommendations [8]–[10] to provide reliable results. The use of subjective scales is necessary for image evaluation. This is due to the correlation between the physical characteristic of the stimulus generated by the video and the stimulus itself. Thus, methods were developed for different purposes. Some of them are absolute category rating (ACR), where the video sequence is presented once at a time, and after each, the subject is requested to rate on a category scale. Absolute category rating with hidden reference (ACR-HR) is similar to the previous method, but it is necessary to include a reference version as any other test stimulus. This is known as a hidden reference and is used to calculate a differential quality score (DMOS) comparing each test sequence with its hidden reference. Degradation category rating (DCR) is a method where the video sequence needs to be presented in pairs. The first one is the source reference, and the second one is the same source but processed by the system that is being tested [8].

Nevertheless, the subjective video quality assessment can be highly laborious and financially expensive. In this context, QoE test crowdsourcing gained some notoriety based on outsourcing tasks via the internet [11]. Even if it solves some issues of subjective assessment with a virtual lab and QoE crowd testing, some new problems arise that are difficult to solve, such as limitations generated by the internet, network

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3101288, IEEE Access

reliability, and following standard recommendations during tests [5], [12].

In order to design a test in which real operating conditions are better imitated, a parallel task can be used [13]. These procedures can reveal the subjective perception differences in multitasking scenarios, which are common both for leisure human activities (e.g., watching video/TV news while exercising in the gym or movie watching while cooking) and for critical professional tasks (online event participants, video-operators in defense and public safety areas, etc.).

The proposed testing procedure was designed mainly for naïve subjects and intended to create a general notion about subjective video QoE.

Section II describes parallel tasks and various types of existing parallel tasks. Section III includes State of the Art methods for video quality testing. The motivation is presented in Section IV. Test preparation, used video materials, and testing procedures are described in Sections V, VI, VII respectively. Section VIII includes preliminary, group, and RMSE data analysis, as well as statistically significant results. In Section IX, the discussion about achieved results is presented. Section X shows the conclusion of the article.

II. PARALLEL TASK

A parallel task is an activity that must be performed simultaneously with the test, and it must be designed to simulate a scenario closer to the real in the controlled environment of the laboratory. Parallel tasks can be divided into mentally oriented tasks, physically oriented tasks, and hybrid tasks [13],[14].

A. MENTALLY ORIENTED TASKS

Mentally oriented tasks are activities that require the subjects to perform mental activities without significantly affecting their physical condition. Some common tasks are logical quizzes, math calculations, and memory-oriented tasks that require the subjects to memorize and subsequently repeat the information.

B. PHYSICALLY ORIENTED TASKS

Physically oriented tasks are activities that require the subjects to perform physical activities without significantly affecting their mental condition. Some common physical tasks are running, cycling, or other sporting activities. It is also possible to measure the effort of the subjects by using sensors as EEG and ECG, and some tests like blood or saliva tests.

C. HYBRID TASKS

Hybrid oriented tasks are activities that require the subjects to perform both physical and mental activities. e.g., driving a car [14], shooting in a laser shooting simulator [15], and also some psycho motors tasks as tasting activities [14].

III. STATE OF THE ART

Various studies about subjective video Quality of Experience (QoE) are available in the scientific literature.

In [16], a subjective QoE was performed to quantify viewer responses to various effects that appear in low bit rate videos. Nineteen subjects in the age range of 18 to 30 participated the test. The peak signal to noise ratio (PSNR) comparison was provided. Various severe artifact effects were analyzed.

Another experiment [17] proposes a new quality metric for subjective testing suitable for mobile video broadcasting applications. The provided experiments showed that for the video rate control purposes, the PSNR does not suit the subjective quality data. The new quality metric accounts for both encoding parameters and intrinsic video sequence characteristics.

The research [18] deals with creating a new mobile video quality database containing 174 videos afflicted with distortions caused by 26 different stalling patterns.

The next study [19] presents a Full HD and Ultra HD video database with consideration of various levels of rate adaptation and rebuffering distortions following with a subjective evaluation of those videos. Next, a QoE evaluation framework comprising a learning-based model during playback and when the network conditions are good and compensate for exponential model during rebuffering was presented. An objective QoE evaluation based on the constructed database was provided.

The experiment [20] deals with subjective and objective quality assessment of compressed 4K UHD videos in an immersive viewing environment using various popular codecs.

In [21], a general classification-based prediction framework for selecting the preferred multidimensional adaptations operations was presented.

Another experiment [22] deals with a web browsing QoE assessment under distractions such as commuting in public transport and watching a video on a TV. Although the authors of the experiment implemented various distractions to the tests, these distractions have an ambient influence on the subjects (public transport). Only watching TV can be considered as mentally oriented tasks -the term “parallel task” is assumed to have an “active” load on subjects as it is described in Section II.

The experiment [23] deals with two multitasking scenarios (chatting and watching irrelevant videos) in various environments while studying a biological video. An EEG analysis was done to measure the working memory, retention, subjective cognitive load, perceived mental effort, and objective cognitive load. In terms of multitasking, as in the previous experiment, the authors used multitasking scenarios, which included exclusively mental activities.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3101288, IEEE Access

IV. MOTIVATION

In many real-life scenarios the video quality can be essential for accomplishing various professional tasks. Distortions and artifacts occurring in the physical channel can strongly impact the users' perception on overall video quality. They also can affect the outcome of numerous video-related tasks.

This paper aims to confirm or reject the following hypothesis: Test persons subjectively evaluate the perceived video quality (using the MOS scale) equally regardless of any other potential cognitive-motor activity that is performed in parallel to the video test.

V. PREPARATION

A. PARALLEL TASK DESIGN

A hybrid-oriented task was chosen as a parallel task, where each subject received three bowls. Two of them were empty, and the other was filled with small, equally shaped objects (non-boiled pasta) in three different colors (natural, dyed red and dyed white). The pasta elements differing in color (only) were chosen to force the test subjects to use their vision to distinguish between different element categories, not being able to sort them using haptic information only.

The subjects were supposed to sort the objects so that each bowl should contain only one color at the end of the reorganization. However, the object amount was chosen so that it would not be possible to complete the task before the end of the test.

The subjective tests were done in accordance with ITU-T P.910 [8] with additional parallel task according to ETSI TR 103 503 [24] recommendation for providing subjective tests with hybrid cognitive-motor tasks.

B. MAIN TASK PREPARATION

The primary task of the test included two videos (Video 1 and Video 2) and a demonstrative video for the subject briefing. The videos were prepared in the open-source editing video software Shotcut [25]. Each video consisted of samples about animals in nature. This neutral content was chosen to obtain a rate given by the subjects influenced exclusively by their subjective video quality perceptions. The content neutrality was verified by expert testing performed prior to the subjective tests. After each sample, a slide containing the sample number was displayed for five seconds. Descriptions of the samples are provided in Table I.

TABLE I
DESCRIPTIONS OF THE VIDEO SAMPLES

Sample	Description
1	A bird eating a berry
2	Monkeys jumping from a wood
3	A lizard on a leaf
4	A bird taking off a tree
5	Close-up capture of birds
6	A bird and a bear
7	Kangaroos in a meadow
8	Ducks swimming in a lake
9	Close-up capture of a bird
10	Ostriches in a field
11	A yawning tiger
12	Another capture of an ostrich
13	A bird and an alligator
14	Antelopes in a field
15	A panther and a hedgehog
16	Meerkats and an elephant
17	Rhinos in a forest
18	A rhino is drinking water
19	A sneaking leopard
20	Birds drinking water

The demonstrative video (Demo) was prepared with only five samples. Each one approximated one of the five groups of video qualities according to the Opinion Scores range shown in Table II. These samples and their distortion types and amounts have been set by an expert pre-test to mimic typic contemporary impairments experienced in videos transmitted by wireless networks (3G, LTE, Wi-Fi). The sample order was random, means not related to the video quality, and the group of each was not indicated.

TABLE II
QUALITY GROUPS

Group	Targeted video quality	Original resolution (pixels)	Amount
1	*Bad*	*256x144* *426x240*	*2* *2*
2	*Poor*	*640x360* *854x480*	*2* *3*
3	*Fair*	*1280x729*	*3*
4	*Good*	*1920x1080* *2560x1440*	*3* *1*
5	*Excellent*	*2560x1440* *3840x2160* *7680x4320*	*1* *2* *1*

The Video 1 and Video 2 were prepared with the same 20 samples but in different random orders. Such number of samples was used to prevent the subjects from remembering the content of the videos, so it does not have an impact on their votes. The original samples had various video resolutions (from 256x144px to 7680x4320px). This way it was ensured that the original videos in fourth and fifth groups are indeed originally high-quality videos. All videos were

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

TABLE III
EFFECTS PER GROUP

Group	Mosaic		Choppy	Old film: Scratches				Blur: Exponential	Glitch				Noise: Keyframes	Noise: Fast
	Width (%)	Height (%)	Number of Frames	Width	Amount	Darkness	Lightness	Amount	Frequency (%)	Block Height (%)	Shift Intensity (%)	Color Intensity (%)	Amount (%)	Amount (%)
1	0,0	0,9	1	2	4	32	40	46.9	0.8	2.0	11.2	2.5	-	-
2	0,7	0,5	2	-	-	-	-	-	-	-	-	-	-	-
3	0,7	0,4	-	-	-	-	-	-	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-	-	-	-	-	4.0	7.0
5	-	-	-	-	-	-	-	-	-	-	-	-	-	-

afterwards exported to 1280x720 pixels to match the tablets screen resolutions.

The selection of video impairments was arbitrary. It mixes resolution changes, and spatial and temporal artifacts to achieve subjectively realistic samples, spanning the quality range approximately uniformly. This approximate uniform span of the quality groups has been determined by informal expert watching tests performed by the experiment authors during the test design phase. Therefore, the further described experiment is not examining differences in influence of different video impairment types but rather the differences between subjective scores under both subjective test conditions (means with and without parallel task.

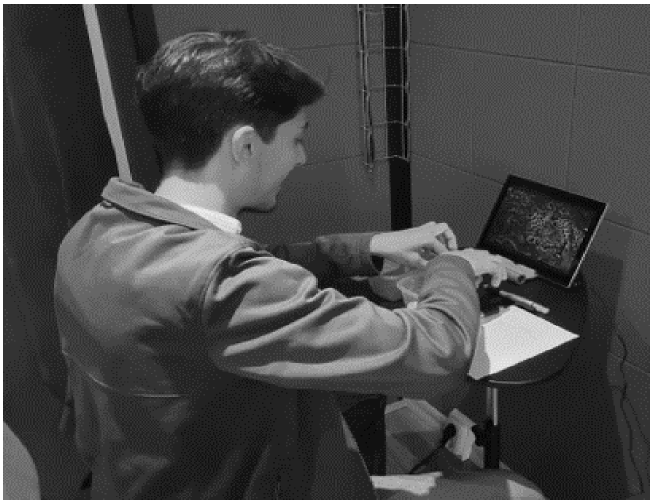

FIGURE 1. Subjective video quality test procedure

Besides the differences among resolutions, distinct distortion types, intensities, and effects were added to the videos according to their correspondent quality groups, as shown in **Error! Reference source not found.**. These video distortions were added by the standard features of Shotcut software (mosaic – for pixelating the samples by a defined percentage, choppy – for removing several frames from the samples, etc.). The full list of features is available in the software documentation. Since the fifth group is the reference group of the best quality, no distortion was added to it.

VI. MATERIAL

PC-tablets with 1280x720 (pixels) resolution were used to present the video material. Their screen brightness was calibrated prior to the test, according to P.910 requirements. Pens and forms, in which each line was corresponding to a sample and each column - to a score from 1 to 5 (1- Bad, 2 - Poor, 3 - Fair, 4 - Good, 5 - Excellent), were used.

During the multitasking test, bowls (and portions of raw pasta in the natural, dyed red, dyed white colors) were used.

VII. PROCEDURE

The first played file was always the demonstration video (Demo). The second and third files (Video 1 and Video 2) were designed for the evaluation. Because of the laboratory capacity limits, the experiment was carried out in three different sessions, with a total of 24 subjects. The satisfactory number of 24 – 32 participants for subjective tests was suggested by [26] and stated in [27].

In order to eliminate the learning effect, the test subjects were divided into two groups. It is important to mention that each of 24 subjects assessed the same 20 samples (24 subjects per condition). In the first group, the subjects were asked to watch the video "Demo" to make them acquaint with the test procedure and to create their internal reference/experience. They were asked to evaluate the perceived video quality on the common scale of five discrete grades, according to TABLE II. Then the subjects were asked to evaluate the second video (Video 1) without any other simultaneous task performed during the watching test. As previously explained, after each sample, a slide was presented asking for the Opinion Score entry related to the previous video clip. The duration of this voting phase was 5 sec. After evaluating all 20 samples of Video 1, there was a short break before starting the last stage. This last part of the experiment consisted of evaluating the third video (Video 2) and simultaneously performing the cognitive-motor parallel task, as explained in Section IV).

The test procedure followed by the second group of subjects performed the parallel task during watching the second video (Video 1), and Video 2 has been watched and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3101288, IEEE Access

evaluated without any activity performed in parallel. The order of video clips in the single-tasking and multitasking was swapped. All other factors, such as the order of the video and explanations, were kept identical. The choice to change this order was based on preventing the results obtained from being masked by other factors (i.e., learning effect) that were not related to the test's purpose.

FIGURE 1 illustrates the testing procedure with the designed parallel task.

VIII. DATA ANALYSIS

The following section is about the data analysis of the subjective test results (MOS values) with and without parallel tasks. The parallel task was designed to occupy the subjects' attention and requires a part of it for all the test duration. Also, it was not necessary to monitor the performance of each subject's parallel task, but only to confirm that they were dealing with both tasks.

After the tests, a database was created containing the number assigned to the individual, the type of test (single-tasking or multitasking), the number of the sample evaluated, and the score assigned to the sample so that the data could be analyzed. Pre-processing was carried out so that the marks attributed by the individuals due to shifting, multiple answers, or other errors were not considered.

A. PRELIMINARY ANALYSIS

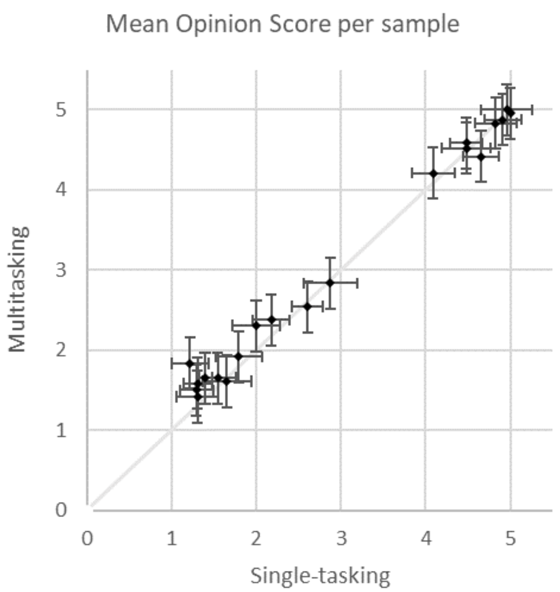

FIGURE 2. Mean Opinion Score per sample (Single-tasking versus multitasking). Error bars show the CI95 confidence intervals.

For an initial analysis, an average of the scores given by all test participants was made for each sample with and without parallel activity. This analysis can be seen in FIGURE 2. The samples above the diagonal line had their averages improved during the parallel task. The samples under the line have their average decreased on the multitasking test. Although this type of analysis shows the variation of grades for each sample, it did not show any significant results. Thus, group analysis was also made.

B. GROUP ANALYSIS

In the group analysis, for each of the groups already specified, an average score was made based on each sample's average scores from the previous analysis.

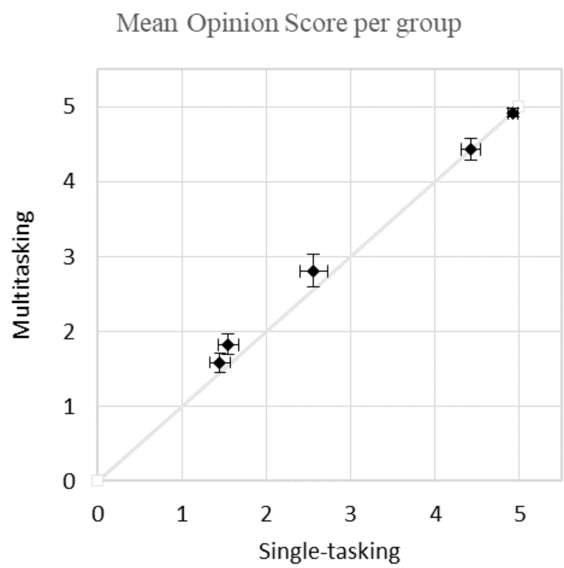

FIGURE 3. Mean Opinion Score per group (Single-tasking versus multitasking). Error bars show the CI95 confidence intervals.

In Table IV, the MOS values with and without parallel task and the CI95 confidence interval for each group are shown.

This second analysis already reveals the expected results obtained in the experiment. In the low-quality video groups (group 1 and group 2), the subjects' perception of quality was more affected by the second task than the groups of average (group 3) and good quality (group 4 and group 5).

TABLE IV
MEAN PER GROUP

Group	MOS with Parallel task ± CI95	MOS without Parallel task ± *CI95*
1	1.447 ± 0.122	1.585 ± 0.129
2	1.543 ± 0.122	1.822 ± 0.136
3	2.557 ± 0.166	2.806 ± 0.218
4	4.424 ± 0.112	4.432 ± 0.141
5	4.924 ± 0.055	4.917 ± 0.056

C. RMSE ANALYSIS

This analysis presents the Root mean square error (RMSE), which is a comparison between the predicted value and the real value, in this case, a comparison between the scores with and without the parallel task.

FIGURE 3 displays the scores with and without the parallel task for each group and also the confidence interval

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3101288, IEEE Access

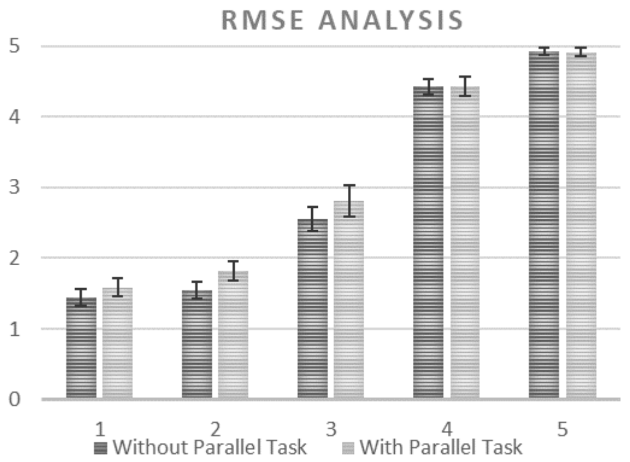

FIGURE 4. **RMSE Analysis per group. Error bars show the CI95 confidence intervals.**

for these values. Error bars show the CI95 confidence intervals. FIGURE 3 is also a scatter plot as in FIGURE 2. However, in this graph, the data are represented by the score mean of each group. FIGURE 4 represents the RMSE analysis per group for both tests. The error bars show the CI95 confidence intervals.

Even though the group 1 and 2 had similar values for the single-tasking test, in the multitasking test is more apparent, the distribution in five distinct value ranges. It is possible to see that the first two groups were the ones in which the scores were more significantly affected by the second task considering the confidence intervals. Nevertheless, only the second and fifth groups had non-overlapping confidence intervals, making it a statistical relevance result. In agreement with [14], the high and medium-quality video subjective perception was not substantially interfered by the second task, while the low-quality scores showed an improvement during multitasking. This can have two explanations. The first one is that high and medium-quality subjective perceptions always are quite the same even if the subject is not entirely focused on the evaluation, whereas the tolerance for bad quality is increased. The second explanation is that some distortions that were not constant over time as the choppy effect (**Error! Reference source not found.**) were barely perceived since the subject's attention to the video had decreased.

D. STATISTICALLY SIGNIFICANT RESULTS

To eliminate the type 1 error in the results and verify the significance of the results, statistical tests were performed in order to reject the null hypothesis. For this purpose, Mann-Whitney U-test was carried out for all the Groups according to [28]. The test is used to compare the means of two independent groups of samples.

For the selected set of data with a consideration of the Confidence Interval 95% (α=0.05), the critical range for rejection of the null hypothesis (z value) is from -1.96 to +1.96.

TABLE V shows the z-values for each quality group and rejected or not rejected the null hypothesis. The data analysis indicates that the observed z score for the quality group 2 is -2.705, which is outside of the critical range of our 0 hypothesis (claiming the groups are identical).

These results once more prove the initial findings that the parallel task can significantly affect provided subjective testing results.

TABLE V
Z-VALUE FOR EACH QUALITY GROUP

Group	z-value	Null hypothesis
1	-1.414	Do Not Reject
2	-2.705	Reject
3	-0.345	Do Not Reject
4	-0.586	Do Not Reject
5	-0.858	Do Not Reject

IX. DISCUSSION

After the analysis of the results, the following statements were made:

- For higher quality samples, no differences were observed; only the variance increases.
- For lower quality samples, there is an evident trend of a higher MOS value for tests with the parallel task. For videos with time impairment (condition 2), this difference is statistically significant, which is demonstrated both by non-overlapping CI95 values and in the subsequent U-test, which in this case rejects the 0-hypothesis.
- It can also be seen that the two conditions, designed to provide quality levels approximately MOS=3 and MOS=4, were evaluated differently by naïve subjects, causing non-uniform spread of samples across quality range (no samples achieved MOS value located between point 3 and 4). This gap, however, has no significant influence to further results analysis, the only effect may be slight increase in the overall test variance (as variance is always lower closer the scale ends), which is not a subject of this test analysis.

The observation shows that the video services users are less dissatisfied with the low-quality videos caused by the time varying impairments, which may be a guide to information for planners of networks dedicated primarily for entertainment.

Of course, the reduced criticality of the viewers does not reduce the risk of losing significant information during professional or critical activities such as telesurgery or special (defense/public safety) operations.

X. CONCLUSION

Two subjective video experiments were carried out, with and without parallel tasks, and their results were compared. Results analysis shows that high-quality samples are evaluated in the same manner, even in the context of a parallel task. In contrast, low-quality samples are evaluated less critically (more positive) with the parallel task. This phenomenon is stronger for video quality impairments that are not permanently present but occasionally only affect the subjective quality. Further studies will provide a more detailed comprehension of how distinct video issues affects the human subjective perception of video differently.

The obtained results are useful for more realistic interpretations of different analyses (e.g., Return-On-Investment analyses versus Quality of Service provided) performed in the context of multitasking users, typically important for mobile operators and VoIP applications and devices developers.

REFERENCES

[1] A. Khan, L. Sun, and E. Ifeachor, "QoE Prediction Model and its Application in Video Quality Adaptation Over UMTS Networks," *IEEE Trans. Multimed.*, vol. 14, no. 2, pp. 431–442, Apr. 2012, doi: 10.1109/TMM.2011.2176324.

[2] Quan Huynh-Thu and M. Ghanbari, "Temporal Aspect of Perceived Quality in Mobile Video Broadcasting," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 641–651, Sep. 2008, doi: 10.1109/TBC.2008.2001246.

[3] M. Li, Z. Chen, and Y.-P. Tan, "On quality of experience of scalable video adaptation," *J. Vis. Commun. Image Represent.*, vol. 24, no. 5, pp. 509–521, Jul. 2013, doi: 10.1016/j.jvcir.2013.03.006.

[4] A. Raake, *Speech Quality of VoIP*. Chichester, UK: John Wiley & Sons, Ltd, 2006.

[5] M. Shahid, "Methods for Objective and Subjective Video Quality Assessment and for Speech Enhancement," 2014.

[6] Z. Wang, "Applications of objective image quality assessment methods," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 137–142, 2011, doi: 10.1109/MSP.2011.942295.

[7] P. Mohammadi, A. Ebrahimi-Moghadam, and S. Shirani, "Subjective and Objective Quality Assessment of Image: A Survey," no. December 2015, 2014, [Online]. Available: http://arxiv.org/abs/1406.7799.

[8] ITU-T Rec. P.910, "Subjective video quality assessment methods for multimedia applications," *Subj. video Qual. Assess. methods Multimed. Appl.*, 2008.

[9] ITU-R Rec. BT.814-1, "Specification and alignment procedures for setting of brightness and contrast of displays," *Int. Telecommun. Union, Recomm. BT.814-1*, pp. 1–5, 2018.

[10] ITU-R Rec. BT.500-14, "Methodologies for the subjective assessment of the quality of television images," *Int. Telecommun. Union, Recomm. BT.500-11*, no. BT.500-14, pp. 1–102, 2019, doi: https://www.itu.int/rec/R-REC-BT.500-14-201910-I/en.

[11] S. Möller, "ITU-T PSTR-CROWDS Subjective evaluation of media quality using a crowdsourcing approach," 2018.

[12] T. Hossfeld *et al.*, "Best practices for qoe crowdtesting: Qoe assessment with crowdsourcing," *IEEE Trans. Multimed.*, vol. 16, no. 2, pp. 541–558, 2014, doi: 10.1109/TMM.2013.2291663.

[13] European Telecommunications Standards Institute, "TS 103 106 - V1.5.1 - Speech and multimedia Transmission Quality (STQ); Speech quality performance in the presence of background noise: Background noise transmission for mobile terminals-objective test methods," 2018.

[14] J. Holub, J. Sula, L. Soares, and O. Slavata, "Subjective audio quality testing, with tasting and car driving as parallel tasks," *IEEE Access*, vol. 6, pp. 1–1, 2018, doi: 10.1109/ACCESS.2018.2873568.

[15] H. Avetisyan and J. Holub, "Subjective speech quality measurement with and without parallel task: Laboratory test results comparison," *PLoS One*, vol. 13, no. 7, p. e0199787, Jul. 2018, doi: 10.1371/journal.pone.0199787.

[16] M. Masry and S. S. Hemami, "An analysis of subjective quality in low bit rate video," in *IEEE International Conference on Image Processing*, 2001, vol. 1, pp. 465–468, doi: 10.1109/icip.2001.959054.

[17] R. Feghali, F. Speranza, D. Wang, and A. Vincent, "Video quality metric for bit rate control via joint adjustment of quantization and frame rate," *IEEE Trans. Broadcast.*, vol. 53, no. 1, pp. 441–446, Mar. 2007, doi: 10.1109/TBC.2007.891700.

[18] D. Ghadiyaram, J. Pan, and A. C. Bovik, "A Subjective and Objective Study of Stalling Events in Mobile Streaming Videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 1, pp. 183–197, Jan. 2019, doi: 10.1109/TCSVT.2017.2768542.

[19] N. Eswara *et al.*, "A Continuous QoE Evaluation Framework for Video Streaming over HTTP," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 11, pp. 3236–3250, Nov. 2018, doi: 10.1109/TCSVT.2017.2742601.

[20] M. Cheon and J. S. Lee, "Subjective and Objective Quality Assessment of Compressed 4K UHD Videos for Immersive Experience," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 7, pp. 1467–1480, Jul. 2018, doi: 10.1109/TCSVT.2017.2683504.

[21] Y. Wang, M. Van Der Schaar, S. F. Chang, and A. C. Loui, "Classification-based multidimensional adaptation prediction for scalable video coding using subjective quality evaluation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 10, pp. 1270–1279, Oct. 2005, doi: 10.1109/TCSVT.2005.854224.

[22] D. Guse, S. Egger, A. Raake, and S. Moller, "Web-QOE under real-world distractions: Two test cases," in *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, Sep. 2014, pp. 220–225, doi: 10.1109/QoMEX.2014.6982322.

[23] Ö. Örün and Y. Akbulut, "Effect of multitasking, physical environment and electroencephalography use on cognitive load and retention," *Comput. Human Behav.*, vol. 92, pp. 216–229, Mar. 2019, doi: 10.1016/j.chb.2018.11.027.

[24] ETSI TR 103 503, "Speech and multimedia Transmission Quality (STQ); Procedures for Multimedia Transmission Quality Testing with Parallel Task including Subjective Testing," *Etsi*, pp. 1–17, 2018, [Online]. Available: https://portal.etsi.org/webapp/WorkProgram/Report_WorkItem.asp?WKI_ID=51341.

[25] "Meltytech, LLC. Shot Cut." https://shotcut.org/.

[26] M. H. Pinson *et al.*, "The influence of subjects and environment on audiovisual subjective tests: An international study," *IEEE J. Sel. Top. Signal Process.*, vol. 6, no. 6, pp. 640–651, 2012, doi: 10.1109/JSTSP.2012.2215306.

[27] ITU-T Rec. P.800, "Methods for subjective determination of transmission quality," *ITU-T Recomm.*, 1996.

[28] K. Black, *Business Statistics for Contemporary Decision Making sixth edition*, vol. 53, no. 9. 2013.

JAN HOLUB received his Ph.D. degree from the Czech Technical University in Prague in 1999 in Measurement and Instrumentation and become the full professor there in 2016. He is currently the head of the Department of Measurement there. His research areas cover digital signal processing, objective and subjective quality measurement and perception modelling methods. Author of several international patents and leading contributor to multiple telecommunication recommendations (ETSI, 3GPP).

HAKOB AVETISYAN received his Ph.D. degree from the Faculty of Electrical Engineering at Czech Technical University in Prague in 2021. He received his master's degree in Communication Systems from the State Engineering University of Armenia in 2014.

PAULO VICTOR LIMA is currently pursuing the master's degree with the Universidade Federal do Rio de Janeiro. His specialization is electronics and computing engineering.