Speech sound coding using linear predictive coding

Working Paper · July 2017

DOI: 10.13140/RG.2.2.24214.04162

1 author:

Oday Kamil
Dijlah University College
5 PUBLICATIONS **0** CITATIONS

SEE PROFILE

ISSN: 2413-6999
Journal of Information, Communication, and Intelligence Systems (JICIS)
Volume 3, Issue 1, May 2017

Speech Sound Coding Using Linear Predictive Coding (LPC)

Oday Kamil Hamid

***Abstract*— The key objective of this research is to estimate the basic speech parameters e.g., pitch formants, spectra, and vocal tract area function these can be used to identify the type of sound being produced, for speech recognition or speaker recognition. Linear Predictive Coding (LPC) technique with order equal sixteen (P=32) was used which for estimating the basic speech parameters (feature sound extraction). The proposed algorithm is tested upon a database consist of (2) speakers (1 male and 1 female). The main goal of the research is for representing speech samples with just 16 point instead of 256 samples The proposed algorithm are examined through theoretical analysis and computer simulation using Matlab version 6 programming language and sound forge 5 as a speech analyzer under Microsoft Windows XP operating system.**

***Index Terms*— LPC, AUTOCORRELATION. DURBINS**

I. INTRODUCTION

This method has become the predominant technique for estimating the basic speech parameters, e.g., pitch, formants, spectra, and vocal tract area functions [1]. These coefficients accurately indicate the instantaneous configuration of the vocal tract. These can then be used to identify the type of sound being produced, for speech recognition or in this, to identify the speaker based on small variations in the parameters [2].

And these coefficients have been dependent in this research for this reasons [3]: -

1) An LP coefficient provides a good model of the speech signal. This is especially true for the quasi steady state voiced regions of speech in which all-pole model of LPC provide a good approximation to the vocal tract spectral envelope. During unvoiced and transient regions of speech, the LPC model is less effective than for voiced regions but it still provides an acceptably useful model for speech-recognition purposes.

2) The way in which LPC is applied to the analysis of speech signals leads to a reasonable source-vocal tract separation; a parsimonious representation of the vocal tract characteristics becomes possible.

***Oday Kamil Hamid, Dept. of Computer Techniques Engineering, Dijlah University College,** (oday.kamil@duc.edu.iq). Baghdad, Iraq.*

3) LPC is an analytically tractable model. The method of Straight forward to implement in either software or hardware.

The computation involved in LPC processing is considerably less than that required for another model. The LPC model works well in recognition applications. Experience has shown that the performance of speech recognizers, based on LPC front ends, is comparable to or better than another model.

II. THEORY

One of the most powerful speech analysis techniques is the method of linear predictive analysis. This method has become the predominant technique for estimating speech parameters, e.g., pitch, formants, spectra, vocal tract area functions, and for representing speech for low bit rate transmission or storage [1]. In recent years the method of linear prediction has been quite successfully used in speech compression systems. A speech synthesis model shown as in block diagram of figure (1):

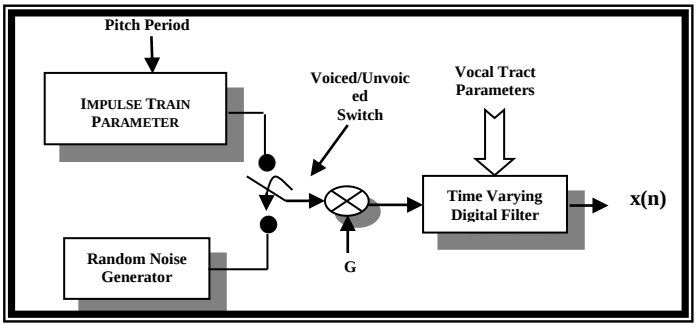


Fig.1.Speech synthesis model based on LPC model

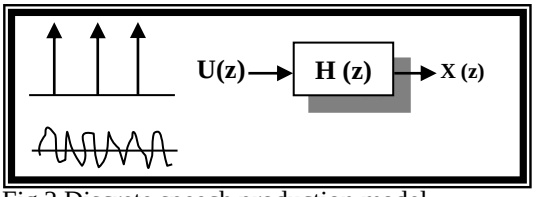


Fig.2.Discrete speech production model

ISSN: 2413-6999
Journal of Information, Communication, and Intelligence Systems (JICIS)
Volume 3, Issue 1, May 2017

The input to the filter is either a sequence of pulses separated by the pitch period for voiced sounds, or a random noise source for unvoiced sounds. Thus, the parameters of this mode are voiced/unvoiced classification, pitch period for voiced speech, gain parameter *G*, and the coefficients (a_k) of the digital filter. These parameters, of course, all vary slowly with time [4]. One of the most powerful models currently in use is that where a signal *x(n)* is considered to be the output of some system with some unknown input u(n) such that the following relation holds [5]:

$$x(n) = \sum_{k=1}^{p} a_k \, x(n-k) + G \sum_{l=0}^{q} b_l \, u(n-l), \quad \begin{matrix} b_0 = 1 \\ a_0 = 1 \end{matrix} \quad \text{.....(1)}$$

Where a_k , $1 \le k \le p$, b_l , $1 \le l \le q$, and the gain *G* are the parameters of the hypothesized system. Equation (1) says that the 'output' *x(n)* is a linear function of past outputs and present and past inputs. That is, the signal *x(n)* is predictable from linear combinations of past outputs and inputs. Hence the name linear prediction. Equation (1) can be specified in the frequency domain by taking the z-transform on both sides of eq. (1) .If *H(z)* is the transfer function of the system, as in figure (2) then we have from eq. (1):

$$H(z) = \frac{X(z)}{U(z)} = G \, \frac{1 + \sum_{l=1}^{q} b_l z^{-l}}{1 - \sum_{k=1}^{p} a_k z^{-k}} \quad \text{.....(2)}$$

Where $X(z) = \sum_{n=-\infty}^{\infty} x(n) \, z^{-n}$ is the z-transform of *x(n)* and *U(z)* is the z-transform of *u(n)* .*H(z)* in eq.(2) is the general pole-zero model. The roots of the numerator and denominator polynomials are the zeros and poles of the model-respectively. There are two special cases of the model that are of interest:

1. All-zero model: $a_k = 0, 1 \le k \le p$
2. All-pole model: $b_l = 0$, $1 \le l \le q$

The major part of this work will be devoted to the all- pole model. This has been, by far, the most widely used model and it is known as the autoregressive model (AR). Then the transfer function [6]:

$$H(z) = \frac{G}{1 - \sum_{k=1}^{P} a_k z^{-k}} \text{.....(3)}$$

3. The basic problem of linear prediction analysis is to determine the set predictor coefficients (a_k), which will be described later [7]. The gain (G) is usually ignored to allow the parameterization to be independent of the signal intensity. The basic idea behind the LPC model is that a given speech sample at time (n), *x(n)*,can be approximated as a linear combination of the past *(p)* speech samples such that [8]:-

$X(n)=a_1x(n+1)+a_2x(n+2)+\ldots.+a_P(n-p)+G(n)$-------(4)

$$x(n) = \sum_{k=1}^{p} a_k \, x(n-k) + Gu(n) \quad \text{.....(5)}$$

The signal *x(n)* can be predicted only approximately from a linearly weighted summation of past samples. Let this approximation of *x(n)* be $\tilde{x}(n)$ where

$$\tilde{x}(n) = \sum_{k=1}^{p} \alpha_k \, x(n-k) \quad \text{.....(6)}$$

α_k: Prediction coefficients

The prediction error *e(n)* is defined as

$$e(n) = x(n) - \tilde{x}(n) = x(n) - \sum_{k=1}^{p} \alpha_k \, x(n-k) \quad \text{....(7)}$$

It can be seen by comparing eqs. (5) and (7) that if the speech signal obeys the model of eq. (5) exactly and if $\alpha_k = a_k$ then $e(n) = Gu(n)$. [6]. The basic approach is to find a set of predictor coefficients (α_1 --- α_p) that minimize the mean-squared prediction error *e(n)*

$$E = \sum_n e(n)^2 = \sum_n \left[x(n) - \sum_{k=1}^{p} \alpha_k x(n-k) \right]^2 \text{.....(8)}$$

leaving the range of summation unspecified for the moment.

To minimize *(E)* by choice of the coefficients α_k, differentiate with respect to each of them and set the resulting derivatives to zero [8]:

$$\frac{dE}{d\alpha_i} = 0 \qquad 1 \le i \le p$$

$$= -2 \sum_n x(n-i) \left[x(n) - \sum_{k=1}^{p} \alpha_k \, x(n-k) \right] = 0$$

$$\sum_{k=1}^{p} \alpha_k \sum_n x(n-i)x(n-k) = \sum_n x(n)x(n-i) \qquad 1 \le i \le p \quad \text{.....(9)}$$

There are many approaches to find the coefficients α_1 --- α_p , from eq. (9) like: -[8]

1-The covariance method.
2-The autocorrelation method.
3-The lattice method.

ISSN: 2413-6999
Journal of Information, Communication, and Intelligence Systems (JICIS)
Volume 3, Issue 1, May 2017

There are many differences in the theoretical formulations of the covariance, autocorrelation, and lattice formulations of the linear predictive analysis equations. Autocorrelation method requires less computation than the other two methods and the filter can be guaranteed to be stable. It should be noted that this theoretical guarantee of stability for the autocorrelation method might not hold in practice if the autocorrelation function is computed without sufficient accuracy. For the covariance method the stability of the filter cannot be guaranteed and no windowing is required. However in practice, if the number of samples in the frame is sufficiently large then the resulting predictor polynomials will almost be stable. For the lattice method the predictor polynomial is guaranteed to be stable and no windowing is required and it needs more computation than the other two methods thus the lattice method is the least computationally efficient method for solving the LPC equation [9]. So for these reasons the autocorrelation method will be used to solve the LPC equation in this research.

III. Autocorrelation Method

Because of the time-varying nature of the speech signal the predictor coefficients must be estimated from short segments of speech signals (10-40msec) where the characteristics of speech signals are constant in this range. The basic approach is to define a set of predictor coefficients that will minimize the mean squared prediction over a short segment of the speech waveform. The resulting parameters are then assumed to be the parameters of the system function *H(z)*, in the model for speech production [15].Then eq. (8) can be written a

$$E_N = \sum_m e_n^2(m) \quad \text{……(10)}$$

One approach to determining the limits on the sums in eq.(9) is to assume that the waveform segment, $x_n(m)$ is identically zero outside the interval $0 \le m \le N-1$ this can be conveniently expressed as:

$$x_n(m) = x(m+n)\, w(m) \quad \text{……..(11)}$$

Where $w(m)$ is a finite length window (e.g. *a Hamming window*) that is identically zero outside the interval $0 \le m \le N-1$.

If $x_n(m)$ is nonzero only for $0 \le m \le N-1$, then the corresponding prediction error, $e_n(m)$ for an P^{th} order predictor will be nonzero over the interval $0 \le m \le N-1+p$.

Thus, for this case E_n is properly expressed as:

$$E_n = \sum_{m=0}^{N+P-1} e_n^2(m) \quad \text{……..(12)}$$

And

$\phi_n(i,k)$ Can be expressed as:

$$\phi_n(i,k) = \sum_{m=0}^{N-1+P} x_n(m-i)\, x_n(m-k) \qquad \begin{matrix} 1 \le i \le p \\ 0 \le k \le p \end{matrix} \quad \text{…..(13)}$$

Or

$$\phi_n(i,k) = \sum_{m=0}^{N-1-(i-k)} x_n(m)\, x_n(m+i-k) \quad \text{……..……(14)}$$

Further more it can be seen that in this case $\phi_n(i,k)$ is identical to the short time autocorrelation function where

$$R_n(k) = \sum_{m=0}^{N-1-k} x_n(m)\, x_n(m+k) \quad \text{……...……(15)}$$

Since $R_n(k)$ is an even function it follows that

$$\phi_n(i,k) = R_n(|i-k|) \quad \text{……..…….(16)}$$

Therefore eq. (2-9) can be expressed as:

$$R_n(i) = \sum_{k=1}^{p} \alpha_k R_n(|i-k|) \qquad 1 \le i \le p \quad \text{…(17)}$$

Similarly the minimum mean squared prediction error takes the form:

$$E_n = R_n(0) - \sum_{k=1}^{p} \alpha_k R_n(k) \quad \text{……..…..(18)}$$

The set of eqs. (17) can be expressed in matrix form of $(p * p)$ autocorrelation value is a *Topplitz* matrix, i.e., it is symmetric and all the elements along a given diagonal are equal.

Several efficient recursive procedures have been devised for solving this system of equations; the most efficient method known for solving this particular system of equations is *Durbin's recursive procedure* [9].

IV. Durbin's Recursive Method:

It is a formal method for converting from autocorrelation coefficients to an *LPC* parameter set in which the set might be *LPC* (a_k) coefficients, reflection coefficients (k_i) and can formally be given as the following algorithm [9]:

$$E^{(0)} = R(0) \quad \text{…..…… (19)}$$

$$k_i = \left\{ R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(|i-j|) \right\} / E^{(i-1)} \qquad 1 \le i \le p \quad \text{…..(20)}$$

$$\alpha_i^{(i)} = k_i \quad \text{..……...…(21)}$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i\, \alpha_{i-j}^{(i-1)} \qquad 1 \le j \le i-1 \quad \text{……..…(22)}$$

$$E^{(i)} = \left(1 - k_i^2\right) E^{(i-1)} \quad \text{.....……...(23)}$$

LPC coefficients $a_m = \alpha_m^p \qquad 1 \le m \le p.$ …..(24)

ISSN: 2413-6999
Journal of Information, Communication, and Intelligence Systems (JICIS)
Volume 3, Issue 1, May 2017

V. DATA BASE

- In High quality microphones used in recording sounds.
- Ideal recording must be used in rooms with little or no background noise or reverberation for both training and testing sessions.
- Collect a large database for many tries.
- Using modern programs in recording because it has a capability for cutting voices or synthesis and it gives a good representation of signal's shape.

Database had been recorded by using high quality microphone to record voices of 2 human (1male and 1female), each speaker speaks **(8)** words in two versions **(ياسين- صباح الخير- مساء الخير- يمين- وفي- لندن)** these words are recorded by using Sound Forge program figure (3.1) shows recorded signal displayed by this programs screen.

Data is sampled at **11.25 KHZ** (sampling rate), with **16-bit** sample value A/D, but unfortunately we did not have the chance of recording in suitable place for such a purpose and that deal of the collected database from each speaker considered as a little for suggested systems.
Figure (3) bellow represent the spoken signal recorded by sound forge program where X-axis=Time (sec), Y-Axis=Volt (v)

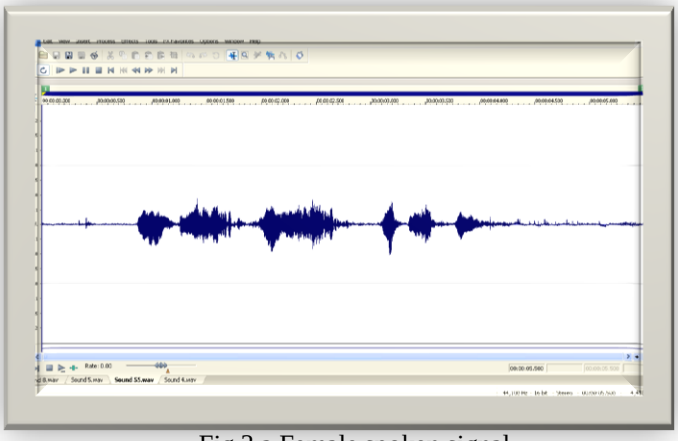


Fig.3.a.Female spoken signal

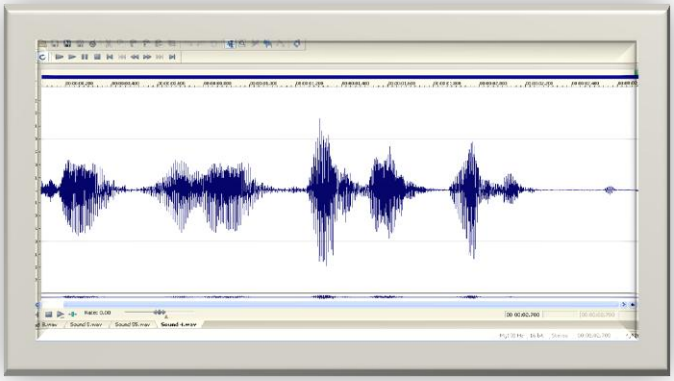


Fig.3.b.Male spoken signal

VI. PRACTICAL WORK

After data passed through preprocessing program in this moment we have a group of vectors each vector contain **(256)** samples. These vectors inters to sounds feature extraction program which represented by linear predictive coding *(LPC)* where each frame of *p+1* autocorrelation converted into an LPC parameter set in which the set might be the *LPC* coefficients, the reflection coefficient, the log area ratio coefficients, the cepstral coefficients, the formal method for this converting is known Durbin's method which described in chapter two. So we get from the output of this program a group of vectors called feature vectors that contain each of them from several coefficients represent order of LPC. The sound features represented in LPC vectors are extracted by special method the researcher did not explain it because it is not the goal of this paper so these vectors are interred to the distance measure program.

The proposed work examined through theoretical analysis and computer simulation using MATLAB version 6 programming language under MICROSOFT WINDOWS XP operating system.

VII. RESULTS

After each speaker spocke the words then we input the digital sound signal into LPC program steps.

Figure (4) show the 256 samples for first speaker before intering it to LPC program.

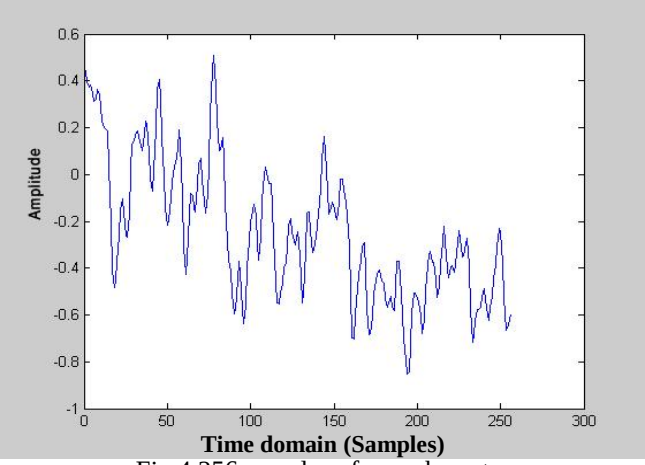
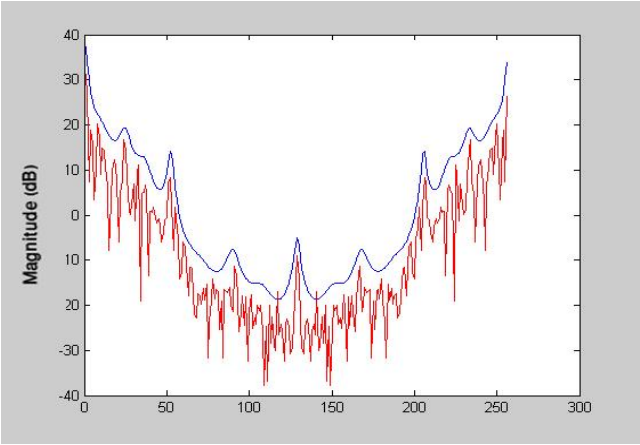


Fig.4.256 samples of speech vector

ISSN: 2413-6999
Journal of Information, Communication, and Intelligence Systems (JICIS)
Volume 3, Issue 1, May 2017

Then linear predictive algorithm was applied with order P=16 for this 256 samples as shown in figure (5)



Frequency domain (Samples)

Fig.5. Spectral shape of feature vector with LPC=16

Figure (5) shows spectral forms for the feature vector of voice by using order of LPC=16. Spectral shape of vector is plotted with the shape that produced from interring sampled vector to Fast Fourier Transform (FFT) to show the similarity between predictive spectral shape and real spectral shape, most important point that must be noted that when using order of *LPC=16(p=16)* a good matching result is achieved.

VIII. CONCLUSION

1) A high efficiency of linear predictive coding in the extraction of the aspects of the voice signal. First, it reduces the size of the memory used. Second, it can easily change the parameters (P, N, M) which controls the extraction of the LPC parameters.
2) The employment of the programming language (Matlab) becomes wealthy in the shorthand on the program volume and the ease of its correction.

IX. SUGGESTIONS FOR FUTURE WORK

1) Testing the system on other sentences to make sure it is more effective
2) Use another parameter instead of LPC like autocorrelation coefficients, log area ratio coefficients, cepstral coefficients.
3) Increasing order of LPC (P) by take it as P=32

REFERENCES

[1] M.Bassaville ,"*Distance Measure for signal processing and pattern recognition*".

[2] A. Buzo, A. H. Gray, R. M. Gray and J. D. Markel, *"Speech Coding Upon Vector Quantization"*, IEEE, Trans. Acoust. , Speech and Signal Process. , Vol. ASSP-28, October 1980.

[3] L. R. Rabiner and B.H. Juang "Fundamentals of speech recognition", M. N. AL-Trfi *"Speaker recognition based"* , prentice-Hell, New Jersey 1993.

[4] A. H. Al-Nakkash, *"A Novel Approach For Speakers Recognition Using Vector Quantization Technique"*, M.Sc. Thesis, University of Technology, Baghdad, 2001.

[5] A. Buzo, A. H. Gray, R. M. Gray and J. D. Markel, *"Speech Coding Upon Vector Quantization"*, IEEE, Trans. Acoust. , Speech and Signal Process. , Vol. ASSP-28, October 1980, pp. 562-574.

[6] M. J. Haider, *"Speech Compression and Recognition Using Wavelet Transform"*, M.Sc. Thesis, College of Engineering, University of Baghdad, Baghdad, 1999.

[7] L. R. Rabinar, M. M. Sondhi and S. E. Levinson, *"A Vector Quantization Combining Energy and LPC Parameters and Its Application To Isolated Word Recognition"*, AT&T Bel Laboratories Technical Journal, Vol. 63, No. 5, June 1984.

[8] W. A. Mahmoud, *"Quantization Techniques For The Classification and Recognition of Speech Signals"*, Ph.D. Thesis, University of Swansea, England, 1986.

[9] F. Itakura, *"Minimum Prediction Residual Principle Applied To SpeechRecognition"*, IEEE Trans. Acoust. , Speech and Signal Process. , Vol. ASSP-23, Feb. 1975, pp. 67-72.