Quality user experience in advanced IP video services

António Pedro Marques Inácio
Instituto Superior Técnico
Av. Prof.Aníbal Cavaco Silva, Oeiras, Portugal
Email: apedro333@gmail.com

***Abstract*—The continuous improvement in the delivery of advanced video services along with the progress in the technical conditions at the client side contributed to the appearance of new methods for the evaluation of QoS and QoE from the user's point of view. The work described in this thesis aimed to develop a model that correlates the relevant QoS parameters and QoE factors with impact on the variation of the user's perception of the quality. To achieve this goal, a process was planned and designed, taking into account all the relevant techniques and recommendations, duly described in the survey of the related work. This process started with a quality assessment test, performed by 40 participants that ranked more than 140 videos. Taking into account all the test results, a detailed analysis was performed and six factors were studied: the bitrate, the resolution, the temporal activity, the spatial activity, the level of interest and the device type. From the analysis of the collected data it was possible to conclude that all these factors had great impact on the perceived quality. To aggregate all these factors in one single model, a linear regression was used in order to join their behavior, associating adequate weights to each one. To validate the model a set of tests was performed, with special attention to the *Pearson Coefficient* in order to analyze the accuracy of the model. The result was encouraging, achieving 99% of accuracy. This model may be considered an interesting non-reference metric to infer the perceived quality, applicable in many contexts and would allow Service Providers to have an estimation of the rank that the observer will probably give to a content, allowing consequently the tuning of features that would maximize the user experience.**

***Index Terms*—Quality of Service; Quality of Experience; Quality of Perception; Streaming Technologies.**

I. INTRODUCTION

The Internet Protocol Television (IPTV) service environment has recently been under a dramatic change, along with the IPTV market growth and subscribers' demanding more requirements. This growth leads to the need of making improvements to the IPTV service delivery, from the point of view of the quality that maximizes the subscriber experience.

The reality of video services delivery is constantly changing and excels by the great competition between Service Providers. Faced with this competition, and taking into account the strong technological developments in the sector, Service Providers need to adopt a different perspective, more focused on customers and their global satisfaction.

The success of a particular service is ultimately determined by what the end-user has experienced from it. In this context, it is essential to take advantage of the various technologies and concepts that are evolving and expanding, using them in a consistent, continuous and orchestrated manner for the improvement of user experience.

With the advances in Internet Protocol (IP) video coding, such as Scalable Video Coding (SVC) [1], and new video transport technologies such as Hypertext Transfer Protocol (HTTP) Streaming of Moving Picture Experts Group (MPEG) media and Rich Internet Application (RIA) technologies, new methods are required for the evaluation of the Quality of Service (QoS)/Quality of Experience (QoE), from the user's point of view.

A. Problem

Researchers have been trying to find models that may explain consistently the relationship between the quality of the service and the quality perceived by the user. However, studies in this area have proven to be quite complex due to the extreme ambiguity surrounding the human factors and the correlation between subjective measure scales and objective parameters.

Each end-user can have a different perception of video quality. This perception can be influenced by the complete end-to-end system effects (network infrastructure) or by user expectations, ambient conditions, psychological factors, application context, etc. While the component related to network infrastructure enables the easy definition of quality metrics, issues related to end-users are difficult to expose in the form of measurable parameters, due to the ambiguity of human factors.

Furthermore,the few developed metrics to solve this problem are constantly based on a video reference, which is compared with many versions of the same video. This mechanism disables the possibility of using the model in real time environments, where no reference is available.

B. Objectives and Contributions

Quality of Perception (QoP) is very subjective in nature and it is very important that a strategy can be defined to measure it as realistically as possible. The ability to evaluate QoP will give to a provider some sense of the intrinsic contribution that the performance of its network gives to the overall level of subscriber satisfaction.

In order to address this type of problem, the work proposed in this paper aims to develop a QoS/QoP correlation model for

[1] an extension to the H.264/MPEG-4 Part 14 (MP4) Advanced Video Coding (AVC) standard

QoE evaluation that will try to characterize the user's perception on multimedia quality through a mapping between QoS parameters and their relevance to user perceived qualifiers. Furthermore, we aim to develop a metric that may be use in real time environments.

Allied to this objective, fit a set of contributions in this scientific area, such as:

- Methods and heuristics for predicting the end-user perceived quality of visual stimulus;
- Mechanisms to adapt the quality in video streaming;
- Real-time rating mechanisms.

II. Related Work

In this section an overview on the quality in advanced video services, traditional and evolved quality assessment models and a survey on streaming technologies and video coding techniques are presented.

A. *Quality in Advanced Video Services*

Network infrastructures and end user devices are constantly evolving and growing. Today, not only Service Providers are able to deliver better services but also end user devices are increasingly able to support higher bit rates and higher processing power. These factors incentive the increase in the demand for real time multimedia streaming services, and Service Providers feel the need to improve the quality of their services to ensure higher customer satisfaction. In this context, for video services it is possible to distinguish three approaches in terms of quality [1]:

Quality of Service. Traditionally, the approach used for quality management was only related to networks and systems QoS metrics, such as jitter, delay, packet loss, throughput or availability. QoS can be considered as a technology-centered concept and is defined in [1] as the *"totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service."*. Measurements of network impairments can be used to estimate the impact on video quality, but do not represent an exhaustive set of metrics that enable an end-to-end quality management.

Quality of Experience. In [2] the authors state that QoE is the *"user's perceived experience of what is being presented by the Application layer, where the Application layer acts as a user interface front-end that presents the overall result of the individual quality of services."*. The majority of the work on QoE uses experimental methods for the evaluation, whereby a set of participants assign a score using a scale of values, meaning that the QoE results are usually presented in terms of user satisfaction.

Quality of Perception. User-centered approach, which is distinguished from QoE by the type of data collected from users [1]. According to ETSI [1] this approach is mainly concerned with the *"detectability of a change in quality or the acceptability of a quality level"* and, as for QoE, is based on the idea that the technical perspective alone is incapable of defining the perceived quality of multimedia video.

B. *Traditional Quality Assessment Models*

To assess video quality it is important to conduct the quality tests efficiently and timely. The assessment of video quality on in this type of systems can be made in two distinct ways: subjectively or objectively. To decide the methodology to evaluate video quality all methods were studied and taken into account.

Objective Methods. Objective evaluation of QoE data include measures of the communication processes and tasks outcome [1]. These methods are instrumental techniques that produce, from several measurements, results that approximate the ratings that would be obtained by using subjective methods. It can be categorized into five different types [3]:

- Media-Layer models focus on the simulation of the characteristics of human vision system using speech or video signals to gather QoE information [3] and can use Full-Reference (FR), No-Reference (NR) or Reduced-Reference (RR) metrics [4].
- Parametric Packet-Layer models predicts QoE only from packet-header information allowing a very light measurement, without handling the media signal itself [3].
- Parametric Planning Models require a priori information about the system, as the input they take consists of the quality parameters for networks and terminals [3].
- Bitstream-Layer models is a new objective quality-assessment approach that use both packet-header, as in parametric packet-layer models, and encoded bitstream (media payload) information [3] [5].
- Hybrid Models [6] are a combination of the previously mentioned technologies. These models aim to obtain the best of those different approaches, exploiting as much information as possible to predict QoE.

Subjective Methods. Subjective assessment models use measurements to anticipate, more directly, user perceptions, using human viewers to rate the video quality [4]. In these kind of experiments, a number of human viewers is asked to watch a set of video clips and rate their quality [7]. The results of the tests are treated statically and the output is often an average of the quality ratings known as Mean Opinion Source (MOS) [8]. The MOS is generated by averaging the results of a set of standard, subjective tests where a number of users rate the quality on a five point scale from 1 (Bad / Very Annoying) to 5 (Excellent / Imperceptible impairments) [9]. In this thesis eight types of subjective methods were explored: The double-stimulus continuous quality-scale (DSCQS), Double-stimulus impairment scale (DSIS), Single-stimulus (SS), Absolute category rating with hidden reference (ACR-HR), Stimulus-comparison (SC), Single stimulus continuous quality evaluation (SSCQE), Simultaneous double stimulus for continuous evaluation (SDSCE) and Subjective Assessment of Multimedia Video Quality (SAMVIQ) that are specified by the International Telecommunication Union (ITU) in their standards International Telecommunication Union - Radio-communication (ITU-R) BT 500-12 [10] and International Telecommunication Union - Telecommunication (ITU-T) P.910 [11] and synthesized in the full thesis document.

C. Evolved Quality Assessment Models

Subjective assessment models show excellent results, but there are factors that are extremely difficult to predict, leading to some unexpected errors in the results. Given this situation, researchers believe that the best way to ensure the validity of the quality assessment models results is through the establishment of a symbiosis between subjective and objective quality assessment models. Recent developments in this area aim to understand the effects of human factors in the quality assessment models and some attempts were already made to define an objective method for measuring the perceptual quality of video. These attempts can be seen in Jong Kim *et al.* [12] model and in Yamagishi *et al.* work [13].

HSV-based Metrics. The existing and standardized metrics measure image fidelity instead of perceived quality. To eliminate this shortcoming, some studies were developed that take into account the perception of end-user. Typical Human Visual System (HVS)-based metrics include the **Perceptual Distortion Model** [14], **Just Noticeable Difference (JND)** [15], **Digital Video Quality (DVQ)** [16] and the **Structural Similarity Index (SSIM)** [17]. These types of metrics have excellent results in the user's perception assessment, however, despite all the advantages, measures based on complex HVS models are computationally expensive and not practical for real-time video applications.

Trends in User's Perception Assessment. In order to solve this problems, many approaches appeared based on the analysis of features with strong correlation with the human perception. In this context, Hekstra *et.al* [18] proposed the Perceptual Video Quality Measure (PVQM), that uses a linear combination of three indicators, namely the loss of edge sharpness, the color error normalized by the saturation, and the temporal variability of the reference video. In order to increase the correlation between subjective and predicted video quality, Oelbaum *et al.* [19] developed an extension of Peak Signal-to-Noise Ratio (PSNR), the **PSNR-plus** method, that operate by estimating the parameters of the linear regression line for each video sequence. Bhat *et al.* [20] developed a new perceptual metric for compressed video, identified by **Predicted MOS (MOSp)**. The aim of this metric is to automatically predict the visual quality of compressed video in real time.

III. Architecture

This section describes the architecture proposed to define a QoP-Aware WebTV model that uses the characteristics of contents and the user preferences to maximize the multimedia experience. An empiric no-reference assessment model is also proposed to correlate QoS parameters with the factors that have impact on the perceived video quality. Through this correlation it will be possible to characterize (in a MOS scale) the quality perceived by each user of any type of content, taking into account a no-reference assessment method.

A. Functional Requirements

To achieve our main goal, it is very important to define some important requirements:

- **MOSp**—This metric should infer the quality perceived by the user without any video reference;
- **MOSp Accuracy**—This metric should infer the quality perceived by the user with a accuracy superior to 97%;

B. Non-Functional Requirements

- **QoE Monitoring**—The client player must constantly analyze the network and host conditions, in order to capture some parameter values, such as, delay, delay jitter, packet loss, latency, available bandwidth, local CPU load, screen resolution.
- **QoP Rating**—The user can assess the level of perceived quality at any time.
- **QoP Monitoring**—The QoP Monitoring is performed exclusively in the client side.

C. Formulation process and Architecture

To formulate a metric that represents, with great accuracy, the quality perceived by the user, it is necessary to define a well-structured process. Throughout this process a set of modules and technologies will be coordinated bearing in mind the ultimate goal.

As can be seen in Figure 1, along the process three technological components (the database, the streaming server and the client user interface) interact with four modules (assessment, data analysis, formulation and feedback) from the analysis of data.

Database. The database (DB) was built to store the classification given by each observer to each video clip.

Streaming Server. The web streaming server stores all the available video files used in the tests.

Clients. The number of client terminals will be as many as the number of observers, previously selected to perform the tests.

Assessment Module. The assessment module will integrate all actions relating to the evaluation of a video, allowing the collection of complete data, which will serve as input to the next module.

Data Analysis Module. This module is responsible for collecting and analyzing all the data sent by the assessment module.

Formulation Module. The feedback module is responsible to, through a mathematical procedure, combine all the analyzed data and infer the quality perceived by the user.

Feedback. The treated data will be sent to the streaming server in order to advertise the service provider to the factors that are influencing the user experience. Through the proposes formulation, it is possible to tune the different factors that may be used to maximize the user experience, which will result in a better service and consequently in more satisfaction for the customers.

D. Features do explore

From several aspects found in related work regarding the features that would eventually influence the user perceived quality, a set, apparently uncorrelated, was considered as the most promising for the analysis. The selected features are

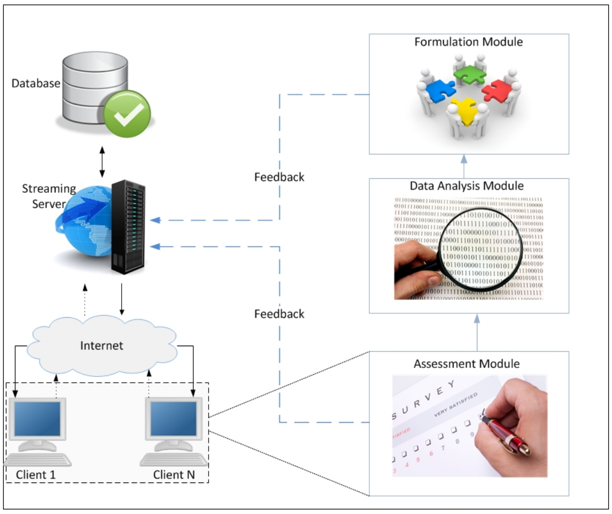

Figure 1: The used architecture structured by modules.

related not just to QoS, but also to human factors, due to human behavioral complexity, and described as follows: **Bitrate**, **Level Of Interest**,**Temporal Activity**, **Spatial Activity**, **Resolution**, **Device Type**.

IV. Subjective Quality Evaluation

The subjective evaluation was performed in order to infer the influence of some specific factors in the overall QoP previously described. This section describes the methodology that has been followed to prepare and setup the subjective test sessions, covering the materials used,the participants selection, the scenario of the tests, the rating method and the architecture used.

A. *Selection of Test Materials*

To perform these subjective tests based on a set of videos with very specific and diverse features, encompassing much of the existing video content, 28 movies were selected and prepared.

The set of videos is highly diverse, containing videos of action and sports with high motion and detail but, also, videos about nature where the detail stands up and the motion is lower. Some videos, whose content is greatly appreciated and viewed on the Internet, such as novels, romantic movies, celebrated dialogs between characters, rally, news, were also chosen, in order to assure a great variety of content.

Thus, and to ensure that each observer would evaluate different types of content, the videos are grouped into four categories of seven videos each. Each group has at least one video with great amount of motion, one video with high detail, one video with low motion, one video with low detail, one video with a normal level of motion and one video with a normal level of detail.

Taking into account that each video has five versions (one for each bitrate), this means that each observer will evaluate a set of 35 videos. Overall, 140 videos were published in the Web server for the evaluation.

B. *Selection of Participants*

In order to reduce possible errors related with the choice of participants, or due to an eventual inability of the observers, a set of preliminary tests was conducted. These preliminary tests, which would define the group of observers were divided in two steps: Initial Survey and Visual Acuity (Snellen chart and Ishihara plates tests).

C. *Scenario*

In the sessions, the panel of observers experienced an environment similar to what they are used to at their homes. This scenario was simulated in the office 1.4.24 at Instituto Superior Técnico (IST) - Tagus Park and was prepared to meet the requirements established by ITU-R [10]. It was taken into account factors as the environmental luminance, the observation angle, the observer posture, the brightness and the contrast, the distance between the observer and the screen and the room environment.

D. *Quality Assessment Method*

After a thorough explanation to the participants regarding the evaluation process that would be performed, with the purpose of assuring the complete adaptation of each observer to the assessment system and to the test methodology, the experimentation phase began. In this phase each observer evaluated a video sequence of 5 videos and clarified any eventual questions regarding the process. After that experimentation phase, the assessment test phase started, with the reproduction of a random set of video sequences.

During the video quality assessment tests, the observers rated the perceived video quality of 35 video clips, grouped in 7 different sequences. As explained above, each video version has a duration of 20 seconds, making a maximum of 15 minutes per session, as recommended in [21]. The assessment system is based on two subjective metrics described in Section 2: SS and SSCQE.

Under these conditions, at every 20 seconds (time of each video version) a ranking window pops-up in the screen and the video sequence is automatically stopped, in order to allow the observers to assess, avoiding any type of pressure, the video they had just seen. The vote is made through a discrete MOS scale (from zero - very bad to ten - excellent). At the end of each sequence the observers were asked about their interest on the content of the previous sequence in order to study the content impact.

V. Data Analysis and Formulation

In this section an analysis of the data collected with the assessment was performed. Throughout the following sections will be expose the analysis and conclusions taken from all the evaluation session.

A. Analysis of Participants' Data

In the end of the test session days, we received on the test room, 43 candidates (31 men and 12 women), that were submitted to the visual acuity test and to the preliminary questionnaire. The major part of observers were students in technological areas with ages between 18 years old and 33 years old.

The participants selection intended to, assure a group of observers fulfill a set of requirements(e.g. occupation, experience and visual acuity). Taking into account the established requirements, three candidates did not fulfill, at least one of them, and were eliminated from the tests.

Two of the set requirements were determinants to refuse three participants: the profession and the Sneelen chart. One of the candidates was refused due to is experience in video and multimedia areas and the two others had some visual incapacities and did not brought the glasses, factor that made it impossible to pass in the visual acuity tests.

B. Assessment results analysis

Taking into account the evaluation made by the observers the following sub-sections explore the impact of each considered feature in the QoP:

MOS as function of Interest Level. This study allow taking conclusions about the human factors, in this case, about the human subconscious. Unconsciously, the observers voted the quality of the videos based on their own preferences,a factor that was considered of extreme importance as also very interesting.

Figure 2 illustrates the impact of the interest level variation in the assessment results. In this figure, these assessment results are represented by MOS level, turning possible to conclude that:

- The interest level had a great importance in the video assessment, regardless of the bitrate;
- For Level 1, the classification difference between the minimum and maximum bitrate is about 3.5 points in the MOS scale. In an extreme situation (Level 5), the classification difference between the minimum and the maximum bitrate is about 6 points in the MOS scale.
- On average, the same video, with a lower bitrate had a rating of one point for a minimum level of interest and 3 points for a maximum level of interest. However, this difference of 2 points in the MOS scale increases to more than 4 points when the bitrate is maximum, ranging from 4.5 point in the Level 1 to 8.8 points in the Level 5.

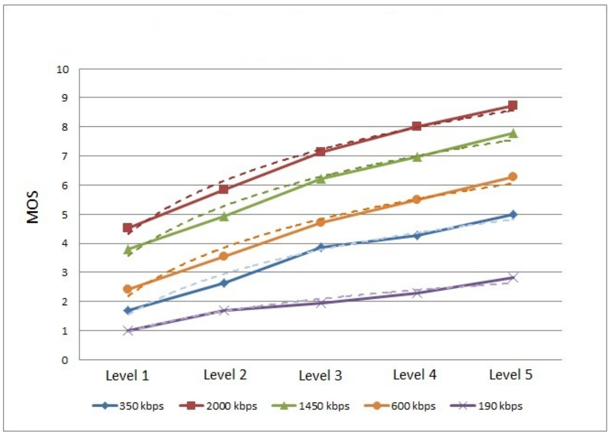

Figure 2: Chart representing MOS as function of interest level for each available bitrate.

Averaging each tendency line, it is possible to present a global expression for the MOS in function of interest level (Formula 1):

$$\textbf{MOS} = 2.1209 \ln InterestLevel + constant \quad (1)$$

MOS as function of Bitrate. It is now possible to represent the impact of the bitrate variation in the assessment results, as illustrated in Figure 3. In this figure, the assessment results are represented by MOS, turning possible to conclude that:

- Keeping the bitrate fixed, it can be observed that the classification increases with the increase of the interest level, as demonstrated previously;
- At each interest level, with the increase in bitrate, the classification in the MOS scale increases logarithmically;
- To minimum bitrates (190 Kbps) the MOS classification is independent from the interest level, staying between Level 2 and Level 3, with an average of 2 points;
- At the Levels 4 and 5, for 190 Kbps, the same situation is verified but with a classification of 3 points;
- For a maximum bitrate (2000 Kbps) it is possible to verify that the classification increases, in average, 1 point at each interest level.

It is possible to present a global expression for the MOS in function of bitrate (formula 2):

$$\textbf{MOS} = 1.98628 \ln Bitrate + constant \quad (2)$$

MOS as function of Temporal Activity. From the treatment of the results, it was possible to represent the impact of the temporal activity (motion) in the assessment results, as illustrated in Figure 4. In this figure, the assessment results are represented by MOS, turning possible to conclude that:

- There is a dependency between the classification of a given video and the temporal activity, or motion. In other

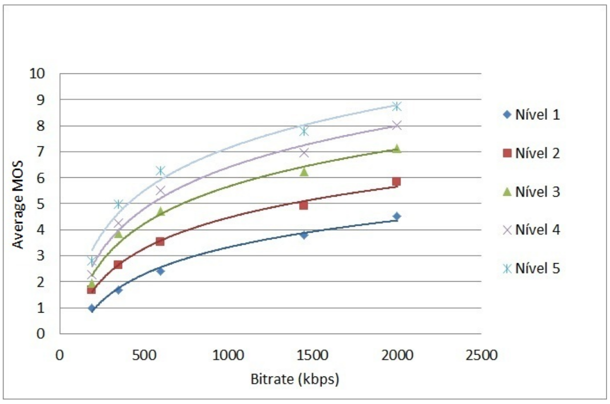

Figure 3: Chart representing MOS as function of Bitrate for each level of interest.

words, it can be noticed that the observers perceive more easily the degradation of videos if they possess a high level of motion. Consequently, the observers lower their classifications to videos with high levels of motion (loss of detail).

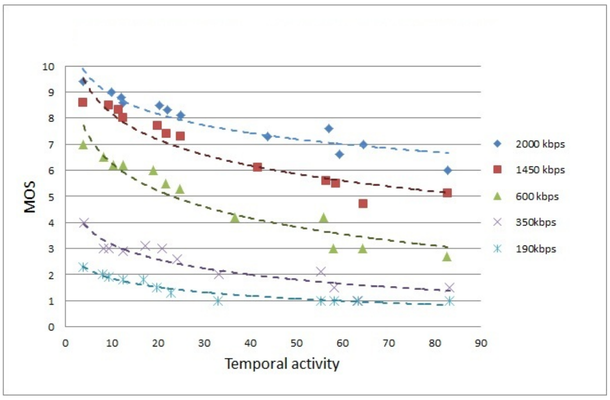

Figure 4: Chart representing MOS as function of temporal activity for each available bitrate.

Thus, we can conclude that, on average, the MOS as a function of the temporal activity, can be calculated by Formula 3:

$$\textbf{MOS} = -1.069 \ln \mathit{TemporalActivity} + \mathit{constant} \qquad (3)$$

MOS as function of Spatial Activity. The impact of the spatial activity (detail) in the assessment results is represented in Figure 5, turning possible to conclude that:

- There is a relationship between the MOS classification and the spatial activity, regardless of the the bitrate.
- Despite the impact of the level of detail in the perceived quality being not as strong as the impact of motion, it is visible that above 600 Kbps the detail variation is more important, resulting in variations in the MOS classification up to 4 points, for a bitrate of 1450 Kbps.
- On the other hand, for a bitrate of 190 Kbps, the detail variation has an impact that results in a maximum classification variation of 2.5 points.

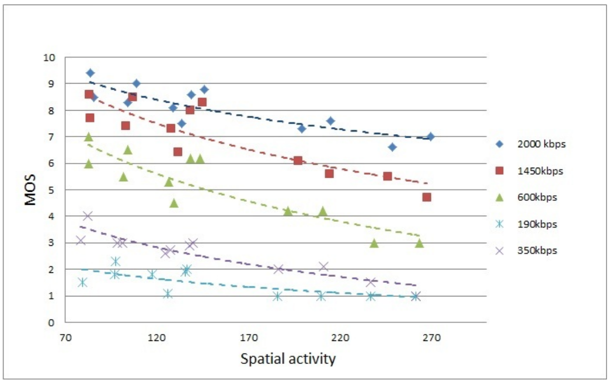

Figure 5: Chart representing MOS as function of spatial activity for each available bitrate.

Thus, we can conclude that, on average, the MOS as a function of the spatial activity, can be calculated by Formula 4:

$$\textbf{MOS} = -2.0674 \ln \mathit{SpatialActivity} + \mathit{constant} \qquad (4)$$

MOS as function of Resolution. The video resolution is a factor extremely important when viewing a video. Therefore, 10 of the test videos where coded in 7 different resolutions: 1920x1080 (16:9), 1080x720 (16:9), 1024x768(4:3), 800x600(4:3), 640x480(4:3), 320x240(4:3) and 160x120(4:3), in order to understand the impact on the perceived quality.

From the assessment results, and analyzing Figure 6, it can be concluded that:

- For each available resolution, the MOS classification presents a logarithmic distribution, with variable bitrate;
- Regardless of the bitrate, the difference between the resolution 1980x1080p and 1080x720p is insignificant.
- It is quite visible a change in the MOS, when aspect ratio is changed from 16:9 to 4:3. The difference between 1080x720p and 1024x768 is of almost 2 points with a bitrate of 2000 Kbps;
- The resolution impact on the perceived quality is more determinant to bitrates above to 600 Kbps;
- To a bitrate below to 600 Kbps the difference between resolutions is almost insignificant;
- To a minimum bitrate the influence of resolution is not significant;
- With the 4:3 aspect ratio, the MOS classification decreases 0.5 points at each resolution decrease, to a bitrate equal of above to 600 Kbps;

MOS as function of Device Type. As in the case of resolution, the device type was also considered for an important factor in the perceived quality. Therefore, three devices were tested: a laptop with screen size of 13.3”, a desktop with a screen size of 19” and a smartphone with a screen size of 3.5”. From the obtained results, represented in Figure 7, it can be concluded that:

- For each available device, the MOS classification presents a logarithmic distribution with variable bitrate;

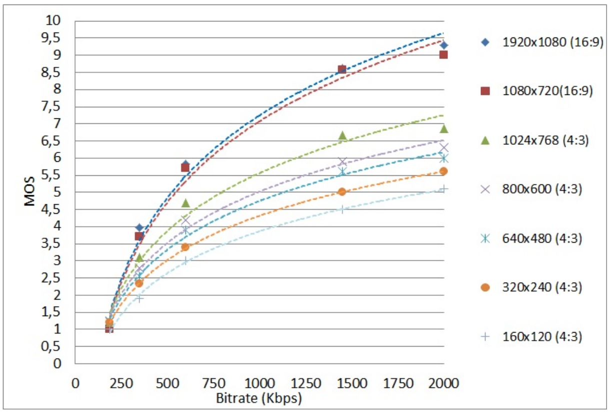

Figure 6: Chart representing MOS for each resolution.

- For each bitrate, the MOS classification for the device with 19" is 1 point higher than the MOS classification for the device with 13.3".
- The MOS classification for the device with 13.3" is 2 point higher than the MOS classification for the device with 3.5", for a bitrate above to 600 Kbps, and 1 point higher for bitrates lower than 600 Kbps.

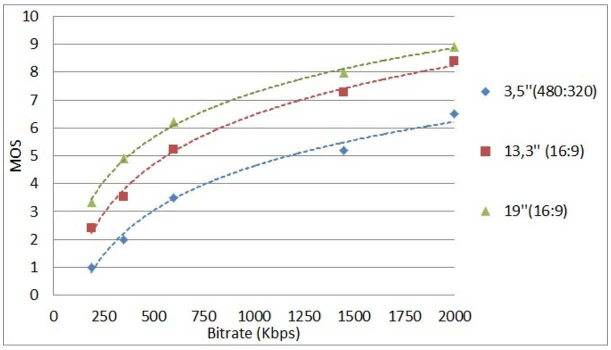

Figure 7: Chart representing MOS for each device type.

VI. Formulation and Test Validation

In the previous section, four different expressions were obtained to characterize the different behaviors of MOS results, when subject to the effect of the features under study: Bitrate, Level of Interest, Temporal Activity and Spatial Activity.This section describes the mathematical procedures to encapsulate the different expressions in a single and global expression.

After the analysis of the relationship between feature values and MOS, it is necessary to combine them for accurately predicting the MOS values. This combination will follow a linear regression model, represented in Formula 5. Specifically, the predicted MOS value can be seen as the dependent variable in a linear equation, modeled as a function of the feature values and their corresponding **linear weights**.

$$MOS_p = \beta_0 + \Sigma_{i=1}^{n}(\beta_i \cdot \chi_i) \tag{5}$$

Where β_0 is the offset value, $[\beta_1 \cdots \beta_i]$ are the linear weights associated to each feature and χ_i is the value of feature i.

It is necessary to find a methodology to allow the weight calculation for each feature and to obtain the final formula that will infer the perceived MOS. One possible method to compute the weights ($\beta's$) is by minimizing the square error between MOS (the true MOS) and MOS_p (the estimated MOS), for a set of training video sequences. The training set is based on K video sequences, with their corresponding MOS values, from which K feature vectors will be extracted for training.

Regarding the ($\beta's$) results, the global prediction expression is achieved, and is represented in 6:

$$\begin{aligned} MOS_p = 0.01 &+ 0.75 \times (1.98628 \ln BR - 8.324) \\ &- 0.9 \times (-2.1209 \ln LI + 2.5097) \\ &+ 0.35 \times (-1.069 \ln TA + 8.138) \\ &+ 0.05 \times (-2.0674 \ln SA + 15.095) \end{aligned} \tag{6}$$

A. Validity Tests

To validate the MOS_p expression, some common tests to determine the consistency, validity, accuracy and usefulness of the expression were performed, namely:

- *Pearson Coefficient* (PC): 0,98 which represents an accuracy of 98%;
- *Root Mean Square Error (RMS)*: 0,54;
- Other tests:
  - Mean error: The mean error of MOSp expression is 0.44 (in a 0-10 scale);
  - Maximum error: The maximum error of MOSp expression is 1.3 (in a 0-10 scale).

To analyze in detail the accuracy of the expression, a chart representing all the results was traced. As can be seen in Figure 8 the ideal distribution between the real and the estimated MOS is represented, and all the points around it show the relation between the real MOS of each result with the estimated MOS obtained by the prediction expression.

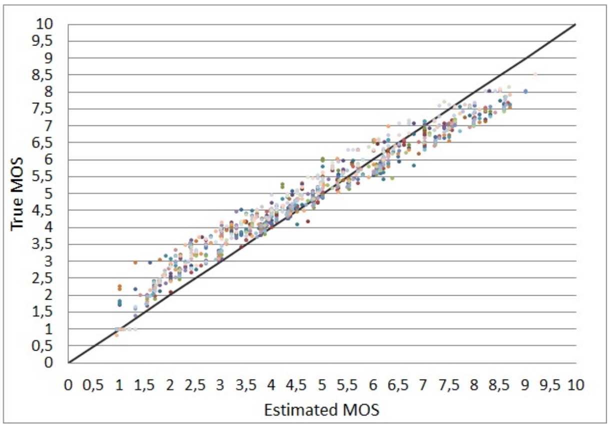

Figure 8: Chart representing the model accuracy.

After a deep analysis of the values a pattern was perceived related with higher MOS values almost always falling below the ideal line, and on the other hand, lower MOS values almost always falling above the ideal line. With the certainty that one additional factor would be forcing this deviation, a refinement of the function was worked out.

To define this function, and after observing the values and the behavior the choice fall on parameterizing the function at different bitrates (as one of the most influential features). As such, each step of the regression model was repeated for each bitrate, computing different weights for each bitrate. The weighs variation were small, due to the similarity of values entered for each bitrate, but helped refining the model.

At the end we obtained the following parametrization:

If the bitrate is superior to 1450 Kbps, the MOSp is given by:

$$\begin{aligned} MOS_p = 0.514 &+ 0.75 \times (1.98628 \ln BR - 8.324) \\ &- 0.9 \times (-2.1209 \ln LI + 2.5097) \\ &+ 0.35 \times (-1.069 \ln TA + 8.138) \\ &+ 0.05 \times (-2.0674 \ln SA + 15.095) \end{aligned} \tag{7}$$

If the bitrate is between to 600 Kbps and 1450 Kbps, the MOSp is given by:

$$\begin{aligned} MOS_p = 0.01 &+ 0.75 \times (1.98628 \ln BR - 8.324) \\ &- 0.9 \times (-2.1209 \ln LI + 2.5097) \\ &+ 0.35 \times (-1.069 \ln TA + 8.138) \\ &+ 0.05 \times (-2.0674 \ln SA + 15.095) \end{aligned} \tag{8}$$

If the bitrate is between to 190 Kbps and 350 Kbps, the MOSp is given by:

$$\begin{aligned} MOS_p = -0.4987 &+ 0.75 \times (1.98628 \ln BR - 8.324) \\ &- 0.9 \times (-2.1209 \ln LI + 2.5097) \\ &+ 0.35 \times (-1.069 \ln TA + 8.138) \\ &+ 0.05 \times (-2.0674 \ln SA + 15.095) \end{aligned} \tag{9}$$

Taking into account this parametrization, the tests were repeated to validate the solution:

- *Pearson Coefficient* (PC):
  Result:0.99, which reveals an amazing accuracy of our model and an increase when compared to the previous test.
- *Root Mean Square Error* (RMS):
  Result:0.281, which is a good indicator of validity in our model. As in the *Pearson Coefficient*, it can be seen and improvement relatively to the previous test.
- Other tests
  - Mean error: The mean error of MOSp expression is 0.24 (in a 0-10 scale);
  - Maximum error: The maximum error of MOSp expression is 0.77 (in a 0-10 scale);

Finally, the MOS/MOS_p comparison (Figure 8) was repeated but with the new results. The optimized results are illustrated in Figure 9. In this figure the deviation problem described in the Figure 8 has been corrected, which, allied with the improvement in the tests performed, is a good indicator to attest the validity of the model.

Given these results, the Model expression can be generally set as:

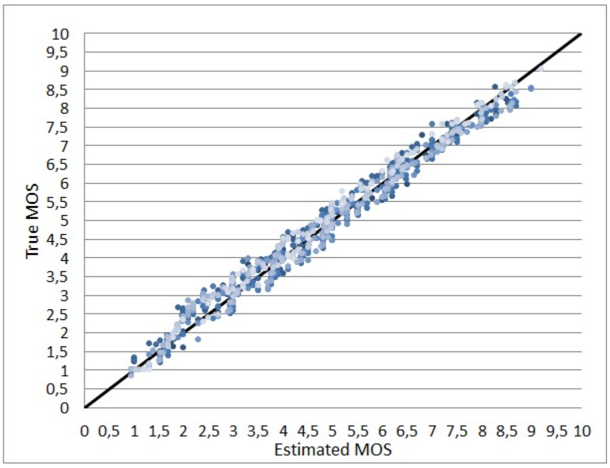

Figure 9: Chart representing the model accuracy after the refinement.

$$MOS_p = \beta_0 + MOS_{factors} \tag{10}$$

Where $MOS_{factors}$ corresponds to:

$$\begin{aligned} MOS_{factors} = {} & 0.75 \times (1.98628 \ln BR - 8.324) \\ &- 0.9 \times (-2.1209 \ln LI + 2.5097) \\ &+ 0.35 \times (-1.069 \ln TA + 8.138) \\ &+ 0.05 \times (-2.0674 \ln SA + 15.095) \end{aligned} \tag{11}$$

The parametric Model comes with the following final expression:

$$MOS_p(BR, LI, TA, SA) = \begin{cases} 0.514 + MOS_{factors} \\ \quad BR > 1450 \text{ Kbps} \\ 0.01 + MOS_{factors} \\ \quad 600 \text{ Kbps} < BR < 1450 \text{ Kbps} \\ -0.4987 + MOS_{factors} \\ \quad 190 \text{ Kbps} < BR < 350 \text{ Kbps} \end{cases}$$

Taking into account the obtained results, it is possible to state that the developed Model has great potential to be considered an accurate no-reference quality assessment method. Using this method it is possible to infer the quality perceived by the user without necessity of a video reference. This allows service providers to predict previously to the video visualization the rank that the observer will give to the content, consequently allowing the prevention of problems, by manipulation of some features in order to maximize the user experience.

VII. Conclusion

This paper addressed the inherent problem to the subjectivity of the QoP in order to define a strategy to characterize this concept about the perceived quality by the user. The major part of the studies in quality of video are focus on the quality of service, forgetting the components related to the human factors.

In this context, it was decided to incorporate in our strategy factors related with the QoS, but also other factors related with the human vision, human preferences and human perceptions. By incorporating factors completely different, the difficulty of our solution increased, also become more accurate and realistic. Other obstacle that we tried to exceed was the usual necessity of a reference video to predict the perceived quality, factor that decreases the range of the solution, making it unusable, for example, in real time system.

Taking into account our goals, the motivation and the context about our work, we developed a extensive study in this area, with particular focus in the concepts about quality, the objective and subjective quality assessment methods, and the evolved quality assessment models that focus on the human sense. After that, we studied some factors that influence the perceived quality to decide the set of features to explore.

Through this complete survey on state of art, we define a strategy and architecture to achieve our goal. We design the process that would culminate with the achievement of the QoP metric. This process was design with two distinct areas: the technological area and the data treatment area. The first area, was designed with 3 components: a database to save all the information, a web streaming server that stores all the videos and the clients. The data treatment is composed by four modules: assessment module responsible to receive all the users' evaluations, the data analysis module, the formulation module where is achieved our main goal, and, finally, the feedback module (responsible to communicate to the provider the obtained results in order to improve the user experience).

After defining the architecture, a set of features to be analyzed were selected: the bitrate, the resolution, the device type, the temporal activity, the spatial activity and the interest level.

The next phase of this work was to defining a set of materials and the tests to the subjective assessment tests. To participate on the evaluation the participants were submitted to some preliminary tests. Finally, it was defined the quality assessment method to proceed the evaluation. The chosen method was a mix between SS and SSCQE.

The evaluation was performed by 40 observers and 4 types of tests, distributed by the motion and detail level. All the collected data was analyzed, and some conclusions about the impact of the studied features in the perceived quality were inferred.

From the studied features, we concluded that the most influent was the bitrate and the level of interest, and the less important was the spatial activity. Taking into account all the conclusions and the results, we defined a parameterized expression to infer the perceived quality, for 3 ranges of bitrates.

Finally, we performed some validity tests to approve our solution. The *Pearson Coefficient* test, assess the accuracy of the model and the result was excellent, with an accuracy of 99% fulfilling the defined requirement. Thus, it is possible concluded that our model was a great non-reference metric to infer the perceived quality, and can be used in a great variety of contexts.

VIII. Future Work

In the middle of this dissertation, we understand that would be very difficult to perform a fair set of tests with seven features. We would need approximately 80 observers to obtain valid results. Taking this into account, and with the conscious that is very difficult to perform an evaluation with this dimension in a university environment, we decided to exclude two features from the evaluation: device type and resolution.

Nevertheless, we perform some tests (with a smaller dimension) with the excluded features and the results prove that this features should be included in the global expression. So, the major failure of our model is the exclusion of these two features. However, this gap should be understood as an opportunity of future work.

All the feedback and results we achieved encourage us to reason and plan on the improvement of the model. So, it would be very interesting to re-define our model recurring to this two features. This process would be very extensive and difficult, but we are confident that it is a great opportunity to increase and improve our results.

References

[1] ETSI, "Human Factors (HF); Quality of Experience (QoE) requirements for real-time communication services," European Telecommunications Standards Institute, TR 102 643, Jan. 2010.

[2] M. Siller and J. Woods, "QoS arbitration for improving the QoE in multimedia transmission," in *Visual Information Engineering, 2003. VIE 2003. International Conference on*, july 2003, pp. 238 – 241.

[3] A. Takahashi, D. Hands, and V. Barriac, "Standardization activities in the ITU for a QoE assessment of IPTV," *Communications Magazine, IEEE*, vol. 46, no. 2, pp. 78 –84, february 2008.

[4] S. Winkler, "Video quality measurement standards x2014; Current status and trends," in *Information, Communications and Signal Processing, 2009. ICICS 2009. 7th International Conference on*, dec. 2009, pp. 1 –5.

[5] K. Watanabe, K. Yamagishi, J. Okamoto, and A. Takahashi, "Proposal of new QoE assessment approach for quality management of IPTV services," in *Proceedings of the 15th IEEE International Conference on Image Processing, ICIP '08.*, Oct. 2008, pp. 2060 –2063.

[6] P. Romaniak, "Hybrid Solution for Quality of Experience Assessment of Video Streams Integrating Network Provider and User Approaches — Introduction to Research," Ph.D. dissertation, AGH University of Science and Technology, Krakow, Poland, 2009.

[7] S. Winkler and P. Mohandas, "The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics," *Broadcasting, IEEE Transactions on*, vol. 54, no. 3, pp. 660 –668, sept. 2008.

[8] M. Sullivan, J. Pratt, and P. Kortum, "Practical issues in subjective video quality evaluation: human factors vs. psychophysical image quality evaluation," in *Proceeding of the 1st international conference on Designing interactive user experiences for TV and video, UXTV '08*. New York, NY, USA: ACM, 2008, pp. 1–4.

[9] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union - Radiocommunication Sector, Recommendation BT.500-11:, June 2002.

[10] ——, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union - Radiocommunication Sector, Recommendation BT.500-12, sept. 2009.

[11] ITU-T, "Subjective video quality assessment methods for multimedia applications," International Telecommunication Union - Telecommunication Standardization Sector, Recommendation P.910 :, April 2008.

[12] H. J. Kim and S. G. Choi, "A study on a QoS/QoE correlation model for QoE evaluation on IPTV service," in *Advanced Communication Technology (ICACT), 2010 The 12th International Conference on*, vol. 2, 2010, pp. 1377–1382.

[13] K. Yamagishi and T. Hayashi, "QRP08-1: Opinion Model for Estimating Video Quality of Videophone Services," in *Global Telecommunications Conference, 2006. GLOBECOM '06. IEEE*, 27 2006-dec. 1 2006, pp. 1 –5.

[14] S. Winkler, "A Perceptual Distortion Metric for Digital Color Video," in *in Proc. SPIE*, 1999, pp. 175–184.

[15] J. Lubin, "A human vision system model for objective picture quality measurements," in *Broadcasting Convention, 1997. International*, sep 1997, pp. 498 –503.

[16] A. B. Watson, J. Hu, and J. F. M. Iii, "DVQ: A digital video quality metric based on human vision," *Journal of Electronic Imaging*, vol. 10, pp. 20–29, 2001.

[17] Z. Wang and A. Bovik, "A universal image quality index," *Signal Processing Letters, IEEE*, vol. 9, no. 3, pp. 81 –84, mar 2002.

[18] A. P. Hekstra, J. G. Beerends, D. Ledermann, F. E. de Caluwe, S. Kohler, R. H. Koenen, S. Rihs, M. Ehrsam, and D. Schlauss, "PVQM - A perceptual video quality measure," *Signal Processing: Image Communication*, vol. 17, no. 10, pp. 781 – 798, 2002.

[19] K. D. Tobias Oelbaum and W. Zia, "A Generic Method to increase the prediction accuracy of visual quality metrics," *PCS*, 2007.

[20] A. Bhat, I. Richardson, and S. Kannangara, "A new perceptual quality metric for compressed video," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, 2009, pp. 933–936.

[21] ITU-T, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," International Telecommunication Union - Telecommunication Standardization Sector, Recommendation J.144 (03/2004), Mar. 2004.