DOGS2010

DOGS2010 is the eighth conference offering an overview of current state and research directions in digital signal processing. Traditional conference topics are digital speech, image and biomedical signal processing.
The eighth DOGS will take place at the hotel Norcev at Iriški venac, one of the mountain tops of Fruška gora, from December 16th to 18th, 2010. The previous conferences took place in Novi Sad (1996), on Fruška gora (1998), in Novi Sad and Sremski Karlovci (2000), in the Dunđerski castle near Bečej (2002), in Sombor (2004), Vršac (2006) and Kelebija (2008).
This conference is unique within the boundaries of former Yugoslavia. Among its topics speech technologies traditionally stand out, being a very challenging problem highly dependent on language. Speech technologies help us to preserve our languages through application in various user services, and they are of immense importance to visually or hearing impaired people.

CONFERENCE TOPICS

Digital speech signal processing: speech generation and perception, phonology and phonetics, speech pathology, speech technologies (speech analysis and synthesis, speech recognition, speaker identification and verification), speech coding and transmission, speech cryptography techniques, noise reduction, analog and digital speech processing systems, implementation and applications (communications, computer telephony, etc.), and others.

Digital image processing: image coding and transmission, image analysis and segmentation, linear and non-linear image filtering and restauration, image modelling and representation, digital transformations, movement detection and estimation, implementation and applications (communications, multimedia, robotics, control, etc.), and others.

Digital Biomedical Signal Processing: linear and non-linear processing of 1-D biomedical (cardiovascular, neural and other) signals, decomposition, transformation, biomedical imaging, medical statistics.

CONFERENCE OVERVIEW

- Introductory invited papers
- Papers presenting an overview of current research (reviews of noteworthy papers published between two DOGSes)
- Original, previously unpublished papers
- Practical demonstrations of digital signal processing applications
- Round table discussions

DOGS is organized by the Faculty of Technical Sciences of Novi Sad, in cooperation with the Faculty of Electrical Engineering in Beograd, the Faculty of Electronics in Niš and Provincial Secretariate for Science and Technological Development of Vojvodina, under the auspices of the Ministry of Science and Technological Development of the Republic of Serbia and the IEEE Section of Serbia and Montenegro.

CONFERENCE COMMITTEE

Dragana Bajić, PhD
FTN Novi Sad

Vladimir Crnojević, PhD
FTN Novi Sad

Vlado Delić, PhD
FTN Novi Sad

Slobodan Jovičić, PhD
ETF Belgrade

Ljiljana Milić, PhD
Mihajlo Pupin Institute
Belgrade

Milan Milosavljević, PhD
ETF Belgrade

Vladimir Milošević, PhD
FTN Novi Sad

Milorad Obradović, PhD
FTN Novi Sad

Miodrag Popović, PhD
ETF Belgrade

Branimir Reljin, PhD
ETF Belgrade

Vidosav Stojanović, PhD
EF Niš

Vojin Šenk, PhD
FTN Novi Sad

Željen Trpovski, PhD
FTN Novi Sad

Jerneja Žganec Gros,

ORGANIZATION COMMITTEE

Dragana Bajić, PhD
FTN Novi Sad
draganab@uns.ac.rs

Vladimir Crnojević, PhD
FTN Novi Sad
crnojevic@uns.ac.rs

Vlado Delić, PhD
FTN Novi Sad
vdelic@uns.ac.rs

Nikša Jakovljević, MSc
FTN Novi Sad
jakovnik@uns.ac.rs

Vladimir Milošević, PhD
FTN Novi Sad
tlk_milos@uns.ac.rs

Milan Sečujski, PhD
FTN Novi Sad
secujski@uns.ac.rs

Vojin Šenk, PhD
FTN Novi Sad
vojin_senk@uns.ac.rs

Željen Trpovski, PhD
FTN Novi Sad
zeljen@uns.ac.rs

PhD
Alpineon, Ljubljana

Milan Sečujski, PhD
FTN Novi Sad

OFFICIAL LANGUAGES

The official languages of the conference will be Serbian and English. Papers must be written and abstracts presented in one of the two.

ATTENDANCE FEE

The conference attendance fee is 5000 RSD, and it covers:

- Participation in programme activities
- Cost of proceedings (hard-copy and CD-ROM)
- Extra-curricular activities

Attendance fees should be paid into the bank account that will be announced shortly.

INSTRUCTIONS FOR WRITING PAPERS

Papers should be submitted abridged (1 to 2 pages) and/or unabridged, according to the instructions for authors (in Serbian orEnglish language), in an electronic form to the e-mail address:

secujski@uns.ac.rs

PROCEEDINGS

Accepted papers will be published in the Proceedings, available as a Book and a CD-ROM. Both will be available at the conference.

DEADLINES

31.07.2010. Submission of abstracts
15.10.2010. Submission of unabridged papers (up to 4 pages; invited up to 8 pages)
15.11.2010. Acceptance information

dogs2010

digitalna obrada govora i slike
iriški venac, decembar 2010.

PROGRAM KONFERENCIJE

Četvrtak, 16. decembar

19.00 Registracija učesnika

20.00 Koktel dobrodošlice

Petak, 17. decembar

9.00 Sala A – Sesija A1 (Govor)

A1.1 ROLE OF PROSODY IN HUMAN-COMPUTER INTERACTION – Milana Bojanić, Vlado Delić

A1.2 PROSODIC CONSTITUENTS IN SERBIAN – Maja Marković, Tanja Milićev

A1.3 TRAJANJE GLASOVA I NAJUTICAJNIJI FAKTORI U SRPSKOM JEZIKU – Sandra Sovilj-Nikić

A1.4 ANALIZA TRAJANJA I INTENZITETA ZVUČNIH GLASOVA /dž, ž, r, l/ U TIPIČNOJ I ATIPIČNOJ PRODUKCIJI – Silvana Punišić, Slobodan Jovičić, Zorka Kašić, Slavica Golubović

A1.5 MODELOVANJE IZGOVORA AFRIKATA /c/ – Milan Vojnović, Miško Subotić

A1.6 MODELOVANJE ATIPIČNOG IZGOVORA AFRIKATA /c/ – Milan Vojnović, Silvana Punišić

9.00 Sala B – Sesija B1 (Biomedicinski signali)

B1.1 IDENTIFIKACIJA GOVORNIH MOŽDANIH ZONA FUNKCIONALNOM MAGNETNOM REZONANCOM – Olivera Šveljo, Katarina Koprivšek, Miloš Lučić, Mladen Prvulović, Branimir Reljin, Milka Ćulić

B1.2 ROBUST FEATURE-BASED REGISTRATION FOR CT-MR IMAGES – Nemir Ahmed Al-Azzawi, Wan Ahmed K. Wan Abdullah

B3.4 KORIŠĆENJE GENETSKOG PROGRAMIRANJA ZA DETEKCIJU POPLAVLJENOG POLJOPRIVREDNOG ZEMLJIŠTA – Predrag Lugonja, Nemanja Petrović, Dubravko Ćulibrk, Vladimir Crnojević

B3.5 KLASIFIKACIJA SLIKA ZASNOVANA NA ADABOOST ALGORITMU I STABLIMA ODLUKE SA HOG I LBP OBELEŽJIMA – Marko Panić, Predrag Lugonja, Dragan Letić, Dubravko Ćulibrk, Vladimir Crnojević

B3.6 PREPOZNAVANJE PEŠAKA ZASNOVANO NA HOG I LBP OBELEŽJIMA PRIMENOM RANDOM FOREST ALGORITMA – Dubravko Ćulibrk, Marko Panić, Dragan Letić, Predrag Lugonja, Vladimir Crnojević

B3.7 AUTOMATIZOVANI VIDEO NADZOR SAOBRAĆAJA KORIŠĆENJEM DETEKCIJE I PRAĆENJA POKRETNIH OBJEKATA – Dragan Letić, Branko Brkljač, Predrag Lugonja, Dubravko Ćulibrk, Vladimir Crnojević

B3.8 EYE LOCALIZATION USING CORRELATION FILTERS – Vitomir Štruc, Jerneja Žganec Gros, Nikola Pavešić

B3.9 APPLICATION OF THE PROGRESSIVE WAVELET CORRELATION TO CONTENT-BASED IMAGE RETRIEVING – Igor Stojanović, Ivan Kraljevski, Slavčo Čungurski

18.00 Sala A – Sesija A4 (Govor)

A4.1 KONSTRUKCIJA DEO PO DEO UNIFORMNOG KVANTIZERA I PRIMENA U KODOVANJU GOVORNOG SIGNALA – Zoran Perić, Jelena Nikolić, Aleksandra Jovanović

A4.2 KOMPRESIJA SA GUBICIMA I BEZ GUBITAKA GOVORNOG SIGNALA VISOKOG KVALITETA – Zoran Perić, Milan Savić, Milan Dinčić

A4.3 UTICAJ KARAKTERISTIKA AMBIJENTA NA KVALITET GOVORNOG SIGNALA – Petar Prokić, Slobodan Jovičić

A4.4 UTICAJ NAČINA UPOTREBE MOBILNOG TELEFONA NA FORMANTNE FREKVENCIJE – Nikola Jovanović, Slobodan Jovičić

A4.5 PRIMENA GOVORNIH TEHNOLOGIJA U ADAPTACIJI RAČUNARSKE IGRE LUGRAM ZA SLEPU I SLABOVIDU DECU – Branko Lučić, Nataša Vujnović Sedlar

A4.6 MODELING MACEDONIAN INTONATION FOR TEXT-TO-SPEECH SYNTHESIS – Branislav Gerazov, Zoran Ivanovski, Ružica Bilibajkić

A4.7 TIME ENCODED SIGNAL PROCESSING FOR SPEECH QUALITY ASSESSMENT – Ivan Kraljevski, Igor Stojanović, Slavčo Čungurski, Sime Arsenovski

A4.8 SPEECH SYNTHESIS OF DISSIMILAR LANGUAGES USING THEIR PHONETIC SUPERSET – Slavčo Čungurski, Ivan Kraljevski, Igor Stojanović, Blerta Prevalla

TIME ENCODED SIGNAL PROCESSING FOR SPEECH QUALITY ASSESSMENT

Ivan Kraljevski[1], Igor Stojanović[2], Slavčo Čungurski[1], Sime Arsenovski[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, chungurski, sime.arsenovski}@fon.edu.mk, igor.stojanovik@ugd.edu.mk

ABSTRACT

In this paper a method for speech quality assessment is described and 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. Link adaptation 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].

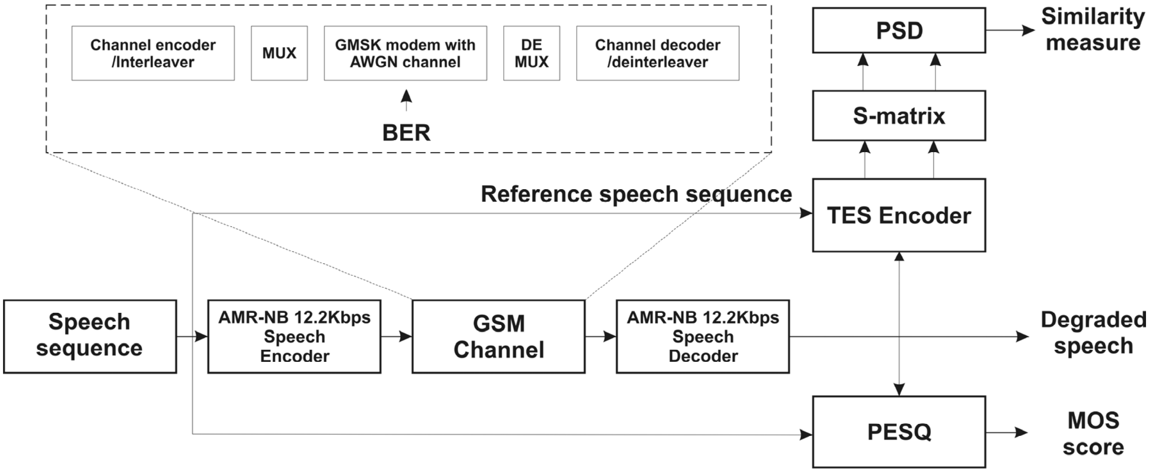
3.2. BER simulation

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.



Channel encoder /Interleaver
MUX
GMSK modem with AWGN channel
DE MUX
Channel decoder /deinterleaver
BER
PSD
Similarity measure
S-matrix
Reference speech sequence
TES Encoder
Speech sequence
AMR-NB 12.2Kbps Speech Encoder
GSM Channel
AMR-NB 12.2Kbps Speech Decoder
Degraded speech
PESQ
MOS score

Figure 1: Simulation system architecture

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) (Figure 2). Thus the signal is transformed into time encoded stream of discrete numerical descriptors – TES symbols.

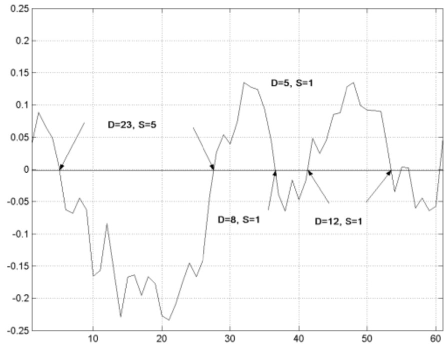


Figure 2: Signal waveform divided in TES epochs

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.

A histogram of the signal array with 28 possible symbolic descriptors can be produced, forming so called S-matrix with fixed dimension 1x28 (Figure 3), which carry information about symbols frequency.

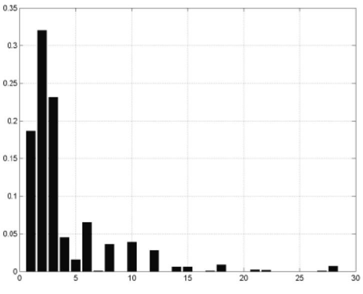


Figure 3: S-matrix of reference speech sequence.

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

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

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 4).

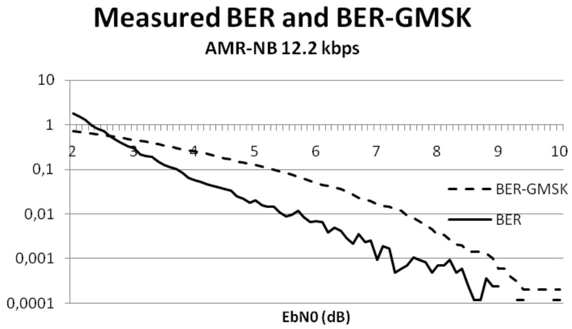


Figure 4: BER vs. EbN0 (dB).

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.

Figure 4 presents the estimated values for measured and introduced BER over AWGN channel with GMSK modem, PESQ-MOS (Figure 5) and TES similarity score (Figure 6) regarding EbN0 (dB).

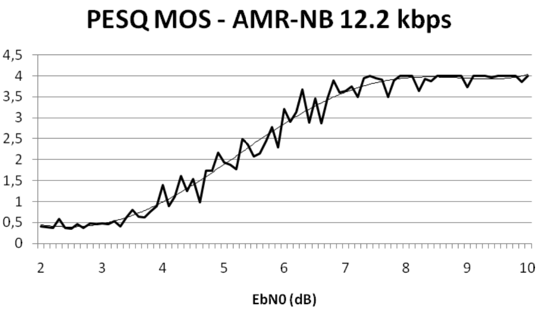


Figure 5: PESQ-MOS values vs. EbN0 (dB).

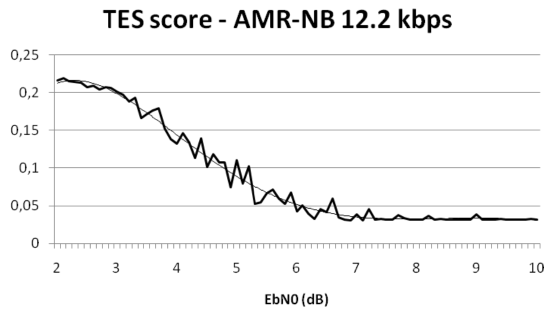


Figure 6: TES similarity score vs. EbN0 (dB).

It could be noticed that the observed values for similarity measure for TES, as well as PESQ, in case of AMR-NB (Figure 5 and 6) 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.

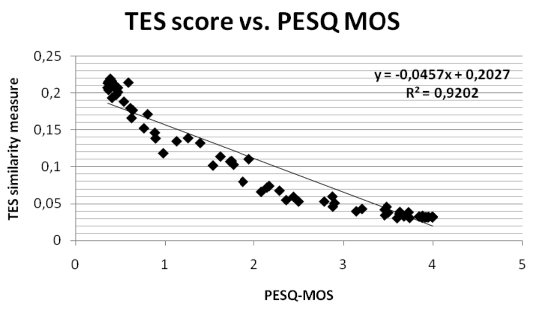


Figure 7: Regression analysis of measured MOS score (horizontal axis) and TES score (vertical axis).

The results show that there is medium correlation between the TES measured similarity score (0,6936), the measured MOS values (-0,5699) 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.

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.

6. REFERENCES

[1] P. Ji, B. Liu, D. Towsley, J. Kurose, “Modeling Frame-level Errors in GSM Wireless Channels”, *IEEE Globecom, Internet Performance Symp.* 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+), UMTS, AMR Speech Codec, 3GPP TS 26.071 Version 6.0.0 R6.

[4] Kraljevski *at al.*: *“Perceived Speech Quality Estimation Using DTW Algorithm“,*Telfor Journal, Vol. 1, No. 1, 2009.

[5] ITU-T Rec. P.862, 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.