Paper

Revising and Improving the ITU-T Recommendation P.912

Mikołaj Leszczuk

AGH University of Science and Technology, Department of Telecommunications, Krakow, Poland

Abstract—It was once thought that high Quality of Service (QoS) performance solves recurrent problems of low-quality multimedia services. Since then, solutions have been proposed to ensure a high level of Quality of Experience (QoE). In this paper, the author attempts to outline an understanding of an accurate meaning of multimedia services quality. Starting from QoS and passing through generalized QoE, the author focuses on subjective aspects and objective quality modeling and optimization of visual performance for Target Recognition Video (TRV) applications (such as video surveillance), to outline the ITU-T standardization path in this area. The revising the ITU-T Recommendation P.912 is proposed to reflect improved subjective test techniques developed since this Recommendation was approved. Also at least some existing errors of reasoning are predicted, which are likely to become evident for the industry in the next decade. Finally, the author invites all researchers working on topics related to TRV to join him in the process of improving P.912.

Keywords—CCTV, ITU-T, P.912, QoE, QoS, TRV.

1. Introduction

A decade ago, the telecommunications industry believed that high-performance Quality of Service (QoS) techniques resolve any recurrent problems of low-quality multimedia services. However, within a few years, it became clear that optimization of QoS parameters such as throughput, packet loss, delay, or jitter is not the best way of improving the quality experienced by users. The problem of low bandwidth can be compensated by more efficient codecs. The impact of packet loss is strongly dependent on their distribution, and the use of redundancy coding and transmission. For many applications, buffering multimedia data streams can alleviate major delays and jitter.

Since discovering that QoS is not a sufficient metric of network quality, most proposals have been suggesting that quality should be measured on the user level. This process was named Quality of Experience (QoE) [1], [2]. Such a measurement calls for special structures (frameworks) of quality of video sequences integrated assessment [3]. These structures are increasingly being filled with solutions that attempt to model the overall quality, operating at the intersection of QoS and QoE [4] or only in QoE. However, it has become obvious that such a general approach simply does not work for many visual applications such as target recognition (utility) applications (video surveillance, telemedicine, remote diagnostics, fire safety, backup cameras, games, etc.) [5], [6].

In fact, QoE – the way of perceiving multimedia services quality – depends on a number of objective and subjective contextual parameters [7]. Only a full understanding, usually only possible with strong area limitations of the QoE modeling application, makes it possible to obtain results consistent with the expectations of service users, and, consequently, to optimize quality [8]. Unfortunately, high numbers of contextual parameters mean this research question is still open.

2. Target Recognition Video

In many visual applications, the quality of the motion picture is not as important as the ability of the visual system to perform specific tasks for which it is created, given the processed video sequences. Such sequences are called Target Recognition Video (TRV). Regardless of the different ways in which the concept of TRV quality is understood, its verification is necessary to perform dedicated quality testing. The basic premise of these tests is to find TRV quality limits for which the task can be performed with the desired probability or accuracy.

Such tests are usually subjective psychophysical experiments with a group of subjects. Unfortunately, due to issue complexity and relatively poor understanding of human cognitive mechanisms, satisfactory results of TRV quality computer modeling have not yet been achieved beyond very limited application areas.

Given the use of TRV, qualitative tests do not focus on the subject's satisfaction with the video sequence quality, but instead they measure how the subject uses TRV to accomplish certain tasks. Purposes of this may include:

- video surveillance – recognition of vehicle license plate numbers,
- telemedicine/remote diagnostics – correct diagnosis,
- fire safety – fire detection,
- rear backup cameras – parking the car,
- games – spotting and correctly reacting to a virtual enemy.

The human factor is a significant influence, therefore it is necessary to ask questions on the procedures to be com-

plied with in order to make a subjective assessment of TRV quality. In particular, questions arise on:

- method of selecting the TRV source from which the test TRV (with degraded quality) arises,
- subjective testing methods and the general manner of conducting the psychophysical experiment
- method of selecting a subjects group in the psychophysical experiment, especially identification of any prior task knowledge,
- training subjects before the start of the experiment,
- conditions in which the test will be carried out,
- methods of statistical analysis and presentation of results.

3. Methods for Subjective Evaluation of TRV

Questions formulated in the previous section are addressed by Recommendation ITU-T P.912 "Subjective Video Quality Assessment Methods for Recognition Tasks", published in 2008 [9]. In addition, Recommendation P.912 organizes terminology related to subjective TRV testing, introducing appropriate definitions for the testing methods (psychophysical experiments).
Unfortunately, Recommendation P.912 is only the first step in the standardization of subjective TRV testing methods. In the author's opinion, based on available research results and observations conducted during numerous experiments with TRV, many claims of Recommendation P.912 are formulated at too high generality level. What's more, selected statements are not supported by research results and are significantly disputable. In this situation, a number of steps have been taken to introduce significant modifications (amendments) to the Recommendation. For this purpose, in order to formalize the procedures, the author has established collaboration with the Polish Ministry of Administration and Digitization, and received a formal nomination as a delegate of the Polish government. The procedure for submitting amendments commenced in 2014. The detailed scope of the proposed amendments to Recommendation P.912 is discussed in the following subsections.

3.1. Source Signal

Introduction: in Clause 5, Recommendation P.912 states:

Test sequences should follow the general principles stated in [10] and [11], which specify that scenes should be consistent with the transmission service under test, and should span the full range of spatial and temporal information. It is critical for the nature of these evaluations that the stimuli used actually reflect the true operational parameters of the conditions under which the video material is collected, and cover the entire range of scenarios possible for the application area that one is identifying. Unlike other subjective assessment methods developed for quality evaluations, this method is directed at the usefulness of the video material to complete a task and not the quality of the video itself.

Unfortunately, in certain cases, data availability is very limited. Let us consider the impact of studying the quality of still images on the accuracy of X-ray diagnosis of bone fractures. It is clear that due to the low frequency of certain types of fractures, the availability of a database of corresponding images is very low.
Another example concerns research on the impact of CCTV recordings on the accuracy of license plate recognition [8]. For the purposes of this study, a special video database was created [12]. The recordings have been created using fixed CCTV cameras, recording cars entering the car park at the AGH University of Science and Technology in Krakow, Lesser Poland (Fig. 1). Again, it is clear that due to the abovementioned conditions of acquisition, recordings represent a particular CCTV camera, its specific location and direction, a specific distance from the object, and lighting conditions. What's more, since the recordings were made in Krakow, most of the license plates have the letter "K" (distinguishing the Lesser Poland province) in the first position on the plate and "R" (distinguishing the Krakow county) in the second position.

Fig. 1. Source signal.

As shown, contrary to Recommendation P.912, it is very difficult to ensure complete coverage of the potential applications of the recordings. Any record database expansion is laborious, time-consuming, or even impossible. This does not mean that the cited studies are useless. However, their applicability must be explicitly limited to the scope of the recordings database. Unfortunately, literature frequently includes attempts to extrapolate the applicability of test results (in particular among less experienced researchers), which the author believes may be due to the fact that issues in Recommendation P.912, which frequently include instructions to carry out tests, are not addressed explicitly.

Proposal: the author proposes the introduction of the following amendments to Clause 5 of Recommendation P.912:

Test sequences should follow the general principles stated in [10] and [11], which specify that scenes should be con-

sistent with the transmission service under test, and should span the full range of spatial and temporal information. It is critical for the nature of these evaluations that the stimuli used actually reflect the true operational parameters of the conditions under which the video material is collected. ***If the stimuli used cannot actually cover*** *the entire range of scenarios possible for the application area that one is identifying,* ***the application description needs to be explicitly limited. For example, the results should not be generalized****. Unlike other subjective assessment methods developed for quality evaluations, this method is directed at the usefulness of the video material to complete a task and not the quality of the video itself.*

3.2. Testing Methods and Experimental Design

For videos used to perform a specific task, it may not be appropriate to rate video quality according to a subjective scale such as Absolute Category Rating (ACR) [10]. The goal of test methods for TRV is to assess the viewer ability to recognize the appropriate information in the video, regardless of his perceived quality of the viewing experience. To assess the quality level of TRV, methods that reduce subjective factors and measure the participant ability to perform a task are useful in that they avoid ambiguity and personal preference.

The TRV application is directly related to the user ability to recognize targets at increasing levels of detail. These levels are referred to as discrimination classes (DCs). When determining the DC for particular scenarios, one must consider that for a set distance from the camera to the object of interest, the DC directly correlates to decreasing video resolution of the target, and therefore the object is represented by fewer cycles per resolution degree. Fewer cycles per resolution degree also means that the object subtends less of the information content of the video, making the target identification more difficult.

Experimental methods should consist of responding to questions related to the content in the image or video. The parameter addressed by the question is the target to be recognized.

3.2.1. Multiple Choice Method

Introduction: in Clause 6.1, Recommendation P.912 states:

The number of choices offered to the viewer will depend on the number of alternative scenes being presented. "Unsure" may be one of the listed choices.

It should be noted that subjects tend to abuse the "unsure" response. This problem has been observed when applying a Comparison Category Rating (Table 1), as defined in Recommendation ITU-T P.800 [13], in which subjects tend to abuse the response "0" ("about the same"). A similar trend was observed independently in author's TRV studies.

Unfortunately, Recommendation P.912 is missing a clear warning against the prudent use of the "unsure" response (Recommendation P.912 even encourages its use).

Table 1
Comparison Category Rating (CCR)

3	Much better
2	Better
1	Slightly better
0	About the same
–1	Slightly worse
–2	Worse
–3	Much worse

Proposal: it is proposed that the entry in Recommendation P.912 should be amended as follows:

The number of choices offered to the viewer will depend on the number of alternative scenes being presented. ***The use of*** *"unsure"* ***as*** *one of the listed choices* ***is discouraged but allowed. The experimenter should be aware that individual subjects tend to overuse the "unsure" choice, leading to contamination of results. Consequently, special care must be taken when "unsure" is one of the listed choices****.*

3.2.2. Single Answer Method

Introduction: in Clause 6.2, Recommendation P.912 states:

If there is a non-ambiguous answer to an identification question, the single answer method may be used. This method is appropriate for alphanumeric character recognition scenarios. A viewer is asked what letter(s) or number(s) was present in a specific area of the video, and the answer can be evaluated as either correct or incorrect.

It should be noted that, contrary to Recommendation P.912, it is also possible to apply fuzzy logic [8]. For scenarios where the recognition result is an alphanumeric string, assistance may come from measuring differences between two strings using the Hamming distance (applicable only for strings of the same length) [14], or Hamming distance's generalization – the Levenshtein distance [15]. Using the experiment shown in Fig. 2 as an example, results containing no more than one error may be regarded as correct [8]. This is because even in the event of a plate being recognized incorrectly, by correlating it with a vehicle database containing the make and vehicle colour, the risk of the vehicle being identified incorrectly is substantially reduced.

Proposal: the author proposes that the description of the single choice method be expanded as follows:

If there is a non-ambiguous answer to an identification question, the single answer method may be used. This method is appropriate for alphanumeric character recognition scenarios. A viewer is asked what letter(s) or number(s) was present in a specific area of the video, and the answer can be evaluated as either correct or incorrect. ***Alternatively, fuzzy logic may be used (e.g. Hamming distance or Levenshtein distance), as shown in [8].***

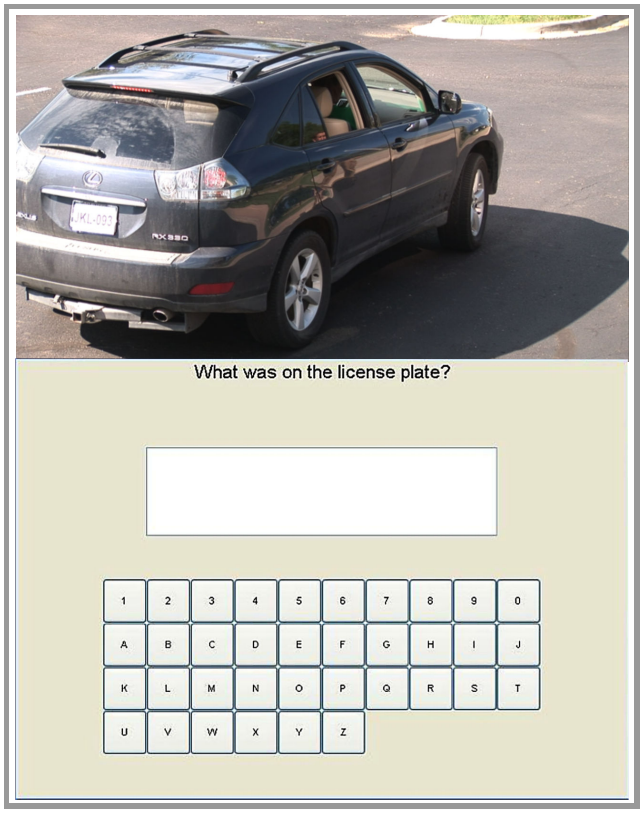

Fig. 2. Single answer method.

3.3. Subjects

Introduction: in Clause 7.3, Recommendation P.912 states:

Subjects who are experts in the application field of the target video recognition should be used. The number of subjects should follow the recommendations of [10].

In order to verify this finding, experiments testing subjects' ability to recognize certain objects (mobile phone, flashlight, gun, mug, radio, aluminum soda can, electric "Taser" stun gun) shown in video sequences were carried out. In the first experiment, the subjects were experts – law enforcement officers [16], [17]. When the experiment was repeated with non-experts, very similar results were obtained, as long as the non-experts were compensated for their time [18].

Proposal: the author proposes an entry introduction which allows the use of non-expert subjects providing they are motivated in an appropriate manner (such as being paid for their time). Naturally, this is only possible for certain areas of testing, since non-experts subjects cannot be used in tests associated with (for example) medical diagnostics.

Subjects who are experts in the application field of the TRV should be used. ***For certain areas of application testing, where neither specific experience nor expertise is required, non-expert subjects may also be used. Such non-experts must be motivated in an appropriate manner (e.g. being paid for their time). The validity of this approach is shown in [18].*** *The number of subjects should follow the recommendations of [10].*

4. Conclusions and Future Work

The discussion of statements contained in ITU-T Recommendation P.912 shows that some of the findings and observations require the certain provisions verification of the Recommendation. The author proposes to revise Recommendation P.912 to reflect improved subjective test techniques developed since this Recommendation was approved. Sufficient justification exists to support a new ITU-T work item, and contributions to this topic have been encouraged by ITU-T.

Ultimately, the amended recommendations should have a broader scope: to expand target testing methods, provide better instruction and training of subjects, improve conditions for testing, statistical analysis and reporting, and extend the applicability of techniques in the field of crowdsourcing for the subjective assessment of the quality of TRV. In cooperation with the US National Telecommunications and Information Administration (NTIA, originator of the Recommendation), there are also plans to expand the Recommendation to include metrics of Video Acuity, created at the NASA Vision Group [19]. The author would like to invite all researchers working on TRV-related topics to join him in the process of improving P.912.

Acknowledgements

This work is funded by the Polish National Centre for Research and Development, contract no. C2013/1-5/MITSU/2/2014 under the EUREKA international programme: Next Generation Multimedia Efficient, Scalable and Robust Delivery.

References

[1] E. Cerqueira, S. Zeadally, M. Leszczuk, M. Curado, and A. Mauthe, "Recent advances in multimedia networking," *Multim. Tools and Appl.*, vol. 54, no. 3, pp. 635–647, 2011.

[2] M. Grega, L. Janowski, M. Leszczuk, P. Romaniak, and Z. Papir, "Quality of experience evaluation for multimedia services", *Telecomm. Rev.*, vol. 81, no. 4, pp. 142–153, 2008.

[3] M. Mu, P. Romaniak, A. Mauthe, M. Leszczuk, L. Janowski, and E. Cerqueira, "Framework for the integrated video quality assessment", *Multim. Tools and Appl.*, vol. 61, no. 3, pp. 787–817, 2012.

[4] M. Leszczuk, L. Janowski, P. Romaniak, and Z. Papir, "Assessing quality of experience for high definition video streaming under diverse packet loss patterns", *Signal Proces.: Image Commun.*, vol. 28, no. 8, pp. 903–916, 2013.

[5] M. Leszczuk, I. Stange, and C. Ford, "Determining image quality requirements for recognition tasks in generalized public safety video applications: Definitions, testing, standardization, and current trends", in *Proc. 2011 IEEE Int. Symp. Broadband Multim. Sys. Broadcast. (BMSB),* , Nürnberg, Germany, 2011, pp. 1–5.

[6] S. Möller and A. Raake, Eds., *Quality of Experience: Advanced Concepts, Applications and Methods*. Cham: Springer, 2014.

[7] K. Brunnström *et al.*, "Qualinet White Paper on Definitions of Quality of Experience", Fifth Qualinet General Meeting, Novi Sad, Rep. of Serbia, Mar. 12, 2013.

[8] M. Leszczuk, "Optimising task-based video quality", *Multim. Tools and Appl.*, vol. 68, no. 1, pp. 41–58, 2014.

[9] "Subjective video quality assessment methods for recognition tasks", ITU-T P.912, 2008.

[10] "Subjective video quality assessment methods for multimedia applications", ITU-T P.910, 1999.

[11] "Digital Transport of Video Teleconferencing/Video Telephony Signals – Video Test Scenes for Subjective and Objective Performance Assessment", ANSI T1.801.01, 1995.

[12] M. Leszczuk and L. Janowski, "Database for video quality assessment in license plate recognition", in *Proc. Sig. Proces.: Algorithms, Architectures, Arrangements, and Applications SPA 2013*, Poznan, Poland, 2013, pp. 51–55.

[13] "Methods for subjective determination of transmission quality", ITU-T P.800, 1996.

[14] R. Hamming, "Error detecting and error correcting codes", *Bell System Tech. J.*, vol. 26, no. 2, pp. 147–160, 1950.

[15] V. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals", *Soviet Physics Doklady*, vol. 10, p. 707, 1966.

[16] VQiPS, "Video quality tests for object recognition applications", Public Safety Communications DHS-TR-PSC-10-09, U.S. Department of Homeland Security's Office for Interoperability and Compatibility, June 2010.

[17] VQiPS, "Recorded-video quality tests for object recognition tasks", Public Safety Communications DHS-TR-PSC-11-01, U.S. Department of Homeland Security's Office for Interoperability and Compatibility, June 2011.

[18] M. Leszczuk, A. Kon, J. Dumke, and L. Janowski, "Redefining ITU-T p.912 recommendation requirements for subjects of quality assessments in recognition tasks", in *Multimedia Communications, Services and Security*, A. Dziech and A. Czyzewski, Eds., *Communications in Computer and Information Science*, vol. 287, pp. 188–199. Berlin-Heidelberg: Springer, 2012.

[19] A. Watson, "Video acuity: A metric to quantify the effective performance of video systems", in *Imaging Systems and Applications ISA 2011*, Toronto, Canada, 2011.

Mikołaj Leszczuk started his professional career in 1996 at Comarch Company as manager of the Multimedia Technology Department. Since 1999 has been employed at the AGH Department of Telecommunications. In 2000 he moved to Spain for a scholarship at the Universidad Carlos III de Madrid. After returning to Poland, he was employed at the Department of Telecommunications as a research and teaching assistant. In 2006 he successfully defended his Ph.D. as an assistant professor. His current research interests are focused on multimedia data analysis and processing systems, with particular emphasis on Quality of Experience. He has authored over 100 scientific publications. He has participated more than 20 major research projects. Between 2009 and 2014, he was the administrator of the major international INDECT research project, dealing with solutions for intelligent surveillance and automatic detection of suspicious behavior and violence in urban environments. He is a member of VQEG (Video Quality Experts Group), IEEE, and GAMA (Gateway to Archives of Media Art).
E-mail: leszczuk@agh.edu.pl
AGH University of Science and Technology
Department of Telecommunications
Mickiewicza st 30
30-059 Krakow, Poland