# ERK 2010

http://www.ieee.si/erk10/

20. - 22. september 2010 GH Metropol, Portorož, Slovenia

**Objava [Call]** | **Organizatorji [Organizers]** | **Zbornik [Proceedings]** | **Program**

IEEE Slovenska sekcija Slovenian Section

## Devetnajsta mednarodna Elektrotehniška in računalniška konferenca
## Nineteenth International Electrotechnical and Computer Science Conference

<table>
<tr><th></th><th colspan="5">Ponedeljek, 20. september, 2010</th><th colspan="5">Torek, 21. september, 2010</th><th colspan="3">Sreda, 22. september, 2010</th></tr>
<tr><td>soba</td><td>9h45</td><td>11h-13h</td><td>13h-14h</td><td>15h-17h</td><td>17h-19h</td><td>8h30-10h30</td><td>10h30-13h</td><td>13h30-15h</td><td>15h-17h</td><td>17h-19h</td><td>8h30-10h30</td><td>10h30-12h30</td><td>12h30-14h</td></tr>
<tr><td>A</td><td rowspan="5">otvoritev</td><td>VP1,2,3</td><td>VP4</td><td>CS1</td><td>CS2</td><td>ME3</td><td>VP5,6,7</td><td>ST</td><td>CS3</td><td>CS4</td><td>CS5</td><td>CS6</td><td rowspan="2"></td></tr>
<tr><td>B</td><td rowspan="4" colspan="2"></td><td>TC1</td><td>TC2</td><td>TC3</td><td rowspan="2" colspan="2"></td><td>ME4</td><td>TC4</td><td>TC5</td><td>TC6</td></tr>
<tr><td>C</td><td>ME1</td><td>ME2</td><td>BM1</td><td>BM2</td><td>BM3</td><td>MT1</td><td>MT2</td><td>MT3</td></tr>
<tr><td>D</td><td>AV1</td><td>AV2</td><td>AV3</td><td></td><td>14h<br>SM1</td><td>SM2</td><td>SM3</td><td>PR1</td><td>PR2</td><td rowspan="2"></td></tr>
<tr><td>E</td><td>RO1</td><td>RO2</td><td>EL1</td><td colspan="2"></td><td>EL2</td><td>EL3</td><td>AI1</td><td>DI1</td></tr>
</table>

### Področja / Topics

**VP** - Vabljena predavanja / Invited Lectures
**EL** - Elektronika / Electronics
**TC** - Telekomunikacije / Telecommunications
**AV** - Avtomatika / Automatic Control
**SM** - Simulacije, identifikacije, modeliranje / Simulations, Identifications, Modelling
**ME** - Močnostna elektrotehnika / Power Engineering
**MT** - Merilna tehnika - ISEMEC 2010 / Measurement - ISEMEC 2010
**CS** - Računalništvo in informatika / Computer and Information Science
**AI** - Umetna inteligenca / Artificial Intelligence
**RO** - Robotika / Robotics
**PR** - Razpoznavanje vzorcev / Pattern Recognition
**BM** - Biomedicinska tehnika / Biomedical Engineering
**DI** - Didaktika / Didactics
**STUD** - Študentski članki / Student Papers

### Kazalo / Table of Contents



Zbornik sta uredila / Proceedings Edited by:
Baldomir Zajc, Andrej Trost

## Sponzor / Sponsor

- Javna agencija za raziskovalno dejavnost RS

## Organizatorji / Organizers

- IEEE Region 8
- Slovenska sekcija IEEE
- Slovenska inženirska zveza – SIZ
- Elektrotehniška zveza Slovenije – EZS
- Fakulteta za elektrotehniko, Univerza v Ljubljani

## Strokovna društva

Pri delu posameznih konferenčnih sekcij sodelujejo naslednja strokovna društva:

- Društvo avtomatikov Slovenije,
- Slovensko društvo za merilno-procesno tehniko (ISEMEC 2010),
- SLOKO–CIGRE,
- Društvo za medicinsko in biološko tehniko Slovenije,
- Društvo robotikov Slovenije,
- Slovensko društvo za umetno inteligenco,
- Slovensko društvo za razpoznavanje vzorcev,
- Slovensko društvo za simulacijo in modeliranje,
- Slovenska inženirska zveza.

# Konferenčni odbori
# Conference Committees

*Predsednik konference / Conference Chairman*
**Baldomir Zajc**

*Podpredsednik konference / Conference Vicechairman*
**Saša Divjak**
**Matej Zajc**

*Predsednik programskega odbora / Program Committee Chairman*
**Jurij Tasič**

*Programski odbor / Programe Committee*
**Tadej Bajd**
**Genevieve Baudoin**
**Saša Divjak**
**Janko Drnovšek**
**Peter Farkaš**
**Matjaž Gams**
**Ferdo Gubina**
**Franc Jager**
**Marko Jagodič**
**Karel Jezernik**
**Drago Matko**
**Damijan Miklavčič**
**Miro Milanovič**
**Josef Modelski**
**Michael Muhr**
**Janez Nastran**
**George Paunović**
**Nikola Pavešić**
**Kurt Richter**
**Franc Solina**
**Aleksander Szabo**
**Fabio Tosato**
**Borut Zupančič**
**Andrej Žemva**

*Predsednik odbora za tisk / Publications Chairman*
**Andrej Žemva**

# Speech Quality Measurement in GSM Networks Using Time Encoded Signal Processing

**Ivan Kraljevski[1], Igor Stojanovic[2], Zoran Gacovski[1], Biljana Spireva[1]**

[1]*Faculty for ICT, FON University, bul. Vojvodina bb, Skopje, Macedonia*
[2]*Faculty of Computer Science, University Goce Delcev, Toso Arsov 14, 2000 Stip, Macedonia*
*E-mail: {ivan.kraljevski, zoran.gacovski, biljana.spireva}@fon.edu.mk, igor.stojanovik@ugd.edu.mk*

## Abstract

*In this paper a method for speech quality estimation was evaluated simulating transmission of AMR-NB encoded speech over noisy GSM channel. The proposed system uses comparison of Time Encoded Signal (TES) processing of speech sequences, where one original and one degraded speech signal were transmitted trough GSM simulation system with AWGN noise channel. Several tests have been made on reference speech sample of single speaker with simulated bit-error loss effects on the perceived speech. The achieved results and the similarity measure scores between two TES speech sequences for various levels of noise channel conditions were compared with measured PESQ MOS values of the used channel and the correlation between them was observed.*

## 1 Introduction

Speech quality measurement is very important factor in the process of providing quality of service for voice telecommunications networks, particularly in wireless communication networks as GSM. In these networks, channel characteristics are very different comparing to those found in wired communications networks. In this case, for speech transmission frames are used, and transmission errors are presented in form of bit or frame error ratio.

Understanding and estimation of these parameters are of great significance for optimization of the telecommunication services and infrastructure [1].

Speech quality is defined by the way how the listeners valued the perceived speech signals on the receiver side of the communication channel. Due to evermore increasing complexity of the communication networks and the number of parameters which characterizes the communication channel, it is much more difficult to establish straightforward relation between transport parameters and the perceived speech quality. Besides that, there is a problem of accurate extraction or estimation of the exact communication parameters, and the perceived speech quality does not always correspond to the measured or estimated transport parameters of the received speech.

In this paper a method for speech quality measure is presented and its usability is assessed. The presented approach uses comparison between TES (*Time Encoded Signal*) coded reference and degraded speech sequences after transmission over GSM mobile communication channels. This method uses transformation and comparison of AMR-NB encoded original and degraded speech into TES S-matrices. TES matrices represent precise mathematical description of speech, where band limited signals (such as human speech) may be completely described by the locations of their real and complex zeros [2].

The recorded speech sequence was AMR-NB coded with 12.2 kb/s (compatible with GSM-EFR) [3], GSM and AMR codecs were employed as the main voice codecs used in mobile networks. The effects of bit-error rate errors were simulated by using complete GSM simulation model. The model simulates all aspects of speech transmission through GSM channel. The BER (*Bit Error Rate*) performance of the simulated transmission channel is estimated by comparing AMR-NB encoded input sequence with the reconstructed sequence on the receiver side.

Different values for similarity measure are observed after comparing the test with the received speech sequences by varying the level of noise in the transmission channel and introducing appropriate BER values. Achieved results were compared with PESQ measured values (P.862 ITU-T) [5] on the transmission channel. They introduce high correlation values which justify the usability of this technique as a simple tool for perceived speech quality measurement in GSM networks.

## 2 Speech quality

The perceived quality of a telephony services can be measured with subjective tests. Humans evaluate the quality of service according to a standardized quality assessment process. The speech quality is described by a mean opinion score (MOS) values (from 1 - bad to 5 - excellent), also called MOS-Listening Quality Subjective (MOS-LQS).

The MOS-LQS test is also called the Absolute Category Rating (ACR) test and it is described in details in ITU Recommendation P.80. Speech quality estimation could be performed by intrusive and non-intrusive methods. Non-intrusive methods monitor the received speech information, where some characteristics are extracted and used for further processing for speech quality estimation. The drawback is the unavailability of the original speech sample for comparison with the received one, and it is possible to oversee some distortion effects of the signal that are not possible to be detected or measured but have significant influence to perceived speech (like e.g. 3SQM P.563 ITU-T).

Intrusive methods for quality estimation use reference speech sequences that are transmitted over the communication channel. The received speech is compared with the test sequence in a similar way as the human speech perception and the quality is graded as the listeners should do in traditional subjective tests (like MOS). An example of one of the most used algorithms for intrusive tests in packet switched and mobile networks is PESQ (*Perceptual Evaluation of Speech Quality)*, defined in P.862 ITU-T, but it introduces some disadvantages regarding computing complexity and it is not possible to use it for data rates below 4 kbps [5].

# 3 Simulation system description

Block diagram of the system that is used for simulation is shown on Figure 1. The system is designed and coded in MATLAB and allows simulation of reference sequence transmission over the GSM network with bit error or frame loss events. At the receiver side, a comparison between the reference and received speech sequence is done. The system consists of speech encoder/decoder, GSM transmitter and receiver (channel encoder/interleaver, multiplexer, GMSK modem with BER simulator, channel demultiplexer and decoder/deinterleaver) and comparator of the TES S-matrices of degraded and the test sequence.

## 3.1 Speech encoder

The evaluated system uses AMR-NB speech coder based on Algebraic Code Excited Linear Prediction (A-CELP). It is standardized by ETSI and widely used in GSM and UMTS [3]. It uses link adaptation to select one of eight different narrowband modes of operation and transmission data rates between of 4.75 and 12.2 Kb/s.

Using AMR requires link adaptation that selects the optimal codec mode to adapt to the channel and capacity requirements. This improves robustness of the network connection in bad channel condition but decreases the perceived speech quality.

In the experiments, the 12.2 Kb/s Full Rate mode compatible with GSM-EFR (3GPP TS 26.071) was used. Speech segments with 20 ms duration (8 KHz, 13 bit PCM) are AMR-NB coded into 31 bytes long frames without VAD *(Voice Activity Detector)* or PLC (*Packet Loss Concealment*) option. The reference speech sample has duration of 13 seconds, and it is recorded by a male speaker in Macedonian language.

Low bit rate (high compression ratio) speech coders used to reduce required bandwidth distort the original waveform significantly before it is even transmitted. The compressed speech produced by such coders is also more sensitive to frame loss [4].

## 3.2 BER simulation

Frame corruption due to introduced BER is a major source of speech impairment in GSM channel. The *Independent Channel Model* was used for the transmission channel. It is very simple and determines if the transmitted bit is false or not, that is, there is a bit error in a frame. The result of this model is obtained using Bernoulli function [6] with parameter $P_{frerr}$ (1).

$$P_{frerr} = 1 - (1 - p_{ber})^L \qquad (1)$$

$L$ is the length (in bits) of the frame and $p_{ber}$ is the Bit Error Rate (BER) probability associated with the channel. BER is estimated by measuring performance of a *Gaussian Minimum Shift Keying* modem for given energy per bit to noise power spectral density ratio (Eb/N0) or normalized signal-to-noise ratio (SNR) over AWGN channel [7].

The impact of degraded frame occurrence on the perceived speech quality depends on several factors, including loss pattern, codec type, and frame loss size [4]. It may also depend on the location of loss within the speech, for example degradation of unvoiced frames has less impact in perceived speech quality, than degradation of voiced frames. Even more, as most real communication channels exhibit burst of frame loss, occurrence of burst of false frames has significant influence over perceived speech.

## 3.3 Time Encoded Signals Processing

TES coding is based on precise mathematical description of waveforms, involving the polynomial theory that shows how band limited signals (such as human speech) may be completely described by the locations of their real and complex zeros [2].

The interval between two adjacent zero-crossings of speech waveform is called an epoch and, for every epoch, three parameters are derived: duration of the epoch (D), shape (S) and the magnitude (M) of the signal. D is the number of samples of one epoch, S is the number of positive or negative local maxima and minima and M is the largest value of samples in the given period. These parameters are encoded with assigning a unique symbol for certain combination of the epoch duration (D) and its shape (S). Thus the signal is transformed into time encoded stream of discrete numerical descriptors – TES symbols.

Using vector quantization and K-means algorithm, generalized code-book was created – TES alphabet. Standard symbol alphabet consists of 28 different symbols, and it has been proved to be quite sufficient for the representation of speech and other band limited signals. These strings of numerical descriptors can be easily converted to TES matrices with fixed dimension.

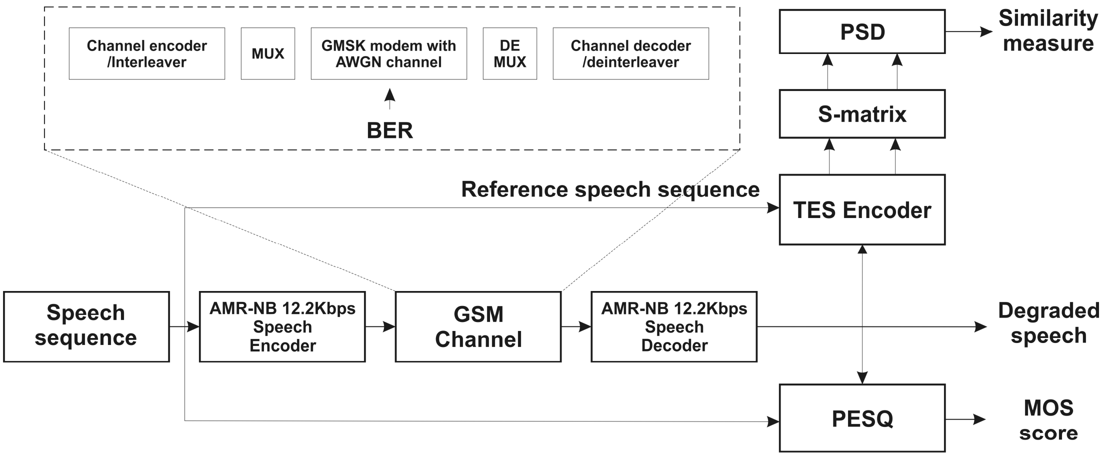


Figure 1: System architecture.

A histogram of the signal array with 28 possible symbolic descriptors can be produced, forming so called S-matrix with fixed dimension 1x28 (Figure 2), which carry information about symbols frequency.

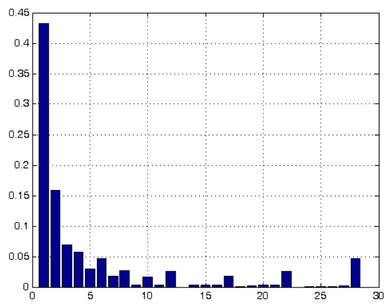

Figure 2: S-matrix of reference speech sequence.

The difference between two S-matrices is calculated using the Perceptual Spectral Distance (PSD) between the $S_1$ and $S_2$ matrices (2):

$$PSD = \sqrt{\sum_{n=1}^{N} [S_1(n) - S_2(n)]^2} \qquad (2)$$

where $S_1$ and $S_2$ represent S-matrices of the original and degraded signals, respectively. N is the length of the both S matrices, typically 28.

The biggest advantage of the S-matrix representation of speech compared to other arrays of speech descriptors in the frequency domain (ex. MFCC) is that, regardless of the signal length they have fixed dimensions. This is significant advantage that allows intrusive measurements with longer speech sequences than is currently feasible with standardized testing algorithms (like PESQ).

## 4 Simulation Results

Speech sequence with duration of 13 seconds of male speaker on Macedonian was AMR-NB coded with data rate of 12.2 Kbps. The speech sequence was processed by channel encoder/interleaver [8] and accepted by the multiplexer that splits the incoming sequence to form a GSM normal burst. After creation of the prescribed GSM normal burst data structure, the MUX returns this to the GMSK modulator. There, occurrence of bit errors takes place regarding given EbN0 value (2-10 dB, with 0.1 dB increments) and on the receiver side the transmitted GSM burst data appears with false bits.

The demodulated and degraded bit sequence is then used as input to the demultiplexer where the bits are split in order to retrieve the actual data bits. As a final operation, to retrieve the estimated transmitted bits - channel decoding and de-interleaving is performed.

Channel coder/decoder succeeds to fix some of the erroneous bits by convolution decoding of Class 1A bits (high subjective importance bits) [3]. Because of that, the introduced BER by AWGN channel is greater compared to the measured BER on the exit of the receiver (Figure 3). The erroneous bits induces frame loss event and the received speech sequence is distorted by the effects of false decoding of corrupted frame. Packet loss concealment methods may be used to minimize the impact of the corrupted frame decoding.

For given transmitted and received speech sequences, averaged difference values (similarity score) of their S-matrices were produced, as well as averaged PESQ MOS values regarding EbN0 as input parameter.

The following graphs present the estimated values for measured and introduced BER over AWGN channel with GMSK modem (Figure 3), PESQ-MOS (Figure 4) and TES similarity score (Figure 5) regarding EbN0 (dB).

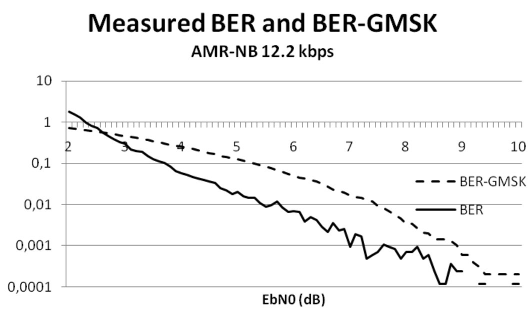


Figure 3: BER vs. EbN0 (dB).

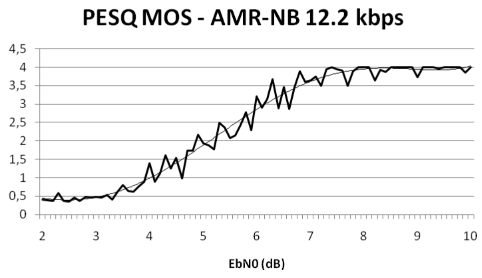


Figure 4: PESQ-MOS values vs. EbN0 (dB).

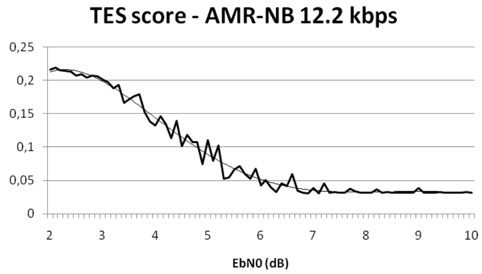


Figure 5: TES similarity score vs. EbN0 (dB).

It could be noticed that the observed values for similarity measure for TES, as well as PESQ, in case of AMR-NB (Figure 4 and 5) differs from their minimal and maximal values respectively even before introducing bit errors in the system. The reason is that AMR-NB is lossy coder and it degrades the signal even before the transmission process.

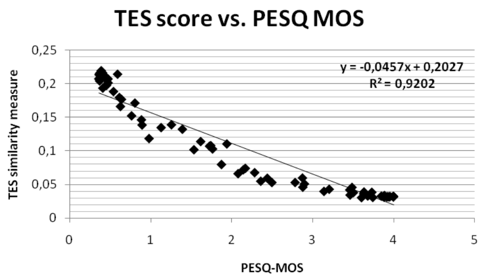


Figure 6: Regression analysis of measured MOS score (horizontal axis) and TES similarity score (vertical axis).

The results presented in Table 1 show that there is medium correlation between the TES measured similarity score, the measured MOS values against the introduced BER. The reason for that is the channel coding of the AMR-NB speech sequence, where in the presence of BER it is still able to fix some bit errors.

On the other side, the achieved correlation coefficient between the TES measured similarity values and the PESQ MOS score (R= -0,95925) is respectably higher compared to other well known objective speech quality measures: SNR, BSD, PAMS, MBSD, EMBSD, and comparable with: PSQM, PSQM+ and MNB2 [9]. That gives us an opportunity to estimate the linear regression parameters for particular codec and to use TES encoding and appropriate matching algorithm (which is much computationally efficient for long test sequences) for perceived speech quality assessment instead of PESQ based system.

| | *TES score* | *PESQ* |
|---|---|---|
| **Correlation with BER** | **0,693568** | **-0,5699** |

Table 1: Correlation coefficients vs. introduced BER.

## 5 Conclusions

This paper presents a method for intrusive procedure for speech quality estimation, where test and reference speech sequence were coded into TES S-matrices and their similarity is measured and scored. Several tests under various channel noise conditions have been made on reference speech sample simulating frame degradation effects on the perceived speech. Different values for similarity scores were produced after the test and received speech sequences comparison.

Achieved results were compared with measured values by PESQ MOS model and it has been shown that the TES-based measurement system correlates very well with MOS score. The measured difference scores justify the use of this technique as a simple and efficient tool for perceived speech quality measurement in GSM networks instead of basic model of Perceptual Evaluation of Speech Quality (PESQ), especially in cases of intrusive measurements with longer speech sequences.

## References


[1] P. Ji, B. Liu, D. Towsley, J. Kurose, "Modeling Frame-level Errors in GSM Wireless Channels", *IEEE Globecom, Internet Performance Symposium* 2002.

[2] King, R. A., "*Waveform Coding Method*". United States Patent No: US 6,748,354B1, June 2004

[3] Digital Cellular Telecommunications System (Phase 2+), Universal Mobile Telecommunications System (UMTS), AMR Speech Codec, 3GPP TS 26.071 Version 6.0.0 R6.

[4] Ivan Kraljevski, Slavcho Chungurski, Zoran Gacovski, Sime Arsenovski, "*Perceived Speech Quality Estimation Using DTW Algorithm*", 16th TELFOR, 25-27 November 2008, Belgrade, Serbia

[5] 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*", ITU Geneva (2001 Feb.)

[6] A. M. Law and W. D. Kelton, "*Simulation Modeling and Analysis*", 3rd ed. New York: McGraw-Hill, 2000

[7] Haykin, S. 2001: "*Communication Systems*", 4th ed. New York, NY. John Wiley & Sons.

[8] 3GPP TS 05.03 V8.9.0 (2005-01), Radio Access Network; Channel coding (Release 1999).

[9] S. Mohamed, G. Rubino, M. Varela. "*A method for quantitative evolution of audio quality over packet networks and its comparison with existing techniques*", in Measurement of Speech and Audio Quality in Networks (MESAQIN), 2004.