Parametric Models for Speech Quality Estimation in GSM Networks

M. Barile, P. Camarda, R. Dell'Aquila and N. Vitti
DEE - Politecnico di Bari, Via E. Orabona, 4 - 70125 Bari (Italy)
e-mail: mibarile@libero.it, {camarda, r.dellaquila}@poliba.it, manivit@aliceposta.it

***Abstract*: Speech quality assessment is an emergent QoS aspect in cellular network monitoring. This contribution deals with the estimation of perceived speech quality in GSM networks, by means of parametric models. Our goal is to provide a model for MOS estimation, based on parameters available in the *Operation and Maintenance Centre* (OMC) and in the measurement reports from mobile stations.**

Specifically, based on MOS values from a large database obtained performing intrusive tests in a real GSM network, and on the correspondent measured transmission parameters, an extensive data analysis has been conducted. The correlation coefficients of GSM parameters with the objective speech quality have been then maximized, and a simple linear model has been proposed. Moreover, handover events have been considered and their impact on perceived speech quality has been proved.

***Index Terms*—QoS, GSM, speech quality, parametric models**

1. INTRODUCTION

The rapid growth of cellular networks in recent years has produced the effect of a fast and tight optimization of GSM networks, so that they presently provide a good and stable service; nevertheless, the extreme competition between cellular operators requires a continuous optimization process, necessary for the survival in the market.

In order to measure and benchmark cellular network performances, different *Key Performance Indicators* (KPIs) are used. These KPIs can be collected either from *Network Management System* (NMS) statistics or from drive tests. NMS statistics represent the overall network performance but they do not include speech quality KPIs. On the other side, drive tests give results about speech quality but they are limited only to the tested areas.

During a drive test, calls are continuously originated and received from one or more mobiles carried in a car running along predetermined routes. A dedicated equipment collects all data and provides detailed reports about measured speech quality. This procedure is obviously both time and cost consuming, so each network area is tested only few times a year.

A more simple procedure may be desirable for speech quality estimation, so that a network operator may quickly perform such a measurement and subsequently determine the locations in which a more reliable test needs to be exploited.

The aim of our work is to develop a model for speech quality estimation, based on parameters available in the *Operation and Maintenance Centre* (OMC) and in the measurement reports from mobile stations. Such a model will immediately provide an estimate of speech quality for any part of the network, since the measurement reports of the parameters are always available for the whole network.

There are few previous work in literature concerning this approach, most of them incomplete (i.e. only considering the EFR codec) and/or based on simulated data [1,2], or on parameters unavailable in standard GSM SACCH messages [1,3], or on algorithms unsuited for cellular networks [3]. Additionally, none of them consider the handover event in the analysis of the resulting perceived speech quality.

Our work is instead based on data obtained from a real GSM network; furthermore, both EFR and HR codec have been considered, as they are equally used in GSM networks; finally, handover events have been included in the analyzed parameters.

The rest of the paper is organized as follows: Section 2 introduces the KPIs used in cellular networks; Section 3 outlines speech quality objective measurement methods; Section 4 presents an extensive analysis of the available data and some results necessary for develop the model; in Section 5 the proposed model is described and its performances are evaluated; Section 6 concludes the paper.

2. NETWORK PERFORMANCE INDICATORS

Traditionally, cellular operators have measured and benchmarked their network QoS performances mostly using the *Dropped Call Rate* (DCR) and the *Bit Error Rate* (BER) to quantify the rate of lost connections and the speech quality, respectively.

In a GSM stable network, however, DCR is almost always near 0%, consequently this KPI is not suitable in order to benchmark the QoS provided to users from different networks. Moreover, BER is not the only factor that produces effects on user perceived speech quality. The most important KPI for speech quality assessment is indeed the *Mean Opinion Score* (MOS), defined as the arithmetic mean of subjective evaluations in listening or conversational tests, which range from 1 (bad) to 5 (excellent) [2]. Figure 1 illustrates the different KPIs and the location in the receiving chain where they are measured [1].

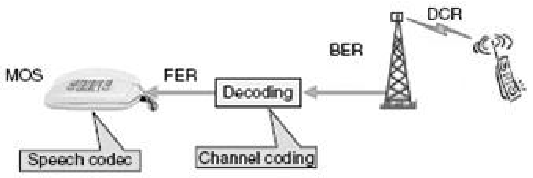

Figure 1 - Main speech KPIs

3. SPEECH QUALITY OBJECTIVE MEASUREMENT

As they have been defined, MOS values can only be measured in a test laboratory environment, since they are obtained from the subjective voice quality perceived from the group of listeners participating in the test. It is worth noticing that they are time consuming to perform and subject to low reproducibility.

In order to estimate perceived speech quality in a telecommunications network, objective measurement methods have been introduced during the last ten years. Objective methods intend to produce an estimate of the MOS score that will be provided by a subjective test, but in an objective and automated way. They can be classified as intrusive or non-intrusive.

Intrusive objective measurement methods provide MOS values comparing speech samples transmitted over the monitored network with the received ones. A typical example of an intrusive algorithm is PESQ [6]. These methods have been developed on the basis of auditory perception and deliver excellent correlations with subjective listening tests; therefore they are implemented in commercial test equipments for network monitoring. However, additional network load is generated by these measurements and they always require an ad hoc test (e.g. a drive test) to be performed on the monitored network. Moreover, they are time-expensive and provide information only about the checked part of the network.

On the other side, non-intrusive objective measurement methods provide MOS values by network passive monitoring. This can be done in two ways: directly analyzing live voice signals being transmitted on the network or by mean of parametric models that provide a MOS estimate based on the values of the monitored parameters.

Non-intrusive objective measurements of course provide less reliable results than the intrusive ones, but the advantage is that an entire network can be monitored at a glance, so giving a first raw estimate of the MOS values in the network.

The aim of our work is therefore to develop a parametric model for non-intrusive speech quality measurement.

4. DATA ANALYSIS

In order to propose a new model, a large collection of real data has been considered. The speech samples have been obtained performing intrusive tests in a real GSM network, during which the correspondent measured transmission parameters have been stored. The MOS values of the output speech samples have been thence evaluated with the SQuad algorithm (an intrusive algorithm like PESQ but optimized for cellular communications [7]) implemented in the adopted equipment.

Totally, 17235 speech samples were acquired, whose MOS distribution is reported in fig. 2. First of all, speech samples have been classified according to the adopted codec (for the most part HR and EFR) and to the speech direction (UL/DL), as reported in table 1, resulting in four data sets. The four correspondent MOS distributions fit the four peaks of the global MOS distribution, as expected in a stable GSM network. Subsequently, for each data set and for each considered parameter, the scatter plots have been drawn. One of these scatter plots is reported in fig. 3.

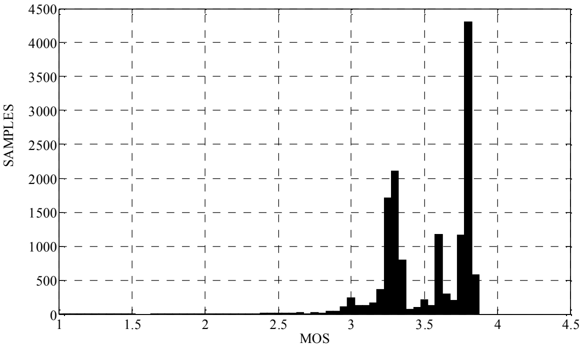

Figure 2 - MOS distribution of all the speech samples

		Codec	
		EFR	HR
Direction	UL	5440	3169
	DL	5459	3167

Table 1 - Speech Samples Distribution

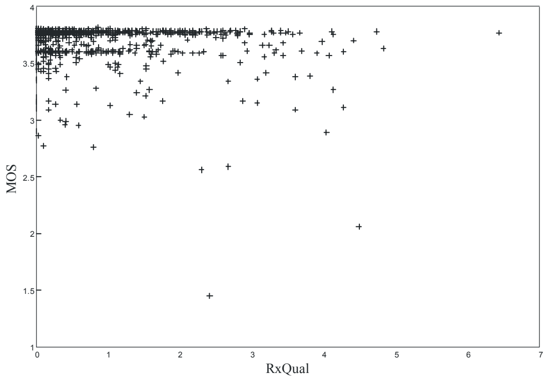

Figure 3 - Relation between MOS and RxQual Sub (EFR, UL)

For our purpose, the following parameters have been initially considered: RxLevel_Full, RxLevel_Sub, TA, RxQual_Full, RxQual_Sub and FER. Handover events, DTX and hopping features have been considered too. The chosen parameters are available every 480 ms on the SACCH channel while the speech samples last 15 seconds, consequently each MOS value corresponds to a set of about 30 measures of each single parameter. In order to calculate the correlations between each parameter and the resulting MOS value, the *Lp*-norm of each set has been considered:

$$L_p(X_i) = \left[\frac{1}{N} \sum_{k=1}^{N} (X_i(k))^p \right]^{1/p} \tag{1}$$

where $X_i(k)$ is the *k*-th measure of the parameter X in the *i*-th set. The reason for using *Lp*-norm is that for each parameter, variations may be perceived in a different way with respect to the resulting speech quality. High values for *p* emphasize parameter variations [1].

For each data set and for each parameter, the correlation coefficient has been calculated for $p \in \{1/10, 1/9, 1/8, ..., 1/2, 1, 2, 3, \ldots, 9, 10\}$ and displayed in a graphic way in order to evaluate the value of *p* that maximize the correlation coefficient. An example is reported in fig. 4. In table 2 the correlation coefficients for $p = 1$ (i.e. for the arithmetic mean of each set of *N* measures) are reported.

As expected, RxLevel (both Full and Sub) is not correlated with MOS value, as this parameter simply represent the received signal power. A low correlation with MOS is obtained also for TA, that is related to the distance between the mobile station and the serving base station. RxQual and FER (both deriving from the BER and thence from the signal to noise ratio) get instead more correlated with MOS. Moreover, RxQual_Sub gets more correlated with MOS than RxQual_Full, because of the predominant use of DTX feature in the considered network. Finally, FER is correlated with MOS only when EFR codec is used, because the limited number of bits used for the CRC in HR transmission can cause a bad frame to be reputed as good.

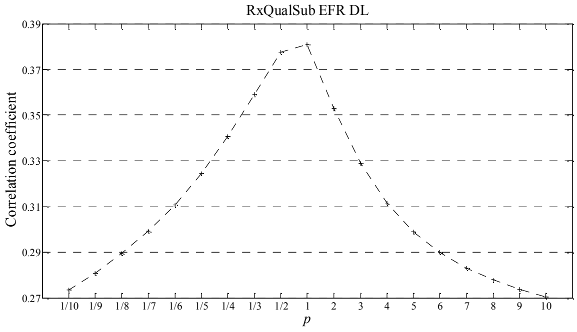

Figure 4 - Correlation coefficient vs. p for RxQual_Sub (EFR, DL)

In order to consider handover events that can cause short interruptions during the communication, the artificial parameter HO has been introduced. For each sample, HO is equal to 1 if a handover occurs during its transmission, otherwise it is equal to 0. This parameter has shown a high correlation with MOS, as reported in table 2.

PARAMETER	EFR		HR	
	UL	DL	UL	DL
RxQual_Sub	-0.372	-0.387	-0.293	-0.117
RxLevel_Sub	0.020	0.114	0.010	0.023
FER	-0.572	-0.517	-0.024	-0.022
TA	-0.093	-0.101	-0.086	-0.104
HO	-0.395	-0.251	-0.248	-0.368

Table 2 - Correlation coefficients of some parameters with MOS for p=1

5. PROPOSED MODEL

The proposed model is a simple linear one, derived by the application of the multiple linear regression method. The independent variables are RxQual_Sub, FER and HO, for the EFR case, and only RxQual_Sub and HO for the HR case. In order to consider the parameter variations during the transmission of a speech sample, for RxQual_Sub and FER the *Lp*-norm with the *p* values that maximize the correlation coefficients have been considered, as explained before. The model expression is:

$$MOS = A_0 + A_1\, Lp_1(RxQ\_S) + A_2\, Lp_2(FER) + A_3\, HO \tag{2}$$

where $A_2 = 0$ for the HR case, and the values for p_1 and p_2 are reported in table 3.

COEFFICIENT	EFR		HR	
	UL	DL	UL	DL
p_1	1/2	1	2	2
p_2	1	1	-	-

Table 3 - p values that maximize the correlation coefficients

In order to obtain the coefficients, the method of least squares has been applied to all the data sets and the coefficients of determination R^2 have been calculated, providing the results summarized in table 4.

COEFFICIENT	EFR		HR	
	UL	DL	UL	DL
A_0	3.772	3.683	3.271	3.257
A_1	-0.009	-0.044	-0.003	-0.014
A_2	-0.023	-0.041	-	-
A_3	-0.115	-0.071	-0.156	-0.241
R^2	0.272	0.256	0.121	0.174

Table 4 - Coefficients of the proposed model

The distribution of MOS values obtained by applying the proposed model with the measured transmission parameters has been evaluated (see fig. 5). This distribution is similar to the original one shown in fig. 2.

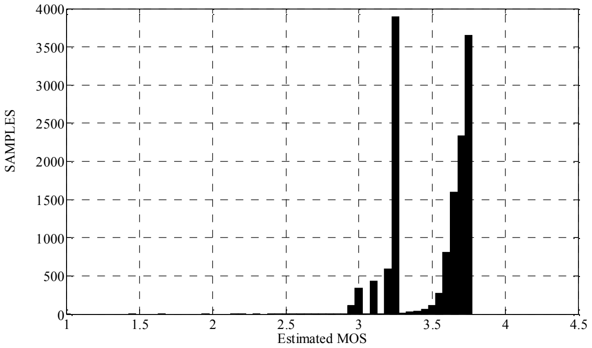

Figure 5 - Estimated MOS distribution

Finally, the cumulative distribution functions of the two distributions have been plotted in fig. 6 and, according to the speech transmission quality categories defined in [8], table 5 reports both the measured and the estimated percentages of samples which MOS fall in that categories.

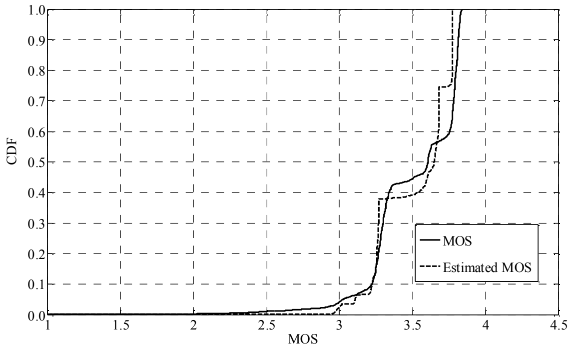

Figure 6 – CDFs of MOS and Estimated MOS

MOS	Speech Transmission Quality Category	% real samples	% estimated
[1.0, 2.6[	Poor	1.3	0.1
[2.6, 3.1[	Low	5.0	3.5
[3.1, 3.6[	Medium	42.8	40.8
[3.6, 4.5]	High/Best	50.9	55.6

Table 5 - Percentages of real and estimated MOS samples for each speech transmission quality category range

From table 5 it is possible to notice that, with our model, the estimated percentages of speech samples for each quality category range are close to the real ones.

6. CONCLUSIONS

An analysis of the influence of some parameters with perceived speech quality has been conduced and a simple model for an estimate of MOS values in a cellular network has been proposed. The impact of handover events on the resulting perceived speech quality has been proved.

The proposed model has shown to be suitable for estimate the MOS distribution in a network area by considering some parameters available in the network operator's OMC.

All the analyzed data derive from measures on a real GSM network without never recur to simulation neither of the speech samples nor of the physical channels, although this has caused a limited variability of the MOS values in the available set of measures.

A future extension of the present work will consider a larger set of measures and a more sophisticated model, in order to obtain a more accurate MOS estimation.

REFERENCES

[1] M. Werner, K. Kamps, U. Tuisel, J.G. Beerends and P. Vary, "Parameter-based speech quality measures for GSM" in *Proc. of IEEE PIMRC*, Sept. 2003.

[2] B. Dolezalova, J. Holub and M. Street, "Mobile network voice transmission quality estimation based on radio path parameters" in *Proc. of IEEE Wireless Telecommunications Symposium*, April 2005.

[3] A. Karlsson, G. Heikkila, T.B. Minde, M. Nordlund and B. Timus, "Radio link parameter based speech quality index-SQI" in *Proc. of IEEE Workshop on Speech Coding*, June 1999.

[4] Int. Telecommun. Union, Methods for subjective determination of transmission quality, ITU-T Recomm. P.800, Aug. 1996.

[5] T. Halonen, J. Romero and J. Melero, Eds., *GSM, GPRS and EDGE Performance*, 2nd ed. Wiley 2003.

[6] Int. Telecommun. Union, Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs, ITU-T Recomm. P.862, Feb. 2001.

[7] Swissqual inc., Cause Analysis of Objective Speech Quality Degradations used in Squad, ITU-T SG12 (Study Period 2001-2004) Contribution 32, Aug. 2001

[8] Int. Telecommun. Union, Definition of categories of speech transmission quality, ITU-T Recomm. G.109, Sept. 1999