

The process model of subjective quality assessment of videoconferencing in enterprise

Jana Filanová^{1,*}

¹Ekonomická univerzita v Bratislave, FPM, Dolnozemska cesta 1/b, 852 35 Bratislava, SR

Abstract. Videoconferencing is one of the most appropriate ways to transmit information online to participants not only during quarantine. This paper describes a novel process of evaluating the quality of videoconference. Time-consuming subjective measurements were supported by models and programs that simplified the preparation, testing, and processing of results. The process of quality assessment can help IT management to decide on the use of videoconferencing as a means of communication in business. This is especially important in times of pandemics and quarantine.

1 Introduction

Videoconferences represent a form of synchronous communication based on audio and video transmission with the possibility to integrate text and other forms of presentation of information at a distance. The quality of this communication is influenced by the used communication technologies and transmission characteristics of communication networks. Videoconferencing is one of the most appropriate ways of online transmission information to participants. The videoconferences could be recorded and it is possible to view the records even in off-line mode [1].

Videoconferencing allows to conduct meetings with several people at the same time. We can talk and chat with people from different parts of the world at the ease of your residence or office. Business owners can connect with clients and employees and hold discussions. Strategic planning and meetings which require several members to be present can be carried out with ease. We get the option of talking to multiple people at the same time and carry out your business strategies with ease. Thus projects get completed faster and deals get finalized within a short period of time.

As information overload becomes a more and more serious topic in modern society, many efficient tools have been designed to overcome information anxiety. Videos, the fastest-growing information carrier, will account for more than 80% of all Internet traffic by 2020 [2]. Video data has become easily accessible and dominant in the present time [3].

The visual perception of people is a highly complex matter that involves several mechanisms. It is influenced by their expectations and their previous experience. The view of the quality is linked to their mechanisms of imagination. In [4] the authors show that semantics has a significant impact on viewers' sensitivity to the quality of a video sequence

* Corresponding author: jana.filanova@euba.sk

for spatially separated parts of the sequence and, more importantly, that this difference in sensitivity can be changed by the presence of an audio signal. This result is important for any testing of subjects' responses to visual material. One example is the subjective assessment of the quality of video in an audio-visual communications system (such as television or videoconferencing) [4].

The results of the research [5] have shown that the presence or absence of audio has a significant impact on the overall subjective perception of the videoconferencing quality. It has also been found that the viewer is more sensitive to the quality of the image in the foreground of the speaking person than to the quality of the image in the background. If there are multiple people in the scene, even not speaking right now, the viewer is likewise more sensitive to the quality of the image of the captured people than to the quality of the image in the background. Video should supplement, not replace, the telephone, for which there is a considerable evidence base from research studies and some guidance [6].

Subjective quality cannot be represented by an exact figure. Due to its inherent subjectivity, it can only be described statistically. Even in psychophysical threshold experiments, where the task of the observer is just to give a yes/no answer, there is a significant variation in contrast sensitivity functions and other critical low-level visual parameters between 50 different video quality observers. When the artefacts become supra-threshold, the observers are bound to apply different weightings to each of them [7].

2 Research methods

The subjective assessment is based on the use of human observers (people) who watch the sequences and score the video quality. It is the most reliable way to determine the video quality and should not be replaced with an objective assessment. The disadvantage of this method is that it is time consuming and human resources are needed [8].

International recommendations for subjective methods of quality testing include specifications on how to implement different types of subjective tests. Some of these test methods are known as "double stimulus" methods where an observer evaluates quality or quality change between two (reference and test) video sequences. There are also "single stimulus" methods where the observer evaluates the quality of just one (test) video sequence [9, 10].

2.1 DSCQS method

The Double Stimulus Continuous Quality Scale (DSCQS) method is suitable for measuring the quality of the system that is related to the reference value as the observer is not familiar with the reference sequence order. DSCQS is quite sensitive to small differences in quality and is thus the preferred method when the quality of the test sequence and reference sequence are similar [9].

2.2 DSIS method

The Double Stimulus Impairment Scale (DSIS) method is suitable for assessing the extent of degradation of the test sequence as compared to the reference one, especially in case of visible/significant degradation. For example, it is used to evaluate the degradation of the sequence during transport. This method is faster than DSCQS since the sequences are displayed only once. Subjects rate the amount of impairment in the test sequence on a discrete five-level scale ranging from "very annoying" to "imperceptible". The DSIS

method is well suited for evaluating clearly visible impairments such as artefacts caused by transmission errors [9].

2.3 ACR method

The Absolute Category Rating (ACR) method is a single stimulus method; viewers only see the video under test, without the reference. They give one rating for its overall quality using a discrete five-level scale from “bad” to “excellent”. The fact that the reference is not shown with every test clip makes ACR a very efficient method compared to DSIS or DSCQS, which take almost 2 or 4 times as long, respectively [10].

2.4 MPEG-4 H.264/AVC compression standard

The latest and today most used compression standard designed for a wide range of applications, ranging from mobile video to HDTV, is MPEG Part 10, called also MPEG-4 H.264/AVC. MPEG-4 H.264/AVC defines the Profiles and Levels, too, but its organization is much simpler than in MPEG-4 Part 2. There are only three Profiles currently defined (Baseline, Main, Extended) [8].

3 Process model of subjective quality assessment of video sequences

In this research, we simulated an environment of a real packet network and tested video sequences that would simulate the diverse content of video calls. We artificially degraded the quality of video sequences by packet loss and jitter. The objective of the test was to compare subjective methods, which was used. We also wanted to show that semantics has a significant impact on viewers' sensitivity to the quality of the video sequence [11]. We describe a novel process of evaluating the quality of videoconference. Time-consuming subjective measurements were supported by models and programs that simplified the preparation, testing, and processing of results.

The process of quality assessment can help IT management to decide on the use of videoconferencing as a means of communication in business. This is especially important in times of pandemics and quarantine [6].

Subjective video quality testing is difficult not only because of the time-consuming nature of testing itself but also due to the complexity of the steps that precede the actual testing. Figure 1 describes five steps of the process model we have designed for subjective evaluation of the quality of video sequences. It is based on the process model that we presented in the article [12].



Fig. 1. The process model of video sequences quality evaluation

3.1 Recording and coding of test sequences

Reference video sequences tests were created based on real video calls. These sequences were recorded using the Logitech C270 web camera with HD resolution of 1270 x 720 pixels,

utilizing the Logitech Webcam Software shipped with the web camera. Due to the purpose of the testing, it was important to create diverse demonstrations with a different emphasis on content, the importance of video or audio capture. Two types of reference video sequences are described below.

In the first test sequence (video sequence No. 1), the intention was to create a preview where the emphasis would be on the picture detail. The manager in this video preview informs their employees that if they have any questions, they can contact him at his e-mail address. The person in the preview does not pronounce this e-mail address but writes it on the board. So the only way this e-mail address information gets to the user of the videoconference is that the image quality will be sufficient to recognize it without difficulty.

In the second test video sequences, the aim was to create a demonstration where an emphasis would be placed on the quality of the audio during static image transfer. In the video sequence No. 2, a woman asks the recipient to contact someone by phone. She dictates her name and phone number.

Each video sequence was encoded, because the video and audio formats used, as well as bit rates, do not match those used in videoconferencing. Recording and coding technical parameters of reference video sequences are described in Table 1.

Table 1. Recording and coding technical parameters

parameter	recording	coding
pixel	1270 x 720	1270 x 720
frame rate	15	30
sequence length [sec]	10	10
video format	WMV2	MPEG-4 AVC
audio format	WMA	AAC
bit rate of video [kbit/sec]	3535	1024

3.2 Degradation of test sequences

To introduce degradations into the reference videoconferencing sequences, it was necessary to emulate the transfer environment through which the sequences were transmitted (Figure 2). Network emulation is a process by which we can control and repeatedly simulate network performance. The changes in network parameters such as latency and packet loss are provided by traffic shapers. They must be controlled according to predefined specifications to simulate the required features of the network.

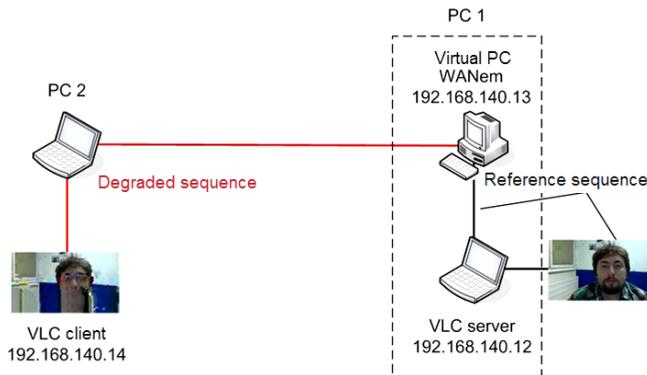


Fig. 2. Network model for creation of degraded sequences

Each of the four reference samples was degraded by packet loss (0.5 %, 1 %, 3 %, 5 %, and 10 %) and jitter (50 ms jitter at 100 ms latency).

3.3 Selection of appropriate methods

Absolute Category Rating (ACR) and Double Stimulus Impairment Scale (DSIS) methods were selected for the subjective evaluation of video samples. The ACR method has the advantage of being fast as the evaluator watches the sample only once and the length of the sample is relatively short (about 10 seconds). The DSIS method was also selected because of its time efficiency and the ability to capture more accurate differences between degraded samples, as we also have a reference sample for this method [9, 10]. The choice of suitable methods was also influenced by the fact that both the ACR's and DSIS's outputs are MOS scores with values ranging from 1 to 5, so the results can easily be compared [13].

3.4 Preparation of test scenarios and selection of respondents

Since the testing was performed within the VLC multimedia player environment, it was necessary to create playlists in which the individual video sequences were arranged appropriately. To prepare the scenarios and the course of the subjective measurements, a program was created in the C# programming language. To play a video sequence the program uses an open-source DmediaPlayer that is a modification of the VLC player [14]. The program consists of two parts: test manager part and tester part (Figure 3). Test manager part is an interface used to create structure of the test. You can choose the type of subjective method, test sequence, reference sequence (if necessary) and enable or disable sound step-by-step.

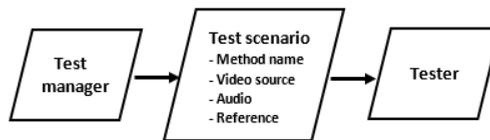


Fig. 3. The proces model of testing and test scenarios preparation program

The ITU-T Recommendations specify that the number of respondents for subjective quality assessment must be greater than 4 and less than 40 [9, 10]. Based on this, we chose 20 respondents (10 women and 10 men), aged 20-51.

The fifth step of the subjective quality evaluation includes testing. The course of testing, evaluation, and comparison of the results are described in the following section.

4 Comparison of subjective evaluation results

Our research aimed to compare subjective methods of real video calls testing and to determine the degree of impact of audio quality on overall video quality with respect to the semantics.

Due to the time-consuming manual processing of results, two programs were created. The first program was written in the C# programming language. The program has two outputs in the form of text files. In the first one, the results are processed according to the evaluation of the individual sequences. In the case of the DSIS method, the format is the reference sequence name and the test sequence name followed by five numbers. In the case of the ACR method, the format is the test sequence name and five numbers. The five numbers correspond to the evaluation scale of the given methods [9, 10]. If it has been

chosen to take the statistical processing into account, the output is in the same file. In the second text file, the results are processed according to respondents who evaluated individual sequences. This output is needed for statistical processing of results for the DSIS method.

The second program is used for statistical processing of measured results. It was created in Matlab version R2008b. The average score \bar{u}_{jkr} is calculated for each test sequence

$$\bar{u}_{jkr} = \frac{1}{N} \sum_{i=1}^N u_{ijkr} \quad (1)$$

where u_{ijkr} is the respondent score i for test condition j , sequence k and number of repetitions r . N is the total number of respondents. One of the output of the program is a text file with the names of the individual respondents who were excluded on the basis of the calculation and comparison of the standard deviation.

Table 2 lists the summary of video sequences quality measurement results. In the case of subjective evaluation, the video sequences were rated by MOS scores that range from 1 to 5 [9, 10]. A video sequence rated by the score of 4 or higher is considered to be of high quality [13, 15].

Table 1. Results of subjective quality evaluation (d.a. – degraded audio, r.a. – reference audio)

Video sequences	Degradation	ACR	ACR	DSIS	DSIS
		d.a.	r.a.	d.a.	r.a.
No. 1	Packet loss 0.5%	4.70	4.65	4.80	4.75
	Packet loss 1%	3.55	3.90	4.15	3.90
	Packet loss 3%	2.25	3.15	3.15	3.30
	Packet loss 5%	2.15	2.65	1.95	2.85
	Packet loss 10%	1.00	1.95	1.05	2.35
	Latency 100 ms, Jitter 50 ms	3.65	x	3.75	x
No. 2	Packet loss 0.5%	2.45	3.25	3.25	3.70
	Packet loss 1%	2.10	3.40	2.30	3.80
	Packet loss 3%	1.55	2.65	2.05	3.20
	Packet loss 5%	1.70	2.20	1.35	2.25
	Packet loss 10%	1.10	1.70	1.00	1.85
	Latency 100 ms, Jitter 50 ms	4.15	x	3.85	x

The *ACR d.a.* and *DSIS d.a.* columns show the results for video sequences degraded by packet loss (0.5 %, 1 %, 3 %, 5 %, and 10 %) and jitter (50 ms jitter in 100 ms latency). The *ACR r.a.* and *DSIS r.a.* columns show the results for video sequences degraded by packet loss and in which the degraded audio track was replaced by the audio track from the reference sequence.

Comparing the evaluations for the ACR and DSIS methods, we found that video sequences were rated by a higher score when the DSIS method was used. This difference can be explained by the fact that in the case of the DSIS method the respondent was influenced by the reference sample. Even at 0.5 % and 1 % packet loss degradation, some video sequences with the reference audio track received higher ratings than those with the original disturbed audio track. The results also imply that, in general, the degradation caused by jitter (50 ms jitter in 100 ms latency) does not affect the quality ratings as much as the degradation due to packet loss.

The results of the subjective quality evaluation have shown that under the ideal conditions in the transmission network (without packet loss and latency) the quality of videoconferencing has been rated as "good" (MOS > 4). Therefore, from the perspective of

the user, the video frame resolution, audio and video bitrate, and the used codecs provide the user with sufficient quality.

Based on the results of the subjective evaluation of the sequences with the original audio and the sequences in which the degraded audio was replaced by the reference, we see (table 2) that for the packet loss of 3 % and 5 % the sequences with the reference audio are rated much higher (often by more than 1 point on the MOS scale). The difference between individual sequence evaluations is much smaller in samples with the reference audio compared to sequences with the original audio track. In our research, we have confirmed that the quality of audio has a great impact on the overall quality of videoconferencing. In future work, we can investigate whether a similar trend is observed when changing the tasks, that is, if we gradually insert different deteriorated audio tracks into the reference video sequence.

5 Conclusion

Increasing bit rates in today's modern networks allow us to provide ever-new services to support enterprise communications such as video conferencing, on-demand streaming or online streaming [16]. Like voice services, moving image services need to be monitored to see if the service is of adequate quality to the customer. This quality monitoring must necessarily be automated, as it would of course be impractical, costly and error prone to employ people for these activities.

In this paper we describe a novel process of evaluating the quality of videoconference. Time-consuming subjective measurements were supported by models and programs that simplified the preparation, testing, and processing of results. The process of quality assessment can help IT management to decide on the use of videoconferencing as a means of communication in business. This is especially important in times of pandemics and quarantine.

Based on the results of the subjective evaluation of sequences with the original audio track and the sequences in which the degraded audio track was replaced by the audio track from the reference sequence, we have confirmed that the quality of the audio has a significant impact on the overall quality of videoconferencing and the understanding of its content. As a result, if any video information is supported by relevant audio information, we can compensate for the loss of video information by improving audio quality. We can also influence the quality of videoconferencing by ensuring correct pronunciation, intelligibility, and articulation.

The paper was elaborated within VEGA No. 1/0388/20 IT Management in Enterprises in Slovakia: International Standards and Norms Versus Individual Business Processes – proportion 100 %.

References

1. P. Bisták, et al.: *Utilisation of Videoconferencing for Education*. 1 st . ed., Elfa Kosice (2005)
2. J. Fajtl, H. S. Sokeh, V. Argyriou, D. Monekosso, P. Remagnino. *Summarizing videos with attention*. In ACCV Workshops (2018)
3. R. Kapoor, O. Mishra, O., M. Tripathi, Human action recognition using descriptor based on selective finite element analysis, *Journal of Electrical Engineering*, **70**, 443-453 (2019)

4. M. R. Frater, J. F. Arnold, A. Vahedian, Impact of audio on subjective assessment of video quality in videoconferencing applications, *IEEE Transactions on Circuits and Systems for Video Technology*, **11**, 9, 1059-1062 (2001)
5. O. O. Andriichenko, O. I. Denysenko, Subjective evaluation of the clarity of the noisy language in the lecture room, *Electronic and Acoustic Engineering*, **2**, 3, 55-60, (2019)
6. T. Greenhalgh, J. Wherton, S. Shaw, C. Morrison. *Video consultations for covid-19 BMJ 2020*, 368 :m998 (2020)
7. T. T. Hai Uyen, O. Kwon, S. Choi, I. Hussain, Subjective Assessment of 360° Image Projection Formats, in *IEEE Access*, **8**, 33588-33599 (2020)
8. M. Uhrina, J. Hlubik, M. Vaculík, Correlation between Objective and Subjective Methods Used for Video Quality Evaluation, *Advances in Electrical and Electronic Engineering*, **11**, (2013)
9. ITU-R Recommendation ITU-R BT.500-11, *Methodology for the subjective assessment of the quality of television pictures*, (2012)
10. ITU-T Recommendation ITU-T P.910, *Subjective video quality assessment methods for multimedia applications*, (2008)
11. S. Han, Z. Xi, Dynamic scene semantics SLAM based on semantic segmentation, *IEEE Access* (2020)
12. J. Filanová, M. Mardiak, Meranie kvality video signálu, *Elektrorevue*, **15**, pp. 32.1-6, (2010)
13. H. Rizek, K. Brunnström, K. Wang, B. Andrén, M. Johanson, Subjective evaluation of a 3D videoconferencing system. *Proc. SPIE 9011, Stereoscopic Displays and Applications XXV*, (2014)
14. W. Jackson, Digital Video: Streaming Video, MediaPlayer, and MediaController Classes. *Android Apps for Absolute Beginners*, 291-337 (2017)
15. S. Katsigiannis, J. Scovell, N. Ramzan, L. Janowski, P. Corriveau, M. Saad, G. Wallendael, Interpreting MOS scores, when can users see a difference? Understanding user experience differences for photo quality. *Quality and User Experience*, **3**, (2018)
16. P. Kuppusamy, Modern Network Analytics Architecture Stack to Enterprise Networks, *International Journal for Research in Applied Science and Engineering Technology*, **7**, (2019)