Speech quality dependence investigation from transmission channel characteristics and voice encoding methods

Snejana G. Pleshkova, Kalina Hr. Peeva

This article examine the dependence of speech quality from transmission channel characteristics and speech encoding methods. As speech quality estimation is proposed to use the objective methods for quality control and evaluation of reproduced speech signals transmitted in voice audio communications systems. Also the models of human auditory perception are applied as a tool for current control and further improvement of the adopted speech quality. Especially, as a basis algorithm for speech quality control simulation in this paper is proposes to use the widespread ITU-R standard for objective measurement of perceived audio quality through models of human auditory system. The simulation model create in this article is the improved version of the existing in the mentioned standard method for objective measurement of perceived audio. The developed algorithm and simulation models can be used in voice communication systems like GSM, VoIP, DVB, etc. In each of these concrete voice communication systems are applied different methods for source and channel coding and they are examined in respect to the perceived audio information quality against negative factors such as the occurrence of signal distortions and noise in the communication channel within the chosen type of voice audio system. Presented are the quantitative experimental results of the analysis for the reliability and the effectiveness of the applied algorithm, obtained during the simulation.

Изследване на зависимостта на качеството на говора от характеристиките на каналите за предаване и методите за кодиране на говор (Снежана Плешкова, Калина Пеева). *Тази статия изследва зависимостта на качеството на говора от параметрите на канала за предаване и методите за кодиране на говор. За оценка на качеството на говора е предложено използването на обективни методи за контрол на качеството и оценка на възпроизведените говорни сигнали, предавани в системите за гласови комуникации. Също така се прилагат модели на слуховото възприятие на човека като средство за текущ контрол и следващо подобряване на качеството на приетия говор. По-точно, като основен алгоритъм за симулация на оценката на качеството на говора е предложено да се използва широко разпространения стандарт ITU-R за обективна оценка на качеството на приетия говор чрез моделиране на слуховата система на човека. Симулационният модел създаден в тази статия е подобрена версия на споменатия по-горе стандартен метод за обективна оценка на качество на приетия говор. Разработеният алгоритъм и симулационен модел може да се използва в системи за гласови комуникации от типа на GSM, VoIP, DVB и др. Във всяка от тези конкретни комуникационни системи се прилагат различни методи за кодиране на източника и канала и те са изследвани по отношение на качеството на приетата говорна информация спрямо отрицателни фактори, като наличие на изкривявания и шумове в комуникационния канал при съответно избран тип на аудио система. В статията са представени в количествена форма форма експериментални резултати от анализа на надежността и ефективността на приложените алгоритми, получени при симулацията.*

I. Introduction

Audio quality is of leading significance in designing and realization of voice communication and multimedia systems [1]. The rapid development of audio systems with different rates of transmission, however, led to the development of methods with different accuracy in evaluation and control of perceived quality. Depending on the nature of speech signals in widespread communication systems are applied two popular methods of quality control, according to the particular encoding method for voice signals, and the specific channel characteristics:

-Subjective Listening tests [2];

- Objective Measurements of perceived audio quality [3].

The specific channel characteristics and coding standards for speech information will be reflected in the simulation model by specifying the appropriate parameters of the communication system during experimental verification of the proposed algorithm and simulation program, regardless of the improved version of objective measurement of perceived audio quality, by model of auditory perception. The purpose of this article is to develop algorithm and simulation models which could be used in the field of professional measurements or in student education to study and improve the quality of speech signals transmitted over communication systems in real time.

One of the goals of the simulation program is the creation of a suitable functional scheme for simulation model, which will by applicable for practical realization in the specific science-oriented Matlab Sumulink space.

II. Auditory perception model for speech quality control in voice communication system

The developed algorithm is focused and describes the preparation of simulation model for analysis of the quality of reproduced speech signals in real time in consistence with the improved version of the established method for objective measurements of perceived audio quality, which apply the model of auditory perception.

In this article, will be used the following generalized block diagram (Fig. 1), including the main blocks of the model of human auditory perception [4].

The given test signals represent a natural speech signals of duration in the range of 5-10 seconds, as the applied Signal Under Test and Referent Signal are synchronized in time. Their length could be limited to a very short interval, since it does not be expected that these signals will be used in subjective listening tests.

III. Output parameters from speech quality estimation with auditory perception model

The output parameters are designed in accordance with ITU-T Recommendation P.835 [5] and methodology, which are designed to achieve the reduction of auditory uncertainty against speech signal distortion, background noise distortions or both, and thus form the basis for overall quality assessment.

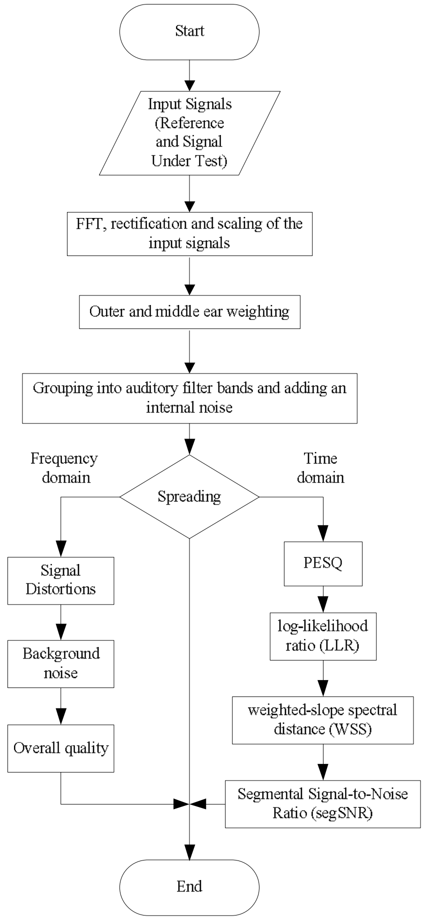
Start
Input Signals (Reference and Signal Under Test)
FFT, rectification and scaling of the input signals
Outer and middle ear weighting
Grouping into auditory filter bands and adding an internal noise
Frequency domain
Spreading
Time domain
Signal Distortions
Background noise
Overall quality
PESQ
log-likelihood ratio (LLR)
weighted-slope spectral distance (WSS)
Segmental Signal-to-Noise Ratio (segSNR)
End

Fig.1. Functional block scheme of model of peripheral auditory perception

The methodology successfully allows assessing and controlling the quality of enhanced speech signals by measurements of the speech signal alone - using a five-point scale of signal distortion (SIG), the background noise alone - using a five-point scale of background intrusiveness (BAK) and by measuring the overall quality using the scale of the mean opinion score (OVRL) [1=bad, 2=poor, 3=fair, 4=good, 5=excellent].

The model used LPC-based objective measure by log-likelihood ratio (LLR), described with the following equation:

$$(1) \qquad d_{LLR}(\mathbf{a}_p, \mathbf{a}_c) = \log(\frac{\mathbf{a}_p \mathrm{R}_c \mathbf{a}_p{}^T}{\mathbf{a}_c \mathrm{R}_c \mathbf{a}_c{}^T}),$$

where $\boldsymbol{a}_c$ is the LPC-based vector of the original signal fragment, $\boldsymbol{a}_p$ is the LPC-based vector of the enhanced speech fragment and R_c is the autocorrelation matrix of the original speech matrix. The segmental LLR values were limited in the range of (0, 2) to further reduce the number of outliers.

The weighted-slope spectral distance (WSS) measure computes the weighted difference between the spectral slopes in each frequency band, obtained as the difference between adjacent spectral magnitudes in dB. The WSS measure evaluated in this paper is defined as:

$$(2) \qquad dwss = \frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{j-1}^{K} W(j,m)(S_c(j,m)-S_p(j,m))^2}{\sum_{j=1}^{K} W(j,m)},$$

where $W(j,m)$ are the weighted spectral slopes for the frequency band at fragment of the clean and processed speech signals. The number of bands was set to K=25. The time-domain segmental Signal-to-Noise Ratio (segSNR) measure, instead of working on the whole signal, computes the average of the SNR values of short segments (15 to 20 ms) and is widely used for evaluating the performance of speech enhancement algorithms. It is described by the equation:

$$(3) \qquad SNRseg = \frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\sum_{i-Nm}^{Nm+N-1}\left(\frac{\sum_{i=1}^{N} x^2(i)}{\sum_{i=1}^{N}(x(i)-y(i))^2}\right),$$

where $x(i)$ and $y(i)$ are the clean and processed speech samples indexed by i, and N and M are the total number of samples and segment length.

The PESQ measure is the most complex in the objective speech measures and is the one recommended by ITU-T [6] for speech quality assessment of 3.2 kHz handset telephony and narrow-band speech codecs. The PESQ score is computed as a linear combination of the average disturbance value D_{ind} and the average asymmetrical disturbance values A_{ind} as:

$$(4) \qquad \mathrm{PESQ} = a_0 + a_1 D_{\mathrm{ind}} a_2 A_{\mathrm{ind}},$$

where a_0 = 4.5, a_1 = - 0.1, a_2 = - 0.0309. The parameters a_0, a_1 and a_2 were optimized for speech processed through networks and not for speech enhanced by noise suppression algorithms.

IV. Transmission and receiving parts of simulation model for speech quality investigation and control

In Fig.2 is presented the transmission part of the proposed simulation model to prepare various scenarios of speech quality analysis, estimation and control in voice communication systems.

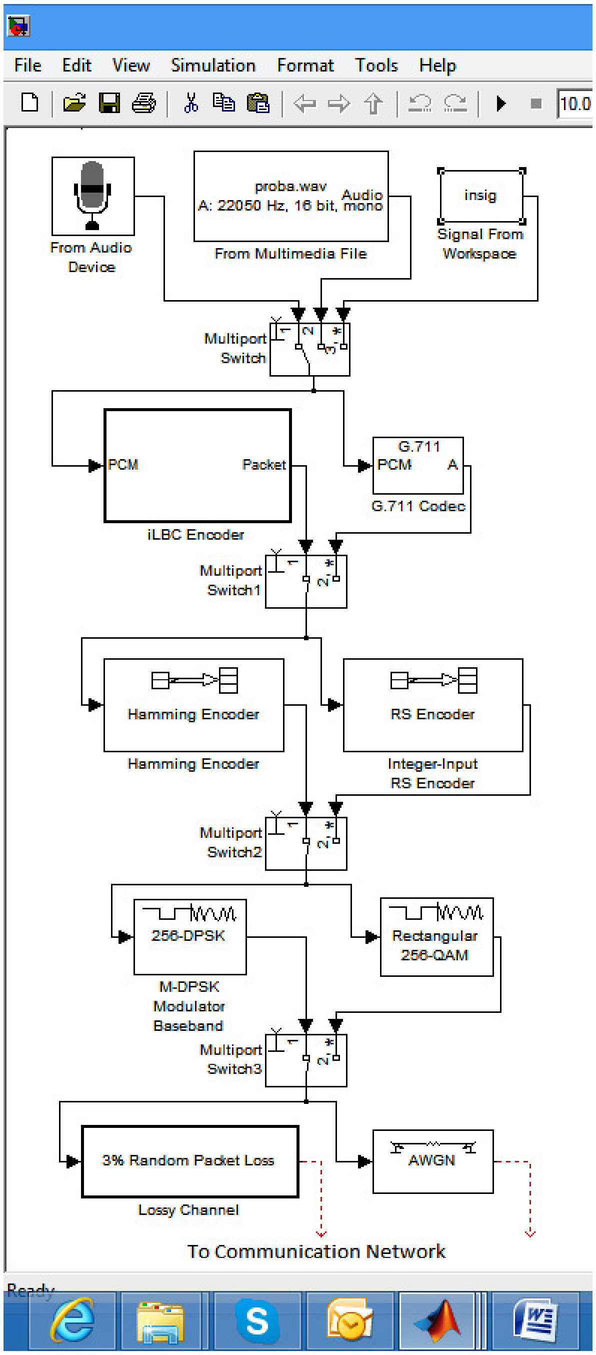

Fig.2. Transmission part of the proposed simulation model to prepare various scenarios of speech quality analysis, estimation and control in voice communication system

Similarly, on Fig.3 is presented the receiving part of the proposed simulation model to prepare various scenarios of speech quality analysis, estimation and control in voice communication systems.

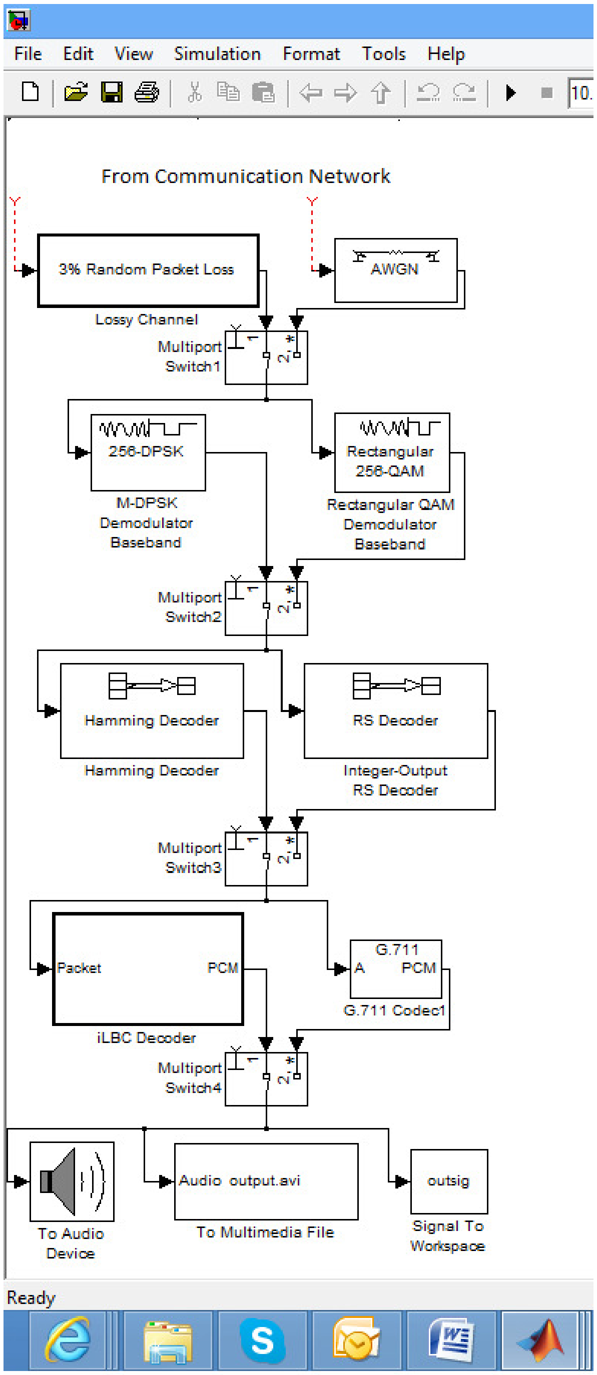

Fig.3. Receiving part of the proposed simulation model to prepare various scenarios of speech quality analysis, estimation and control in voice communication systems.

There are many cases of using the simulation models for modeling and testing the appropriate systems. The main advantages of the proposed simulation model are that it offers the possibilities to carry out the simulations choosing different encoding methods and channel characteristics, shown in Fig.2 and Fig.3 as appropriate functional blocks existing in Matlab Simulink Communications System Toolbox. Along with this, the transmission and receiving part of this simulation model provided the ability to select the speech signal sources, type of modulation, the communication channel type and noise characteristics, etc.

Here also there are offered the numerous kinds of choosing different decoding and demodulation methods, like functional blocks from Matlab Simulink Communications System Toolbox, corresponding to the encoding and modulation methods.

The receiving part of the simulation model provided the ability to select three different methods of treating the received and decoded speech signals, before their quality analysis or estimation: listening, saving in an audio file or exporting the speech signal to Matlab Workspace.

The proposed on Fig.2 and Fig.3 simulation models of transmission and receiving parts of a voice communication system is executed for simulation of speech transmission using various selections of encoding/decoding methods and channel character-ristics. The results of the received speech signals are collected in the appropriate audio files. These files are used in the next step for objective comparison with the corresponding original speech files to prepare the suitable quality estimation of the received speech signals using the above described auditory perception model (Fig.1.) for speech quality control in voice communication systems.

V. Experimental results from simulation

With the described above voice communication system simulation model (Fig.2 and Fig.3) are prepared numerous experiments using the presented in Fig.1 functional block scheme of model of peripheral auditory perception as a tool for received speech quality estimation and analysis. In this article are presented only a part of these experiments to show the main results achieved with the proposed objective method for speech quality estimation in voice communication systems.

The experimental results from the control of the quality of speech audio signals, transmitted over the proposed model of voice communication channel (Fig.2 and Fig.3) and affected by different levels of channel packet loss or channel noise, are illustrated ellow in the following tables (Tab.1 to Tab.5), where the values of the output parameters from speech quality estimation with auditory perception model

(Fig.1.) PESQ_MOS, LLR, SNRseg, WSS, PESQ, Csig, Cbak, Covl are used.

The original speech signals under tests are submitted using different variants for source and channel coding methods and also different models of communication channels. It is possible to set and change the channel distortions and error levels, in order to observe different scenarios with the following defined real channel characteristics:

- known packet loss (3%, 10%, 15%, 30%) in the first case (Tab.1 and Tab.2) and for two speech coding methods Internet Low Bitrate Codec (iLBC - Tab.1) [6] and G.711 Codec (Tab.2) [7];
- known error to noise ratio *Eb/No* values (10dB, 20dB, 30dB, 40dB) for a chosen communication channel simulation model (AWGN) and for G.711 Codec (Tab.3).

Table 1

Experimental results of speech quality estimation in voice communication system with iLBR speech coding method

iLBR Codec	Channel packet loss 3%	Channel packet loss 10%	Channel packet loss 15%	Channel packet loss 30%
PESQ_MOS	0.449	0.419	0.428	0.399
LLR	0.222711	0.222711	0.236564	0.281204
SNRseg	-1.813491	-1.820839	-1.859493	-1.932785
WSS	33.75907	37.56839	39.24336	48.93606
PESQ	0.449458	0.419038	0.428075	0.399475
Csig	2.8603	2.7784	2.7545	2.6041
Cbak	1.498	1.4566	1.4468	1.3606
Covl	1.6200	1.5543	1.5428	1.4290

Table 2

Experimental results of speech quality estimation in voice communication system with G.711 speech coding method

G.711 Codec	Channel packet loss 3%	Channel packet loss 10%	Channel packet loss 15%	Channel packet loss 30%
PESQ_MOS	0.183	0.163	0.131	0.097
LLR	0.251760	0.27396	0.304029	0.38377
SNRseg	13.067487	12.48326	10.82266	7.28556
WSS	20.024764	22.29775	28.20228	28.2022
PESQ	0.182779	0.163043	0.130813	0.09683
Csig	2.7639	2.7087	2.6052	2.3632
Cbak	2.4044	2.3423	2.1809	1.8334
Covl	1.4721	1.4289	1.3462	1.1696

The results from experiments of speech quality estimation with various methods of modulation used in voice communication systems are presented here in Tab.4 and Tab.5 with applying the most popular base band modulation models QAM and QPSK, respectively.

Table 3

Experimental results of speech quality estimation in voice communication system with G.711 speech coding method and AWGN channel

G.711 Codec and AWGN channel	Channel packet loss 3%	Channel packet loss 10%	Channel packet loss 15%	Channel packet loss 30%
PESQ_MOS	2.275	0.033	0.469	0.085
LLR	3.942798	2.615075	2.192947	1.238560
SNRseg	-1.038056	0.296460	1.130196	4.791768
WSS	93.84945	53.62838	45.79661	29.429517
PESQ	2.275327	0.033121	0.469261	0.085274
Csig	-0.4368	-0.0606	0.7073	1.6051
Cbak	1.9993	1.2931	1.6089	1.7706
Covl	0.7500	-0.0937	0.5284	0.8225

Table 4

Experimental results of speech quality estimation in voice communication system with QPSK modulation and AWGN channel

QPSK modulation and AWGN channel	Signal to Noise Ratio (SNR) 10 dB	Signal to Noise Ratio (SNR) 20 dB	Signal to Noise Ratio (SNR) 30 dB	Signal to Noise Ratio (SNR) 40 dB
PESQ_MOS	-0.620	2.524	0.284	0.284
LLR	3.20919	4.13938	4.14872	4.14872
SNRseg	-10.0000	-10.000	-10.0000	-10.0000
WSS	77.3486	88.9560	89.1453	89.1453
PESQ	-0.6203	2.5238	0.2838	0.2838
Csig	-1.2795	-0.4452	-1.8072	-1.8072
Cbak	0.1660	1.5877	0.5156	0.5156
Covl	-1.0899	0.8836	-0.9257	-0.9257

The values of composite measures Csig, Cbak and Covl in all tables (Tab.1 to Tab.5) present the absolute difference between the two signals original and signal under test, where the ‘0’ value mean a full accordance and the positive values indicate all other transmission differences in the estimated signal components.

There are also calculated and listed bellow:
PESQ_MOS = 4.500; LLR = 0.0000;
SNRseg = 35.0000; WSS = 0.0000; PESQ = 4.5000;
Csig = 5.8065; Cbak = 5.9900; Covl = 5.2165,
the same objective characteristics PESQ_MOS, LLR, SNRseg, WSS, PESQ, Csig, Cbak, Covl used in the experiments, but for ideal case, where the original and the referent signal are the same. These values are used as reference values for comparison with all other cases presented as results in tables (Tab.1 to Tab.5) of

received speech signals and for analysis and conclusion about speech quality estimation.

Table 5

Experimental results of speech quality estimation in voice communication system with QAM modulation and AWGN channel

QAM modulation and AWGN channel	Signal to Noise Ratio (SNR) 10 dB	Signal to Noise Ratio (SNR) 20 dB	Signal to Noise Ratio (SNR) 30 dB	Signal to Noise Ratio (SNR) 40 dB
PESQ_MOS	0.636	1.317	1.552	2.085
LLR	3.96079	3.25434	3.25434	0.63729
SNRseg	3.83427	20.5648	11.3469	31.3387
WSS	62.61185	42.40131	42.40131	27.46816
PESQ	0.635728	1.551920	1.551920	2.085474
Csig	-1.1628	1.7777	1.7777	3.4475
Cbak	1.7412	3.3746	3.3746	4.4129
Covl	-0.3604	1.6163	1.6163	2.7542

VI. Conclusion

In conclusion it is possible to summarize the main results in the following way:

- the values of the objective quality parameter WSS is biger for encoding methods G.711;
- method has little value of WSS in comparison of encoding method iLBR, which means the best speech quality of encoding methods G.711;
- the values of the objective quality parameter LLR have the small difference from the zero reference is smaller for the encoding method iLBR in comparison of encoding methods G.711;
- the method G.711 is more efficient than the method iLBR, because of the achieved better quality of speech audio signals;
- it is possible to use the maximal values of WSS parameters as criteria of the degradations in the received speech signals;
- the small or minimal values of LLR parameter, also can be used for small and acceptable criteria degradation.

In comparative evaluation of speech quality, obtained after applying an appropriate QPSK and QAM modulations, signal transmitted through a noisy AWGN channel is observed more efficiency and higher degree of accuracy in QAM modulation. As a result of QPSK modulation, the most degraded variation is observed in the parameters WSS and SNRseg, while the highest similarity with the absolute value - in LLR parameters.

The values of the experimental measurements in the QAM modulation are mainly similar, where the most degraded value is the WSS parameter, and the most similar to the absolute value - LLR and PESQ_MOS.

Acknowledgements

This paper is supported by Technical University – Sofia inner program for PhD research projects under Contract 132 PD 0025-07: “Development of algorithms for quality estimation of audio-visual information in multimedia computer systems and networks”.

REFERENCES

[1] Sat B., B. W. Wah, “Playout scheduling and loss concealments in VoIP for optimizing conversational voice communication quality,” in Proc. ACM Multimedia, Augsburg, Germany, Sept. 2007, pp. 137– 146.

[2] http://www.itu.int/rec/T-REC-P.800-199608-I/

[3] ITU-T Recommendation P.861, "Objective quality measurement of telephone-band (300–3400 Hz) speech codecs," International Telecommunications Union, Geneva, Switzerland,1996.

[4] Rix A., M. Hollier, A. Hekstra, and J. Beerends, “Perceptual evaluation of speech quality (PESQ),” J. Audio Engineering Society, 2002, vol. 50, pp. 755–764.

[5] http://www.itu.int/rec/T-REC-P.835

[6] http://www.ilbcfreeware.org/

[7] http://www.itu.int/rec/T-REC-G.711

Assoc. Prof. Dr. Snejana G. Pleshkova *- was born in 1964, Dobritch, Bulgaria. She received M. S. degree in Electronic and Automatic engineering from the Technical University-Sofia, Bulgaria in 1989. She is associate professor in the University from 2009. She published over 80 research papers. Her research interests are in the field of Neural Networks, Signal Processing and Pattern Recognitions, Digital Signal Processors for Image and Audio.*

tel.: 0035929653300 e-mail: snegpl@tu-sofia.bg

Eng. Kalina Hr.Peeva *was born in 1980, Varna, Bulgaria. She received M. S. degree in Computer engineering from the French Department of Information Technologies in Technical University, Sofia, in 2005. Her scientific research interests are in the field of Video and Audio Signal Processing, Digital Television, Communication and Internet Networks, Programming Language Matlab.*

tel.: 0035929653300 e-mail: kala_peeva@yahoo.com