Quality-of-Experience Perception for Video Streaming Services: Preliminary Subjective and Objective Results

Khalil ur Rehman Laghari[1], Omneya Issa[2], Filippo Speranza[2], and Tiago H. Falk[1]
[1]Institut National de la Recherche Scientifique (EMT-INRS), Montreal, QC, Canada
[2] Communications Research Centre Canada, Ottawa, ON, Canada

Abstract—**Quality-of-Experience (QoE) is a human centric notion that produces the blue print of human perception, feelings, needs and intentions while Quality-of-Service (QoS) is a technology centric metric used to assess the performance of a multimedia application and/or network. To ensure superior video QoE, it is important to understand the relationship between QoE and QoS. To achieve this goal, we conducted a pilot subjective user study simulating a video streaming service over a broadband network with varying distortion scenarios, namely packet losses (0, 0.5, 1, 3,7, and 15%), packet reorder (0, 1, 5, 10, 20, and 30%), and coding bit rates (100, 400, 600, and 800 Kbps). Users were asked to rate their experience using a subjective quantitative metric (termed Perceived Video Quality, PVQ) and qualitative indicators of "experience." Simulation results suggest a) an exponential relationship between PVQ and packet loss and between PVQ and packet reorder, and b) a logarithmic relationship between PVQ and video bit rate. Similar trends were observed with the qualitative indicators. Exploratory analysis with two objective video quality metrics suggests that trends similar to those obtained with the subjective ratings were obtained, particularly with a full-reference metric.**

I. Introduction

Today, we are witnessing continuous growth in multimedia services and applications. For example, video streaming has become a prominent method of media exchange with applications in video conferencing, video-on-demand, and e-learning. It is known that the quality of the streamed video can be affected by various network-dependent, application-specific, content-based, business- and context-oriented factors. Commonly, network-related parameters (e.g., packet loss rates) are grouped into a Quality-of-Service (QoS) rating. For multimedia service providers today, however, understanding the degree of influence of various QoS parameters on user quality of experience has become a priority. The burgeoning demand for understanding the end-user's quality requirements has led to the Quality-of-Experience (QoE) terminology [1].

QoE provides an assessment of human expectations, feelings, perceptions, cognition and satisfaction with respect to a particular product, service or application [1]. With video streaming, for example, services are commonly based on perceptual experiences and users are known to make aesthetic judgments quickly. A user's quality of experience of a service can evoke a wide range of emotions and attitudes. These emotions and perceptions impact the user's attitude towards the quality, content, advertised products, and price.

The greatest challenge today is to be able to measure and analyze QoE factors for different multimedia services with precision and accuracy. However, it is quite complex to capture QoE metrics considering the influence of multiple confounding factors, including technical, economic, social, and human (e.g., psychological, physiological). In addition to these, there are other prominent issues related to QoE measurement and analysis, such as:

Human Subjectivity: The important challenge is related to variability and complexity of human behavior; not all humans have similar preferences, feelings or perceptions about a particular service and furthermore, their perceptions and preferences continuously change over the time. The challenge is on how to capture human subjectivity and how to transform it into meaningful data. Human physiological and QoS parameters are easily monitored and engineered due to their quantitative nature but as human perceptions and feelings are inherently subjective and the levels of expectation vary between users, it is hard to quantify and measure QoE with complete accuracy.

Laboratory test vs. field test: Should user studies be conducted in controlled or in "living-lab" environments? Some experts believe that controlled setups do not provide a sense of a real environment, thus may limit the elicitation of human perceptions or feelings [2]. On the other hand, controlled testing provides better control over several influencing factors.

Subjective vs. objective methods: Subjective methods rely on human participants to provide useful and reliable QoE feedback about a particular multimedia service. Subjective testing, however, is expensive and time-consuming. Objective methods may involve purely technical factors and/or human factors. The former include objective metrics, which attempt to predict human behavior using a mathematical model/formula. Unfortunately, there are still no objective metrics that can fully capture the complexity of QoE. The existing metrics are generally limited to only some aspects, e.g., the picture quality, which are part of the QoE framework (and hence related to it), but disregard influential factors, such as contextual, economy, and user expectations, which are gathered via surveys and user studies. An example of such metrics is the peak signal-to-noise ratio (PSNR) metric which infers information about video quality (i.e., as related to picture quality) based on pixel differences between original and network-transmitted distorted coded video sequences. Objective human factors are related to the human physiological and cognitive systems [3]. These objective factors are difficult to obtain and interpret, but could provide useful insights into human behavior and cognition.

In this paper, we present preliminary results of a simulated video streaming service to a) better understand the impact of different QoS parameters on both subjective quantitative and qualitative QoE factors, as well as b) investigate the efficacy of existing objective metrics as correlates of subjective QoE parameters. Here, we explore two objective metrics, namely PSNR (a full-reference method which requires access to the original and degraded video files) and VisualMPEG (a no-reference method which only requires the degraded video file).

The remainder of the paper is organized as follows. Section II provides a brief overview of different QoE assessment methods. Sections III-IV presents the experimental setup and data analysis results and discussion, respectively. Lastly, Section V presents the conclusions.

II. Quality-of-Experience Background

QoE is based on several psychological and cognitive factors such as such as habits, moods, expectations, and needs. For service providers, it is important to quantify QoE and measure it with accuracy. Quantifying QoE means translating user perception and performance into statistical and interpretable values. There are two main methods for measuring and analyzing QoE, as depicted by Figure 1.

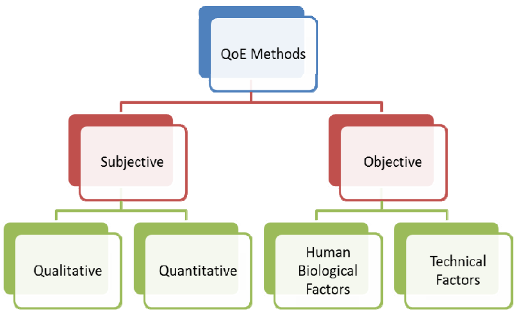

Figure 1. QoE Assessment Methods

A. *Subjective QoE Assessment Methods*

Subjective assessment methods are commonly based on surveys, interviews and statistical sampling of users and customers to analyze their perceptions and needs vis-à-vis service and network quality. There are two broader techniques for conducting subjective studies: (i) Qualitative techniques (ii) Quantitative techniques.

A.1 Qualitative Technique: Qualitative data represent verbal behavior and consist of words and observations, not numbers [4]. They produce individual's interpretation of events. Qualitative techniques capture human perceptions, feelings and opinions through verbal behavior. Open-ended survey questions, customer interviews, testimonials, comments on blogs, and social media produce the bulk of qualitative data. All of these methods produce a wealth of qualitative data, in the form of researcher notes, transcripts from interviews and focus groups, participant journals, photographs and more. The most meaningful metric for the analysis of verbal behaviors is the ratio of positive to negative comments [5] and it is also commonly known as CCA (catalog, categorize, analyze) framework. CCA categories the ratio of positive to negative comments and produces results in histogram formats and is explored here as a qualitative analysis technique. Other representative social science methods for the evaluation of qualitative data include bottom-up coding, which allows themes to emerge from the data, or top-down coding to identify existing constructs in a particular data set [6].

A.2 Quantitative Techniques: Quantitative factors are in the form of numbers and statistics. Surveys and user studies are normally conducted either in laboratory environment or in natural environment to measure human perceptions, feelings and intentions. These methods typically involve the construction of questionnaire with rating scales to produce quantitative data. These methods produce precise measurement and analysis of target concepts. The International Telecommunications Union (ITU) has produced various subjective study guidelines, such as ITU-T Recommendation P.910 [7] for video quality, P.800 [8] for speech quality, and G.1030 [9] for web traffic quality.

Another quantitative approach to evaluating the experience of technology usage builds upon existing psychological models. Technology acceptance models were first developed in the business arena for customer surveys. Technology acceptance models are used to predict a subject's likelihood of using a particular technology over time based on several measures. The most popular models such as the Technology Acceptance Model (TAM) [10], the Theory of Planned Behavior (TPB) [11] and the Unified Theory of Acceptance and Use of Technology (UTAUT) [12] consider "intention" as the main driving factor for technology acceptance or rejection. Intentions are normally triggered by some motivational factors, which influence human behavior. Knowledge of user acceptance and adoption trends for particular services and/or products is invaluable for service providers.

B. *Objective QoE Assessment Methods*

Traditionally, there are two classes of objective assessment methods, namely QoS/technology centric and human physiological/ cognitive-based techniques, as detailed below.

B.1 QoS/Technology centric techniques: In these approaches, QoE is predicted from QoS data using some mathematical estimation techniques and tools rather than getting direct feedback from end-users. The most popular are objective methods for the measurement of picture quality. These methods can be classified as Full Reference (FR), No Reference (NR), or Reduced Reference (RR) methods. FR methods compute the quality difference between an original (i.e., unprocessed) version of the image/video/audio signal and its distorted (i.e., processed) counterpart. NR methods estimate the quality of the signal using only the distorted version. Finally, RR methods have access to partial information (e.g., features) about the clean original signal in order to estimate the quality of its degraded counterpart.

The so-called peak signal-to-noise ratio (PSNR) is the simplest and most common FR objective measurement method; it is mostly used with encoded signals and it is based on the mathematical difference between every pixel of the processed video and the original video [13]. The Structural SIMilarity (SSIM) is another popular FR method, which compares information about luminance, contrast and structural similarity between the original and processed images [14]. Another well-known FR method is the Video Quality Metric (VQM), which was developed by the Institute for Telecommunications Sciences, National Telecommunications and Information Administration (TIS/ NTIA) [15]. This method has been standardized by the American National Standards Institute and adopted in two ITU Recommendations, namely ITU-T J.144 and ITU-R BT.1683. Current FR methods have reached a high level of development and generally exhibit good predictive performance. On the other hand, no-reference and reduced-reference methods are still in their infancy. For example, a few RR methods for standard- and high-definition digital broadcast television (ITU-R BT. 1885 and ITU-R BT.1908, respectively) have been standardized only recently. There are no standard NR methods. Nonetheless, there are a few commercial algorithms that have shown good correlations with subjective ratings. One such example is the VisualMPEG algorithm developed by BeSchuur (Germany). The algorithm provides several indices related to video quality, such as blurriness, blockiness, frame freezing, and packet loss effects, which are used to derive a measure for the quality as a whole [16]. In the VisualMPEG implementation used in our experiments, a thresholding rule is used and the outputs of the algorithm are five labels which coincide with the labels used in the absolute category rating scale described by ITU-T Recommendation P.800, i.e., 'bad', 'poor', 'fair', 'good', and 'excellent'.

B.2 Physiological and cognitive techniques: The human brain is the epicenter of every human action. Understanding human physiology, especially neurophysiology and cognitive science, may pave the way for a better understanding of the human behavior and its relationship with QoE. Neurophysiological insights may be obtained via neuroimaging techniques, such as electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and near-infrared spectroscopy (NIRS), as well as body area sensors and networks. EEG/MEG relies on measuring the electrical/magnetic activity in the brain. And fMRI and NIRS are based on tracking blood flow (and correlates) that accompanies neuronal activity. While EEG and MEG provide data with high time resolution (in the order of milliseconds), they provide limited spatial resolution. In contrast, fMRI and NIRS provide good spatial resolution but relatively poor temporal resolution [17, 18]. Table I provides a comparison of the different neuroimaging technologies.

TABLE I. COMPARISION OF DIFFERENT NEUROIMAGING TECHNOLOGIES

Method	Neuronal Activity	Hemoglobin Dynamics	Time resolution	Spatial Resolution	Subjects Mobility
EEG	Yes	No	ms	cm	Yes
MEG	Yes	No	ms	cm	Limited
fMRI	Yes	Yes	s	1mm	Limited
NIRS	Yes	Yes	ms	10mm	Yes

Each method has its own strengths and weaknesses, and no single method is best suited for all experimental conditions. Because of the limitations of individual techniques, researchers are increasingly trying to combine existing techniques in order to synthesize the strengths belonging to each method. For example, the authors in [19] have recently reported the use of NIRS and physiological bio signal sensors to characterize subjective image preferences; in [20] EEG was used to characterize videos of varying quality. In short, available neurophysiological/cognitive tools and techniques provide precise quantitative data about human cognition and behavior thus has sparked interests in the telecommunication community, particularly in their use to provide valuable insight about human multimedia quality perception.

In order to combine neurophysiological recording with perceptual quality assessment, care has to be exercised. For example, current guidelines for subjective quality assessment usually involve audio-visual stimuli of 8-12 seconds in duration and varying content; furthermore, users are normally free to move around their chair during the experimental protocol. With neurophysiological experiments, on the other hand, short-duration stimuli are commonly presented (in the order of 100-1000 ms), content variability is kept to a minimum (with the exception of odd-ball paradigms where varying content is used to trigger a mismatch negativity effect), and users are instructed to stay as still as possible, as to minimize movement artifacts in the collected data. Given these constraints, it is likely that new subjective study guidelines will have to be developed, thus allowing for accurate quantitative information about human behavior, emotions and cognition to be gathered and analyzed for different multimedia services. Only then will complete objective QoE measurement be made possible.

III. EXPERIMENTAL SETUP

In this section, we describe the testbed setup for evaluating the impact of different QoS parameters on video streaming QoE. First, a private local area network (LAN) was created using two laptops connected to a gateway via a switch. The open source media player VLC Player [21] was used for streaming the video and then receiving it at the receiver side. Figure 2 depicts the testbed setup used in the experiments described herein. Both laptops had 14 inch screen size and they were running Microsoft Windows operating system. For the gateway, Ubuntu was used to emulate the varying network conditions with 'NetEm' [22], an emulator which comes in most standard newer Linux distributions.

This testbed was used to emulate wireless environment in order to analyze the effects of varying network conditions on video streaming QoE; specifically we considered three QoS network parameters: packet loss (PLR), packet re-order (PRR), and video bit rate (VBR). The video test clip used was the popular fast moving rugby clip with CIF resolution and 8-second duration (see clip at [23]). Media content was encoded using the H.264/MPEG-4 video coder and streamed using UDP protocol over a broadband network with 100 Mbps bandwidth.

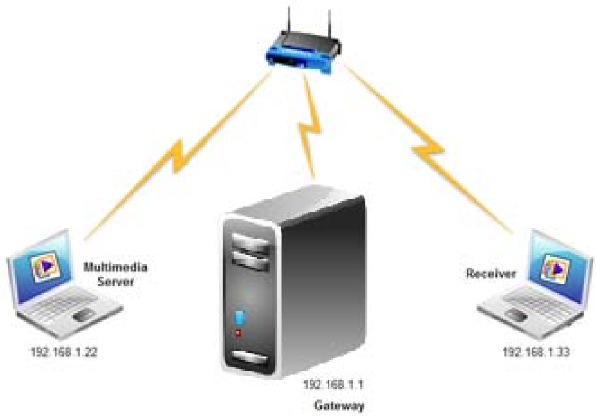

Figure 2. High Level Diagram of the Testbed Setup

A user experiment was conducted with 33 subjects (25 male and 8 female), their mean age was 26.2 with standard deviation of 3.75. All subjects were university students, researchers, or staff not affiliated with the study. Subjects were provided with a questionnaire and asked to provide their profile information and feedback about the perceived video quality (PVQ) using a 5-point scale, where a label '1' corresponded to "Worse/Strongly dissatisfied" and a label '5' to "Excellent/Strongly satisfied". Table II presents the different test setup parameters used in the experiments.

TABLE II: TEST SET UP PARAMETERS

Name	Packet Loss Rate (PLR) %	Packet Reorder rate (PRR) %	Video Bit Rate (VBR) kbps
PVQ (Ref)	0	0	800
PVQ (PLR)	0.5, 1, 3, 7, 15	0	800
PVQ (PRR)	0	1, 5, 10, 20, 30	800
PVQ (VBR)	0	0	100, 400, 600

IV. RESULTS AND DISCUSSION

A. Subjective Quantitative Analysis

In order to investigate the effects of different QoS parameters on subjective video streaming QoE perception, different statistical methods were utilized. First, descriptive statistics (e.g., mean, standard deviation, confidence interval) were calculated for the PVQ ratings to provide intuitive first-hand information about variables and their possible trends. Second, non-linear relationships between the different QoS parameters and the average PVQ ratings were explored, together with "goodness of fit" scores computed based on an alpha level of 0.05 (95%) [24].

A.1 QoE and Packet Loss Rate (PLR): Packet loss is the failure of one or more transmitted packets to arrive at their destination. Packet loss can be caused by variety of factors such as network congestion, network element failure, inadequate signal strength, low layer bit error rate, excessive system noise, hardware failure and software corruption. Packet loss creates the artifacts in the video sequence, thus negatively impacting the user's quality of experience. A wireless environment is emulated in this experiment to better understand the effects of PLR on QoE. Moreover, 7% and 15% PLR are included to emulate worst case scenario.

At the reference settings (see 1st row of Table II), a mean PVQ score (average over the 33 participants) of 4.06 ±0.22 was observed. As the packet loss rate (PLR) increased to 0.5% and 1%, significant reductions in PVQ scores were observed from 3.12±0.22 to 2.72±0.15, respectively. At a PLR=15%, an average PVQ of 1.12±0.11 was observed, thus suggesting a non-linear relationship between PLR and subjective PVQ. An exponential curve fit (depicted as a dotted line in Figure 3) showed a goodness of fit of R^2=0.83 and a correlation r=-0.91. For comparison purposes, a linear regression fit resulted in r=-0.83 and R^2=0.69.

As can be seen from the plot, PVQ degrades quickly with packet loss rates up to 7% and more subtly from 7% to 15%. This is likely due to two factors: (1) The predictive coding strategies employed in MPEG-4, as it introduces temporal dependencies into the video data in order to improve compression ratios; this can result in greater error propagation in the event of a packet loss or its late arrival [25], and (2) since the content used for study was fast-moving rugby clip, lost packets caused blocking and pixelization artifacts to be introduced in fast-motion spatial areas [26]. In both scenarios, the QoE degradation is more modest once a relatively poor experience is achieved around 7% PLR.

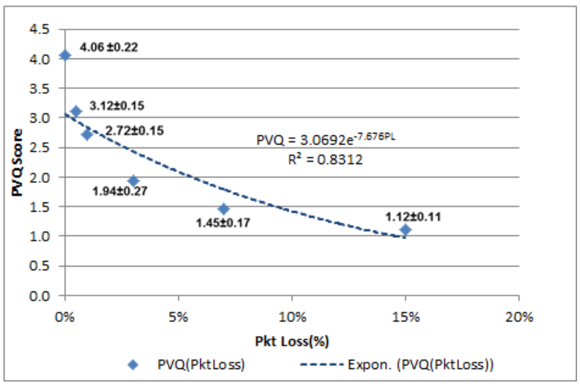

Figure 3. Scatter plot of PVQ ratings vs. Packet Loss

A.2 QoE and Packet reorder: Packet reorder is also an important QoS aspect which may degrade video quality and it is characterized as having varying delays that could cause out-of-order packets. Packet reorder can cause apparent loss of data in real time flows, such as voice over Internet protocol (VoIP) and video, and also causes TCP to use available bandwidth less effectively [25]. An application might be able to handle delay and jitter by using an appropriate buffer size; however, packet reordering might be more difficult to deal with at the application layer and hence may cause significant QoE degradations. Although in digital subscriber line based systems, it is assumed that the amount of reordered packets is not relevant; it was emulated here as a wireless environment to better understand the effect of packet reorder on QoE.

As mentioned previously, at the reference settings, a mean PVQ score of 4.06 ±0.22 was observed. When 1% packet reorder was introduced, the average PVQ score dropped to 3.75±0.17. At 30% reorder, the average PVQ score dropped to 1.09±0.10, thus suggesting a quasi-linear relationship between packet reorder rate (PRR) and subjective PVQ. Figure 4 depicts the obtained results along with an exponential fitted curve (dotted line). As can be seen, the exponential fit resulted in r=-0.98 and R^2=0.96. A linear fitted curve, on the other hand, resulted in r=-0.95 and R^2=0.90. Both functions showed a statistically significant relationship between PRR and PVQ, with the exponential fit resulting in slightly better results

A.3 QoE and Video Bit Rate: The default video bit rate (VBR) for H.264/MPEG-4 encoded video was 800 Kbps; at this rate, an average PVQ score of 4.06 ±0.22 was obtained. At 600, 400, and 100 Kbps, the mean PVQ scores dropped to 3.87±0.18, 3.60±0.17, and 2.79 ±0.2, respectively. The drop in perceived quality was not substantial, even at half the default rate, thus corroborating claims that H.264/MPEG-4 may provide better video quality relative to its predecessor (MPEG-2) even at half the bit rate [27].

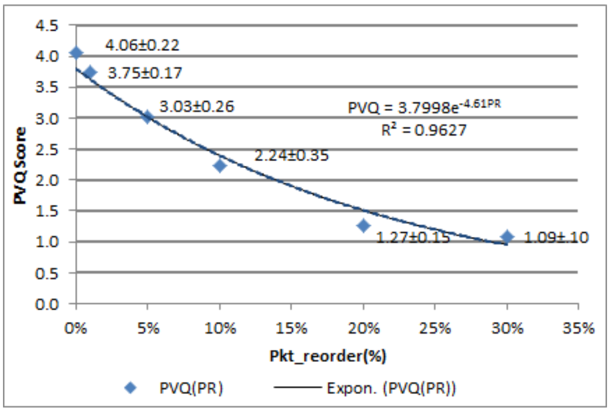

Figure 4. Scatter plot of PVQ ratings vs. Packet Reorder

To establish the relationship between perceived video quality and video bit rate, r and R^2 values were computed based on linear, exponential and logarithmic functions. Unlike the two previous scenarios, video bit rate and PVQ showed a strong positive and statistically significant correlation (r=0.977) only with the logarithmic function. The goodness of fit (R^2) values for the respective fitting functions were 0.95, 0.93, and 0.99. Figure 5 depicts the fitted logarithmic function. Table III summarizes the relationships obtained between PVQ and the various QoS parameters, as well as the correlation and goodness of fit values.

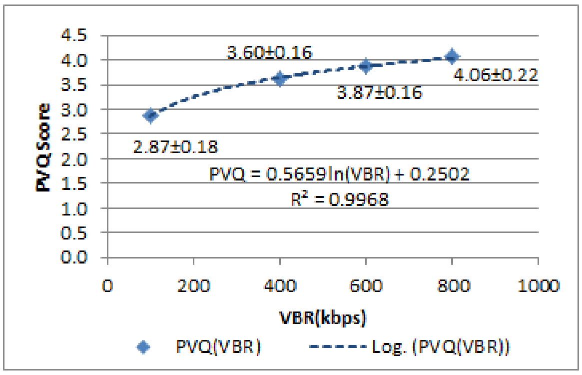

Figure 5. Scatter plot of PVQ ratings vs. Video Bit Rate

TABLE III SUMMARY OF BEST FITTING FUNCTIONS

QoS Parameter	QoE Parameter	r value	R^2 value	Relationship
Packet Loss (%)	PVQ	- 0.91	0.8312	Exponential Function
Packet Reorder (%)	PVQ	-0.978	0.9583	Exponential Function
Video Bit Rate (Kbps)	PVQ	0.9993	0.99	Logarithmic Function

B. *Qualitative Analysis*

In modern research, most researchers tend to adopt a combination of qualitative and quantitative approaches, which allow statistically reliable information obtained from numerical measurement to be backed up and enriched by information concerning participants' explanations [28]. The most meaningful metric related to verbal behaviors is the ratio of positive to negative comments [5]. It follows the so-called CCA framework depicted by Figure 6.

Figure 6. Preliminary Steps for Qualitative Data Analysis

During user experimentation, we asked subjects one open-ended question to give their comments and opinions about the overall video quality. Out of the 33 subjects, only 25 subjects preferred to answer this open-ended question. During experimentation, QoS parameters were varied 13 times, thus this question was posed 13 times. It is important to mention that not all subjects responded to all 13 inquiries. As a first step, all user comments and opinions were catalogued. Second, comments were categorized as positive, neutral, or negative. Positive comments reflect user satisfaction with video quality. Comments such as "*Excellent*", "*Very satisfactory*", *"I am happy with video quality now*" were classified as positive. To better capture the users' opinions, it was decided to further classify negative comments as "negative-suggestive" and "purely negative". The former represents user complaints or problem descriptions, such as "*Video freezes or pauses*", "*Video is slower in the start and then stops in the middle*", or "*It is blurry*". The latter, in turn, reflects the user's annoyance, dissatisfaction, and anger. Comments such as "*Catastrophic*", "*Terrible*", "*Video has very bad quality*", "*I'll never buy such type of VoD service*", "*Strongly dislike with -2 score*" were classified in this category. Neutral comments, in turn, were reflected by comments such as "*Normal quality*" or "*nearly fair quality*". Lastly, the third step consists of histogram analysis, as will be shown in the subsections to follow. For comparison purposes, the reference (i.e., the unprocessed original) video clip received 74% positive comments and 5% negative-suggestive comments, mostly due to the lower resolution CIF video used in the experiment; comments such as *"it was not HD like experience"* were seen with the reference video clip.

B.1 QoE and Packet Loss: For the reference video clip, users gave 74% positive comments, while with the introduction of only 0.5% packet loss, positive comments dropped to 28% and at they disappeared after 1% packet loss. The highest peak of neutral comments was achieved at a PLR of 0.5%, after which it decreased to 5% at PLR= 7%. An interesting observation found was the trend of negative-suggestive to negative comments. As packet loss rates increased, the percentage of purely negative comments rose and the percentage of negative-suggestive comments gradually decreased. Figure 7 depicts the user comment histograms as a function of PLR. As it can be seen, subjects tended to be neutral or negative-suggestive at first but once quality continued to degrade, subjects became more negative and harsher in their comments. Overall, user comments reflected a similar trend seen with the quantitative data, but with the advantage of providing a richer pool of user behavior/opinion information.

B.2 QoE and Packet Reorder: In this qualitative scenario, we noticed the highest percentage of positive comments (58%) at 1% packet reorder. At 5% packet reorder, 36% positive comments were seen, which were further reduced to 18% with a PRR of 10%. Negative-suggestive comments rose as packet reorder rates increased to 10% at which point they started to decrease.

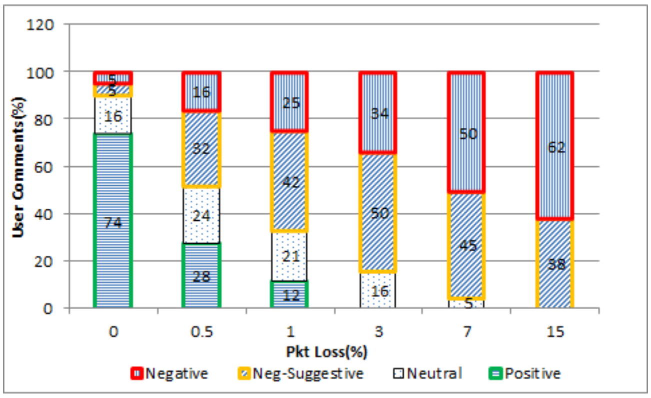

Figure 7. User Comment histogram as a function of PLR

Negative comments, in turn, gradually increased until attaining a peak value of 63% at PRR=30%. Neutral comments increased from 16% to 23% when packet reorder values changed from 0% to 5%. At this point, neutral comments decreased and reached zero at PRR= 30%. Figure 8 depicts the user comment histograms as a function of packet reorder rate. As expected, as quality impairments increase, positive comments decrease. While impairments are still at acceptable values, neutral or negative-suggestive comments rise, but reach a turning point during which negative and harsher comments arise.

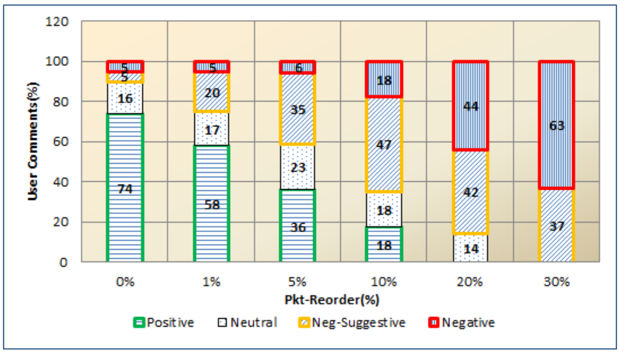

Figure 8. User comment histogram as a function of PRR

B.3 QoE and Video Bit Rate: The trend suggests that by lowering video bit rate from 800 kbps to 400kbps, positive comments decreased from 74% to 50%, while neutral comments increased from 16 to 25%, (see histogram depicted by Figure 9), thus suggesting that even at half the default VBR, users perceive only minor quality degradations in the video clip. However at 100 kbps, negative comments reached a maximum value of 34% and positive comments reached 26%. These qualitative results exhibit similar trends as those found in the quantitative analysis.

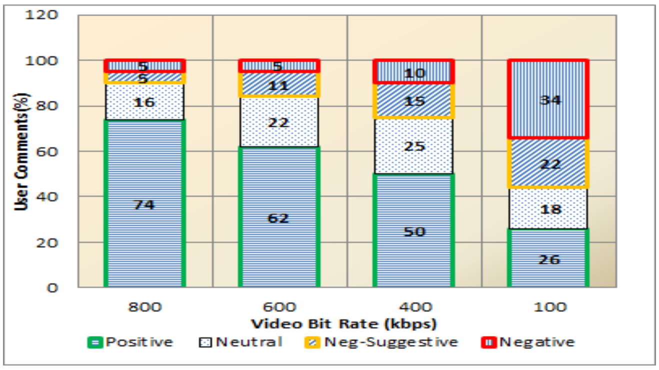

Figure 9. User Comment histogram as a function of video bit rate

C. *Objective Assessment Techniques: Preliminary Findings*

With this rich pool of subjective QoE information in hands, it is useful to gauge the effectiveness of existing objective quality metrics in predicting user QoE. To this end, we used the popular PSNR FR metric and the VisualMPEG NR algorithm. Results are presented in Table IV for the VisualMPEG metric only and in Table V for all metrics across all 14 degradation conditions.

C.1 PSNR and QoS Parameters: The relationships between PSNR and the three QoS parameters, namely, packet loss rate, packet reorder rate, and video bit rate, were examined; they are depicted in Figures 10-12, respectively. Linear, exponential, and logarithmic fitting functions were explored, but only the exponential curves are depicted (dotted line in plots), as they resulted in better "goodness of fit" scores. As can be seen, PSNR scores followed somewhat similar trends as those observed with subjective PVQ ratings, with the exception of video bit rates, which showed a quasi-linear fitting, as opposed to a logarithmic fitting with PVQ.

C2. VisualMPEG and QoS Parameters: Since the VisualMPEG implementation used in our experiments only provide quality 'labels', we assigned values 1-5 to the quality labels and used the Spearman rank correlation to explore the usefulness of the NR metric in predicting subjective QoE parameters. Correlation results are presented in Table IV.

As can be seen, the NR objective metric resulted in a rank correlation of r_s=-0.78 with PLR and PRR; this, however, was not shown to be statistically significant at a 95% level. With relation to VBR, the rank correlation decreased to r_s=0.5.

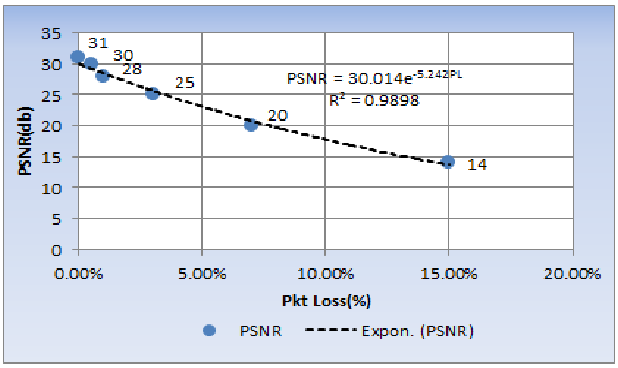

Figure 10. PSNR and Packet Loss relationship

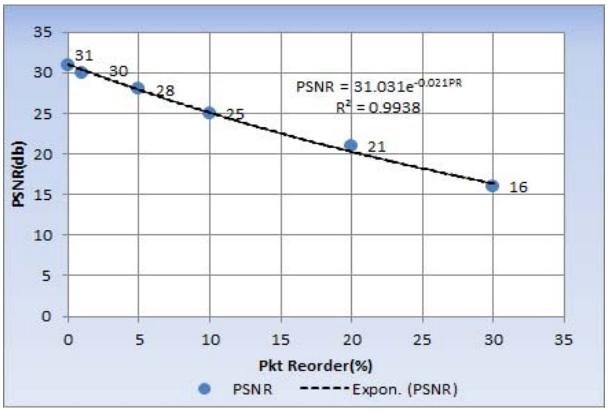

Figure 11. PSNR and Packet Reorder relationship

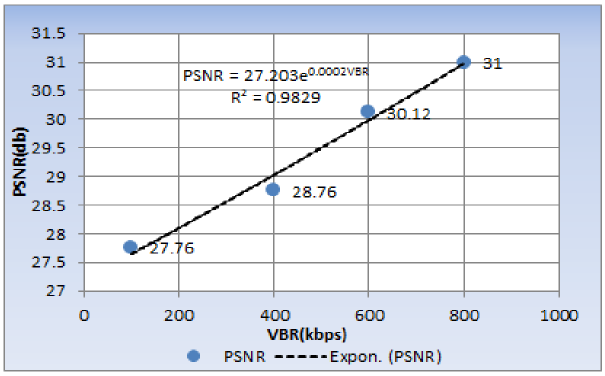

Figure 12. PSNR and Video Bit rate.

TABLE IV SPEARMAN RANK CORRELATION OBTAINED BETWEEN VISUALMPEG SCORES AND SUBJECTIVE PVQ SCORES

Relationship with	**Spearman rank correlation (r_s)**	**P value**
Packet loss rate	-078	0.064
Packet reorder rate	-078	0.064
Video bit rate	0.5	0.5

C2. Objective vs. Subjective Relationships: Lastly, we explore the relationship between the objective metrics and the subjective PVQ ratings across all 14 conditions. Plots in Figures 13 and 14 exhibit this relationship, respectively. From Figure 13, it can be observed that PSNR and PVQ follow an exponential relationship, thus suggesting that an exponential mapping function between PSNR and PVQ. From Figure 14, in turn, several observations can be made. For example, VisualMPEG scores follow similar trends as those observed by the PVQ scores in the following conditions: (1) PLR>3% (i.e., conditions 3-6), (2) across all packet reorder conditions, and (3) during high video bit rates, thus corroborating the correlation results reported in Table IV.

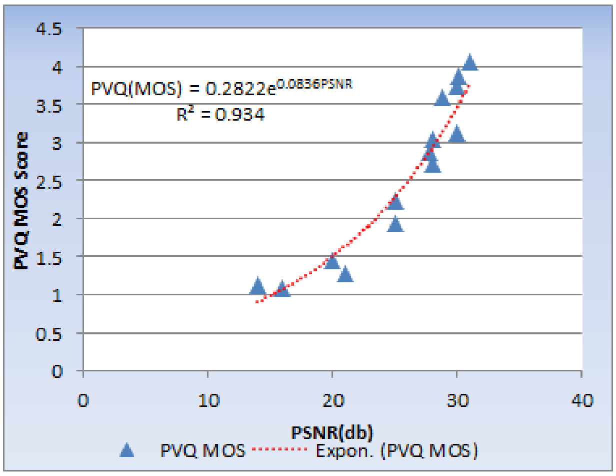

Figure 13. Scatter plot of PSNR vs. PVQ score

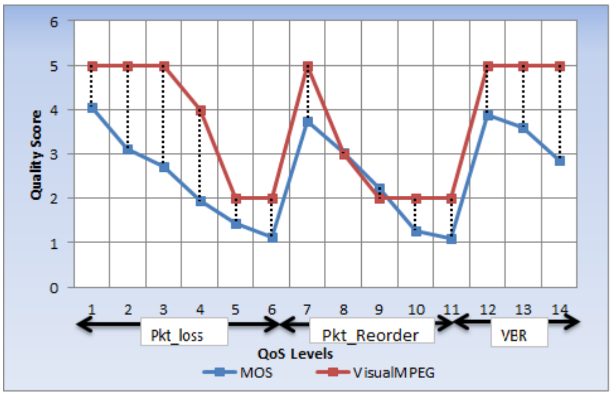

Figure 14. VisualMPEG and PVQ scores across al 14 conditions (see Table V for a description of each degradation condition)

V. CONCLUSIONS

In this paper, preliminary QoE results of a simulated video streaming service were presented. Data were analyzed using two subjective techniques and two objective metrics and results are summarized in Table V. It was observed that the relationship between packet loss rate and human perceived video quality (PVQ) followed an exponential curve, as was the case with packet reorder rates. Video bit rates, in turn, followed a logarithmic curve with PVQ. As per Table V, PSNR follows PVQ scores in packet reordering and VBR, but not in packet loss. Viewers were more critically responsive to packet loss especially with loss rates less than 7%. By comparing No-reference metric VisualMPEG with subjective metrics, we can observe (cf. Table V) that except test no.1, 5, 6, 10, 11, in all other tests, VisualMPEG does not follow subjective metrics.

Qualitative user comments exhibit similar trends as those found in the quantitative subjective analysis. Nevertheless, they provide clearer picture about viewer's feelings and perceptions in the format of their own words.

QoE is multi-disciplinary field; there is burgeoning interest within the community to bring neurophysiological aspects into the equation. Our ongoing studies are focusing on bringing human cognitive insights, obtained via neuroimaging techniques, into QoE monitoring.

ACKNOWLEDGMENT

This work was made possible due to funds from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Ministère du Développement économique, de l'Innovation et de l'Exportation du Quebec (MDEIE).

REFERENCES

[1] Khalil Ur Rehman Laghari, N. Crespi.; Molina, B.; Palau, C.E.; , "QoE Aware Service Delivery in Distributed Environment," Advanced Information Networking and Applications (WAINA), 2011 IEEE Workshops of International Conference on , pp.837-842, 22-25 March 2011.

[2] SmartPsych.co.uk. [Online]. http://www.smartpsych.co.uk/how-science-works-criminological

[3] Khalil Ur Rehman Laghari, N. Crespi, and K Connelly, " Towards Total Quality of Experience: A QoE for multimedia services in communication ecosystem," IEEE Communication Magazine, April 2012.

[4] Taylor-Powell Ellen and Marcus Renner, "Analyzing Qualitative Data," University of Wisconsin, Online Report 2003.

[5] Tullis Tom and Albert Bill, Measuring the user experience. MK Publishers by Elsevier Inc., 2008.

[6] William J. Gibson & Andrew Brown, "Working with Qualitative Data",: 2009 | DOI:10.4135/9780857029041, Print ISBN: 9781412945721 | Online ISBN: 9780857029041.

[7] ITU-T, "Subjective video quality assessment methods for multimedia applications," International Telecommunication Union, Geneva, Switzerland, Recommendation P.910, 2008.

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

[9] ITU-T Recommendation G.1030 (2005), "Estimating end-to-end performance in IP networks for data applications."

[10] Davis, F. D. (1986). A technology acceptance model for empirically testing new end-user information systems: Theory and results. (Doctoral dissertation, Sloan School of Management, Massachusetts Institute of Technology).

[11] Icek Ajzen, The Theory of Planned Behavior- Organizational Behavior and Human Decision Processes, 1991, 179-211.

[12] K.I. Al-Qeisi, "Analyzing the Use of UTAUT Model in Explaining an Online Behaviour: Internet Banking Adoption," Department of Marketing and Branding, Brunel University.2009, Thesis Report (2009).

[13] T. Oelbaum and K. Diepold, "Building a reduced reference video quhttp://www.mpeg-analyzer.com/ality metric with very low overhead using multivariate data analysis," in in Proceedings of the 4th International Conference on Cybernetics and Information Technologies, Systems and Application, 2007.

[14] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli:, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, April 2004.

[15] Stephen Wolf, Margaret H.Pinson, "Application of the NTIA General Video Quality Metric (VQM) to HDTV Quality Monitoring," Third International Workshop on Video

Processing and Quality Metrics for Consumer Electronics (VPQM-07), Scottsdale, Arizona, January 25-26, 2007
[16] No-Reference Video Quality Metrics, VisualMPEG. DOI=http://www.mpeg-analyzer.com/
[17] Luigi Landini, Vincenzo Positano and Maria Filomena Santarelli (eds), Advanced Image Processing in Magnetic Resonance Imaging, Dekker, book series on Signal Processing and Communications, ISBN 0824725425, 2005.
[18] Ferrari M, Mottola L, and Quaresima V (2004): Principles, Techniques, and Limitations of Near Infrared Spectroscopy. Can. J. Appl. Physiol. 29(4) 463-487.
[19] T. Falk, Y. Pomerantz, K. Laghari, S. Moller, and T. Chau, Preliminary Findings on Image Preference Characterization based on Neurophysiological Signal Analysis: Towards Objective QoE Modelling, *4th International Workshop on Quality of Multimedia Experience (QoMEX2012), July 5-7, Yarra Valley, Australia.*
[20] Arndt, S., Antons, J.N., Schleicher, R., Möller, S., & Curio. G. (2012) Perception of Low-Quality Video Analyzed by Means of Electroencephalography. *4th International Workshop on Quality of Multimedia Experience, July 5-7, Yarra Valley, Australia.*
[21] Video Lan VLC player. [Online]. http://www.videolan.org/vlc/
[22] NetEmhttp://www.linuxfoundation.org/collaborate/workgroups/networking/netem
[23] Video Test Clips. DOI=http://media.xiph.org/video/derf/.
[24] Chapter 2:"The Correlation Coefficient", accessed at 14th June 2012.DOI=http://www.biddle.com/documents/bcg_comp_chapter2.pdf
[25] C.M. Arthur, D Girma, D Harle, and A Lehane, "The effects of packet reordering in a wireless multimedia environment," in *1st International Symposium on Wireless Communication Systems,* 20-22 Sept. 2004 .
[26]] J. Greengrass, J. Evans, and A. C. Begen, "Not All Packets Are Equal, Part II: The Impact of Network Packet Loss on Video Quality", IEEE Internet Computing, vol. 13(2), pp.74–82. March 2009.
[27] Margaret H. Pinson, Wolf Stephen, and Cermak Gregory, "HDTV Subjective Quality of H.264 vs. MPEG-2, with and without Packet Loss," *IEEE Transactions on Broadcasting,* vol. 56 , March 2010 .
[28] Khalil Ur Rehman Laghari, Imran Khan, and N. Crespi, "Quantitative and Qualitative Assessment of QoE for Multimedia Services in Wireless Environment"," in MoVid Workshop, ACM Multimedia Systems Conference, Chapel Hill NC USA, 22-24 February 2012.
[29] Korhonen, Jari; You, Junyong. Improving Objective Video Quality Assessment with Content Analysis. In Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM'10). Scottsdale, AZ, USA, Jan, 2010.

TABLE V. SUMMARY OF ALL SUBJECTIVE AND OBJECTIVE METRICS

TEST NO.	**QOS PARAMETERS**	**PVQ MOS SCORE**	**PVQ LEVELS**	**QUALITATIVE SCORES (TOP 2)**	**PSNR**	**NR QUALITY REFERENCE**
1.REFERENCE	PL=0%, PR=0 %, VBR=800	4.06±0.22	EXCELLENT	1.POS=74%, 2. NEU=16	31	EXCELLENT
2	PL=0.5%	3.12±0.18	GOOD	1.NEG_SUG=32% 2.POS=28%	30	EXCELLENT
3	PL=1%	2.72±0.15	FAIR	1.NEG_SUG=42% 2.NEG=25	28	EXCELLENT
4	PL=3%	1.94±0.27	POOR	1.NEG_SUG=50 2.NEG=34%	25	GOOD
5	PL=7%	1.45±0.17	POOR	1.NEG=50 % 2.NEG_SUG=45%	20	POOR
6	PL=15%	1.12±0.11	POOR	1.NEG=62, % 2.NEG_SUG=38%	14	POOR
7	PR=1%	3.75±0.17	GOOD	1.POS=58%, 2.NEG_SUG=20%	30	EXCELLENT
8	PR=5%	3.03±0.26	GOOD	1.POS=36%, 2.NEG_SUG=35%	28	FAIR
9	PR=10%	2.24±0.35	FAIR	1.NEG_SUG=47% 2.POS=18%	25	POOR
10	PR=20%	1.27±0.15	POOR	1.NEG=44%,2.NEG_SUG=42%	21	POOR
11	PR=30%	1.09±0.10	POOR	1.NEG=63%, 2.NEG_SUG=37%	16	POOR
12	VBR=600	3.87±0.16	GOOD	1.POS=62%, 2.NEU=22%	30	EXCELLENT
13	VBR=400	3.60±0.16	GOOD	1.POS=50%, 2.NEU=25%	29	EXCELLENT
14	VBR=100	2.87±0.18	FAIR	1.NEG=34%, 2.POS=26%	28	EXCELLENT