Identifying Quality of Experience (QoE) in 3G/4G Radio Networks based on Quality of Service (QoS) Metrics

Vera Pedras
vera.pedras@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

November 2017

Abstract

Two novel Quality of Experience (QoE) prediction models were developed, one for 3G voice calls and other for web browsing in 4G networks, based on real Universal Mobile Telecommunications System (UMTS) and Long Term Evolution (LTE) data, respectively. These models estimate the user perceived quality, in a Mean Opinion Score (MOS) scale, by evaluating several Radio Frequency (RF) channel measurements and Quality of Service (QoS) metrics. Real Drive Tests (DTs) data and MOS measurements have been used as reference data, in order to produce the new QoE prediction models, using machine learning techniques. The Support Vector Regression (SVR) algorithm was used to map the QoS metrics in MOS. The new developed models enable the application of QoE as a more realistic network optimization criteria. The QoE model for voice calls presented a Root Mean Square Error (RMSE) of 11% and a correlation of 62%, when comparing the predicted MOS to the one that was measured. The web browsing model showed an higher correlation (of 92%) and a lower RMSE (of 10%).

Keywords: LTE, UMTS, QoE, QoS, Web Browsing, Voice Calls

1. Introduction

According to research results, for every costumer that complains about a certain provided service there are 29 others who will not complain. In fact, 90% of the customers will simply leave the service once they become unsatisfied [1]. For these and other reasons, it is very important for the operators to do an estimation of the user satisfaction of a service, in order to adjust the service quality according to the users needs. The user experience depends on several factors, some of them are network related and others depend on the type of service being used, the end device features and the user expectation, among others.

A network can be assessed objectively in terms of Quality of Service (QoS), which depends on network parameters like throughput, packet loss, delay and jitter. This measurement is done in the network side and does not take into account the type of service and the user characteristics. The service quality may also be assessed at the application level - the so called application level QoS - with parameters that are application specific; as an example, for a video streaming application the parameters to be assessed may be the waiting time before the start of the video or the frequency of video stallings.

However, a good application QoS and network QoS does not necessarily mean that the end user is satisfied with the provided service, since his satisfaction depends on other factors. Thus, in order to measure the user satisfaction one needs to define the Quality of Experience (QoE), which takes into account factors like expectation, requirements and perception of the end user, content type provided by the service, users device features, network QoS and the context in which the user is using the service, like the access type, movement (mobile or stationary) and location [2].

This work enforces the paradigm shift from the typical network centric QoS indicators to the user centric QoE domain, by proposing two models that map the QoS indicators in a QoE metric for voice and web browsing services, respectively.

The QoE is measured in terms of Mean Opinion Score (MOS) [3], that represents the user's opinion about a service using a scale from 1 to 5, being 5 - Excellent, 4 - Good, 3 - Fair, 2 - Poor and 1 - Bad.

Generally, QoE models can be either subjective or objective. In the first case, the perceived quality may be assessed through a crowd-sourcing approach, where a group of people uses a particular service and fill out a survey qualifying their experience [4]. In the second case, the estimation of QoE is based on network measurements or quality

indicator parameters collected through Drive Tests (DTs) or network-side passive monitoring.

Within objective QoE evaluation methods, there are full reference models such as the Perceptual Objective Listening Quality Assessment (POLQA) [5] and no reference models. The first ones need the original transmitted signal, in order to estimate the MOS value; therefore, these can only be applied for dedicated live network tests (*e.g.* DT campaigns). Regarding the no reference models, these may depend on several network and transmission parameters, as in the E-model [6] defined by the International Telecommunication Union (ITU).

In this document, two novel objective and no reference QoE models for 3rd Generation (3G) voice calls and web browsing in 4th Generation (4G) networks are presented. Based on QoS metrics, for a given time period, the models output the QoE in terms of MOS. This model allows a direct assessment of the perceived QoE for any subscriber, given a specific QoS context.

This document is organized as follows. Section 2 describes the used methodology to derive the new QoE assessment model. The new voice QoE model and the new web browsing QoE model are described, respectively, in sections 3 and 4. The estimations results of both models are presented in section 5. Finally, in section 6, conclusions are drawn.

2. Methodology

The proposed models aim to predict an output y (MOS) through a mathematical expression that takes as input n parameters, called features, $x_j, j = 1, ..., n$ (QoS parameters). The output is achieved through a learning algorithm and the features are obtained from DT data using Test Mobile System (TEMS®) [7], which is an active, end-to-end testing solution, used to verify, optimize and troubleshoot Radio Access Network (RAN) services. During a given service utilization, it continuously measures the corresponding QoS metrics, including the Radio Frequency (RF) ones. These measurements are performed in the time domain, thus constituting several time series. Several statistics from the time series data are calculated and the more significant ones are retained to be included in the models.

The data preprocessing and data cleaning eliminates the less significant features in the MOS estimation; nonetheless, the remaining features still provide room to test different combinations of features, forming several hypotheses. In order to select the best set of features, each hypothesis was evaluated.

Thus, to model each one of the hypothesis, the used machine learning algorithm takes as input a training set $(x_1^{(i)}, ..., x_n^{(i)}, y^{(i)}), i = 1, ..., m$, composed by m training examples $(x_1, ..x_n, y)$, accounting 60% of all available data. The remaining 40% of the data is equally divided between the Validation and Test sets. On one hand, the first one is used to assess each hypothesis and to choose the best one. On the other hand, the second one is used to determine the final performance of the model.

The metrics used to assess the performance of each hypothesis and of the final model are the Root Mean Square Error (RMSE), the Pearson Correlation and the Spearman Correlation. The RMSE measures the square root of the average of the squared differences between the predicted and the original values. The Pearson coefficient assesses the linear correlation between the original and the estimated MOS. The Spearman coefficient evaluates the monotonicity of the estimated MOS, relatively to the original one.

Two algorithms were considered for the development of the models, the Multivariate Linear Regression algorithm [8] and the Support Vector Regression (SVR) algorithm [9].

2.1. Multivariate Linear Regression

The main goal of a Multivariate Linear Regression is to achieve a linear expression $h_\theta(\cdot)$, given by (1), where the $\theta_j, j = 1, ..., n$ are optimized in order to fit the training set.

$$h_\theta(x_1, ..., x_n) = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + ... + \theta_n \cdot x_n \quad (1)$$

The value of the $\theta_j, j = 1, ..., n$ can be calculated using the Regularized Normal Equation model, given by:

$$\boldsymbol{\theta} = (X^T X + \lambda \cdot L)^{-1} X^T \boldsymbol{y}, \quad (2)$$

where λ is the regularization parameter and L is a $(n+1) \times (n+1)$ matrix similar to the identity matrix but differs in the first value of the diagonal which is 0 in this case.

2.2. Support Vector Regression

The SVR algorithm aims to optimize the estimation of a value y through a function $f(\cdot)$. This algorithm uses k support vectors which allow to transform the data into a multidimensional plane, by applying them in a Kernel Function. The function to be optimized is given by:

$$f(\boldsymbol{x}) = \boldsymbol{w} \cdot K(\boldsymbol{x}, \boldsymbol{SV}) - \rho, \quad (3)$$

where $\rho \in \mathbb{R}$, $\boldsymbol{w} \in \mathbb{R}^k$, $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{SV} \in \mathbb{R}^n \times \mathbb{R}^k$. The $\boldsymbol{SV}$ is a matrix with k support vectors ($k \geq n$) that are used to transform the data. $K(\cdot)$ is the Kernel function that can take various expressions. The result of applying this function is an array of dimension k, since it is applied to $\boldsymbol{x}$ and each $\boldsymbol{SV}$ line individually.

The Kernel function and the support vectors allow to transform a non-linear problem in a linear one. This function can take many forms, the most common are the following:

- Linear:

$$K(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u} \cdot \boldsymbol{v} \tag{4}$$

- Polynomial of degree $\boldsymbol{p}$:

$$K(\boldsymbol{u}, \boldsymbol{v}) = (\gamma \boldsymbol{u} \cdot \boldsymbol{v} + c_0)^p \tag{5}$$

- Radial Basis Function (RBF):

$$K(\boldsymbol{u}, \boldsymbol{v}) = e^{-\gamma \|\boldsymbol{u} - \boldsymbol{v}\|^2} \tag{6}$$

- Sigmoid:

$$K(\boldsymbol{u}, \boldsymbol{v}) = tanh(\gamma \boldsymbol{u} \cdot \boldsymbol{v} + c_0) \tag{7}$$

The choice of the Kernel function and its parameters has to be done according to the problem being analyzed.

The γ, $\boldsymbol{SV}$, $\boldsymbol{w}$ and ρ parameters are determined through machine learning techniques in order find out the ones that best fit the training set.

3. 3G Voice Calls QoE Model

TEMS® itself evaluates the QoE of a voice call using the POLQA [5] algorithm which estimates the MOS by comparing the received audio signal with the original one. In general, this algorithm has a low prediction error when compared with the QoE assessments done by large groups of people. TEMS® evaluates the MOS value in periods of five seconds, during a call. Within this interval, Universal Mobile Telecommunications System (UMTS) QoS parameters are registered creating a time series for each one, as stated. The new voice call QoE model should determine the relationship between these QoS parameters and the measured MOS (using the POLQA algorithm) thus creating a low level, objective and no reference QoE model.

3.1. Feature Selection

TEMS® provides several QoS parameters, but for the voice QoE model the objective was to focus on RF parameters to understand and measure it's influence on voice calls QoE. Therefore, the following RF parameters were considered, as they are typically used to assess the radio channel quality:

- **Signal-to-Interference Ratio (SIR) [dB]** - ratio between the average received modulated carrier power and the average received co-channel interference power.
- **SIR Target [dB]** - reference SIR set by the outer loop power control.
- **Active Set (AS) $\mathbf{E_c/N_0}$ [dB]** - AS best received chip energy to noise spectral density ratio.
- **AS Received Signal Code Power (RSCP) [dBm]** - AS best power measured in the Common Pilot Channel (CPICH).
- **Received Signal Strength Indicator (RSSI) [dBm]** - metric that takes into account the RSCP and the received chip energy to interference level ratio (E_c/I_0).

From the data preprocessing and exploratory analysis, it has stand out that the creation of a new time series which accounts the difference between the SIR and SIR Target could be an important MOS estimator. This is supported by the fact that it measures how the actual interference levels differ from the desired ones.

From the parameters time series, the following statistical metrics were extracted: the mean, the maximum, the minimum, the standard deviation (SD), the skewness [10] and the kurtosis [11]. The Skewness measures the symmetry of the distribution of the time series. It takes a value of zero when the distribution is symmetric. Otherwise, it takes positive or negative values for a right side tail greater than left and vice versa, respectively. In regard to kurtosis, it measures the thickness or heaviness of the tails of the time series distributions. Taking always positive values, it will be as close to zero as the tail thickness.

With all these available metrics, a feature preselection was conducted using the correlation between the measured MOS and each feature. The ones that showed the correlations higher than a threshold are presented in Table 1.

The selected threshold is problem dependent as it depends on the initial correlations between the features and the estimated variable. In this case it was select a minimum correlation of 25%.

In order to exclude the highly correlated features, and thus redundant, the correlations between each other were determined (see Figure 1).

From Figure 1 it can be seen that as more correlated are two features the darker the color and the greater the circle size are. For instance, it can be stated that both the RSSI mean and the RSSI max are highly correlated, thus using both features is redundant. For the purpose of defining features highly correlated it was considered a minimum correlation threshold of 75% between two features. For those the one with the highest correlation with MOS prevails, and the remaining are discarded. Therefore, the final selected features are the following:

- AS E_c/N_0 Maximum;

Table 1: Correlation of the pre-selected features with the measured MOS.

Parameter	Statistic Operation	Pearson Correlation
RSSI	Mean	29.36%
	Maximum	28.99%
	Minimum	33.38%
SIR	Minimum	27.43%
	SD	38.73%
SIR Target	Mean	25.71%
	Maximum	31.89%
	SD	39.43%
AS RSCP	Mean	33.71%
AS E_c/N_0	Mean	38.06%
	Maximum	39.60%
	Minimum	32.18%
SIR − SIR Target	Minimum	38.50%
	SD	39.57%

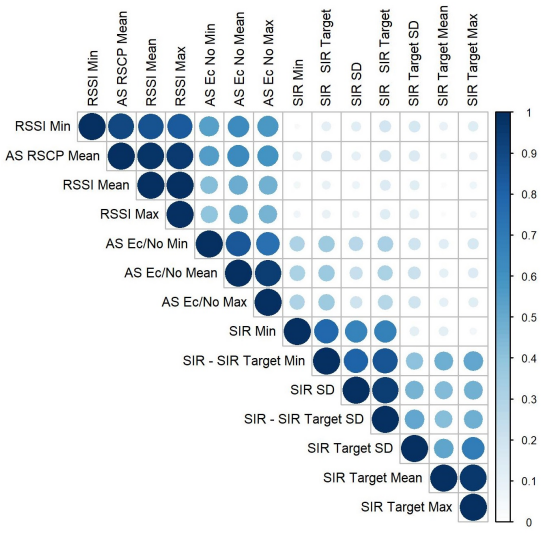

Figure 1: Features Correlation for the Voice Model.

- SIR Minimum;
- SIR Target SD;
- SIR Target Maximum;
- AS RSCP Mean;
- SIR − SIR Target SD.

3.2. Model Selection

The multivariate linear regression algorithm was first applied to a set of hypotheses that corresponded to different combinations of the selected features. These hypotheses were trained using the training set and then the assessment metrics were computed using the validation set. The hypothesis that showed the best performance, according to those metrics, was composed by the following features: E_c/N_0 Maximum; SIR − SIR Target SD; SIR Target SD; AS RSCP Mean; and SIR Target Maximum. The results obtained, with the validation set, for this hypothesis are represented in Table 2.

Table 2: RMSE and correlations for the bets hypothesis using the linear regression algorithm for the voice model.

RMSE	**Pearson Correlation**	**Spearman Correlation**
11.50%	66.81%	57.07%

The SVR algorithm was then applied, for which three different hypotheses were tested, see Table 3. The hypotheses were trained with the same training set which was composed by 207 training examples.

Table 3: Hypotheses considered for the Voice Model.

Features	H1	H2	H3
SIR Minimum			×
SIR Target Maximum		×	
SIR Target SD	×	×	×
AS E_c/N_0 Maximum	×	×	×
AS RSCP Mean	×	×	×
SIR−SIR Target SD	×	×	×

The performance of each hypothesis was then assessed with the validation set, using the previous mentioned metrics (RMSE and Pearson and Spearman Correlation). The obtained results are represented in Table 4.

Table 4: RMSE and correlations for each voice model hypothesis.

	RMSE	Pearson Correlation	Spearman Correlation
H1	11.07%	57.48%	56.29%
H2	11.05%	58.40%	58.61%
H3	10.46%	64.00%	62.35%

The hypothesis with higher correlation and lower RMSE is the 3rd one, thus being the selected one.

Section 5 presents a more detailed analysis to both approaches (the linear regression and the SVR).

4. Web Browsing Model

The QoE model proposed in this section was developed using Long Term Evolution (LTE) data allowing to determine the user perceived quality when web browsing in a LTE network. The data was collected through drive-testing in real mobile networks, using TEMS®.

The QoE was measured using an existing model which takes into account the download time of a web page, and estimates the QoE as MOS. The authors in [12], showed that the model does a good estimation of the perceived quality.

Once again, the challenge was to develop a new QoE model which translates the dependence on QoS metrics for web browsing.

4.1. Feature Selection

Since the web browsing service belongs to the packet switch domain, besides the RF parameters, some additional QoS parameters were considered. The more relevant parameters for MOS estimation include besides the RF radio channel parameters the corresponding modulation and coding parameters. These allow to infer the network resources availability. Hence, the following were considered:

- Reference Signal Received Power (RSRP) [dBm] - average power of resource elements that carry cell specific reference signals over the entire bandwidth.
- Reference Signal Received Quality (RSRQ) [dB] - Indicates the quality of the received reference signal.
- RSSI [dBm] - Total received wide-band power (measure in all symbols) including all interference and thermal noise.
- Physical Downlink Shared Channel (PDSCH) Modulation and Coding Scheme (MCS) - Index that defines the modulation and the size of the transport blocks to be used.
- Block Error Rate (BLER) [%] - percentage of discarded blocks duo to error.
- Channel Quality Indicator (CQI) - Index corresponding to a modulation scheme and coding rate adapted to the radio channel quality.
- Number of used Transport Blocks (TBs) - number of TBs being used that depends on the transmission mode.
- PDSCH Resource Blocks (RBs) [%] - Percentage of the maximum number of PDSCH RBs.

The web page size being accessed is also considered as a feature, since the perceived quality is a function of the downloading time [12], being this metric also dependent on the web page size.

For the above-mentioned parameters the same statistical metrics were calculated as mentioned in section 3. Additionally, in the data exploratory analysis it stand out a possible new feature. Within the MOS evaluation time span the absence of variability of some parameters tended to be associated with higher MOS values. Therefore, this new possible feature, called constant flag, is 1 when the time series is constant and 0 otherwise. The features with higher correlation (higher than 40%) are presented in Table 5.

Table 5: Correlation of the best features with the measured MOS.

Parameter	**Statistic Operation**	**Pearson Correlation**
RSRP	Mean	54.98%
	Maximum	50.25%
	Minimum	61.13%
RSRQ	Mean	63.97%
	Maximum	46.44%
	Minimum	69.44%
RSSI	Mean	47.36%
	Maximum	43.57%
	Minimum	50.81%
MCS	Minimum	43.73%
	Constant Flag	63.28%
	Kurtosis	45.82%
	SD	44.16%
BLER	Mean	44.46%
	Skewness	52.99%
	Kurtosis	54.59%
	Constant Flag	47.13%
CQI	Mean	65.65%
	Maximum	54.95%
	Minimum	52.74%
Web page size		26.57

Note that the web page size feature is also selected, in spite of its MOS correlation (of 26.57%)

being bellow the defined threshold. This feature is independent of the network related features, which justifies its selection. Furthermore, this feature would allow to estimate the QoE for different web page's sizes.

The correlation between each feature of Table 5 were determined in order to avoid redundancy between them. Thus, the features that showed an high correlation were analyzed in order to select just the one that is most correlated with MOS. Figure 2 presents the correlation between features.

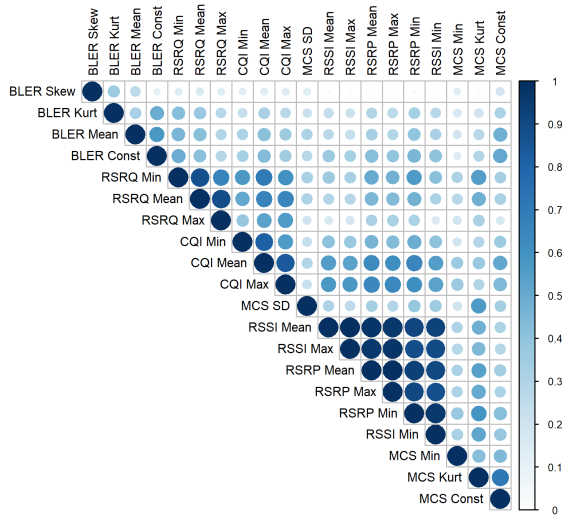

Figure 2: Features Correlation for Web Browsing Model.

After excluding the highly correlated features (greater than 75%), the set of selected features was the following:

- RSRP Minimum;
- RSRQ Minimum;
- MCS Flag Constant;
- BLER Mean;
- BLER Skewness;
- BLER Kurtosis;
- CQI Mean;
- Web Page Size

4.2. Model Selection

Similar to the voice model, first the multivariate linear regression algorithm was considered and the selected features were combined in different hypotheses. These hypotheses were trained using the training set and the validation set was used to determine the RMSE and Pearson and Spearman correlations. According to this metrics, the hypothesis that showed the best performance used the following features: RSRP Minimum; RSRQ Minimum; MCS Constant Flag; BLER Mean; BLER Kurtosis; and CQI Mean. The assessment metrics obtained for this hypothesis using the validation set are represented in Table 6.

Table 6: RMSE and correlations for the bets hypothesis using the linear regression algorithm for the web browsing model.

RMSE	**Pearson Correlation**	**Spearman Correlation**
14.13%	89.99%	83.92%

The SVR was then used in order to assess a possible non linear relation between the selected features and the measured MOS. The selected features were organized in different hypothesis, in Table 7 are represented four of them. These hypotheses were trained following the methodology presented in section 2. The training set was composed by 115 training examples.

Table 7: Hypotheses considered for the Web Browsing Model.

Features	H1	H2	H3	H4
RSRP Minimum	×	×	×	×
RSRQ Minimum	×	×	×	×
MCS Constant Flag	×	×	×	×
BLER Mean			×	×
BLER Skewness		×		×
BLER Kurtosis	×	×	×	×
CQI Mean	×	×	×	×

After training each one of the hypotheses, their performance was assessed using the validation set. Again, the RMSE, the Pearson and Spearman Correlations that resulted from that evaluation are present in Table 8.

The hypotheses that top performed were the 3rd (H3) and 4^{th}(H4) hypotheses which presented similar errors and correlations. Taking into account the similarity of the obtained results for these two hypotheses, the 3rd hypothesis was the chosen one, since the number of features is lower than in the 4th hypothesis. The next section (section 5) presents a more detailed analysis of the selected hypothesis.

5. Results

5.1. Voice Model

The QoE model presented in section 3 was developed using DT data, in dedicated mode, which ac-

Table 8: RMSE and correlations for each web browsing model hypothesis.

	RMSE	Pearson Correlation	Spearman Correlation
H1	10.82%	87.77%	90.48%
H2	9.94%	89.87%	90.04%
H3	9.35%	91.60%	91.51%
H4	9.53%	92.17%	89.92%

counted a total of 87 3G phone calls. In parallel, for each phone call, the TEMS®' QoE model (POLQA) evaluated, on average, 4 times the perceived phone call quality. Overall, 348 entries of MOS measurements were collected. Each entry contains not only the MOS value but also the evolution of the RF parameters throughout each MOS measurement.

To select between the linear regression (LR) model and the SVR model, the assessment metrics were also determined for the test set. The use of this set aims to evaluate the models with brand new data that was not used in the model development process. Table 9 presents the results of these metrics for both models.

Table 9: RMSE and correlations calculated through the test set for both Linear regression and SVR voice models.

	RMSE	Pearson Correlation	Spearman Correlation
LR model	12.01%	50.84%	48.04%
SVR model	10.92%	62.22%	55.27%

The SVR performed with a lower RMSE and higher correlation than the linear regression one. Therefore, the final 3G voice calls model corresponds to the one obtained with the SVR algorithm, being described by an expression similar to (3), where $f(x)$ is the estimated MOS and the Kernel function ($K(\cdot)$) is the RBF. The $\boldsymbol{x}$ corresponds to an array with all the model features (AS E_c/N_0 Maximum, SIR Minimum, SIR Target SD, AS RSCP Mean and SIR − SIR Target SD).

Results show that the model has a good accuracy (see Table 9). The Pearson correlation is above 60% and the RMSE is bellow 11%.

Figure 3 presents the mapping between predicted and measured MOS, for the Test set. The distance between each point and the diagonal line sets the prediction errors.

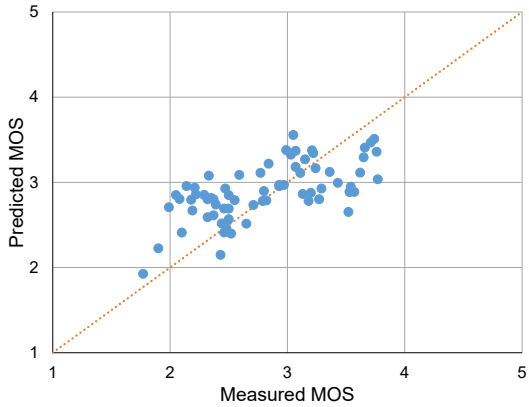

Figure 3: Relation between the Predicted MOS and the Measured MOS (Target) for the Voice Model.

From Figure 3 it can be noticed that the model tends to struggle in differentiate between similar MOS measurements. As the model depends only on RF metrics, the accuracy is limited, since the used metrics are strongly conditioned by the radio channel variability caused, for instance, by multipath fading phenomenon.

5.2. Web Browsing Model

The web browsing model was developed using LTE data also collected through DTs. QoE was assessed using the time that a web page took to be loaded [12]. Therefore, each MOS measurement corresponds to one web page access.

As in the voice model, to select between the two models, obtained using the two machine learning algorithms presented in section 2, the test set was used. The evaluation metrics for both models are present in Table 10.

Table 10: RMSE and correlations calculated through the test set for both Linear regression and SVR web browsing models.

	RMSE	Pearson Correlation	Spearman Correlation
LR model	14.23%	83.17%	86.81%
SVR model	9.79%	91.96%	92.15%

The SVR model performed better that the linear regression one, with a RMSE almost 5% lower and correlations at least 5% higher.

The size of the web page being accessed is considered in all the previously proposed models, therefore, to test the influence of this parameters on the

final results a new model was trained, which considered all the features of the selected model with the exception of the web page size. Thus, this new model takes as input parameters the following features: RSRP Minimum; RSRQ Minimum; MCS Constant Flag; BLER Mean; BLER Kurtosis; CQI Mean.

The model was trained using the SVR learning algorithm. To compare this new model to the previously proposed one, its application to the validation set was assessed. The evaluation metrics computed are represented in Table 11.

Table 11: RMSE and correlations for the bets hypothesis using the linear regression algorithm for the web browsing model.

RMSE	Pearson Correlation	Spearman Correlation
12.39%	81.19%	78.00%

The new model performed with an higher RMSE and lower Pearson and Spearman correlations, which indicates that the model considering the web page size is a better one. However, this new model also shows that with the exclusion of this feature the model still performs with correlations above 75% and with a RMSE of 12.39%. These results indicate that the web page size as a model feature, in spite of improving the model performance, is not a core feature.

If the proposed model had a big dependence on the web page size it would mean that the model did not presented a big advantage relatively to the one used as reference, since the reference model takes as input the download time of a web page which depends on the web page size.

Hence, the final web browsing model is the one obtained with the SVR algorithm that take into account the web page size. It is formulated by (3), being $f(x)$ the estimated MOS and $\boldsymbol{x}$ an array with all the model features (RSRP Minimum, RSRQ Minimum, MCS Constant Flag, BLER Mean and Kurtosis, CQI Mean and Web Page Size). The Kernel function used was the RBF.

The model showed high correlation between the estimated and target MOS (greater than 90%) and the RMSE was bellow 10%. Figure 4 presents the mapping between predicted and measured MOS.

The model being proposed is a function of network QoS parameters, whereas other models proposed in the literature focus on application level QoS, *e.g.* the web page downloading time [12]. The exclusive usage of network QoS parameters enables a direct estimation of QoE, moreover it can be used as an optimization criteria, by allowing to tune the

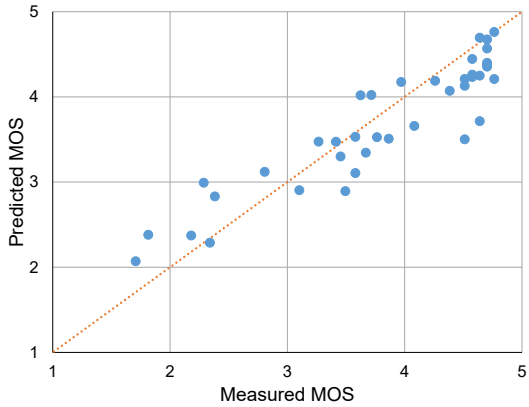

Figure 4: Relation between the Predicted MOS and the Measured MOS (Target) for the Web Browsing Model.

QoS parameters to achieve higher QoE. In future work, a network could be planned using the proposed models by determining the minimum QoS that grants a given user QoE threshold.

6. Conclusions

Two novel QoE models are proposed in this project, one for 3G voice calls and and other for web browsing services, based on real UMTS and LTE data, respectively. Regarding the UMTS voice call QoE estimation, the proposed model obtained a RMSE of 10.92% and a Pearson Correlation of 62.22%. The well known radio channel variability induces uncertainty in the RF parameters, preventing higher correlations and lower RMSE. Nevertheless, the RF data is a widely available resource to Mobile Network Operator (MNO), providing an immediate QoE estimation without any specific setup, by applying the proposed model.

As known, the perceived quality by an end user, when web browsing, is strongly dependent on the web page load time [13]. Hence, the proposed web browsing QoE model besides incorporating QoS metrics, also takes into account the web page size, since its loading time is dependent on this metric. Nonetheless, the proposed model is mainly based on QoS parameters. The developed web browsing model achieved a RMSE of 9.79% and a Pearson Correlation of 91.96%.

QoE modeling has gone through a growing interest among the wireless research community. In that matter, the contribution of this work, which bridges the network QoS domain with the QoE domain, is an early attempt to upgrade from network centric QoS to user centric QoE based radio planning and optimization.

References

[1] NOKIA. Quality of Experience (QoE) of mobile services: Can it be measured and improved? *White Paper*, 2004.

[2] Vasilios A. Siris, Konstantinos Balampekos, and Mahesh K. Marina. Mobile Quality of Experience: Recent Advances and Challenges. *The Sixth International Workshop on Information Quality and Quality of Service for Pervasive Computing*, pages 425–430, 2014.

[3] ITU-T. *Mean Opinion Score (MOS) terminology.* Recommendation ITU-T P.800.1, 2006.

[4] Vasilios A. Siris, Konstantinos Balampekos, and Mahesh K. Marina. Mobile quality of experience: Recent advances and challenges. In *2014 IEEE International Conference on Pervasive Computing and Communication Workshops, PerCom 2014 Workshops, Budapest, Hungary, March 24-28, 2014*, pages 425–430, 2014.

[5] ITU-T. *Perceptual objective listening quality assessment.* Recommendation ITU-T P.863, 2014.

[6] ITU-T. *The E-model: a computational model for use in transmission planning.* Recommendation ITU-T G.107, 2015.

[7] Test Mobile System (TEMS). http://www.tems.com/products-for-radio-and-core-networks/radio-network-engineering/ran-optimization-troubleshooting. Accessed: 2017-03-12.

[8] Andrew Ng. Machine Learning course. https://www.coursera.org/learn/machine-learning/home/welcome. Stanford University.

[9] Alex J. Smola and Bernhard Schlkopf. A Tutorial on Support Vector Regression. Technical report, STATISTICS AND COMPUTING, 2003.

[10] Jushan Bai and Serena Ng. Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics*, 23(1):49–60, 2005.

[11] Lawrence T. Decarlo. On the meaning and use of kurtosis. *Psychological Methods*, pages 292–307, 1997.

[12] Pablo Ameigeiras, Juan J. Ramos-Munoz, Jorge Navarro-Ortiz, Preben Mogensen, and Juan M. Lopez-Soler. QoE Oriented Cross-layer Design of a Resource Allocation Algorithm in Beyond 3G Systems. *Comput. Commun.*, 33(5):571–582, March 2010.

[13] ITU-T. *QoE factors in web-browsing.* Recommendation ITU-T G.1031, 2014.