Annales UMCS Informatica AI 5 (2006) 315-321

Annales UMCS
Informatica
Lublin-Polonia
Sectio AI

http://www.annales.umcs.lublin.pl/

Digital signals analysis with the LPC method

Ireneusz Codello*, Wiesława Kuniszyk-Jóźkowiak

Maria Curie-Skłodowska University, Pl. M.Curie-Skłodowskiej 1, 20-031 Lublin, Poland

Abstract

This paper concerns the issue of signal analysis by the linear predictive method. It is not only about frequency analysis, that is obtaining the LPC spectrum and comparing it with the Fourier one, but also about analysing prediction coefficients themselves. We will also discuss the program made by the authors of this paper, which beside foregoing functionalities, enables vocal tract visualization, modelled on the basis of LPC coefficients.

1. Introduction

The LPC method [1-4] is based on the fact that consecutive samples of speech signal are not random, but they change smoothly. In connection with this, we can approximate a consecutive sample (that is the next value of sound file) with the previous samples

$$v(n)=\sum_{k=1}^{p}\alpha_k s(n-k),$$

where: $s(n)$ n-th input sample, $v(n)$ approximation of n-th sample, α prediction coefficients, p prediction order.

As it can be seen, the goal of this method is to find a set of α coefficients which gives the best approximation. Therefore a prediction error

$$e(n)=s(n)-v(n)$$

should be the smallest. We have to remember about the prediction order (which is equivalent to the sample count used in approximation), which is a very important element. Usually its value lies in the <10, 25> range. Another thing we have to remember about is that the signal has to be stationary – otherwise there is no point in using this method. Since human speech does not meet this condition, we have to split a signal into 20-40 ms pieces. In such short time our speech is almost invariable – quasi-stationary.

*Corresponding author: *e-mail address*: irek.codello@gmail.com

2. Application

The first, most obvious application is speech synthesis. As an example we can mention LPC-10 format, where having prediction coefficient, gain parameter and knowledge if current speech piece is voiced or unvoiced, we are able to reconstruct the signal.

Another application, directly resulting from the first one, is data compression. What is a point in keeping the original wave file, which consists of 220-440 sampled windows (it corresponds to 20-40 ms window by 11 kHz sampling frequency) if we can use its 12-26 numbered representation (it depends on the LPC coefficient count on which we will decide).

LPC parameters can be used in frequency analysis. A spectrum, created with this method, is very smooth in comparison to its Fourier equivalent, which makes it more useful.

In addition, on the basis of LPC coefficients we can model the shape of the vocal tract.

3. Linear Prediction Algorithm

The goal of the LPC algorithm [1-5] is to minimize prediction error E for each window, where

$$E=\sum_m e^2(m)=\sum_m\big(s(m)-v(m)\big)^2=\sum_m\left(s(m)-\sum_{k=1}^{p}\alpha_k s(m-k)\right).$$

We can achieve this by computing partial derivatives for consecutive α coefficients and comparing them to zero.

Of a few available algorithms we chose the autocorrelation method. It assumes that samples outside the range <0, N-1> equals zero. In order to achieve this, we can apply windowing with the rectangular window. Better results give more complicated windows like Hamming or Hann window. Using the autocorrelation function

$$R(k)=\frac{1}{N-p}\sum_{n=0}^{N-1-k}s(n)s(n+k)$$

which we put in the afore-mentioned set of partial derivatives equations, we obtain a set of equations which can be expressed in a matrix form:

$$\begin{bmatrix} R(0) & R(1) & R(2) & \dots & R(p-1)\\ R(1) & R(0) & R(1) & \dots & R(p-2)\\ R(2) & R(1) & R(0) & \dots & R(p-3)\\ \dots & \dots & \dots & \dots & \dots\\ R(p-1) & R(p-2) & R(p-3) & \dots & R(0)\end{bmatrix}\begin{bmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\\ \\ \alpha_p\end{bmatrix}=\begin{bmatrix}R(1)\\ R(2)\\ R(3)\\ \\ R(p)\end{bmatrix}.$$

Because the matrix of autocorrelation values is a symmetric Toeplitz matrix, we can take advantage of the effective Levinson-Durbin algorithm:

$$E^0 = R(0)$$
$$k_i = \left(R(i) - \sum_{j=1}^{i-1} \alpha_j^{i-1} R(i-j) \right) / E^{i-1}, \qquad 1 \le i \le p$$
$$\alpha_i^i = k_i$$
$$\alpha_j^i = \alpha_j^{i-1} - k_i \alpha_{i-j}^{i-1}, \qquad 1 \le j \le i-1$$
$$E^i = \left(1 - k_i^2\right) E^{i-1},$$

where the upper index represents iterations from 1 to p and a result is defined as

$$\alpha_j = \alpha_j^{(p)}, \qquad 1 \le i \le p,$$

E and k_i parameters, used in this algorithm, are important as well. E coefficient is the afore-mentioned prediction error, whereas k_i are called PARCOR coefficients – we will discuss them further in this article.

To ensure that our results are correct, we took advantage of the following dependencies [1]:

– R, α → E

$$E = R(0) - \sum_{k=1}^{p} \alpha_k R(k)$$

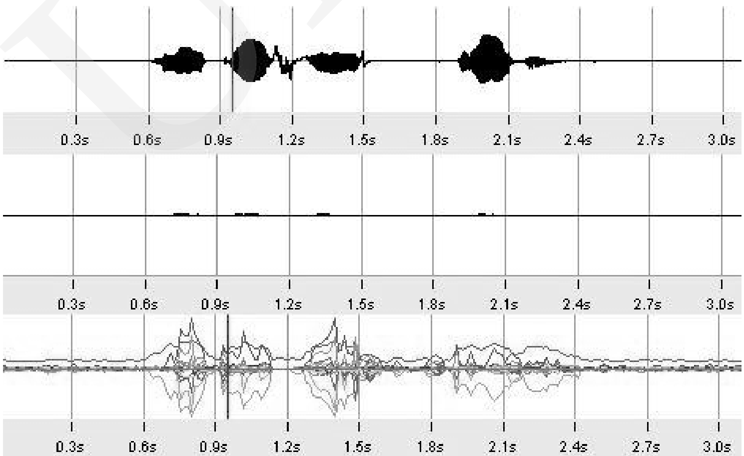

Fig. 1. From the top: input signal, prediction error e(n), changes of 20 α coefficients for the sample „mleko jest zdrowe". Window witdh – 29 ms. The screenshot comes from a program created by the authors of this paper

– α → k

$$k_i = a_i^{(i)}$$
$$a_j^{(i-1)} = \frac{a_j^{(i)} + a_i^{(i)} a_{i-j}^{(i)}}{1 - k_i^2}, \qquad i \le j \le i-1$$

where the upper index represents iterations from p down to 1 and initially we set

$$a_j^{(p)} = \alpha_j, \qquad 1 \le j \le p .$$

– k → α

$$a_i^{(i)} = k_i$$

$$a_j^{(i)} = a_j^{(i-1)} + k_i \, a_{i-j}^{(i-1)}, \qquad 1 \le j \le i-1$$

where the upper index represents iterations from 1 to p and a result is defined as

$$\alpha_j = a_j^{(p)}, \qquad 1 \le i \le p .$$

4. Spectrum

Transfer function representing our system is of the form [1]

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^{p} \alpha_k z^{-k}},$$

where *α* prediction coefficients, *p* prediction error, *G* gain parameter.

If we set prediction order (i.e. 15) and we compute prediction coefficients α_k, we only have to obtain the gain parameter G from formula [1]

$$G = \sqrt{E} .$$

Frequency characteristics can be obtained from a modulus of the transfer function, when we use the arguments of the form $e^{j\omega}$, which gives

$$\left|H\left(e^{j\omega}\right)\right| = \left|H\left(e^{j2\pi f/F}\right)\right|,$$

where *j* imaginary unit, *F* sampling frequency, *f* interesting frequency.

The precision of this spectrum depends on the prediction order. It can be seen by comparing, in decibel scale, the Fourier spectrum – $20\log_{10}|S(e^{j\omega})|$ with the LPC spectrum – $20\log_{10}|H(e^{j\omega})|$.

On these graphs we can see a frayed Fourier spectrum and a smooth LPC spectrum, whose shape depends on the prediction order p. Due to this smoothness, it is much easier to find local maxima, which, in turn, are essential in forma detecting.

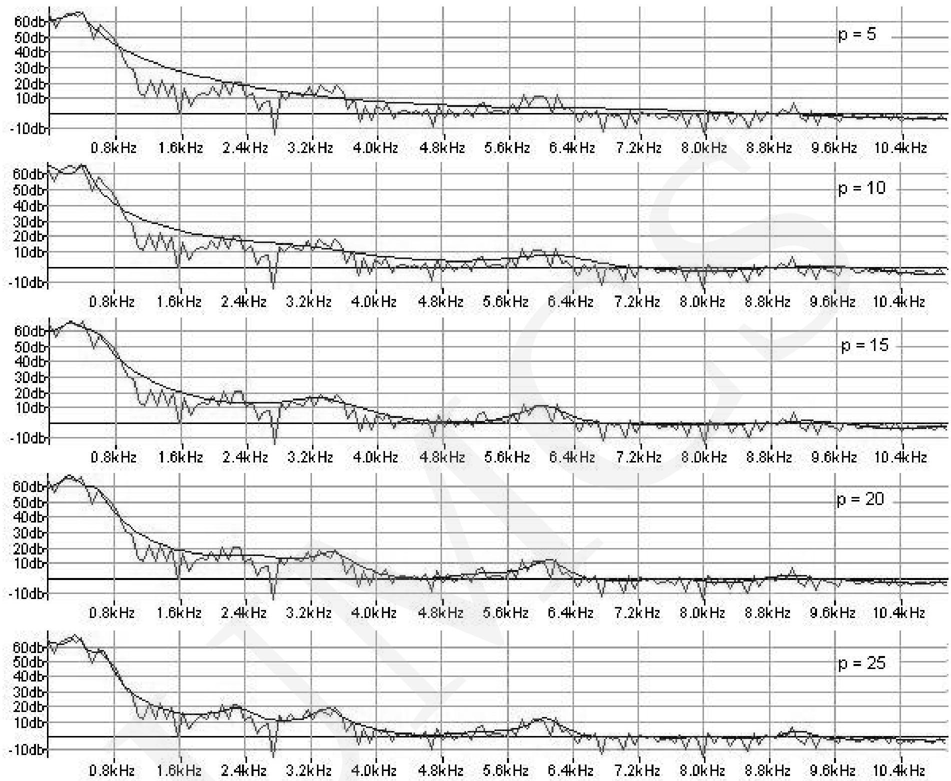

Fig. 2. Comparison of LPC spectrum, for various prediction orders p, with the Fourier spectrum. In both cases we used Hamming window of identical width. Screenshots come from a program created by the authors of this paper

5. Vocal tract

Human speech consists of three characteristics [1]:

- type of glottal excitation: voiced(period impulses) or unvoiced,
- vocal tract shape: modulates excitation signal,
- Set of parameters affecting emission of the sound wave i.e. mouth radiation, environment density.

A vocal tract is simply the section from larynx to lips. Its shape is quite complicated – it depends on the shape of the tongue, lips, pharynx, and teeth dilation. Additionally walls of the tract have absorption properties and it forms an arc with the dilation angle near 90°.

We have used a simplified model of the vocal tract, which consists of a row of various diameter cylinders connected with each other [1,3,4]. They form a lossless tube with varying cross sections.

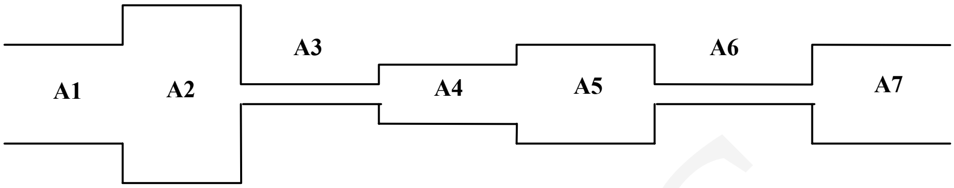

Fig. 3. Vocal tract model. A signal propagates from left to right. A1-A7 are consecutive values of cross sections of the tube

The left side of the tube in figure 3 represents the larynx, while the right side represents lips. On a junction of every pair of cylinders a partial reflection of a sound wave appears. It turns out that reflection coefficients on a joint of two cross sections, represented by the formula $r_i = \left(A_{i+1} - A_i\right) / \left(A_{i+1} + A_i\right)$, are related to PARCOR coefficients through a simple relation $r_i = -k$ [1]. From these two equations, it is simple to show that

$$A_{i+1} = \left(\frac{1-k_i}{1+k_i}\right) A_i .$$

Therefore we obtained the relation between successive cross sections (as we can see, the number of these sections is greater by one from the prediction order). We are left with the issue of choosing the value of the first cross section – A_1 or the last – A_{p+1}, and deciding how long a vocal tract should be. In this paper we assumed, that A_{p+1} (lips area) will be equal to 1 cm^2 and the vocal tract length will have an average length of a full grown man – that is 17.5 cm [1].

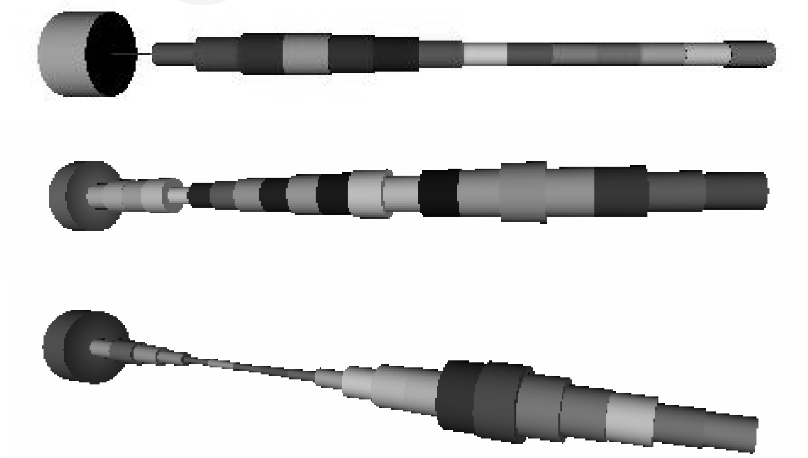

Fig. 4. 3D vocal tract model. The right side represents the lips. From the top: vowel ‘u’, p=15, vowel ‘e’, p=20, vowel ‘i’, p=20. Screenshots come from the program created by the authors of this paper

Conclusions

Our main goal was to create a tool to analyse wave sound files by the LPC method and to obtain – via that tool – the largest set of information. Those results, that is LPC, PARCOR coefficients and spectrum, will be obtained from fluent and dysfunctional speech and then analysed. We hope to distinguish, owing to this research, the characteristic features in various types of dysfunctionalties. We are also planning further research, based on these analyses i.e. speech and speakers recognition.

Acknowledgements

Scientific work partially financed from the grant of Vice-Rector of Maria Curie-Skłodowska Univeristy.

The authors thank Natalia Fedan for language corrections.

References

[1] Rabiner L.R., Schafer R.W., *Digital Processing of Speech Signals*, Prentice-Hall, Inc., New Jarsey, (1978).

[2] Zieliński T.P. *Od teorii do cyfrowego przetwarzania sygnałów*, Wydział EAIiE AGH, Kraków, (2002), in Polish.

[3] Wakita H., *Direct Estimation of the Vocal Tract Shape by Inverse Filtering of Acoustic Speech Waveforms*, IEE Transactions on Audio and Electroacoustics, AU-21(5) (1973).

[4] Ananthakrishnan K.S., *Computer aided pronunciation system (CAPS)*, University of South Australia, (2003), http://www.itr.unisa.edu.au/rd/pubs/thesis/ksa.pdf

[5] Komaee A., Sepehri A., *Linear Prediction and Synthesis of Speech Signals*, Department of Electrical and Computer Engineering, University of Maryland, http://www.enee.umd.edu/~afshin/adsp2/proj2.pdf