Adaptive Streaming: A subjective catalog to assess the performance of objective QoE metrics

Alberto Álvarez, Laura Pozueco, Sergio Cabrero, Xabiel G. Pañeda, Roberto García, David Melendi and Gabriel Díaz*

Informatics Department, University of Oviedo, Gijón, Spain.

* Electrical and Computing department, Spanish University for Distance Education (U.N.E.D), Madrid, Spain

aalvarezglez@gmail.com, {pozuecolaura.uo, cabrerosergio, xabiel, garciaroberto, melendi}@uniovi.es, *gdiaz@ieec.uned.es

Abstract

Scalable streaming has emerged as a feasible solution to resolve users' heterogeneity problems. Scalable Video Streaming (SVC) is the technology that has served as the definitive impulse for the growth of streaming adaptive systems. Systems need to incorporate user's Quality of Experience (QoE) in the layer switching algorithms. This paper characterizes the relation between human decisions and objective metrics that could be mapped into systems. We have performed extensive subjective experiments to corroborate the preference towards adaptive systems when compared to traditional non-adaptive systems. The resulting subjective scores are correlated with most relevant Full Reference (FR) objective metrics. We obtain an exponential relationship between human decisions and the same decisions expressed as a difference of objective metrics. A strong correlation with subjective scores validates objective metrics to be used as aid in the adaptive decision taking algorithms to improve overall systems performance. Results show that, among the evaluated objective metrics, PSNR is the metric that provide worse results in terms of reproducing the human decisions.

Keywords: Adaptive streaming, Scalable Video Coding (SVC), objective video quality metrics, subjective quality assessment.

1. Introduction

Scalable video technology and the concern about its subjective effects have been present for several years [1]. However, Scalable Video Coding (SVC) technology has finally enabled a wide range of applications that exploit the benefits of layered encoding and, in particular, applications that leverage layer encoding to perform adaptive streaming. Adaptive streaming is based on the premise of delivering at each moment the right combination of layers that maximize Quality of Experience (QoE) while coping with network technology constraints.

The question is how to analyze which combinations of layers and adaptations give the best QoE results at the end user location. The answer is clear, ask the user. We previously performed a subjective evaluation of an adaptive system, comparing in different scenarios the adaptive streaming system with a regular non-adaptive streaming service [2]. In that work we answered a set of key questions regarding adaptation decisions and provided a subjective database for others to continue reinforcing our results.

Carrying out subjective tests is rather costly. For that reason, objective metrics are continuously being investigated to be able to fully substitute subjective measures, with sufficient conviction. Using objective QoE metrics is the path for QoE driven automatic adaptation methods. They can be incorporated into adaptive systems to fine tune adaptation decisions on the server or proxy side. However, for that method to serve its purpose, objective metrics have to be strongly correlated with subjective scores.

The most used objective metrics, normally Full Reference (FR), have been Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM). More recently, the Video Quality Metric (VQM) has been adopted by ANSI as the objective metric standard.

The aim of this paper is to corroborate if objective metrics can be correlated with subjective scores. We carried out a subjective study with 75 participants and 48 pairs of video sequences. The participants had to assess their preference towards an adaptive streaming system or a traditional, i.e. non-adaptive, system. The subjective results of the experiment, expressed in Comparison Mean Opinion Scale (CMOS) values, are compared with the objective evaluation of each pair of sequences.

Results show that the evaluated objective metrics have an exponential relationship with CMOS values. Moreover, we have found that, among the evaluated objective metrics, PSNR has the worst behavior while VQM has the best correlation with subjective assessment.

The rest of the paper is organized as follows: Section 2 reviews the most important related work. Section 3 summarizes the subjective experiments performed. Section 4 discusses results and finally, Section 5 contains the conclusions and outlines future work.

2. Related Work

The paradigm of QoE is an important subject of study which has received a great deal of focus in recent years. Seeking an improved performance in terms of users' perception, many

works have focused on understanding the implications of all the factors influencing QoE. The analysis of Mean Opinion Score (MOS) has evolved from codec evaluation to be used to assess QoE studies in general. For practical reasons (related to time consumption), subjective studies need to be complemented by objective measures. Their close relationship is an ongoing work.

The more simple approach to video quality evaluation is adopted by objective quality metrics. Objective QoE metrics rely only on video content characteristics to evaluate the theoretical user perception. PNSR was the first, with the simple application of SNR measure to the received video with respect to the original sequence. The work presented in [3] has tried to improve the PSNR model. The SSIM metric includes structural aspects on the video, as an attempt to improve the correlation with HVS (human vision system). MOVIE index [4] is believed to provide good subjective correlation, but its computational cost is immeasurable. The Video Quality Metric (VQM) by NTIA/ITS has provided excellent results. Due to its outstanding performance, VQM has been adopted as ANSI and ITU standards. Authors in [5] provide a full review with a performance comparison of metrics and classify them according to their approaches for measuring video quality.

To ease the evaluation of objective proposals, subjectively evaluated databases are being released. The work [6] provides a study and a resulting database, known as the LIVE Video Quality Database, where 150 distorted video sequences obtained from 10 different video content sources were subjectively evaluated by 38 human observers. The subjective evaluation was performed using a single stimulus paradigm. They also provide the performance of several freely available objective, FR algorithms on the LIVE Video Quality Database. PSNR is shown to perform very poorly in correlating with human subjective judgments and, according to the authors, is clearly an unreliable predictor of quality in any application where the end user of the video is a human observer. MOVIE index had the best performance of the algorithms tested.

The scalable technology and SVC in particular is an important subject in this field. The survey [7] gives a good review of the state of the art of subjective and objective studies, with scalable streaming in general, including SVC technology. Most of the studies involving SVC are evolutions of codec evaluation studies, applied to SVC to evaluate different layering strategies and their QoE [8]. The direct outcome that SVC technology has enabled is the upsurge of adaptive streaming services. The importance of a QoE driven adaptive system has been experienced in the context of a P2P application [9]. The authors directly use objective QoE metrics of each layer in the scaling decision taking algorithm.

The work performed in [2] and [10] contributes to the SVC adaptive subjective evaluations realm, enlightening several key factors about layer selection on adaptive systems. Both works alter the mainstream approach by evaluating the human decisions when comparing SVC technology with traditional non-layer and non-adaptive streaming schemes.

As the study of new and improved objective algorithms increases, there is a growing number of works seeking to obtain a direct relationship between easily measurable Quality of

Service (QoS) parameters and QoE. The authors of [11] show that they can accurately estimate QoE for SVC video streams through a random neural network (RNN) module. However, that RNN appears to be trained for specific scenario/conditions. Calyam et al. proposed a "GAP-model" based framework for online VoIP QoE measurement [12] which is an offline model of QoE expressed as a function of measurable network factors such as bandwidth, delay, jitter and loss. Using the GAP-model, they claim to be able to provide VVoIP QoE estimates in terms of "Good", "Acceptable", or "Poor" (GAP) grades of perceptual quality solely from the on-line measured network conditions.

It is already assumed that the best fit model for QoS and QoE relationship is an exponential approach (IQX method) [13]. However, this quest is unfinished and there are many flaw points still to be addressed. For example, not all factors influencing QoE are totally understood yet. Parameters rarely investigated include privacy concerns, interaction of audio/video, user interface, user's awareness of quality, cost, last mile equipment/environment or quantization of video coding.

In our paper, we focus on evaluating the correlation between objective QoE metrics and QoE subjectively regarding human decisions, abstracting the QoS factors. When facing the choice between adaptive and non-adaptive systems and their performance in a double stimulus experiment, human decisions can depend on varied factors such as content characteristics or the nature of the adaptation. We use three of the most well known objective metrics, PSNR, SSIM and VQM, to evaluate their ability to model and reproduce human decisions. The correlation of subjective observations and the three objective metrics (PSNR, SSIM, VQM) was briefly introduced by Vranjes et al. in [14]. With two sequences and 12 participants, their main observation was that with a given bitrate, objective metrics produced different results for different content while subjective MOS was very similar. We include the correlation of objective results with the subjective CMOS analyzed from 75 participants responding to 48 choices (96 sequences) and unveil an exponential model characterizing that relationship.

3. Subjective experiments.

The potential gain in the QoE resulting from streaming adaptive services is not stated. For that reason, subjective experiments are needed to directly compare user's choice between an adaptive or traditional non-adaptive streaming service. In our subjective experiments [10], different situations are evaluated to seek the user's limits of tolerance towards any of the systems in chosen scenarios.

The following sub-sections summarize our subjective experiments [10].

In the different network scenarios selected, the adaptive system reacts to network availability. The SVC based system adapts the content being delivered, increasing or reducing layers, according to network conditions. Meanwhile, the non-adaptive system tries to maintain the same delivered layers (i.e. quality) as initially chosen.

3.1 Methodology (double stimulus)

Through *Stimulus Comparison Adjectival Categorical Judgement (SCACJ)* methodology, 75 users are presented with a pair of sequences corresponding to the result of video streaming in a given network situation with the two systems, adaptive and non-adaptive, Fig.1. After the visualization, the users gave grades in a categorical scale, Fig.2. The proposed methodology assumes seven assessment categories to indicate the preferred stimuli, from -3 to +3. Fig.3 shows the proposed scale. Results are given in CMOS. Positive values of CMOS indicated the preference for the adaptive system, while the preference for the non-adaptive system is translated into negative values.

A software tool is used to present the content to the users following the guidelines of ITU BT.500 and to store electronic scores given at the end of each test.
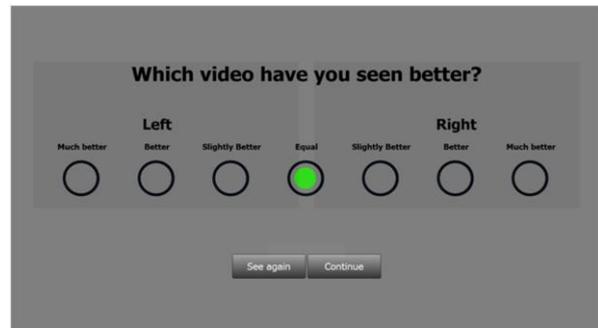

Figure 1. Stimulus evaluation screen

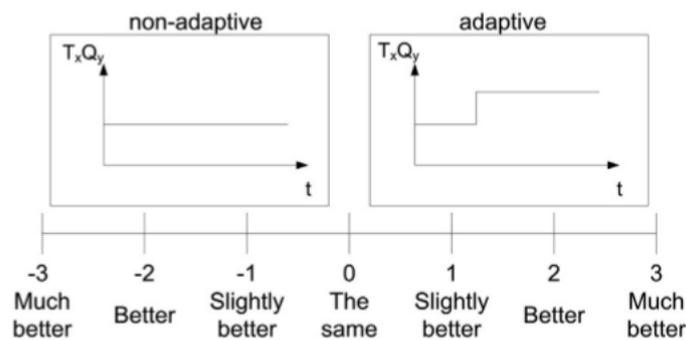**Figure 2. Voting interface**

Figure 3. Modified categories in CMOS scale

3.2 Targets, as planned for subjective corroboration of preference towards regular system

Sequences created for the subjective experiment were grouped into targets. Targets are intended to answer individual questions in a single scalability dimension.

<u>Target I</u>: in severe congestion situations, do users value significant quality and temporal

reductions in quality to avoid high loss rates?

Target II: in light congestion situations, what is the adequate losses threshold to trigger adaptation?

Target III: in severe congestion situations, which is better, a drastic temporal adaptation or keeping up with high loss rate?

Target IV: in bandwidth availability situations, do users value quality layers increases?

Target V: in bandwidth availability situations, do users value temporal layer increases?

3.3 Material and environment

The content is encoded using H.264/SVC using JSVM reference software. Four varied content (*factory, tractor, marathon and touchdown*) 10 second long sequences were generated with 5 temporal levels and 4 quality levels, using Coarse Grain Scale (CGS) scheme, Table 1. Using source content and a Bernoulli loss model implemented in the SVEF framework [15], different scenarios were combined with source sequences to provide a total of 48 pairs of sequences. The reference sequence corresponds to the output of a non-adaptive system. In the reference sequence, for each scenario an initial layer combination is chosen and maintained throughout the duration of the sequence, regardless of the Packet Loss Ratio (PLR) or exceeding throughput availability. In contrast, in the adaptive sequence, after the first 5 seconds the content is adapted to compensate or adapt to network scenario conditions. The nomenclature followed (TxQy->Tx'Qy') indicates the initial layer combination (temporal layer x and quality layer y and the layers after adaptation, temporal x' and quality y'). Finally, sequence pairs were linked to targets defined in order to assess the users' perception of the preferred system, Table 2.

FACTORY		TRACTOR		MARATHON		TOUCHDOWN	
T0Q3	1.875fps, 1183 kbps	T0Q3	1.562 fps, 1504kbps	T1Q3	6.25 fps, 4043 kbps	T2Q3	7.5 fps, 3215 kbps
T1Q3	3.75 fps, 1821 kbps	T1Q3	3.125 fps, 2102kbps	T2Q3	12.5 fps, 5688 kbps	T3Q3	15 fps, 4188 kbps
T2Q3	7.5 fps, 2673 kbps	T2Q3	6.25 fps, 2721 kbps	T3Q3	25 fps, 7474 kbps	T4Q1	30 fps, 1287.5 kbps
T3Q3	15 fps, 3679 kbps	T3Q3	12.5 fps, 3458 kbps	T4Q0	50 fps, 820 kbps	T4Q2	30 fps, 2787 kbps
T4Q0	30 fps, 712 kbps	T4Q0	25 fps, 517.4 kbps	T4Q2	50 fps, 4609 kbps	T4Q3	30 fps, 5212 kbps
T4Q1	30 fps, 1299.3 kbps	T4Q2	25 fps, 2348 kbps	T4Q3	50 fps, 8735 kbps		
T4Q2	30 fps, 2708 kbps	T4Q3	25 fps, 4341 kbps				
T4Q3	30 fps, 4716 kbps						

Table 1 Characteristics of the videos used in the test

Pair	Target	Content	Reference sequence	Adaptive sequence	Pair	Target	Content	Reference sequence	Adaptive sequence
1	I	*Tractor*	T4Q3, 40%PLR	T4Q3->T4Q2	24	III	*Tractor*	T4Q3, 30%PLR	T4Q3->T2Q3
2, 25	I,III	*Tractor*	T4Q3, 40%PLR	T4Q3->T1Q3	27	III	*Factory*	T4Q3, 70%PLR	T4Q3->T0Q3
3	I	*Factory*	T4Q3, 40%PLR	T4Q3->T4Q2	28	III	*Tractor*	T4Q3, 60%PLR	T4Q3->T0Q3

4, 23	I,III	*Factory*	T4Q3, 40%PLR	T4Q3->T2Q3	29	IV	*Touchdown*	T4Q1	T4Q1->T4Q2
5	I	*Marathon*	T4Q3, 40%PLR	T4Q3->T4Q2	30	IV	*Factory*	T4Q1	T4Q1->T4Q2
6, 26	I,III	*Marathon*	T4Q3, 40%PLR	T4Q3->T1Q3	31	IV	*Marathon*	T4Q0	T4Q0->T4Q2
7	II	*Tractor*	T4Q3, 1%PLR	T4Q3->T3Q3	32	IV	*Tractor*	T4Q0	T4Q0->T4Q2
8	II	*Marathon*	T4Q3, 1%PLR	T4Q3->T3Q3	33	IV	*Factory*	T4Q1	T4Q1->T4Q3
9	II	*Touchdown*	T4Q3, 3%PLR	T4Q3->T3Q3	34	IV	*Touchdown*	T4Q1	T4Q1->T4Q3
10	II	*Marathon*	T4Q3, 3%PLR	T4Q3->T3Q3	35	IV	*Factory*	T4Q0	T4Q0->T4Q3
11	II	*Touchdown*	T4Q3, 5%PLR	T4Q3->T3Q3	36	IV	*Tractor*	T4Q0	T4Q0->T4Q3
12	II	*Tractor*	T4Q3, 5%PLR	T4Q3->T3Q3	37	V	*Tractor*	T1Q3	T1Q3->T2Q3
13	II	*Tractor*	T4Q3, 10%PLR	T4Q3->T3Q3	38	V	*Marathon*	T1Q3	T1Q3->T2Q3
14	II	*Factory*	T4Q3, 10%PLR	T4Q3->T3Q3	39	V	*Touchdown*	T2Q3	T2Q3->T3Q3
15	II	*Touchdown*	T3Q3, 1%PLR	T3Q3->T2Q3	40	V	*Factory*	T2Q3	T2Q3->T3Q3
16	II	*Factory*	T3Q3, 1%PLR	T3Q3->T2Q3	41	V	*Marathon*	T3Q3	T3Q3->T4Q3
17	II	*Tractor*	T3Q3, 3%PLR	T3Q3->T2Q3	42	V	*Touchdown*	T3Q3	T3Q3->T4Q3
18	II	*Factory*	T3Q3, 3%PLR	T3Q3->T2Q3	43	V	*Marathon*	T1Q3	T1Q3->T3Q3
19	II	*Touchdown*	T3Q3, 5%PLR	T3Q3->T2Q3	44	V	*Tractor*	T1Q3	T1Q3->T3Q3
20	II	*Factory*	T3Q3, 5%PLR	T3Q3->T2Q3	45	V	*Tractor*	T2Q3	T2Q3->T4Q3
21	II	*Tractor*	T3Q3, 10%PLR	T3Q3->T2Q3	46	V	*Touchdown*	T2Q3	T2Q3->T4Q3
22	II	*Factory*	T3Q3, 10%PLR	T3Q3->T2Q3	47	V	*Factory*	T1Q3	T1Q3->T4Q3
					48	V	*Marathon*	T1Q3	T1Q3->T4Q3

Table 2 Details of the video pairs

3.4 Results of the experiments

Measured as the comparative MOS between adaptive and non-adaptive system, mean CMOS values processed for the total amount of 75 participants were mainly positive, which implies a preference towards the adaptive system. According to age, dividing the population in half (under or over 35 years of age), we observed that the younger group is more critical with their subjective evaluations while the older subset provided CMOS values more centered towards zero (indifference).

As a contribution to the distribution of pairs across targets, we are able to provide answers to the target questions. For Target I, general results show that users' tolerance towards quality changes is greater than to temporal layer changes. Using a two sample t-test we compare the CMOS values of temporal and quality changes scenarios. Resulting p-values indicate that, except for the factory content, all means are truly different and always pointing out quality changes preference. The p-value (>0.01) in the factory scenario indicates that CMOS means are not significantly distinguishable. Indeed, content influence in subjective results is still an open field. Users' tolerance threshold towards packet loss is around 3%, answering Target II. Grouping sequences of 3%PLR for a Kruskal-Wallies rank sum test reveal that all mean CMOS values are statistically different. Below that PLR, participants showed either indifference or even preference towards the non-adaptive system. Exceptionally, if initial temporal quality of sequence is poor, the losses threshold can be set at higher rates, according to users' votes. Target III results should be considered on account of

the concealing algorithm at reception (frame copy). For this reason, users' choice between a severely damaged sequence and a sequence which reduces several temporal layers is uncertain. As an example, a t-test performed on tractor sequence (pair 28) scores reveals that CMOS mean is not statistically different from hypothesized mean 0. The relation of these results with different concealing algorithms should be addressed. For Target IV, CMOS values were kept low, although statistically different from 0, indicating only a slight preference towards the adaptive system. A two sample t-test comparison between quality jumps of 1 and two levels for the same sequence reveals that the means are not statistically different. This raises the question of whether quality increases are truly profitable, pondering subjective opinion and cost of increased bandwidth consumption. For Target V, the skewness factors on the histogram of scores provided for each pair indicate that most of them are highly left-tailed. The subjective opinion of temporal increases is therefore markedly positive, which gives the preference towards increasing temporal layers when choosing the way of incrementing QoE if bandwidth is available.

For more details on the subjective studies and their outcome refer to [10].

4. Subjective results and objective metrics correlation.

Following the structure of the subjective tests performed, we have computed the objective FR metrics for all sequences grouped in pairs used in the subjective experiment. In our subjective experiment, participants were given the choice of which sequence in the pair they liked the most (according to the double stimulus subjective methodology). As a result, we obtain CMOS which indicate the direction and amplitude of the users' preferences. As explained in the previous section, a positive CMOS value for a pair is indicative of the users' preference towards the adaptive system. Correspondingly, a negative CMOS, is a sign of users opting for the reference (non-adaptive) sequence.

To map objective results with CMOS scores, we obtain the difference between the objective metric calculated for the adapted and the reference sequence in the same pair. Then, PSNR and SSIM positive differences indicate that objectively speaking, the adapted sequence is better than the reference sequence. For VQM, since its scale is reversed with respect to the PSNR or SSIM, positive VQM differences indicate objective preference towards the reference sequence.

Next, we provide the graphic relationship between individual objective metric differences and subjective CMOS values. To solve the problem of negative values, a linear transformation is applied on the objective metric series. Then, the exponential relationship can be calculated as a linear relationship between the subjective scores and logarithm of the objective metric series.

Figs. 4, 5 and 6 show the results of the regression applying a logarithmic transformation, previously mentioned, on the objective values of PSNR, SSIM and VQM. Observing the quantile plot (figures on the right) for the three metrics, the PSNR model is the least accurate of the three because PSNR samples are deviated from the linear distribution. Nonetheless,

PSNR has been criticized throughout its history for not serving well as a QoE metric [16][17]. Yet, the exponential model is not entirely inaccurate, as the observation of residuals reveals. The Residuals of the three models (figures on the left) show no discernable pattern. Residual medians of the three models are close to zero.

Both VQM and SSIM present better fitting results, according to the quantile plot (top right). However, both models point out an outlier in observation number 5. This observation corresponds to experiment pair number 5. This pair (Marathon T4Q3_T4Q2) is a quality adaptation of the sequence "marathon" compared to a sequence affected by high loss rate (40%). According to the objective metrics, the quality adaptation performed by the adaptive system improves the quality of the adapted sequence notably. The same content, with the same loss pattern but a temporal adaptation instead of a quality one (6th pair) improves objective quality in a smaller quantity, according to objective metrics used. Other content with similar spatial complexity (Factory sequence, 3rd pair) does not stand out either with a high objective difference between adapted and non-adapted sequences. We argue that the scene change present only in the Marathon sequence is affecting the objective measure of the adapted sequence in the case of quality adaptation. The study of the effects of scene changes on quality and temporal adaptation from both a subjective and objective perspective should be addressed more deeply.

Multiple R-squared statistics for each model are shown in Table 3. Values are low since the predictions of human behavior normally produce lower fitting values. However, observing the residual and quantile plots, the model fitting is reasonably good.

Metric	Multiple R-squared
PSNR	0.2055
SSIM	0.3461
VQM	0.2489

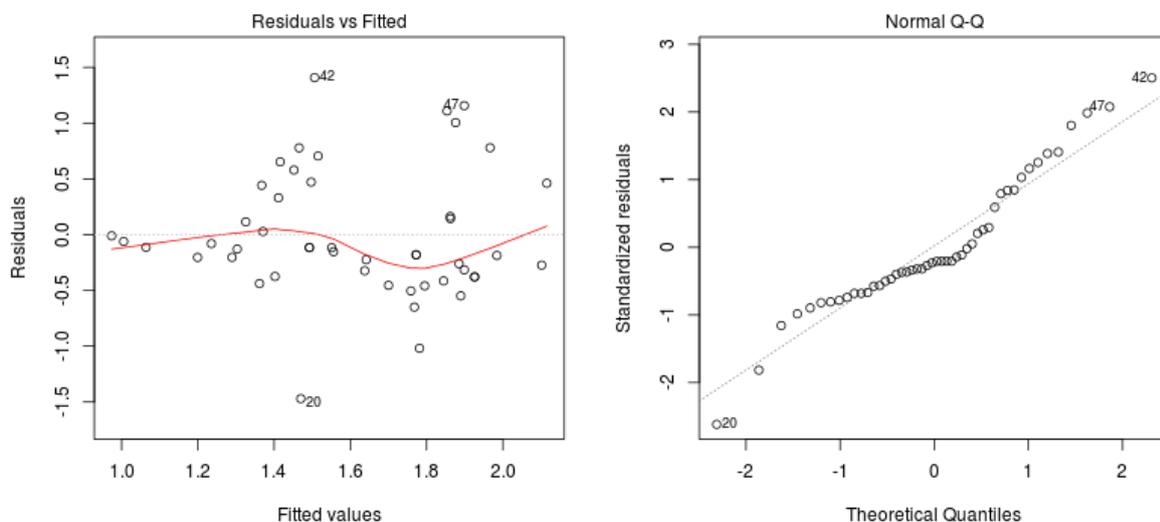**Table 3 Multiple R-squared statistics**

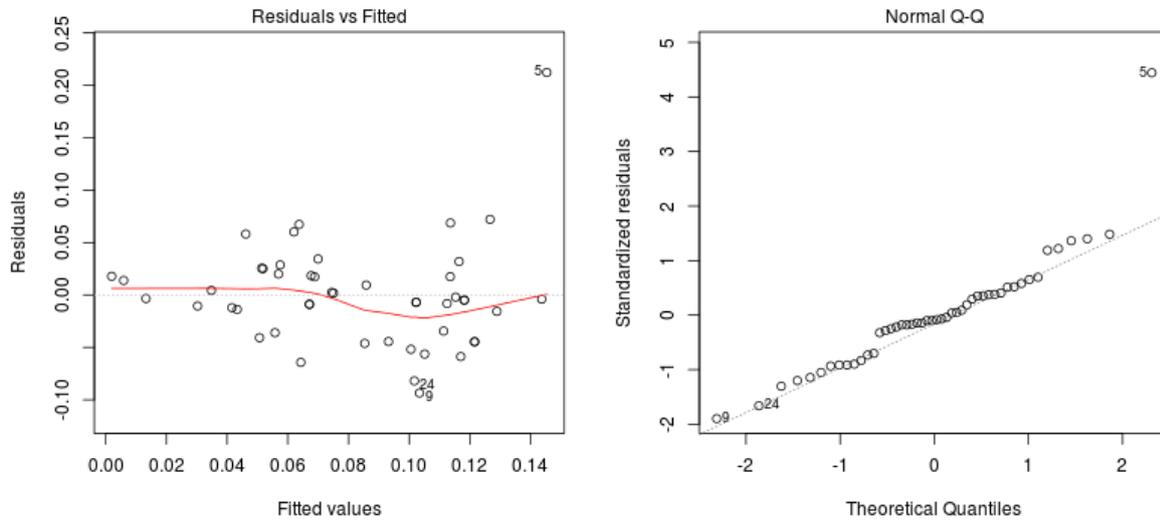**Figure 4 – Graphical summary of the regression for the PSNR metric**

Figure 5 Graphical summary of the regression for the SSIM metric
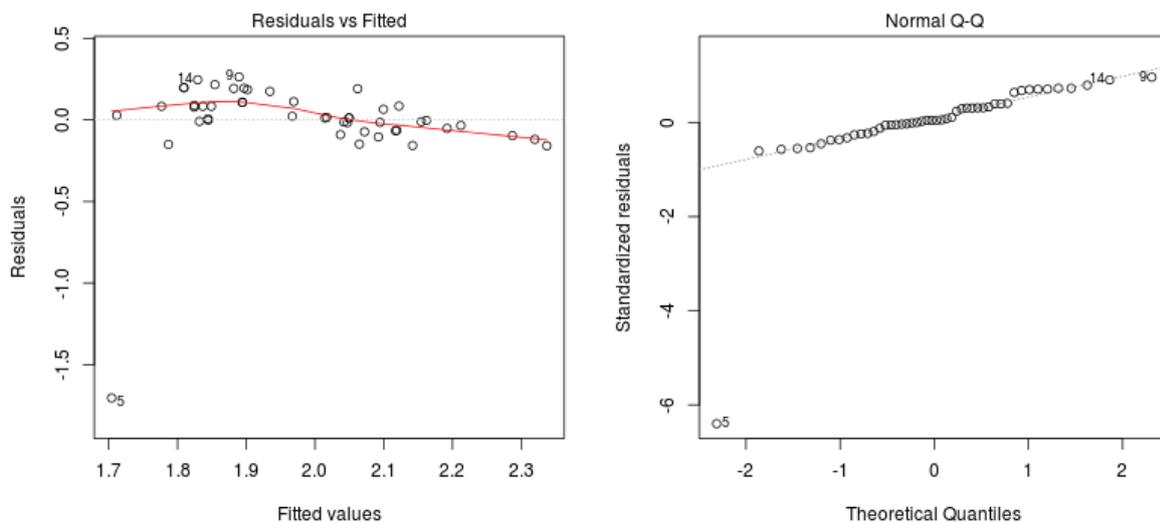

Figure 6 Graphical summary of the regression for the VQM metric

The evaluated objective metrics have demonstrated a good exponential relationship with the subjective CMOS scores. Then we extract exponential correlations with subjective decisions between adaptive and non-adaptive by target, Table 4. The best correlation results are given by VQM while PSNR provided the worst correlation indexes. The result of the correlation with the logarithmic transformation of objective results improves global results of a simple linear correlation by only 0.76%. The model fits better, but when classified by targets, some targets with a small number of sequences are difficult to correlate. The characteristics of sequences in the targets also profile the results of correlations.

For Target I, the outlier case, pair 5 reduces VQM and SSIM correlation index,

increasing to 0.81 and -0.79 correlation indexes for SSIM and VQM respectively without including this pair in the correlation.

SSIM correlation in Target III is weak, mainly due to the influence in SSIM metric of frame copy concealing effect needed in several layer temporal adaptation processes. The subjective preference is strong towards adapted sequence but SSIM gives smaller to adapted sequences, especially for the "Tractor" sequence (with lower temporal and complexity indexes).

Target II is weakly correlated for SSIM and VQM metrics. The scatter plot seen in Fig. 7 shows the relation of subjective CMOS and the logarithm of objective difference (Delta SSIM and Delta VQM). Pair numbers are included in the plot. Assessing the loss tolerance threshold (Target II) through objective measures is even more difficult than through subjective measures. Subjective evaluations showed that threshold definition relies on many unaccounted factors, such as content type or starting conditions (initial layers). Modeling that behavior through objective measures is complex.

PSNR is badly correlated in Target IV. Because PSNR does not consider inter-frame relations, the subjective improvement of increasing temporal layers is not being reflected on the objective results.

Group (pairs by objective)	PSNR subjective-objective	SSIM subjective -objective	VQM subjective - objective
Target I	0.7391	0.627604	-0.5476
Target II	0.459	0.6306	-0.6607
Target III	0.63123	0.51276	-0.85571
Target IV	0.2008	0.8728	-0.7988
Target V	0.5328	0.68202	-0.7534

Table 4 Values of the exponential correlation between objective metrics and subjective results
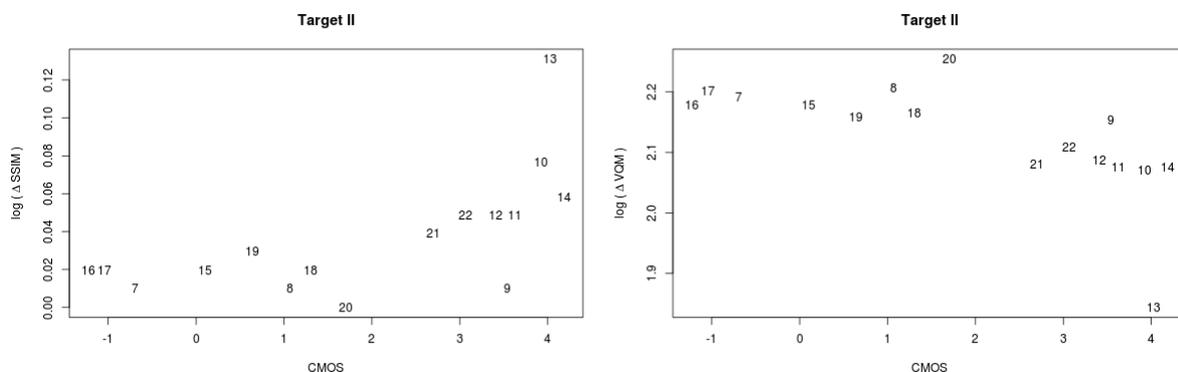

Figure 7 Target II VQM and SSIM

5. Conclusion and future work

In this paper, we have evaluated the relation between human decision and objective FR metrics. This work allows evaluating the readiness of the objective FR metrics to be applied to decision taking algorithms in the adaptive streaming systems. The large subjective studies have allowed performing an expansive study correlating subjective metrics with objective video quality algorithms. The particular nature of the subjective methodology makes it possible to corroborate if objective measures are able to mimic human decisions when choosing between different streaming systems.

Several authors postulated an exponential relationship between QoE and QoS parameters [13]. In our evaluation, we follow a different approach and we find the same relationship between the subjective assessment and the objective metric. We also find that subjective decisions between adaptive and non-adaptive systems have an exponential correlation with the difference between the objective metric results of both systems. We have also found that VQM has the best correlation with subjective preferences. Although the model fits well, it is not possible to obtain a high correlation with subjective data because there are many variables that cannot be taken into account, such as, among others, the characteristics of the video contents, illumination or the intention of the users to perform the test.

Our future work includes the subjective evaluation of more different contents and the study of the correlation of the subjective results with other objective metrics, including reducing reference and no-reference metrics, as well as full-reference metrics. Also, the use of objective metrics in adaptation streaming algorithms has to be analyzed.

References

[1] Zink M., Künzel O., Schmitt J., Steinmetz R., «Subjective Impression of Variations in Layer Encoded Videos», 11th International Conference on Quality of Service (IWQoS 2003). Eds. Springer Berlin Heidelberg, 2003, pp. 137-154. http://dx.doi.org/10.1007/3-540-44884-5_8

[2] Álvarez A., Pozueco L., Cabrero S., García Pañeda X., García R., Melendi D., Díaz Orueta G., «Evaluaciones subjetivas de servicios streaming adaptativos vs no-adaptativos.», Jornadas de Ingeniería Telemática (JITEL 2013). Granada, España, 2013, pp. 197-204

[3] Sohn H., Yoo H., De Neve W., Kim C. S., Man Ro Y., «Full-Reference Video Quality Metric for Fully Scalable and Mobile SVC Content». IEEE Transactions on Broadcasting, vol. 56, n.o 3, pp. 269-280. Sep. 2010. http://dx.doi.org/10.1109/TBC.2010.2050628

[4] Seshadrinathan, K., Bovik, A. C., «Motion-based perceptual quality assessment of video», SPIE 7240, Human Vision and Electronic Imaging XIV, 2009, p. 72400X-72400X-12. http://dx.doi.org/10.1117/12.811817

[5] Chikkerur S., Sundaram V., Reisslein M., Karam L. J., «Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison». IEEE Transactions on Broadcasting, vol. 57, n.o 2, pp. 165-182. Jun. 2011. http://dx.doi.org/10.1109/TBC.2011.2104671

[6] Seshadrinathan K., Soundararajan R., Bovik A. C., Cormack L. K., «A subjective study to evaluate video quality assessment algorithms», SPIE 7527, Human Vision and Electronic Imaging XV, 2010, vol.

7527, p. 75270H-75270H-10. http://dx.doi.org/10.1117/12.845382

[7]   Alreshoodi M., Woods J., «Survey on QoE\QoS Correlation Models For Multimedia Services». International Journal of Distributed and Parallel Systems (IJDPS), vol. 4, n.o 3. May. 2013.

[8]   Lee J. S., De Simone F., Ramzan N., Zhao Z., Kurutepe E., Sikora T., Ostermann J., Izquierdo E., Ebrahimi T., «Subjective evaluation of scalable video coding for content distribution», Proceedings of the International Conference on Multimedia. Firenze, Italy, Oct. 25–29, 2010. http://dx.doi.org/10.1145/1873951.1873981

[9]   Rückert J., Abboud O., Zinner T., Steinmetz R., Hausheer D., «Quality Adaptation in P2P Video Streaming Based on Objective QoE Metrics», in NETWORKING 2012, Vol. 7290, pp. 1-14. 2012.

[10] Alvarez A., Pozueco L., Cabrero S., Pañeda X. G., García R., Melendi D., and Orueta G. D., «Subjective evaluation of critical success factors for a QoE aware adaptive system», Computer Communications, vol. 36, n.o 15-16, pp. 1608-1620. Sep. 2013. http://dx.doi.org/10.1016/j.comcom.2013.07.005

[11] Singh K. D., Ksentini A., Marienval B., «Quality of Experience Measurement Tool for SVC Video Coding», IEEE International Conference on Communications (ICC 2011). 2011, pp. 1-5. http://dx.doi.org/10.1109/icc.2011.5963252

[12] Calyam P., Ekici E., Lee C.-G., Haffner M., Howes N., «A GAP-model based framework for online VVoIP QoE measurement». Journal of Communications and Networks, vol. 9, n.o 4, pp. 446-456. Dec. 2007. http://dx.doi.org/10.1109/JCN.2007.6182880

[13] Fiedler M., Hossfeld T., Tran-Gia P., «A generic quantitative relationship between quality of experience and quality of service». IEEE Network, vol. 24, n.o 2, pp. 36-41. Mar. 2010. http://dx.doi.org/10.1109/MNET.2010.5430142

[14] Vranješ M., Rimac-Drlje S., Žagar D., «Objective video quality metrics», 49th International Symposium ELMAR-2007. Croatia, Sep 12-14, 2007.

[15] Detti A., Bianchi G., Pisa C., Proto F. S., Loreti P., Kellerer W., Thakolsri S., Widmer J., «SVEF: an open-source experimental evaluation framework for H.264 scalable video streaming», IEEE Symposium on Computers and Communications (ISCC 2009). Sousse, Jul. 5-8, 2009. http://dx.doi.org/10.1109/ISCC.2009.5202390

[16] Huynh-Thu Q., Ghanbari M., «Scope of validity of PSNR in image/video quality assessment», Electronics Letters, vol. 44, n.o 13, pp. 800-801, 2008. http://dx.doi.org/10.1049/el:20080522

[17] Wulf S., Zolzer U., «About the imperfection of objective quality metrics on high-definition video content», IEEE 15th International Workshop on Multimedia Signal Processing (MMSP 2013). 2013. http://dx.doi.org/ 10.1109/MMSP.2013.6659319

Copyright Disclaimer