Speech Quality Assessment in Mobile Phones Using a Reduced- complexity Algorithm

Khalid Al-Mashouq
Electrical Engineering Department
King Saudi University
Riyadh, Saudi Arabia
mashouq@ksu.edu.sa

Akram Aburas
Chief Executive Officer
ACES
Riyadh, Saudi Arabia
akram@aces-co.com

Musharraf Maqbool
Communication Head
ACES
Riyadh, Saudi Arabia
khalil@aces-com

***Abstract*—In this paper, we present a reduced-complexity algorithm to assess the quality of speech as perceived by the mobile user. This algorithm utilizes the channel parameters, as measured by the mobile handset, to estimate the speech quality. We used two estimation models; a linear model and a neural network-based model. We compared our estimation with the standard International Telecommunication Union, ITU, objective speech quality measure, perceptual evaluation of speech quality. We found that the linear model can achieve up to eighty four percent correlations with perceptual evaluation of speech quality. Moreover, the neural network-based model can achieve more than ninety percent correlations with perceptual evaluation of speech quality.**

Keywords-Speech Quality Measurement; Perceptual Evaluation of Speech Quality (PESQ); Signal Strength; Bit Error Rate (BER); Frame Erasure Rate (FER); Neural network.

I. INTRODUCTION

Mobile operators are competing to gain customer satisfaction, subsequently reducing the churn rate. Today, voice service is the dominant one among mobile services. Maintaining high speech quality will contribute to better customer satisfaction. The continuous monitoring and assessment of speech quality is essential.

In general, speech quality assessment is performed using subjective or objective methods. A subjective method is based on a group of "good" listeners who can rate the speech signal from 1 (bad) to 5 (excellent). The average score is then taken, which is called mean opinion score, MOS. For obvious reasons, this method cannot be used for the continuous assessment of speech in mobile network.

A subjective method is based on exchanging a "reference" speech segment between a mobile phone to another, preferably, fixed one. The received, possibly noisy, speech segment is then compared with the original "clean" one. International Telecommunication Union, ITU, adopted a standard objective algorithm to process and compares the two speech segments and calculates the quality score, PESQ, or perceptual evaluation of speech quality [1].

Depending on sending a reference speech segment will limit the usability of this method in real-time speech quality assessment. Many researchers investigated other approaches utilizing only the received speech signal, which is called output-based (or non-intrusive) speech quality assessment [2-4]. One approach is to exploit the Markovian structure of speech to detect noise or impairments [2,3]. Another approach is to incorporate some aspects of the human auditory perception mechanism [4].

In general, these approaches are computationally intensive and could affect the processing power of mobile digital signal processor as well as the battery. Moreover, they are generic for any environment and not customized to mobile networks. Distortion in received signal is mainly due to background noise, vocoder imperfection, and/or radio channel noise. Optimization engineer has control only to the later cause. Therefore, our focus is to measure the channel parameters and utilize them to give prediction on the speech quality as affected by the channel impairments.

In this paper, we are collecting a large number of speech samples from a live GSM network using Qvoice (from ASCOM) benchmarking speech tools [5]. Qvoice is used to give PESQ score for all speech samples. In the next section, we describe in details our setup in collecting speech data. Section 3 explains the two different prediction models. They are applied on the collected data and compared with the PESQ score. We outline our conclusions in Section 4.

II. DATA COLLECTION AND PRE-PROCESSING

In this section, we highlight our work which focuses on obtaining speech samples and score them based on the objective speech quality measure PESQ. To facilitate such a testing, we utilized the network benchmarking tool Q-Voice. Random streets from Riyadh city were selected and speech testing was undertaken.

As illustrated in Figure 1, the two major components of the testing system (Q-Voice) are the server and the companion. Pre-recorded speech samples are always stored on the server connected to a PSTN network. The speech samples were carefully selected as "phonetically balanced" to represent normal telephone conversations. The companion who hosts the mobile phones calls the server from the selected streets using the GSM network and once the call is setup, the speech samples are transmitted from the companion to the server. The server upon receipt of these speech samples carries out a comparison and assigns a speech quality score, PESQ, to it.

ISBN: 978-1-61208-229-5

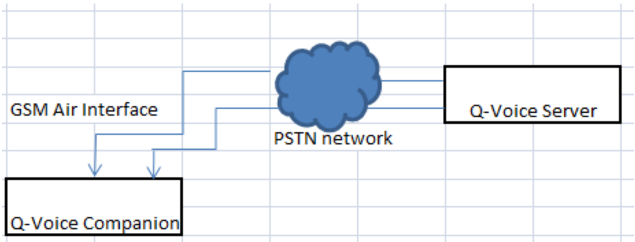

Figure 1. Network Testing Equipment Block Diagram

Upon receipt of these samples at the server end, it is possible to obtain the received speech samples in audio format and verify the degradation level. Speech samples that were transmitted during an active conversation were obtained and parameters associated with each speech sample such as PESQ score, received signal level, RxL, Bit Error Rate, BER, Frame Erasure Rate, FER, and carrier over interference, C/I, were also obtained for the same samples, respectively. Please note that BER is usually mapped under RxQ.

The speech samples obtained from an active conversation were given to listeners with normal hearing. Listeners were trained to familiarize them with the different versions of samples that were received either in excellent or distorted form. Scoring of speech samples by listeners was undertaken in which after hearing each speech sample, the listeners had to grade the speech sample they had listened. This step is needed to randomly verify the machine scoring.

TABLE I. SPEECH SAMPLE GRADING

Score	Classification
1	Bad
2	Poor
3	Fair
4	Good
5	Excellent

Table 1 above outlines the grading or PESQ score of the speech samples by the system that was used in obtaining the speech samples and the associated parameters mentioned earlier. Listeners after been trained on some speech samples were then asked to classify the samples as per the table above.

We see it vital to mention here that the algorithm used by the network testing equipment Q-Voice is a reference based algorithm that utilizes the received signal and then extracts associated signal parameters form the speech sample and tries to evaluate to what degree the distortions in the received signal will be audible to the human ear. Every speech sample obtained or used in this exercise constitutes a 5 second voice transmission.

The approach of this paper is systematically aimed at attempting to correlate and successfully link the correlation between the obtained signal parameters and the speech samples graded Qvoice. Table 2 shows and extract of our collected data. Our data contains 759 speech samples each of which is 5 seconds long.

TABLE II. SAMPLE OF DATA OBTAINED FROM QVOICE

RxL	Rx. Qual	FER	C/I	PESQ
-86.56	0.71428	1.28	20	3.9
-83.31	0	1.5	20	3.9
-78.33	0	1	20	3.9
-81.97	0	1.333	20	3.9
-82.76	0.57142	0.9411	20	3.9
-83.10	0	1.2	20	3.9

We move on further to briefly mention about the parameters of importance to us and their impact on a network. RxL is basically an indicator of the coverage been provided by a network operator and is represented as –dBm. RxQ is one signal parameter that is basically a mapping of time averaged bit errors over a scale of 0 to 7 which gives a rough indication on the speech quality. During the analysis, it was found that for every speech sample of 5 second duration numerous RxQ values were obtained, which is quite logical given the duration of the transmission and the fact that the measuring system undertakes the measurement many times. The RxQ values (or the BER values) were averaged out to gain an average for the whole transmitted sample.

Carrier to Interference ratio, C/I, helps in determining the level of interference the subjected signal has undergone. A High C/I will indicate a good signal and yield good communication. Whereas, a low C/I will result in degraded signal quality.

From our pre-processing results, we have also noted that there is a difference on occasions when the human grading differed from the system grading. We therefore, see that there stands a substantial needs for a modified real time network assessment to enable overcome these gaps.

Keeping in view the above parameters and their significance, extensive simulations were carried out on all the samples to feed our estimator. This is shown in greater detail in the following section.

III. PREDICTION MODELS AND RESULTS

We used a liner model to estimate PSEQ score using the following four parameters RxL, RxQ, C/I and FER. They are combined using least square method for optimal weighting. This model is then tested with real data and compared with PSEQ.

The estimated quality score, q, will be

$$q = \sum_{i=0}^{4} a_i \; w_i$$

where w_i is the weighing factor of the ith parameter

a_1 is RxL

a_2 is RxQ

a_3 is FER

a_4 is C/I

The standard least square solution to this problem is given by [6]:

$$\underline{w} = (A^T A)^{-1} A^T \underline{C}$$

where; $\underline{w} = (w_1 \; w_2 \; w_3 \; w_4)^T$

and A=($\underline{\alpha}_1$ $\underline{\alpha}_2$ $\underline{\alpha}_3$ $\underline{\alpha}_4$) is the measurements matrix. Here, ($\underline{\alpha}_1$ $\underline{\alpha}_2$ $\underline{\alpha}_3$ $\underline{\alpha}_4$) corresponds to the measurement column vector (of length N) of RxL, RxQ, FER and C/I respectively. Each vector contains N samples corresponding to N speech samples. For, the "PESQ" vector $\underline{C}$ = (C_1 C_2 ….. C_N $)^T$; C_i corresponds to the i^{th} sample of PESQ measurement. We used the whole 759 samples to obtain the optimum linear combination waiting vector $\underline{w}$. The correlation between the predicted PESQ, q, and the actual PESQ is 84%.

The second model is a 2-layer back-propagation neural network [x]. The number of hidden layers is varied between 3 and 30. We used Matlab® neural toolbox for our simulations. Figure 2 shows simulations results when the number of hidden units is 5. The four graphs show the scatter diagram between q and PESQ for training, validation, testing and overall data, respectively. The corresponding correlation coefficients are 0.937, 0.908, 0.942 and 0.932, respectively. It is apparent that the neural model yields good improvement in the prediction ability.

IV. CONCLUSTIONS AND FUTURE WORKS

We have addressed the problem of continuous assessment of speech quality in mobile networks. The purpose is to help operators capture the actual impression of their customers. This should complement the network monitoring and operation centers. We relied on the measured channel parameters to estimate the speech quality of service. We used a linear prediction model, which yielded 84% correlation with the PESQ. A 2-layer neural network is also used, after proper training, to predict the speech quality. The predicted quality score achieved higher correlation, which reaches more than 90%, with the PESQ.

Our approach has reduced complexity compared with Markovain-based ones. This supports its usage within the mobile handset as it would not have significant impact on the processing power and battery life.

The ease of this prediction model can pave the road for several applications. Examples of these applications are

- Customer automated evaluation of the network.
- Quality scores can be relayed to the operator to help in optimizing the network
- Can be used by the operator to give a new tariff procedure or compensation for calls, with bad quality

To make the optimal and robust prediction model, one needs to have much more samples collected over wide spectrum of wireless networks. This should include various geographical locations, different operators and network types.

REFERENCES

[1] ITU-T Rec. P.862, "Perceptual Evaluation of Speech Quality (PESQ), An objective method for end to end speech quality assessment of narrowband telephone networks and speech codecs," 2001.

[2] Khalid A. Al-Mashouq, Mohammed S. Al-Shaye. "Output-Based Speech Quality Assessment with Application to CTIMIT Database." Proceedings of the ISCA 17th International Conference Computers and Their Applications, April 4-6, 2002, Canterbury Hotel, San Francisco, California, USA 2002

[3] Chiyi Jin and R. Kubichek, "Vector Quantization techniques for Output-Based Objective Speech Quality," IEEE Int'l Conference on Acoustics, Speech, and Signal Processing (ICASSP) 1996.

[4] Kartik Audhkhasi and Arun Kumar, *"Two-scale auditory feature based non-intrusive speech quality evaluation"*, IETE Journal of Research, vol. 56, no. 2, pp. 111-118, March-April 2010.

[5] http://www.ascom.ch/ch-en/tems-symphony-60-datasheet.pdf [Oct 13, 2012]

[6] Lang, Serge, *Linear algebra*, Berlin, New York: Springer-Verlag, ISBN 978-0-387-96412-6. 1987.

.

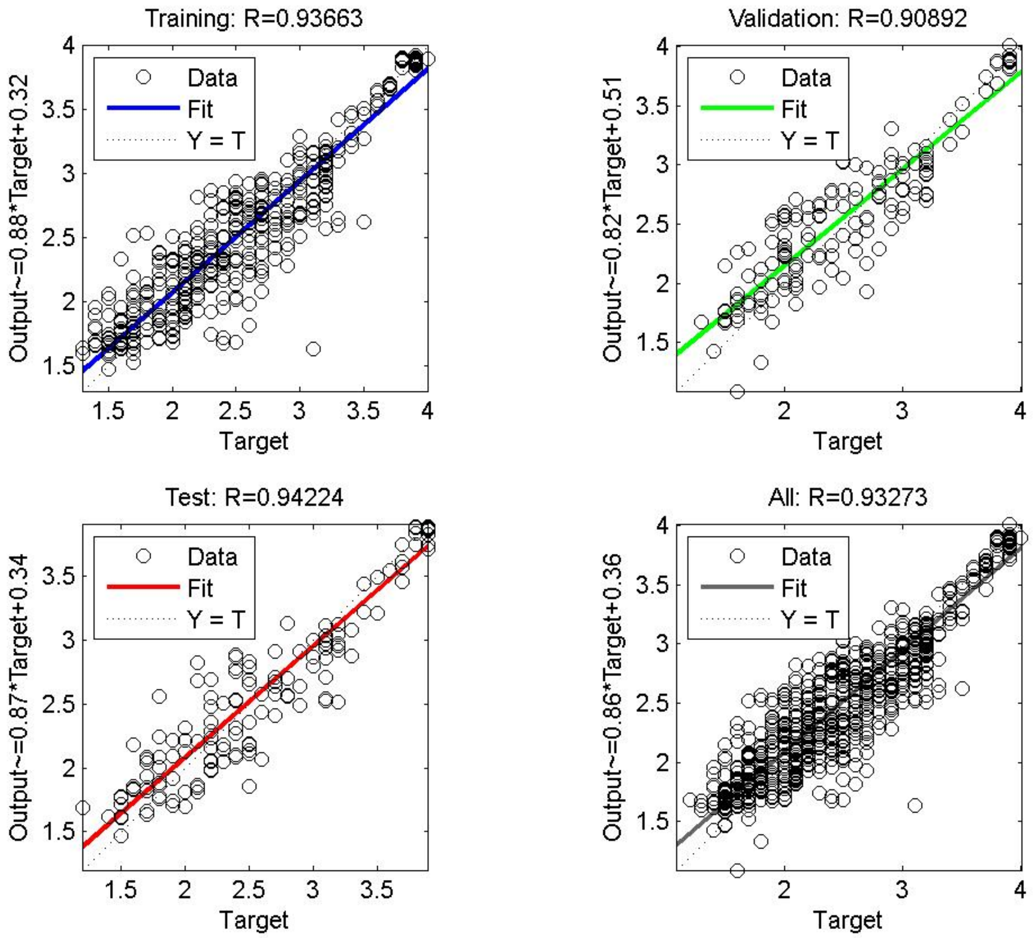

Figure 2. Result of a 2-layer neural network estimator for training (upper left corner), validation (upper right corner), and testing (lower left corner) and over all data (lower right corner). Target stands for PESQ and Output is the estimated PESQ, or q and R is the correlation coeffcient between PESQ and q.