A TOOL FOR TEACHING LINEAR PREDICTIVE CODING

Branislav Gerazov[1], Venceslav Kafedziski[2], Goce Shutinoski[1]

1) Department of Electronics, 2) Department of Telecommunications
Faculty of Electrical Engineering and Information Technologies, University of "Ss. Cyril and Methodius",
Karpos II B.B., 1000 Skopje, Republic of Macedonia, phone: +389 2 3062 224,
E-mail: {gerazov ; kafedzi; sugo}@feit.ukim.edu.mk

We introduce an educational tool for teaching undergraduate students Linear Predictive Coding (LPC). The tool is a fully functional coder-decoder with an intuitive graphical user interface. Students can easily change the parameters involved in LPC and hear the results immediately. Plots of some of the parameters are also given to further illustrate the encoding/decoding process. The inclusion of the tool in laboratory classes resulted in an increased interest for the subject of LPC, providing students with hands-on experience by tuning the LPC codec.

Keywords: Linear Predictive Coding, Audio, Education, Laboratory

1. INTRODUCTION

Linear Predictive Coding (LPC) is one of the most important concepts in the coding, analysis and synthesis of digital speech. The technology offers high quality low bit-rate encoding and is in use in many current audio transmission / storage systems, most notably the GSM (Groupe Spécial Mobile) system. As such it has a special place in teaching digital speech and audio processing, [1 - 4].

It is unfortunate that the interest in electrical engineering education is slowly declining. It is necessary for the courses to adapt to this new situation and use new teaching methodologies, [5]. The students nowadays are largely computer literate, graphically oriented, possess refined hand-to-eye coordination, and, most importantly, are used to immediate feedback or results, [6]. This is why, in order to capture their imagination, modern courses must be accompanied with more attractive educational tools. This trend has led to creation of many web based DSP materials such as tutorials, assignments and web-based tools for courses in electrical engineering, [7 - 9].

In this paper we present an educational tool in the form of a codec, developed for teaching LPC to undergraduate and graduate students in courses dealing with the processing of digital audio and multimedia signals. Similar work has been carried out earlier as presented in [10], where most LPC-based codecs were implemented in Matlab for educational purposes. We focused our work on a single LPC-10 based codec tool with an intuitive graphical user interface (GUI) that was especially adapted for students at the undergraduate level. With the developed GUI, students can easily modify various parameters of the LPC process, and identify their influence on the speech quality. In this way students can get a hands-on experience and undertsand the significance of each parameter used in LPC.

The developed tool was used in laboratory exercises for the course Digital Audio Systems and the course Processing and Transmission of Multimedia Signals, and was well received by the students. It was of great value in illustrating the concepts and the particularities of LPC.

In Section 2 we give a short description of linear predictive coding. In Section 3, we introduce our codec tool and in Section 4 we describe its use in a typical laboratory exercise. We present student feedback in Section 5 and conclude the paper in Section 6.

2. LPC – The Basics

LPC is based on the source-filter model of human speech production. In this model the generation of speech is represented with a system consisting of a source and a filter which correspond to the vocal cords and the vocal tract, respectively.

The source operates in two modes. In the first mode an impulse train is generated, mimicking the pulse excitation from the vocal cords. In the second mode white Gaussian noise is generated, mimicking the noise-like excitation used by humans to generate unvoiced speech.

The filter models the transfer function (impulse response) of the vocal tract. Its role is to shape the signal generated by the source and give it phonetic character. An autoregressive model is used for this modeling. The filter parameters are extracted from the speech input by the use of linear predictive analysis, [1]. This gives the coding scheme its name – linear predictive coding. The speech production model is shown in Fig. 1.

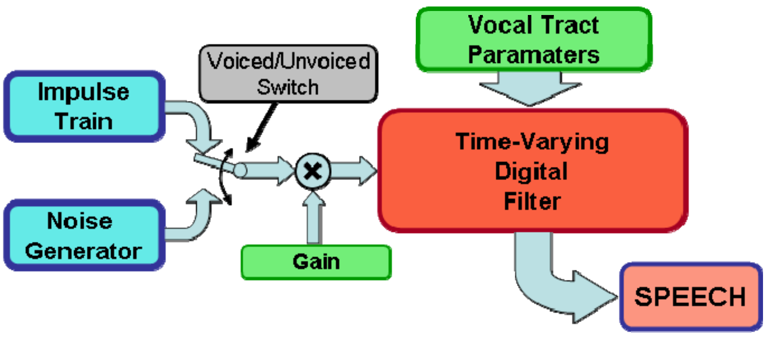

Fig. 1 – The speech production model used in LPC

Using the source-filter model, speech can be transmitted/stored without the use of audio samples, but instead by using the parameters of the model computed on a block (frame) basis. This way, a close approximation of the speech signal can be generated at the decoder using the transmitted/stored parameters.

The input speech is analyzed in frames. For each frame the following set of parameters is extracted:

- Voicing: to discriminate whether the speech signal is voiced or unvoiced
- Gain: the energy level of the signal in the frame
- Filter coefficients: refer to the transfer function of the vocal tract
- Pitch period: base period of voiced frames

The main disadvantage of using this simple model of human speech production is the inappropriate representation of phones and inter-phone transitions that cannot be characterized as purely voiced or unvoiced, [2]. This disadvantage has been overcome by more sophisticated models, namely CELP (Code-Exited Linear Prediction) which is used in GSM telephony. The CELP model uses an excitation codebook in place of the simple voiced/unvoiced source model.

3. THE LPC-10 CODEC TOOL

At the core of the tool is a modified version of the original Federal Standard (FS) 1015 coder [11]. It is a low bit-rate 2.4 kb/s LP-based codec developed originally by the US military and subsequently by NATO. The output of the coder is notably synthetic, however its intelligibility is excellent. Because of the use of a filter of order 10, the codec is also known as LPC-10.

The codec has been adapted for use by students during laboratory classes. The parameters of the model can be easily changed and experimented with. This can be done on two levels. The first level is to change them in the developed GUI. The second level is to change them in the source code itself. For laboratory exercises only the first level is used, while the second level is more suitable to projects or graduate level courses.

The GUI is presented in Fig. 2. There are three main fields in the GUI. The first contains text boxes in which the student can input the parameters of the codec. The second field contains switches which control the algorithms used in the coding process. The third field contains the load, execute, plot and play buttons.

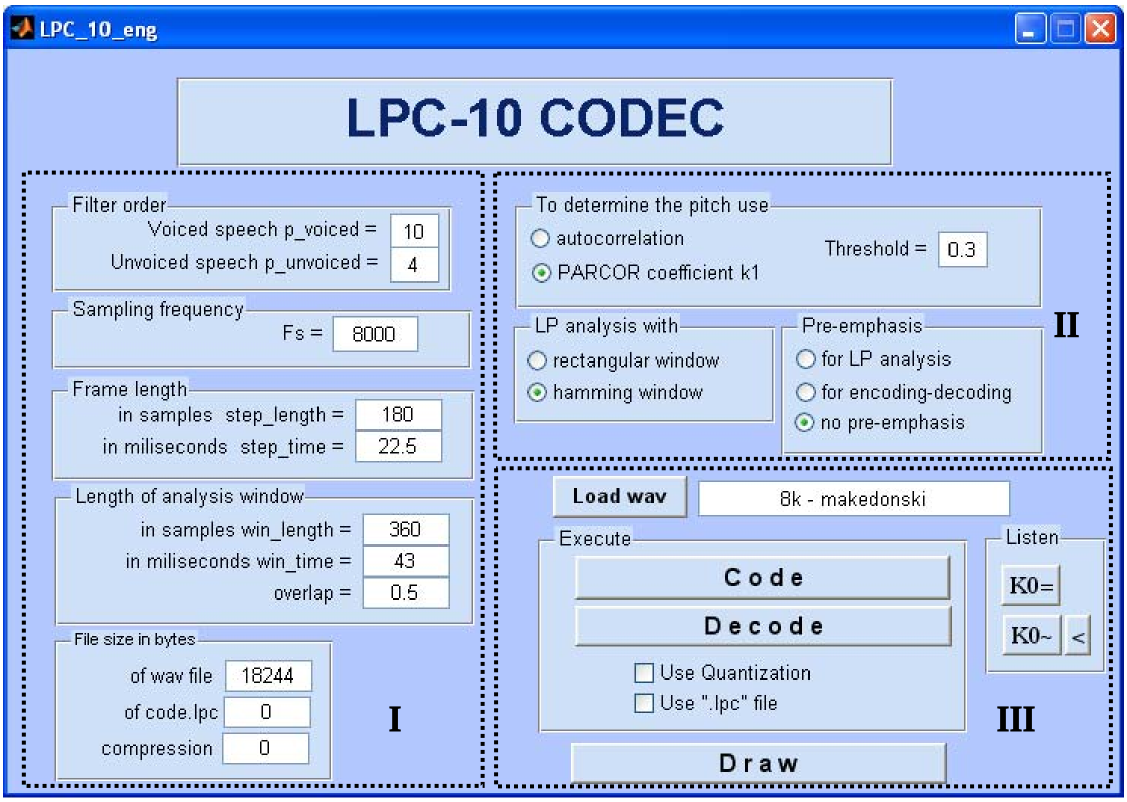

Fig. 2 – The Graphical User Interface of the LPC-10

4. Typical Laboratory Class

The laboratory class was organized as follows. First, students start Matlab and the LPC-10 codec tool. Then, they load one of the audio files and perform the encoding / decoding. They listen to the decoded output and compare it with the original, trying to identify the artifacts. They also view the plots of the original and synthesized speech (Fig.3), comparing the amplitude envelopes of the two signals.

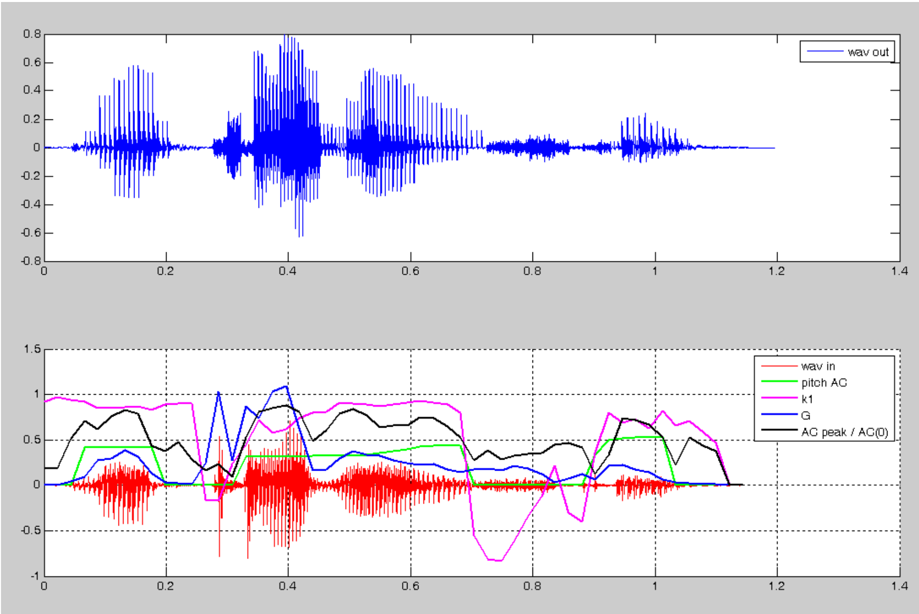

Fig. 3 – Plots of the original and the synthesized speech signal with parameters of the LP analysis

4.1 Changing the filter order

In the next stage students change the filter order. From the default value of 4, they increase the filter order of the unvoiced parts to 10, and try to discern improvement in the output quality. After this they decrease the filter order to 1 to figure out if there is any impact. Increasing filter order bears little improvement in quality, while decreasing it to 1 results in noticeable degradation. For the voiced filter order they use the following values: 1, 2, 4, 6, the default value of 10, and 20. The first three values introduce intelligibility slowly in the resulting speech.

4.2 Changing frame and analysis window size

In this part students increase the frame size from 22.5 *ms* to 100 *ms* and listen how the speech becomes smudged and phones merge with each other. The decrease of frame size to 5 *ms*, on the other hand, results in an increased quality, joined by an increase in bit-rate. Increasing the analysis window size by choosing the overlap to be 0.9, instead of the default 0.5, has similar effects on the speech signal as increasing the frame size (phone merging).

4.3 Adjusting voicing determination

The most illustrative part for the students is to see how the voicing determinaton works. Two algorithms can be used for this. The first one uses the ratio of the pitch autocorrelation peak with the power of the speech signal over the current frame. The second algorithm uses the value of the first PARCOR (Partial Correlation) coefficient k_1. Each algorithm compares uses a decision threshold. Figure 4 shows the effects of threshold variation for the first algorithm in the range from 0.3 to 0.5 for the utterance "makedonski".

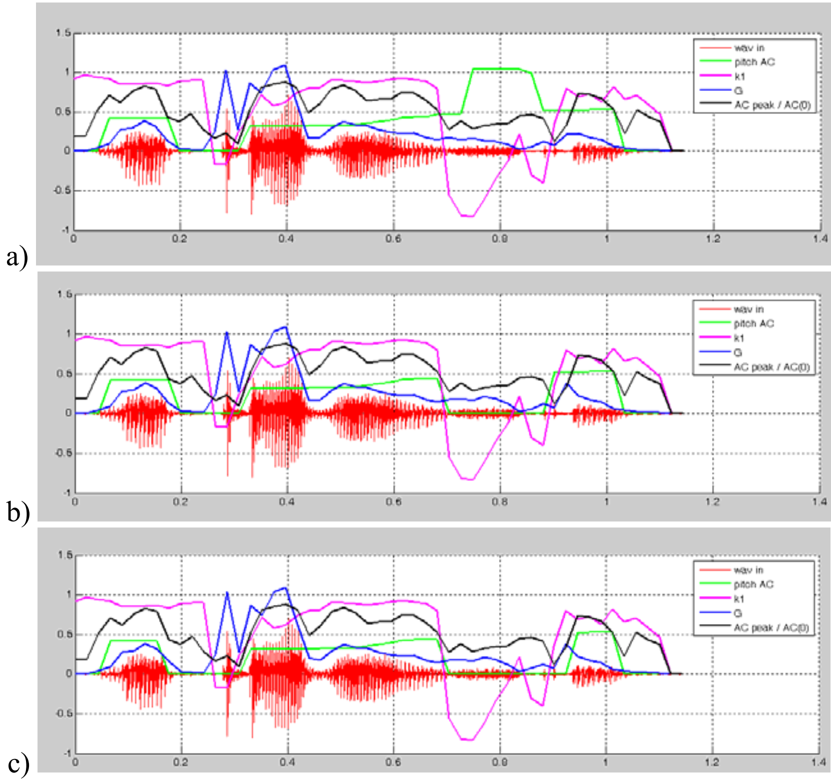

Fig. 4 – Influence of the threshold on the voicing decision when using the autocorrelation algorithm: a) 0.3; b) 0.4; c) 0.5.

It can be seen that when the threshold is 0.3 there is a large error around 0.8 *s* into the signal, where the phone "s" is determined as voiced. The problem is eliminated when using a threshold of 0.4. Still, the second phone "k" ("makedonsKi") is determined as voiced. Increase of the threshold to 0.5 will eliminate this problem as well. Extremely low and extremely high thresholds are also tried, as well as the use of different voice types: female, child, elderly, and a music passage. A key point is to hear that the whole LPC-10 concept breaks down when trying to encode music.

5. Student Feedback

Students were enthusiastic during the laboratory classes. What was most entertaining to them was decreasing the voicing detection threshold to an extreme value, e.g. to 0.01. This forces the speech to be processed and synthesized as entirely voiceless, i.e. with a whisper. Another interesting effect was doing the opposite, forcing the entire speech to be synthesized as voiced. It was obvious that they wanted to continue to experiment on their own and they were encouraged to take the source code home and try it with their own recordings.

The increase in students' interest for the field was evident as a result of using our tool. This is a definite pointer that such tools should be included in the exercises in the areas of digital audio and multimedia signal processing.

6. CONCLUSION

LPC is a very important and powerful concept in the analysis and synthesis of digital speech. It is an important part of digital audio processing and multimedia signal processing courses for undergraduate students. The introduction of the LPC-10 codec in the laboratory classes has helped the students grasp the particularities of LPC in an intuitive way. It has also increased their interest in this area to the extent that they start experimenting on their own.

The LPC-10 codec can be further improved. A graduate level task might be to implement the codec according to the exact specifications of the FS-1015 standard, and, also, implement a CELP coder at a following stage. Further development of similar tools for other areas in digital audio processing and multimedia signal processing is envisioned.

7. REFERENCES

[1] Rabiner L.R., R.W. Schafer, *Digital Processing of Speech Signals*, Prentice Hall, 1978

[2] Chu W., *Speech Coding Algorithms - foundation and evolution of of standardized coders*, Wiley, 2003

[3] Spanias A., T. Painter, V. Atti, Audio Signal Processing and Coding, John Wiley & Sons, 2007

[4] Sayood K., *Introduction to Data Compression*, *Second Edition*, Morgan Kaufmann**,** 2000

[5] Taskovski D., V. Kitanovski, S. Bogdanov, *Developing Collaborative Learning Model in DSP Courses*, 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services, June 2007

[6] Morrow M., Welch T., Wright C., *An Inexpensive Software Tool For Teaching Real-Time DSP*, First Signal Processing Education Workshop, Hunt, Texas, 2000

[7] Taskovski D., V. Kitanovski, S. Bogdanova, Z. Cucej, *New web-based tools for DSP education*, Int. Conf. on Interactive Computer Aided Blended Learning, Brazil, May, 2007

[8] Marín S., F. García, R. Torres, S. Vázquez, A. Moreno, *Implementation of a Web-Based Educational Tool for DSP Teaching Using theTechnological Acceptance Model,* IEEE Transactions on Education, 2005

[9] Rahkila M., M. Karjalainen, *Considerations of computer based education in acoustics and signalprocessing*, Frontiers in Education Conference, 1998

[10] Painter E., A. Spanias, *A Matlab Software Tool For The Introduction Of Speech Coding Fundamentals In A DSP Course,* Frontiers in Education Conference '96, Nov. 1996

[11] Tremain T. E., The Government Standard Linear Predictive Coding Algorithm,LPC-10, Speech Technology, April. 1982, pp. 40–49