Standardizing Quality Measurement for Video Services

Stefan Winkler, Symmetricom, USA
swinkler@symmetricom.com

The industry needs accurate and reliable objective video quality measurement tools to assess and monitor the multitude of digital video services such as mobile broadcasting, Internet video, and IPTV. Overviews of such tools and algorithms can be found elsewhere [1,2].

Standards address a growing number of issues related to quality measurement for these applications, including definitions, terms of reference, requirements, recommended practices, test plans, and more. Here I will focus specifically on work related to objective video quality measurement methods [3]. More details on this topic can be found in a forthcoming paper by the author [4].

I use the term quality to signify media quality from the perspective of the viewer or consumer, with a focus on the perceived quality of the content (or sometimes more comprehensively, the overall user experience).

As a result of this definition of quality, subjective tests with actual viewers form the benchmark for objective quality measurement. Commonly used standards for subjective experiments on video quality include ITU-R Rec. BT.500 and ITU-T Rec. P.910, which suggest viewing conditions, criteria for the selection of observers and test material, assessment procedures, and data analysis methods. The quality ratings obtained from viewers in a subjective experiment are then averaged for each test clip into Mean Opinion Scores (MOS).

The standardization of objective quality measurement has three main objectives, namely meaningful, reliable, and reproducible prediction of MOS. Various standards bodies, industry forums and other groups are working on video quality assessment in one form or another.

The Video Quality Experts Group (VQEG) is perhaps best known for evaluating objective quality metrics [5]. It brings together algorithm developers and users to plan and execute validation tests of objective perceptual quality metrics with the help of independent labs. VQEG's approach to validation testing typically includes the creation of a database of video sequences, which are largely unknown to the model developers, and whose subjective ratings are to be estimated by objective video quality models. Model evaluation is then based on prediction performance according to a number of statistical criteria.

So far, VQEG tests have resulted in a number of ITU recommendations, including ITU-T Rec. J.144 and ITU-R Rec. BT.1683 for full-reference TV applications, as well as ITU-T Rec. J.247 and ITU-T Rec. J.246 for full-reference and reduced-reference multimedia applications, respectively. The evaluation of reduced-reference models for TV applications has been concluded recently, and an HDTV model test is currently being conducted. Tests of hybrid and audio-visual models are in the planning stage.

ITU-T Study Group 12 is working on a non-intrusive parametric model for the assessment of multimedia streaming (P.NAMS for short), which uses packet and codec information as inputs, but explicitly excludes any payload information. A follow-up project called P.NBAMS (B for bitstream) has similar goals, but will allow models to use payload information.

The QoS Metrics (QoSM) committee of the ATIS IPTV Interoperability Forum (IIF) deals with quality issues for IPTV. One of their current work items is the creation of a test plan and a test process for the evaluation of quality metrics. The test plan is intended as a basis for anyone who conducts an evaluation of objective quality models by subjective tests. The test process complements the test plan and proposes a standardized process for the evaluation of quality metrics, rather than the standardization of the quality measurement algorithms. The premise is that it is sufficient for an algorithm to be validated and to perform well, but the algorithm itself does not need to be standardized. The process enables on-demand validation of metrics at any time, thus encouraging innovation and more rapid advancement of the state of the art.

Finally, the Video Services Forum (VSF) started a Quality of Experience (QoE) Activity Group last year, which is working on expanding the IP-layer metrics defined in its existing report to payload-based metrics for video and audio content.

The increasing number of video quality metrics, from free algorithms to commercial products, is in fact becoming somewhat of an issue. Even standards such as the ITU Recommendations discussed above specify more than one algorithm, adding to the confusion of users. A more flexible on-demand validation process, like the one in preparation by ATIS, may further increase the number of available choices in objective quality measurement. While this is positive in terms of competition, it is important to assure comparability of quality measurement results across platform and metric choices, in order to maintain the primary standards objectives of meaningful and reproducible MOS values. This can be achieved by translating (or cross-calibrating) the output of one quality metric to another [6].

One successful outcome of standards work has been the collection of sequences from the VQEG FR-TV Phase I test, which still represents the only public database of video clips annotated with subjective ratings to date. VQEG also plans to release a subset of the HDTV database after test completion. There are other efforts outside the standards groups to make quality metrics more easily comparable through open databases of annotated video sequences. Two examples are the Laboratory for Image & Video Engineering (LIVE) of the University of Texas at Austin, which just released an annotated video quality database, and the MMSPG-PoliMI Subjective Video Quality Assessment Database created by the Politecnico di Milano, Italy, and the Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, both based on their own subjective experiments. The Consumer Digital Video Library (CDVL) (still under construction) has similar goals.

References

[1] S. Winkler, *Digital Video Quality – Vision Models and Metrics*. John Wiley & Sons, Jan. 2005.

[2] A. R. Reibman, "Is IPTV a quality experience? Quality monitoring in an IPTV network," *IEEE COMSOC MMTC E-Letter*, vol. 4, no. 5, June 2009.

[3] S. Winkler, P. Mohandas, "The evolution of video quality measurement: From PSNR to hybrid metrics," *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 660–668, Sept. 2008, invited paper.

[4] S. Winkler, "Video quality measurement standards – current status and trends," *Proc. International Conference on Information, Communications, and Signal Processing (ICICS)*, Macau, Dec. 7-10, 2009.

[5] K. Brunnström, D. Hands, F. Speranza, A. Webster, "VQEG validation and ITU standardization of objective perceptual video quality metrics," *IEEE Signal Processing Magazine*, vol. 26, no. 3, pp. 96–101, May 2009.

[6] M. H. Brill, J. Lubin, P. Costa, S. Wolf, J. Pearson, "Accuracy and cross-calibration of video quality metrics: New methods from ATIS/T1A1," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 101–107, Feb. 2004.

Stefan Winkler holds an M.Sc. degree in Electrical Engineering from the University of Technology in Vienna, Austria, and a Ph.D. degree from the Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland. He is currently Principal Technologist at Symmetricom. Prior to that, he was Chief Scientist of Genista, which he co-founded in 2001. He has also held assistant professor positions at the National University of Singapore (NUS) and the University of Lausanne, Switzerland.

Dr. Winkler has published more than 50 papers and is the author of the book "Digital Video Quality". He also serves as an Associate Editor for the IEEE Transactions on Image Processing. He has been an active contributor to the Video Quality Experts Group (VQEG) since it was founded in 1997 and is currently co-chair of the QoE Metrics Activity Group of the Video Services Forum (VSF).