

Managing Quality of Experience on a Commercial Mobile TV Platform

Vlado Menkovski, Georgios Exarchakos,
Antonio Liotta

Electrical Engineering Department
Eindhoven University of Technology
Eindhoven, The Netherlands
{v.menkovski, g.exarchakos, a.liotta}@tue.nl

Antonio Cuadra-Sánchez

Indra
Parque Tecnológico de Boecillo
47151 Valladolid, Spain
acuadra@indra.es

Abstract – The user perceived quality or Quality of Experience (QoE) is of significant importance to multimedia service providers because of its relevance for efficient management of provided services. However, due to its subjective nature, QoE is difficult to estimate. Subjective methods are costly and impractical, while objective methods do not correlate precisely with the subjective perception. In addition to the challenges in estimating QoE, further challenges are presented in determining the means of managing the QoE in today's complex and varied multimedia distribution systems. As a result of the high number of components and parameters that affect the perceived quality, from content creation to delivery and presentation, the QoE aware management in these highly versatile environments becomes increasingly difficult. We present a method that uses limited initial subjective tests to develop prediction models for QoE as perceived by the viewers. This minimizes the complexities associated with subjective methods while maintaining the accuracy. Further we present a method of calculating the QoE remedies for managing the QoE per stream, based on the QoE prediction models.

Keywords - *Quality of Experience, QoE, Machine Learning, Subjective Testing, Monitoring, QoE management*

I. INTRODUCTION

Multimedia content broadcasting is commonly implemented in a highly diverse and varying environment. In addition to the diversity, the lack of standards for quality assessment in this domain makes estimation of the perceived service quality particularly difficult. The challenges with the estimation of the quality of experience (QoE) make it hard to know whether the delivered service meets the customers' expectations and brings user satisfaction.

The variability of the system is particularly high on the end-user terminal devices. The screen sizes, mode of use and computing capabilities of these devices highly affect the end user QoE. However, due to the technical difficulties and the lack of understanding of the QoE the common approach with service providers is to use an 'average' setup with regards to the multimedia parameters; that is a trade-off between the power of the average device and the quality of the parameters. In addition to this, the service providers need to take into account dimensioning of their own resources in a way that will deliver a functioning service while still being commercially viable.

The down side to the 'average' (or one-size-fits-all) approach is two faceted: i) the service provider remains in

the dark regarding the value of the service to its customers and ii) its resources are not optimally used. These two are usually in competition; any management technique should target the balance of that trade-off. This, however, would imply understanding of the customers' QoE.

The providers are not completely oblivious to the factors that affect the QoE, on the contrary they have access to a plethora of parameters that affect the QoE. Some of which are the QoS parameters, such as network conditions and performance, also the content encoding parameters and characteristics. Nevertheless, a gap exists between all these factors and the QoE itself. This gap is the major motivation of this work. The overall goal reached by this work is to develop a methodology that will bridge this gap and deliver accurate information about the perceived user experience looking at the application QoS (AQoS) and network QoS (NQoS) parameters.

Part of this work has been published in the MMEDIA 2010 conference [1], which focuses on building QoE models for a commercial IPTV platform. In this paper we have extended this work to include calculating the remedies for the QoE per stream that enable QoE management decisions.

In the following sections, we present a methodology and an implementation of a QoE assessment platform developed for a service provider. The designed platform estimates QoE of mobile TV services based on existing QoS monitoring data using QoE prediction models. The prediction models are built using Machine Learning (ML) techniques from subjective data acquired by a limited-size initial subjective test. The work here provides evidence for the efficiency of this methodology and elaborates on the QoE software platform. The QoE value produced by that platform enriches the system's monitoring tools [2] and will be further used for managing the service and dimensioning the resources.

To provide for QoE enabled management a remedies algorithm [3] is described and implemented, which calculates the parameters and amounts that need to be changed to reach the desired QoE in particular streams. The remedies algorithm is in a way extension of the QoE assessment method because it derives the remedies based on the QoE prediction models.

The paper continues with Section II discussing related work that deals with estimation of perceived quality of multimedia. In Section III, we present an overall description of the mobile TV probe-based monitoring solution that measures the QoS of the system and the QoE prediction

platform that delivers QoE values. Section IV discusses the method used for the QoE prediction platform, expanding to the subjective tests and machine learning algorithms. The results from the subjective tests and the ML prediction models are analysed in Section V. Section VI presents the remedies algorithm used to improve the QoE per multimedia stream. Section VII presents results from the remedies used in the particular IPTV platform. Finally, Section VIII sums up with the conclusions and future work.

II. RELATED WORK

Perception of quality for streaming video has been a lively field of research. There are many efforts mostly looking into objective methodologies. Some have also executed subjective tests either to estimate quality or to compare the accuracy of different objective approaches.

Due to a wide diversification of models, it is difficult to select the best model for video quality. In [4] the International Telecommunication Union (ITU) presents a classification of the different objective quality assessment models. Its authors have classified the models into media layer, parametric, bit-stream and hybrid models. This classification is based on the model's focus. The media layer models focus on the media signal and they use knowledge of the Human Visual System (HVS) to predict the subjective quality of video. The parametric ones look at the protocol information and statistics through non intrusive probes to predict the quality. The bit-stream models derive the quality via analysing content characteristics collected from the coded bit-stream information. In [5], a survey of different video quality methods is presented. The paper concludes that there are many different methods and algorithms for video quality estimation; thus, there is need for a standardized way to compare them. There are different international standardization bodies working in this area and they have delivered progress, as given in [4]. However, there is still lack of a comprehensive method for comparing the video quality assessment models with a subjective database that can bring a common reference point. The Peak Signal to Noise Ratio (PSNR) and Mean Squared Error (MSE) are purely computational metrics for comparison of models used by many publications in this research field. However, PSNR and MSE deliver unsatisfactory results as they lack of understanding of the HVS [6]. Therefore, as a common practice, a wide variety of published work in video quality uses subjective tests as a relevant comparison method, much of which follows the standard for subjective studies as given in [7].

As we have seen from the ITU standardization of the models, some models focus on the content, some on encoding and other on the transport of the multimedia content.

The authors of [8] give an analysis of the dependency of the perceived quality on the values of the video frame rate and encoding quantization. They have concluded that

traditional encoding schemes for frame-rate and quantization step are not optimal from the perspective of perceived quality. Keeping the frame rate at a lower value allows for higher budget of bits per picture for coding and produces higher quality overall.

The authors of [9] have developed a utility function for each of the following network QoS parameters: delay, jitter, packet loss rate and bandwidth of the video stream. They used generic utility functions for the parameters and derived the constants from results of executed subjective tests. They claim that managing the multimedia streams with the utility function approach is more effective than reservation protocols in today's converged network environments. Their work is still limited because of the fact that they have not considered the interdependency between the NQoS parameters but considered them independently.

On the perception of loss of data during transport, the work done in [10] presents a methodology that focuses on the stream, to determine the effects of loss on the video. First they estimate the artefacts in the video due to the loss of data. Then, they try to study the visibility of those artefacts and their correlation with the perceived quality. The paper discusses a comprehensive analysis of the error handling schemes of H.264 video codec in order to predict the video artefacts. Then it continues to analyse the artefacts from the point of view of magnitude (spatial inconsistency and special extent), special priority (region of interest) and temporal duration. The results show that this approach can sometimes follow the trend of Mean Opinion Score (MOS) of the subjective study better than the PSNR values but the method is still not accurate sometimes even less than PSNR.

The methods so far are all looking at particular factors that affect the QoE but none of them takes a holistic approach and look at all of the factors. The concept of quality does not have a single dimension but more than one [11], such as qualitative, emotional, and communicative. QoE also encompasses the expectations that the viewers have. So, for accurate assessment, subjective feedback is necessary.

The work done in [12] builds on subjective tests and delivers prediction models using discriminant analysis. Furthermore, in [13] and [14], improvements to the accuracy of the prediction models are given. The work here shows the multiple benefits of optimizing the QoE instead of targeting specific QoS metrics.

Understanding the significance of QoE in determining the value of the service and establishing the optimal balance between resources and quality we propose this method that builds up on previous work done in the area. More particularly we are focused on QoE models induced from subjective tests using ML techniques as an optimal balance between the complexities of subjective studies and accuracy of ML subjective models.

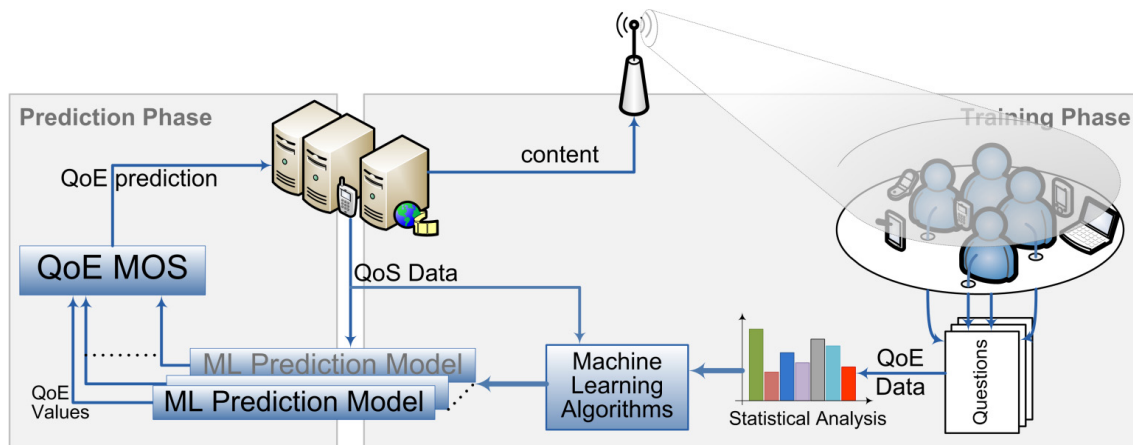


Figure 1: Overall schema for predicting QoE in a commercial system

III. BRIDGING THE QoS TO QoE GAP

Mobile TV is a service for mobile devices where customers can experience video-streaming content. They can select a set of available multicast channels or video contents with fixed quality settings. Some services might offer different streams for the same channel with different quality settings or adapted to specific devices. Offering multilayer video is not common due to the computational complexity. Many providers find it more efficient to maintain more than one stream with different qualities than a single multilayer stream due to compatibility issues with end user devices. In order to monitor the service quality, a probe-based network monitoring system is in place gathering information from the MobileTV content distribution platform. The probes collect information for each stream such as type of device, name of channel, stream, duration of the connection; these are captured in an Internet Protocol Detailed Record (IPDR) format [15]. In addition to this information, using RTCP in conjunction with RTSP helps the collection of QoS statistical values including: number of packets, packet loss ratio for audio, packet loss ratio for video, average delay, maximum delay, and jitter. In a nutshell, there is a deployed system fetching the AQoS and NQoS data from the system in real-time [2]. The AQoS involves application level QoS parameters as Video and Audio Bitrate and Video Frame Rate. The NQoS represents the network QoS parameters some of which are Packet loss, Jitter, and Delay.

The deployed system gives a good overview of the network conditions providing useful information for dimensioning the resources and managing the parameters of the content encoding. However, it cannot give any information as to how the service is perceived by the end-user. The metric QoE is a conglomerate of all the conditions that affect the perception of quality including the AQoS and NQoS as well as external factors such as the terminal type, the content itself and the expectations of the viewers. As QoE is directly linked with the value that the customers perceive it is useful for a service provider to be aware of the QoE instead of only looking at QoS.

The methodology developed here presents a mechanism for mapping QoS to QoE by executing limited initial subjective studies. It relies on Machine Learning techniques to build prediction models that accurately estimate the value of QoE based on the AQoS and NQoS parameters. The method is executed in two phases (Figure 1). The training phase uses input data from IPDR records of QoS parameters and QoE values as captured by surveys on customers' opinions. Its output is a set of prediction models for each question of the survey. In the prediction phase, these models fed with IPDR records and combined with a weighting scheme can predict the final overall QoE of a service. The weighting scheme is implemented using a SVM Regression Model [16].

During the subjective studies, the system records the IPDR values for the specific content provision used in the studies. Then, each queried customer fills in a questionnaire. All these values from the surveys are aligned with the IPDR records by selecting the ones that correspond to each content provision only. After this one-to-one mapping of QoS and QoE values, the Machine Learning algorithms build models that know how to predict the latter from the former; there is one model per question. As long as there is no radical change to the environment (e.g. new device or user group) these models are expected to perform accurate predictions of a subset of QoE values.

The QoE prediction models are plugged in a QoE prediction platform for online use. A statistical analysis on the results of the subjective study about the QoE of a service shows how the service, as a whole, is perceived from the perspective of its quality. Putting more emphasis on specific content attributes, via a different analysis, we can draw conclusions e.g. on a per content type basis. Correlating the subjective test data with the data from the network probes (AQoS and NQoS) we can create a set of training data for the ML algorithms of the prediction models. These ML algorithms, in a supervised learning mode, can develop classifiers (prediction models), which will be further used to

predict the QoE on unobserved cases in the production environment.

In the prediction phase, these models get as input the IPDR values and produce the QoE ones for each question. The statistical analysis of the subjective study results also gives a set of weights for each question. These weights are used to produce the Mean Opinion Score (overall QoE) out of the predicted values via appropriately weighting the output of each model.

IV. SUBJECTIVE STUDIES

Dealing with subjective metrics, such as QoE, requires subjective studies mostly because it is hard to identify the impact of objective metrics, such as QoS, on the perceived quality. In addition, the QoE evaluation will also demonstrate the level of expectations that the customers have.

Subjective studies by themselves pose a number of challenges as they rely on user sampling and their results need to reflect the real preferences of the whole user group. The more representative this sample and controlled the testing conditions are the more accurate the subjective studies are. Meeting all these constraints adds to the expenses and complexities associated with executing subjective studies. We have designed a targeted subjective study with typical users of the service to establish their preferences for quality and measure their QoE for the varied services.

A. The questionnaire

We carried out these subjective studies with the use of a questionnaire. Instead of asking a simple question where people rate the perceived quality from 1 to 5, we devised a questionnaire of ten different questions each of them bringing more subtle differences of the perceived quality of the viewers (Figure 2).

<i>Question 1:</i>	What is the type of content viewed?
<i>Question 2:</i>	Amount of delay before the video started?
<i>Question 3:</i>	Frozen images or interruptions?
<i>Question 4:</i>	Interruptions in the audio
<i>Question 5:</i>	Pixelation artifacts in the video
<i>Question 6:</i>	Noise or distortions in the audio
<i>Question 7:</i>	Audio and video synchronization
<i>Question 8:</i>	Quality of colours
<i>Question 9:</i>	Sharpness of video
<i>Question 10:</i>	Overall quality

Figure 2: Questions of the subjective study given to customers

The first question characterizes the content in seven different categories: News, Music Videos, Entertainment, Documentary, Movie or TV Series, Cartoon and Sports. This question is important because we want to observe i) the different expectations of quality for the different types of content ii) the different user expectations for different type of content [17] and iii) how the dynamics in the video affect the compression ratio also associated to the type.

Questions from two to nine are all of subjective nature with the given four levels of perceived problems from ‘none’ to ‘excessive’. The last question is the overall perception of the QoE. Instead of only asking the last question we have now more information from the previous questions and can map different network and application QoS conditions to answers given the questions Q2 to Q9. While Q1 answers is mapped to the name of the channel in the IPDR records.

B. Statistical Analysis

The results from the subjective study are captured in Figure 3 and Figure 4 and give a clear view as to how users perceive the quality. While the former gives the number of customers per content type the latter presents their opinion per question.

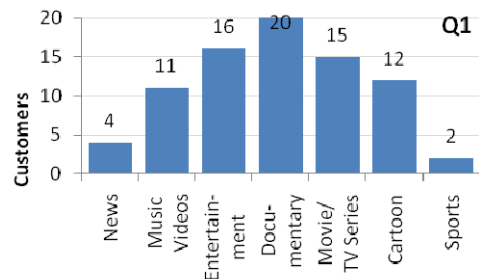


Figure 3: Distribution of queried users (customers) over the different types of viewed content.

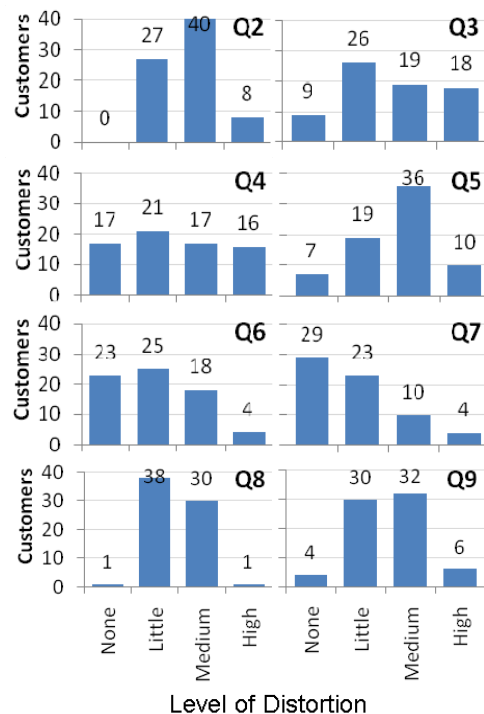


Figure 4: Distribution of queried users based on their answers to questions 2-9.

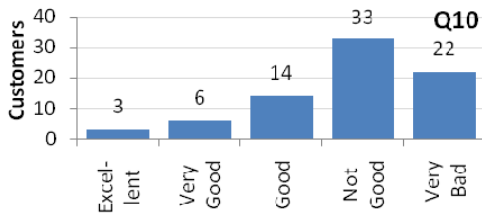


Figure 5: Overall quality as perceived by the users and captured through questionnaires.

Customers' opinion on the overall quality appears in Figure 5. This figure presents the distribution of the queried users over the different levels of perceived quality with high concentration in mediocre or lower quality levels.

A clearer view on the overall user perception of the content quality appears in Figure 6, which presents 8 pie charts each presenting the percentages of users with the same opinion per content type.

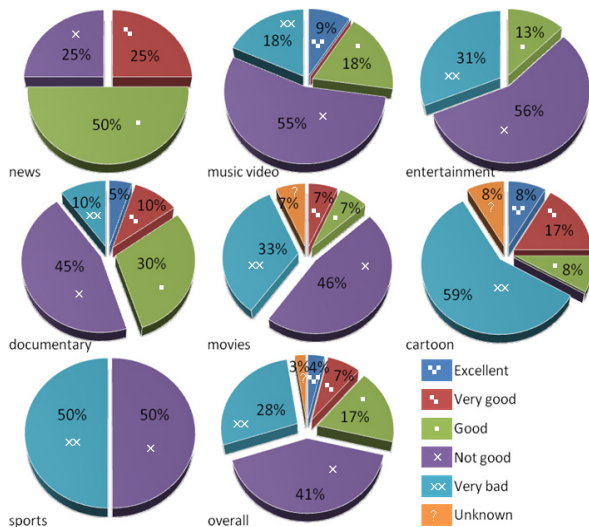


Figure 6: Customers' opinion as distributed per content type

V. PREDICTION MODELS

The first step into building prediction models is to prepare the training data. This data is the input to the prediction models and consists of two sets: the objective data (NQoS and AQoS) from the network monitoring system and the subjective ones from the questionnaires. The role of the devised prediction models is to map the objective to subjective values. In other words, based on the viewing conditions we want to predict the perceived quality.

Initially, we built a prediction model for each question (one through nine) to estimate the QoE value from the subjective questionnaires. In a final step, we built a final prediction model that correlates the answers of questions one through nine in a final answer for question ten. In addition, we provide the confidence of the classification based on errors during the training phase.

Related work that explores ML techniques with subjective data results shows that decision tree (DT) algorithms [13] achieve particularly good results with subjective datasets. In this work, we used decision trees induction algorithm C4.5 implementation (J48) part of the Weka ML platform [18] in combination with an ensemble classifier.

We have developed nine prediction models, one per each question. For the final one we used SVM regression [16] to build a regression model that combines all predictions from the previous questions with different weights to produce the final QoE value (see Figure 7 and Figure 8). The weights are application and system specific and administrators are supposed to set them as parameters of their network.

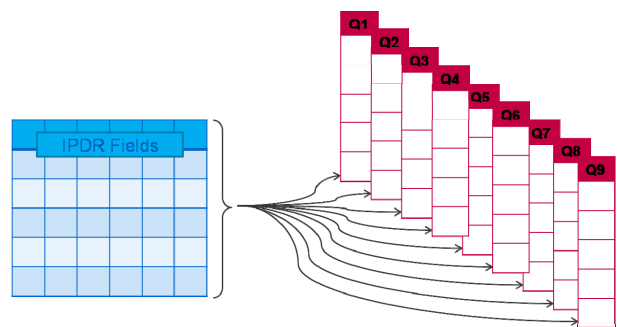


Figure 7: Correlation of QoS metrics from IPDR fields with customers' opinion from the subjective studies

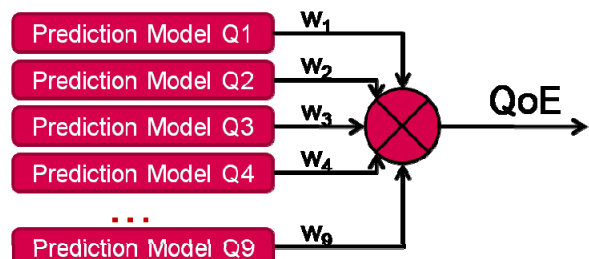


Figure 8: Combination of prediction models for each question to produce final QoE predicted value

The SVM Regression algorithm develops a regression function trained on the mapping of Q1 to Q9 with the QoE value. The results from the prediction models are given in Figure 9. The accuracy of the prediction models is calculated using 10-fold cross-validation [19].

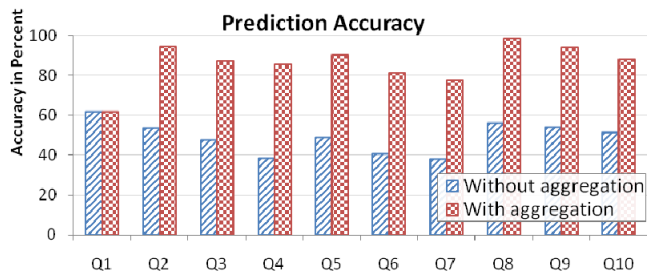


Figure 9: Prediction accuracy with and without output aggregation

The DT classifier uses categorical labels or outputs, which make them easy for humans to read and understand, for example “Excellent” or “Not Good”. But from a ML point of view these labels are not ordered, they are considered the same as we would consider the labels “Red” and “Blue”. So when we are calculating the prediction accuracy, only the exact predictions are taken into account. We do not know how many near misses we have. Most of these near misses would provide for good management tips. For instance, a prediction of “Very Bad” and “Not Good” might lead to the same management decision since both cases are not satisfactory. If we take this into account and also tolerate a small error rate for the output, such as errors with a distance of one or less from the actual value, the accuracy of the models significantly increases. For graphical representation we can look at the confusion matrix in Figure 10; the main diagonal represents the accurate cases (actual value row and predicted value column). If we add the values in the two adjacent diagonals, we can get the new accuracy with tolerance of ± 1 and thus get a higher effective accuracy of our classifier (Figure 9).

a	b	c	d	<-- classified
0	0	0	0	a = None
0	0	25	2	b = Little
0	0	40	0	c = Medium
0	2	6	0	d = High

Figure 10: Confusion Matrix for high accuracy with tolerance ± 1

For the final value of the QoE, as mentioned before, we use the weights from the regression model. The QoE is calculated as a sum of the Q_1-Q_9 answers each multiplied by its weight (w_1-w_9). As Q_1 is categorical, w_1 has different value for each type of content. We can look at these weights as a metric of the influence/importance of each question on the final answer (Figure 11). Question 3 has the most significant influence on QoE, based on the weights from the figure followed by 7 and 9. This information can be useful in improving the service as well as improving the subjective studies for feature iterations.

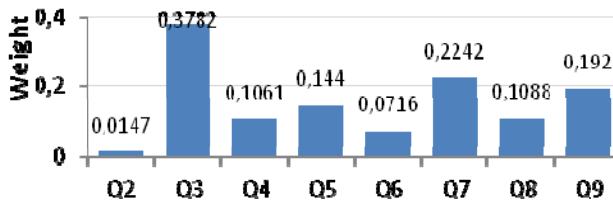


Figure 11: Question weights on the final overall QoE

The models built from the training data are now part of a QoE prediction platform built around them. This platform can load QoS data and feed it into the prediction models, thus, producing the QoE as a final result.

VI. IMPROVING QOE

Estimating the QoE is a crucial step in QoE-aware network management. However for a complete implementation of the management loop we need to be able to maintain a target QoE value for each stream. Maintaining a target QoE involves determining the desired conditions that need to be achieved or the needed changes to the parameters that will achieve the target QoE. To accomplish this task we use an algorithm [3] that based on the QoE prediction model estimates the minimum needed changes in the measured stream parameters to improve the QoE.

This technique is enabled by the DT prediction models we use for estimating the QoE. One of the strengths of DT compared to other ML prediction models is their intelligibility. A DT in a way represents a set of rules stacked in a hierarchical way. Simple decision trees commonly define just a few rules that are deduced from the data and used for classification, but when the number of rules grows the size of the DT also grows, and with that, it loses its intelligibility. This algorithm represents a QoE prediction DT model in the geometric space, defined by the dataset parameters. It considers each of the dataset parameters as a dimension in a hyperspace. Each of the datapoints from the dataset can be represented as a point in this hyperspace. The DT is represented by hyper regions formed by the leaves of the DT (Figure 112). Each node of the DT represents a binary split (for binary trees) that maps into a hyperplane in data hyperspace. At the bottom, the leaves of the tree, carve out hyper regions. These hyper regions, according to the appropriate leaf are associated with a class label membership. Every datapoint in the dataset falls on a leaf from the DT, therefore each corresponding point in the hyperspace falls into one of the hyper regions, and as such is classified with the corresponding class label. In our particular case the hyper regions are associated class labels that are the QoE estimates.

The algorithm (Figure 13) that represents the DT in the hyperspace as follows:

This algorithm implements the DT representation in the dataset’s hyperspace by generating a set of hyper regions that represent the tree leaves. Each hyper region contains a set of split rules that define the hyper-surface, which carves out the hyper region. The split rules are either representing an inequality of the type $Parameter1 \geq Value1$ or of the type $Parameter1 = Value1$ depending on whether $Parameter1$ is continual or categorical. If the leaf is on the left side of a continual $Parameter1$ split then the split inequality will be ‘more than or equal to’, if it is on the right side the split inequality will be ‘less than’.

Having a list of *HyperRegion-s* we can easily determine where each datapoint from the dataset belongs to, by testing the datapoint on the split rules of each hyper region. The hyper region is associated with the same class label as the leaf it represents, so all datapoints that belong to that region are classified as such.

In order to improve the QoE estimation of a particular stream, we need to look at the datapoint that was generated by the monitoring system for that stream. If the datapoint is

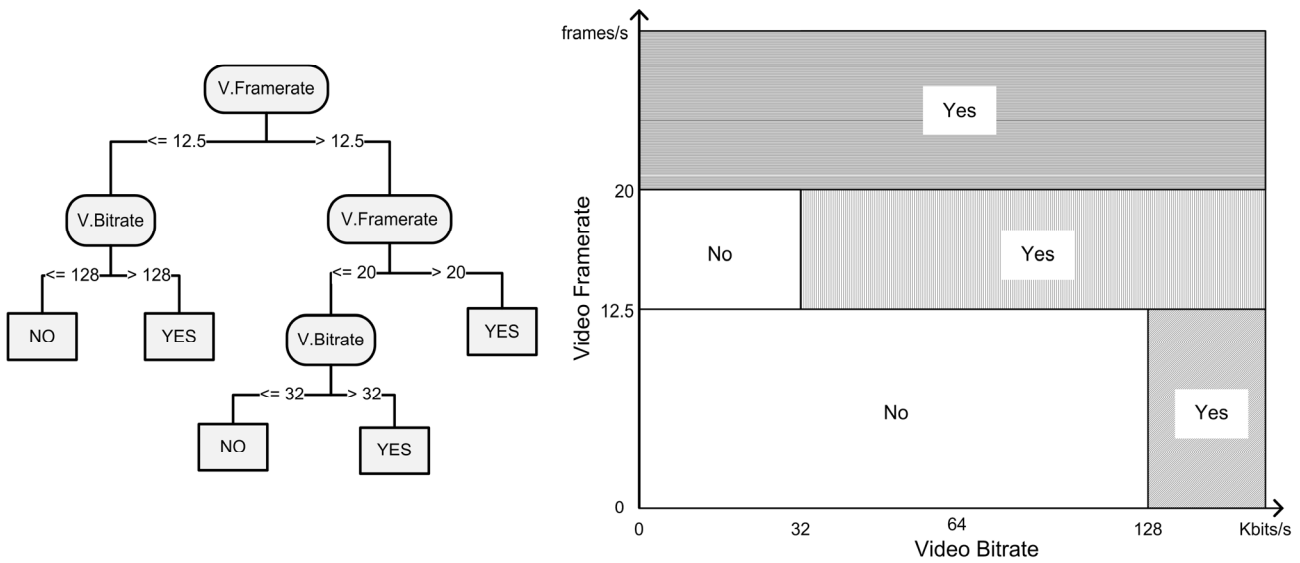


Figure 112. Simple decision tree in 2D space

classified with a QoE value that is not satisfactory, we look at the distance to a set of hyper regions $\bar{\Phi}$ that are associated with a satisfactory QoE value. The distance to each of the desired regions is the difference in parameter values that are needed in order to move the datapoint to the desired regions.

Start from the root node and call a recursive method *FindLeaves*

FindLeaves:

- 1) If the node has children
 - a) Call *FindLeaves* on each child
 - b) Add the SplitRule on each of the Hyper Regions ($\bar{\Phi}$) that are returned
 - i) If the leaf split is categorical add a Split Rule: Attribute = 'value'
 - ii) If the leaf split is continual add on the leaves from the left side SplitRule: Attribute < value, and on the leaves from the right side Attribute > value
 - c) Return the set of Hyper Regions ($\bar{\Phi}$)
- 2) Else, you are in a leaf
 - a) Create an Hyper Region object
 - i) Assign the class of the leaf to the $\bar{\Phi}$
 - ii) Return $\bar{\Phi}$

Figure 13. DT to Hyper Region algorithm

The output of the algorithm is a set of distance vectors, which define the parameters that need to be changed and their change values.

To illustrate the matter better we can take an example from the laptop dataset from [13]. The prediction model

built from this dataset is given in Figure 112. If we look at the datapoint given in Table 1 we can see that this datapoint will be classified by the model as QoE = No ('Not Acceptable'). Since the V. Framerate is less than 12.5 and the V.Bitrate is less than 32 the datapoint reaches a leaf with 'Not Acceptable' class associated with it.

Now, what is the best way to improve the QoE of this stream?

First of all there are parameters that characterize the type of the content such as the Video SI and the Video TI and cannot be changed. In this dataset structure we are looking into increasing the V.Bitrate and V.Framerate. If we increase the V.Bitrate for this particular datapoint by one step to 64kbits/s we can see that the datapoint goes now down the DT to one of the bottom leaves, but it is still classified as QoE Acceptable = No. On another hand if we increase the V.Framerate to 15f/s we can see that the datapoint is classified as QoE Acceptable = Yes without adding more bandwidth.

TABLE I. EXAMPLE DATAPPOINT

Video SI	Video TI	V. Bitrate	V. Framerate
67	70	32	10

We can deduce a rule from the model that a video with these characteristics needs to have higher V.Framerate for it to be perceived with high quality. However, this rule is not easily evident from only looking at the model. We can also imagine a system with large number of attributes that we can change where tuning this attributes the right way becomes an increasing problem. Further down this line of reasoning, if we want to make a system-wise improvement that will increase the QoE of most streams we cannot easily derive which parameters are best to be increased and by how much.

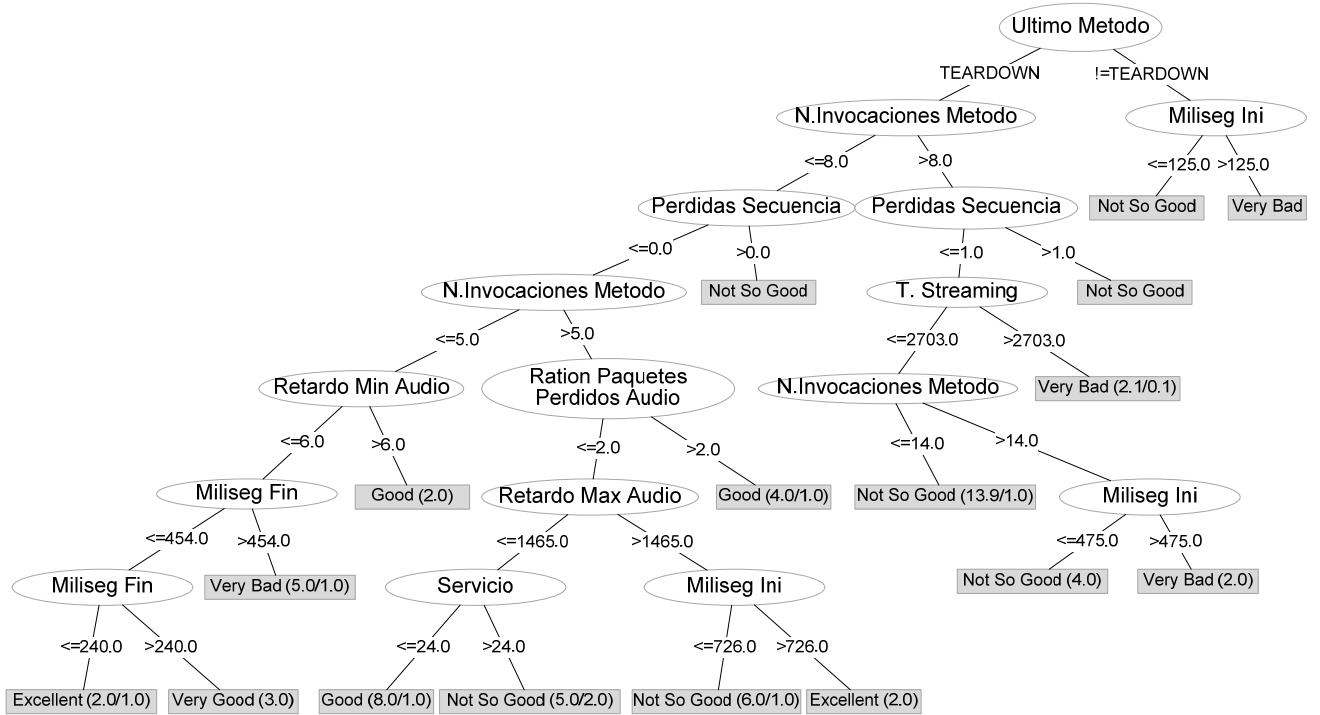


Figure 14. Simple decision tree in 2D space

In the case of the example datapoint the algorithm returns the two possible paths:

- Increasing the Framerate to above 12.5f/s
- Increasing the V. Bitrate to above 32kbts/s and the Video TI to above 87

Since we know that increasing the Video TI is not an option, because this is defining the type of content we can see, then the only option is to increase the frame rate. In a general case, there can be many different paths to a hyper region with the desired class.

To automate the process we can assign cost functions to the change of the attribute values and automatically calculate the cheapest way to reach the desired QoE. In this manner attributes that are not changeable, such as the Video TI, can have infinite value of the cost function.

Given a datapoint and a target label the algorithm produces a set of change vectors. Each of the change vectors applied to the datapoint moves the datapoint to a hyper-region classified with the target label. In other words, each change vector is one possible fix for the datapoint.

$$\bar{\Phi} = FindLeaves(DT, QoE) \quad (1)$$

$$\overline{\Delta\varphi}_i = Distance(\Phi_i, \bar{d}) \quad (2)$$

$$\Delta\varphi_{optimum} = \min_i (Cost(\Delta\varphi_i)) \quad (3)$$

In (1), $\bar{\Phi}$ is a set of regions with a targeted QoE value. The distance function in (2) calculates the vector of distances for each attribute to the target region in $\Delta\varphi$. The optimal distance vector is the one with minimal cost (3) for the given input datapoint \bar{d} . The Cost function in (3) is dependent on the application. Each system has explicit and implicit costs associated with changes of specific parameters.

VII. APPLICATION OF THE REMEDIES IN MOBILE IPTV

The remedies algorithm has been implemented by extending the Weka [20] platform, so that algorithms like J48 [21] that induce decision trees can be used to calculate the hyper regions. Furthermore, we can now measure the distance of any datapoint classified by the DT to the desired hyper-region.

The decision tree built from the data of the subjective study is given in Fig. 14. The boxed nodes represent the leaves and map to the hyper-regions as we have seen in Fig. 2. There are 17 hyper-regions, out of which, only two are with excellent QoE value. The algorithm generates the remedy output specific for each particular broadcasting system. A target QoE values needs to be defined, and a specific cost for changing a parameter needs to be given as well. If the target value if excellent QoE the algorithm will calculate the minimum cost of changing specific parameters so that the datapoint falls in one of the two Excellent hyper-regions. Of course some parameters are not changeable, such as the type of video. For these an infinite cost is

assigned so that the algorithm does not propose these absurd remedies.

A more elaborate QoE improvement is also possible where not all datapoints are targeted for the excellent regions, but the management is executed based on the utility of improving a QoE of a stream in regards to the costs. Then multiple levels of remedies can be suggested by the algorithm with varying costs, and the provider can choose to apply mechanisms to implement the remedies based on their utility to the customers.

VIII. CONCLUSIONS AND FUTURE WORKS

We have presented a method for estimating the QoE which circumvents some pitfalls of exhaustive subjective testing while still resulting in accurate estimation on QoE. We have discussed the importance of QoE as a metric to define the value of a multimedia service provided to the customers. We also presented an algorithm that can generate suggested remedies per multimedia stream for streams with a lower than desired QoE. To implement this method we relied on ML techniques that were successful in building prediction models that accurately predict the QoE from a small training dataset of the subjective tests. We also presented a QoE platform that makes use of the prediction models to bring QoE estimations in real-time based on data from network probes. This platform is currently part of a mobile TV system where it estimates the QoE of the streaming multimedia content and proposes remedies for different streams. The necessity for this kind of platform arises from the need for multimedia service providers to estimate the experienced quality by their customers, to diagnose the reasons for lower than desired QoE and for the remedies that they can implement.

In conclusion this methodology presents a pragmatic solution for estimating and maintaining QoE with a wide range of applicability. Its success and usability depends on the quality of the prediction models, while as architecture it is flexible enough to be used in many different environments. To make use of its full potential a more elaborate subjective study in better controlled conditions will yield in more precise prediction models and better effectiveness overall.

This platform can be extended with Online Learning techniques that will provide continuous improvements in the prediction models and further reduce the load of the initial subjective tests. The online learning approach will also provide the ability of the system to adapt the models to the ever changing conditions of the production environment, such as introduction of new content, new terminal devices etc. In addition to Online Learning techniques, some Active Learning approaches will be of benefit to improve the gain of asking the customer for feedback intelligently as opposed to randomly selecting for feedback.

ACKNOWLEDGMENTS

The work included in this article has been supported by Telefonica R&D and Indra (Spain). The authors thank Maria

del Mar Cutanda, head of division at Indra, for providing guidance and feedback.

REFERENCES

- [1] V. Menkovski, G. Exarchakos, A. Liotta, and A. Sánchez, "Measuring Quality of Experience on a Commercial Mobile TV Platform," in *Advances in Multimedia (MMEDIA), 2010 Second International Conferences on*, pp. 33-38, 2010.
- [2] A. Cuadra-Sanchez and C. Casas-Caballero, "End-to-End Quality of Service Monitoring in Convergent IPTV Platforms," in *Proceedings of the 2009 Third International Conference on Next Generation Mobile Applications, Services and Technologies*, pp. 303-308, 2009.
- [3] V. Menkovski, G. Exarchakos, A. Liotta, and A. Sanchez, "Estimations and Remedies for Quality of Experience in Multimedia Streaming," in *Advances in Human-Oriented and Personalized Mechanisms, Technologies and Services (CENTRIC), 2010 Third International Conference on*, pp. 11-15, 2010.
- [4] 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, 2008.
- [5] S. Winkler, "Video Quality Measurement Standards - Current Status and Trends," in *Proceedings of ICICS 2009*, 2009.
- [6] S. Winkler, *Video Quality and Beyond*. Symmetricom, 2007.
- [7] R. I. ITU-T, "910," *Subjective video quality assessment methods for multimedia applications*, 1999.
- [8] Quan Huynh-Thu and M. Ghanbari, "Temporal Aspect of Perceived Quality in Mobile Video Broadcasting," *Broadcasting, IEEE Transactions on*, vol. 54, no. 3, pp. 641-651, 2008.
- [9] Mu Mu, A. Mauthe, and F. Garcia, "A Utility-Based QoS Model for Emerging Multimedia Applications," in *Next Generation Mobile Applications, Services and Technologies, 2008. NGMAST '08. The Second International Conference on*, pp. 521-528, 2008.
- [10] M. Mu, R. Gostner, A. Mauthe, G. Tyson, and F. Garcia, "Visibility of Individual Packet Loss on H. 264 Encoded Video Stream—A User Study on the Impact of Packet Loss on Perceived Video Quality," *Annual Multimedia Computing and Networking, San Jose, USA*, 2009.
- [11] M. T. Virtanen, N. Gleiss, and M. Goldstein, "On the use of evaluative category scales in telecommunications," in *Proceedings of Human Factors in Telecommunications*, vol. 95, 1995.
- [12] F. Agboma and A. Liotta, "QoE-aware QoS management," in *Proceedings of the 6th International Conference on Advances in Mobile Computing and Multimedia*, pp. 111-116, 2008.

- [13] V. Menkovski, A. Oredope, A. Liotta, and A. Cuadra Sánchez, "Predicting Quality of Experience in Multimedia Streaming," in *Proceedings of the 7th International Conference on Advances in Mobile Computing and Multimedia*, pp. 52-59, 2009.
- [14] V. Menkovski, A. Oredope, A. Liotta, and A. Cuadra Sánchez, "Optimized online learning for QoE prediction," in *Proceedings of the 21st Benelux Conference on Artificial Intelligence*, 2009.
- [15] "TM Forum - TM Forum IPDR Program." [Online]. Available: <http://www.tmforum.org/ipdr>. [Accessed: 23-Mar-2010].
- [16] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.
- [17] F. Agboma and A. Liotta, "Addressing user expectations in mobile content delivery," *Mobile Information Systems*, vol. 3, no. 3, pp. 153-164, Jan. 2007.
- [18] I. H. Witten and E. Frank, *Data mining*. Morgan Kaufmann, 2005.
- [19] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," in *IJCAI*, pp. 1137–1145, 1995.
- [20] "Weka 3 - Data Mining with Open Source Machine Learning Software in Java." [Online]. Available: <http://www.cs.waikato.ac.nz/ml/weka/>. [Accessed: 02-Jun-2009].
- [21] S. L. Salzberg, "C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993," *Machine Learning*, vol. 16, no. 3, pp. 235-240, 1994.