Matlab learning system for speech coding education

Snejana G. Pleshkova-Bekyarska[1] and Michail B. Momcshedjikov[2]

Abstract— **The student speech coding education cover the theoretical study of some of the most important methods in coding of speech signals and also the practical works in which the students obtain the real idea of the characteristics, advantages, disadvantages, applicability the quality of restored speech signals etc., for the learned methods. The complexity of these methods not always permit to describe of the students the whole real working system for speech coding. In the practical works it is possibly to simulate some of the most important parts of each method. This is very suitable because it give to the students the possibility to work personally and to prove and test the theoretical knowledge for the speech coding methods , designing the simulating programs for all or for the chosen part of the algorithm of the speech coding method.**

This article describe the practical education of students in speech coding. It is chosen as a simulation environment Matlab, the wide spread system for mathematical simulation in the science and education. It is very easy in Matlab to describe in natural mathematics expressions the algorithm for each method and to use the embedded functions (Toolbox) for signal processing. This is very important for the students, because of easy modeling of speech coding methods.

Keywords— **Speech analysis, speech coding, speech synthesis, speech coding distortions**

I. Introduction

The Matlab modeling of the speech coding methods give the possibility to have the information for the current and the finale results in each step of algorithm. This give to the students the impression for the quality and the values of the real characteristics of the speech coding method, for the precision of restored speech signal, the errors noise etc. All of these possibilities can be done in Matlab, which have also the system of graphic functions allowing to visualization of speech coded signals and their characteristics.

The system Matlab is build mostly for modeling and it is not possibly to guarantee, that the coding and decoding of speech signals can be done in real time. Therefore the students design their programs for speech coding working with the preliminary saved in the format WAVE speech signals. It is possible to save also the intermediate results in a file and the end coded or restored speech signals, to use the post processing as a visualization or filtering or to listen the original, coded, restored speech in real time for the comparing or estimating subjectively the quality of the restored speech signals.

[1]Snejana G. Pleshkova-Bekyarska is with the Faculty of Communications and Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: snegpl@vmei.acad.bg

[2]Michael B. Momschedjikov is with the Faculty of Communications and Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: snegpl@vmei.acad.bg

The subject of each practical work for the students is in concordance with the theoretical course for speech coding methods, and the practical works are placed in the time and sequence in a manner, that the students became at the end of the practical works with the full working algorithm for coding and decoding of speech signal. The number of practical works is in concordance with the plan of education. It is possible to make modification, to add some new methods and algorithms for speech signal coding. It it possible to use there practical works as a basis for other courses in the learning plan, for example for speech signals recognition.

The fundamental content of the practical works for simulation of speech coding methods is some characteristics of speech signals: energy, segmentation of speech coding methods is some basis characteristics of speech signals: energy segmentation of speech signals on voice/unvoice parts, the pitch detection etc. In the next practical works the student work with the embedded Matlab functions for making the programs for some basis methods in speech coding such as linear prediction coding (LPC) [1], short time Fourier transform (STFT) [2].

The programs for the basic characteristics of speech signals are used in the next practical works for modeling of a concrete method for speech coding. On of these method is the Multi Band Excitation (MBE). In this method it is used for coding of voiced parts of the speech the pitch frequency, the harmonics of pitch frequency and the envelope of the speech signal spectrum. In the next practical work the students used the results from the coding - the coding characteristics and design the program for synthesis (decoding) of voiced speech signal.

In the places, where there are the unvoiced speech segments, the method used for coding is linear prediction filter. The separation of speech signal on voiced/unvoiced parts is made from the logical signal defined in the practical works for basic characteristics of speech signal. The restored speech signal can be represented in form of graphic for visual appreciation or can be saved as a file in WAVE format. This file can be reproducing in real time for listening estimation of restored speech quality.

II. Brief description of some of the practical works

Each practical work can be represented in briefly as mathematical expression of algorithm for modeling of the corresponding method or as graphics of some more important results.

The knowledge of the energy of speech signal is very important for all speech coding methods. The modeling in Matlab in first practical work is made for the mathematical expression of energy for non stationary signals:

$$E(n) = \sum_{m=0}^{N-1} \left[W(m)\,x(n-m)\right]^2 , \qquad (1)$$

where:

W(m) is the window function for m samples of a speech frame;

x(n-m) – current values of speech signal in the frame;

Fig.1 show the results of Matlab program for energy of speech signal calculation with expression (1).

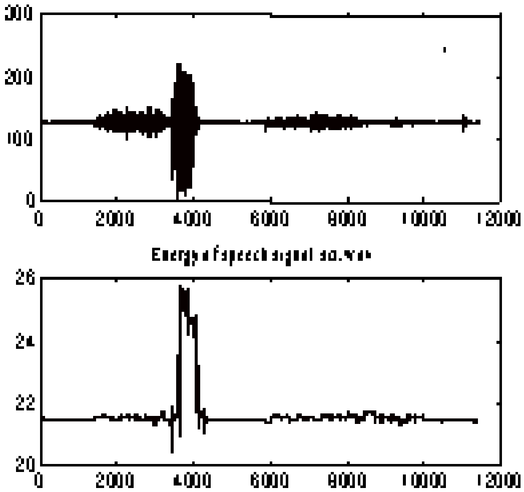

Fig. 1. Speech signal six.wav and the energy of speech signal six.wav.

The values of energy E(n) can be used for voiced/unvoiced segmentation of speech signal. For this reason the values of energy are compared with a preliminarily defined threshold *tvnv* and it is build a logic signal *vnv* equal to logic one, when the current value of energy is above the threshold or equal to logic zero, when the current value of energy is below the threshold:

$$vnv(n) = \begin{cases} 1 & for \quad E(n) \geq tvnv \\ 0 & for \quad E(n) < tvnv \end{cases} \tag{2}$$

The pitch frequency of voiced speech signals one of the basic characteristics of speech signal, which is used in most of the speech coding methods. Therefore the second practical work is for modeling with Matlab of some auto correlation algorithms in time or frequency domain for pitch frequency determination:

$$R(\tau) = \sum_{n=0}^{N-1} S(n) S(n+\tau), \tag{3}$$

where:

R(τ) is auto correlation function of speech signal S(n) in time domain.

There is a similar expression for auto correlation function of speech signal in frequency domain.

The results of modeling are shown in Fig.2.

The methods of Linear Prediction Coding (LPC) and Short Time Fourier Transform (STFT) are also very useful in many of speech coding methods. With these methods it is possible to find the short time prediction and short time spectrum of a frame a speech signal, where it is possible to suppose that the speech signal is stationary. The linear prediction coefficients and the envelope of the spectrum are closely joined with the characteristics of the vocal tract. In the third practical work, which is for

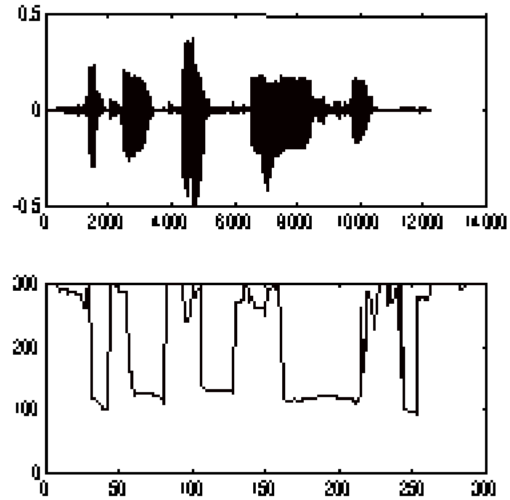

Fig. 2. Speech signal sent.wav and the Pitch frequency in Hz.

modeling of these characteristics, are used the following expressions respectively for linear prediction and short-time Fourier transform:

$$_n = S_n + \sum_{i=1}^{n} a_i S_{n-i} \tag{4}$$

where:

e_n is the residual signal;

S_n and S_{n-i} – current and preceding values of speech signal;

a_i – the linear prediction coefficients;

$$S(n,\omega) = \sum_{m=-\infty}^{+\infty} S(m).W(n-m).\ell^{-j\omega m}, \tag{5}$$

where:

n - the number of harmonic in Fourier transform;

W(n-m) is the window function, which define the short time interval of analysis;

S(m) – the current speech samples.

Fig.3 and Fig.4 show the results from the Matlab modeling of linear prediction of speech signal and short-time Fourier transform spectrum of a frame of speech signal.

The next practical work is for the Multi Band Exitation coding (MBE) [3]. This is a method of separating the speech signal spectrum in some numbers of frequency bands, each of which is determined and set as voiced or unvoiced. The voiced frequency band are represent and restored with the pitch frequency harmonics values. The unvoiced frequency band are synthesized in decoder with the random noise signals. The students used the Matlab functions and the program for Short Time Fourier Transform (STFT) and work for modeling of multi band exitation (MBE) method of speech coding. The results of this modeling as graphical representation of input signal, spectrum and restored speech signal is shown in Fig.5.

The multi band exitation (MBE) can the use the spectral envelope as a characteristics to improve the quality of the restored

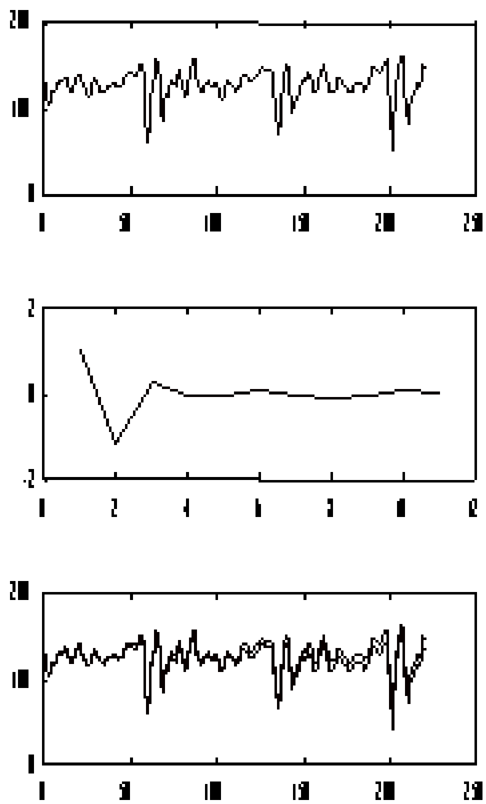

Fig. 3. Current fragment of input signal sa.wav, LPC coefficients for current fragment and the input and restored fragment.

speech signal. The five practical work is for modeling in Matlab of the algorithm for spectral envelope determination and the usage of the spectral envelope to a more precise speech signal restoration. The Fig.6 show the input speech signal and the simulations representation of spectrum of input signal and spectrum of restored speech signal with or without spectral envelope.

The graphic, which is most closed to the spectrum of input signal is for restored speech signal with spectral envelope. It is shown also in Fig.6 with the circles the pitch harmonic amplitudes.

The next practical work is a final modeling of basic characteristics pitch, pitch harmonics and spectral envelope to make the synthesis of speech signal and to compare the results of the different methods (with and without using the spectral envelope) and with the different degree of quantization of speech characteristics. Fig.7 shown the graphic representation of input speech signal and the restored voiced part of the speech signal.

These is a option in the program for this practical work for saving in a WAVE format the synthesized (restored) speech signal. The student can compare the listening quality of the original and the restored speech signal.

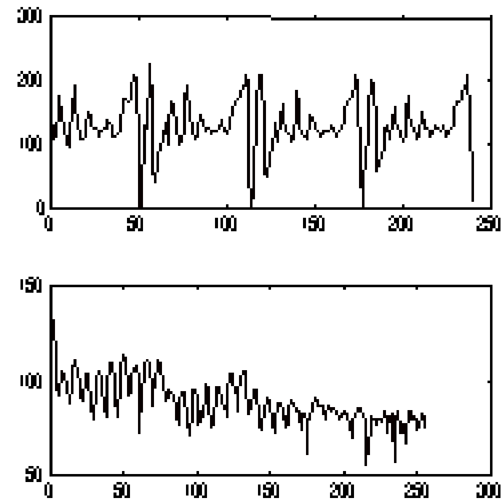

Fig. 4. Current fragment of input signal sa.wav and short time spectrum of current fragment.

III. Conclusion

It is shown in this article that the practical works for a student learning system in speech coding can be made as a simulation in Matlab of the most useful speech coding methods and to give to the students a real imagination of the advantages and difficulties in speech coding.

References

[1] Kondoz A.M, *Digital Speech*, John Wiley and Sons LTD, New Jork, 1994.

[2] Griffin D.W. and J.S.Lim, "Multi-Band Exitation Vocoder", *IEEE Transactions an ASSP*, 36(**??**) August 1988, pp.664-678.

[3] Spanias A.S, "Perceptual Coding of digital Audio", *Proceedings of the IEEE*, Volume 88, No4, April 2000, pp. 449-513.

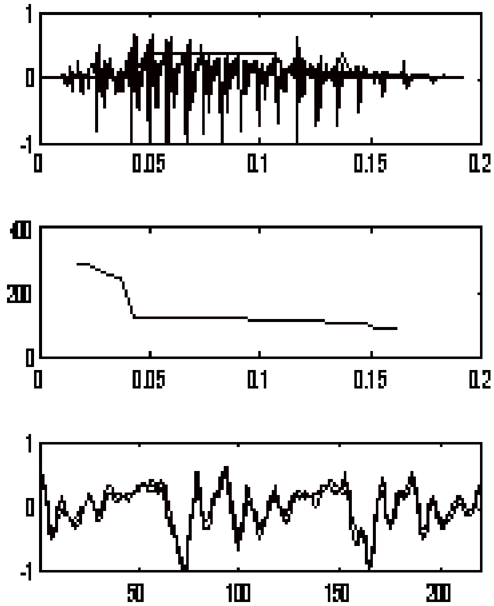

Fig. 5. Input signal sa.wav, spectrum and restored speech signal.

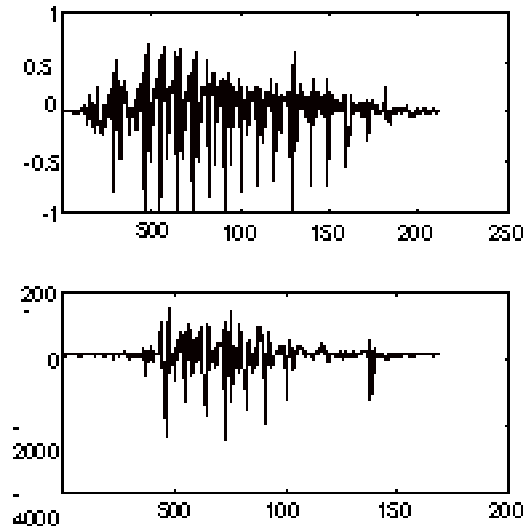

Fig. 7. Input signal and the voiced part of restored signal.

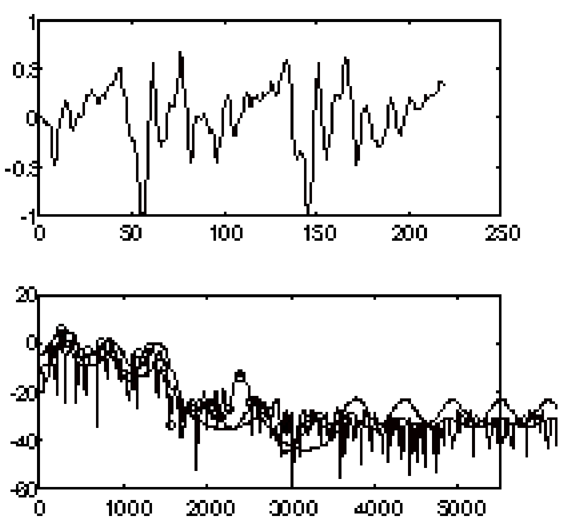

Fig. 6. Input signal and simulations spectrum.