

To Train or Not To Train?

Nicolas Staelens, Wendy Van den Broeck and Filip De Turck

The way standards do

As part of assessing (audio)visual quality by means of subjective experiments, specific instructions are provided on how to evaluate and rate the different video sequences. Also, a training session is used to familiarize the observers with the experiment and the type (and range) of impairments they can expect. As such, observers ‘know’ what to look for and what to expect. However, what about the influence of context and user expectations on quality perception?

It is generally known that subjective (audio)visual quality assessment experiments need to be conducted in stringent controlled environments, as detailed in ITU-T Rec. P.910 and ITU-R Rec. BT.500. This facilitates experiment repeatability, enables comparing results obtained from different experiments conducted at different locations, and minimizes the influence of contextual factors during quality evaluation.

Several subjective testing methodologies have already long been standardized. Notwithstanding their specific application domains (e.g. video, speech, conferencing, recognition), they all share common ground and require, amongst other things, that test subjects be properly informed about the experiment and the task at hand.

Prior to the start of the experiment, detailed instructions are provided to the observers explaining the intended application under test, the overall trial structure, and the quality rating mechanism. Furthermore, a training phase is incorporated in preliminary trials in order to illustrate the type and range of quality impairments that can occur during the experiment. Consequently, observers are primarily focused on active (audio)visual quality evaluation.

These methodologies are widely used in video quality research and ongoing VQEG projects to measure the video's *technical* quality as perceived by the test subjects.

What about Quality of Experience?

But what about measuring Quality of Experience (QoE), a buzzword associated with terms like *delight, user expectations, enjoyment, personality, service, content, and context of use*? To what extent are user expectations and context of use taken into account during standardized subjective quality assessment? Can the existing subjective quality assessment methodologies be used to measure QoE?

“Quality of Experience is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the user’s personality and current state.”

Le Callet P., et al (2012), "Qualinet White Paper on Definitions of Quality of Experience," European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland, Version 1.2, March 2013.

The mandatory training phase prepares subjects for the experiment and informs them about what to expect. By informing them, their

focus is aimed directly at evaluating the video quality as such. So, when we want to assess QoE, should we at all incorporate a training phase as part of our experiments? Or should we try to mimic realistic viewing behavior as much as possible?

Contextualized subjective experiments

With respect to our QoE research, we have conducted several contextualized subjective quality assessment experiments by integrating the everyday life context (Staelens et al., 2012; Van den Broeck et al., 2012). These experiments were conducted in complement to controlled lab tests in order to enable results comparison and study the influence of more ecologically valid

testing environments on quality perception. These studies have highlighted the importance of immersion and primary focus during subjective video quality assessment.

In one of our studies, users were asked to watch a full length DVD movie in their most natural environment, i.e. at home on

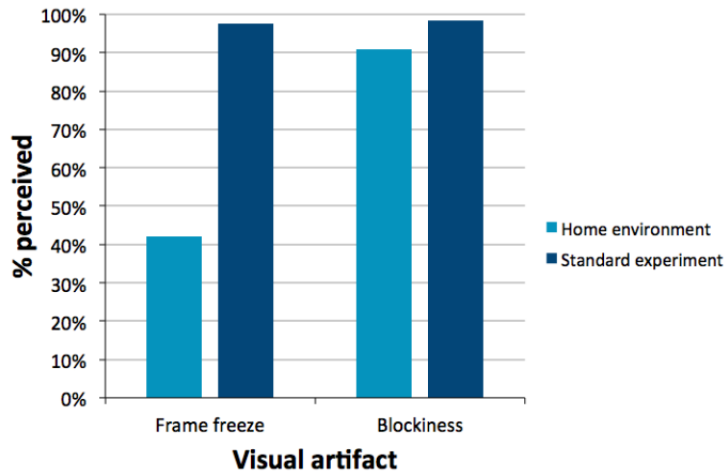


Figure 1. Influence of primary focus on impairment visibility.

their own device (Staelens et al., 2010). Users were not informed about the possible presence of visual artifacts during playback. Hence, primary focus shifted to watching the actual content of the movie. Compared to controlled lab experiments, impairments were less noticed (see Figure 1). However, despite the fact that blockiness is more easily detected

compared to frame freezes, subjects indicated that freezes are more disturbing during DVD playback. Freezes tend to break the natural flow of the movie and users feel their *immersive experience* is hampered. It is important to note that, due to the restrictions imposed by the ITU recommendations, the feeling of immersion cannot be (re)produced during controlled lab experiments, also because the duration of the video sequences is limited.

Controlled lab or real-life?

Matulin and Mrvelj (2013) also state that the most accurate QoE evaluations include real-life experiments in the typical environments where the services are used, without subjects actively being focused on (audio)visual quality assessment. Based on a comprehensive summary of QoE experiments conducted in real-life environments, the authors conclude that there are substantial differences between the results obtained in controlled labs and real-life environments. In general, users



Nicolas Staelens obtained his PhD degree in Computer Science Engineering in 2013 from Ghent University, Belgium. His research activities focus on assessing the influence of network impairments on perceived audiovisual quality. As part of his research, he has gained experience in conducting subjective experiments in real-life environments.



Wendy Van den Broeck is a senior researcher at iMinds-SMIT, Brussels, Belgium. She is active in the user Empowerment research unit and conducted research in different projects concerning the domestication of new media technologies in a home context. Her recent projects are related to multi-screen user practices, QoE research on video streaming, and research related to iDTV, 3D-TV, second screen applications and HDTV.



Filip De Turck leads the network and service management research group at the Department of Information Technology of the Ghent University, Belgium and iMinds (Interdisciplinary Research Institute in Flanders). His main research interests include scalable software architectures for network and service management, design and performance evaluation of novel QoE-aware multimedia delivery systems.

are more forgiving of quality degradations in real-life environments.

Thus, conducting experiments in real-life environments without really informing test subjects might yield more representative results in the case of investigating end-users' QoE.

In this respect, implementing methodologies like the Experience Sampling Method (ESM) (Hektner et al., 2007) might be the way to go in order to get a better understanding of QoE in real-life.

So ...

... *"To train or not to train (test subjects), that is the question."*

And for sure, the answer will depend on what we really want to assess.

References

Hektner, J.M., Schmidt J.A., Csikszentmihalyi M. (2007), "Experience Sampling Method: Measuring the Quality of Everyday Life", California: Sage Publications Inc.

Matulin M., Mrvelj Š. (2013), "State-of-the-Practice in Evaluation of Quality of Experience in Real-Life Environments", Promet – Traffic & Transportation, Vol. 25, No. 3, pp. 255-263.

Staelens N., Moens S., Van den Broeck W., Mariën I., Vermeulen B., Lambert P., et al. (2010), "Assessing Quality of Experience of IPTV and Video on Demand Services in Real-life Environments", IEEE Transactions on Broadcasting, Vol. 56, Issue 4, pp. 458-466.

Staelens N., Van den Broeck W., Pitrey Y., Vermeulen B., Demeester P. (2012), "Lessons Learned during Real-life QoE Assessment", in EuroITV – 10th European Conference on Interactive TV.

Van den Broeck W., Jacobs A., Staelens N. (2012), "Integrating the Everyday-life Context in Subjective Video Quality Experiments", Fourth International Workshop on Quality of Multimedia Experience (QoMEX).