IMPROVING LINEAR PREDICTION ANALYSIS OF NOISY SPEECH BY PREDICTIVE NOISE CANCELLATION

by

Steven F. Boll

UUCS 77-101

IMPROVING LINEAR PREDICTION ANALYSIS OF NOISY SPEECH BY PREDICTIVE NOISE CANCELLATION

Steven F. Boll

Computer Science Department
University of Utah
Salt Lake City, Utah 84112

Abstract

The analysis of speech using Linear Prediction is reformulated to account for the presence of acoustically added noise and a technique is presented for reducing its effect on parameter estimation. The method, called Predictive Noise Cancellation (PNC), modifies the noisy speech autocorrelations using an estimate of present background noise which is adaptively updated from an average all-pole noise spectrum. The all-pole noise spectrum is calculated by averaging autocorrelations during non-speech activity. The method uses procedures which are already available to the LPC analyzer, and thus is well suited for real time analysis of noisy speech. Preliminary results show signal to noise improvements on the order of 10 to 20 db.

Introduction

As noise is acoustically added to speech, the resulting intelligibility and quality of the LPC synthesis degrades [1], [4]. This paper presents a technique which accounts for the noise present and modifies the noisy speech autocorrelations in order to suppress it. The method is based upon the simple observation that if $x(k)=s(k)+n(k)$, where $s(k)$ is clean speech, $n(k)$ is the added noise, and $x(k)$ their sum, and if the noise signal $n(k)$ were known exactly, then the desired speech autocorrelations, $R_{ss}(m)$ can be recovered from the noisy speech, $x(k)$ by computing:

$$R_{ss}(m) = R_{xx}(m) - R_{xn}(m) - R_{nx}(m) + R_{nn}(m) \qquad (1)$$

where

$$R_{xn}(m) = \sum_k x(k)n(k+m) = R_{nx}(-m)$$

$$R_{xx}(m) = \sum_k x(k)x(k+m)$$

$$R_{ss}(m) = \sum_k s(k)s(k+m)$$

Of course the noise is not known within any given analysis frame and must be approximated. A method for estimating it is the subject of this paper. Once an estimate for the local noise component is determined, Equation (1) can be used to calculate the autocorrelations of the estimated speech spectrum from which the LPC parameters can be obtained.

Constraints

Since the noise cancellation is to be integrated into the LPC analysis, it was decided that the estimation of the present noise component be done using algorithms already available to the LPC analyzer. In addition, noise characterization and estimation should depend only upon the actual background environment as recorded by the microphone.

Plan

To satisfy these constraints the noise environment is modeled by an all-pole spectrum. It is estimated by averaging autocorrelations during an initial period of non-speech activity. These averaged noise autocorrelations are then used to estimate the present frame noise component. The local noise component is estimated by convolving the average noise autocorrelations with a correlation filter whose impulse response is estimated for each frame to minimize the mean square error between the average noise and the local signal. Thus the method can be described as that of adaptively filtering past noise to approximate present noise.

Method

There are four phases to the process of Predictive Noise Cancellation. They are: (1) estimation of average background noise using LPC; (2) estimation of noise-signal correlation filter; (3) modification of noisy speech autocorrelations; and (4) calculation of final LPC parameters.

Background Noise Estimation

During the startup or a calibration period when just background noise is recorded by the microphone, the first M+1 autocorrelations representing just noise are computed and averaged together. Define:

$$R_{\bar{n}\bar{n}}(m) = \frac{1}{N_c}\sum_{I=1}^{N_c} R_{xx}^{(I)}(m) \quad m=0,1,\ldots,M \tag{2}$$

where

$$R_{xx}^{(I)}(m) = \sum_{k=0}^{N-1} x(k)x(k+m),$$

is the mth autocorrelation during the Ith frame to be averaged.

$x(k)$ = noise signal ($s(k)=0$)

N_c = number of frames to be averaged (normally 1/2 sec)

M = order of all-pole noise spectrum (set to 10)

At the completion of the calibration period, predictor coefficients representing the noise $a_n(k)$ are computed using Levinson's recursion. Finally since it will be necessary to compute crosscorrelation between the average noise $\bar{n}(k)$ and the noisy signal, $x(k)$, the first N values of the minimum phase impulse response $\bar{n}(k)$ defined from $a_n(k)$ are computed as:

$$\bar{n}(k) = -\sum_{i=1}^{M} a_n(i)\bar{n}(k-i)+G_{\bar{n}}\delta_{k,0} \tag{3}$$

$$k = 0,1,\ldots,N-1.$$

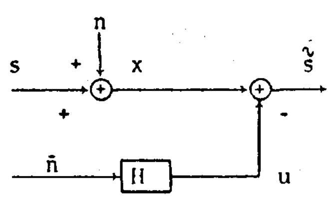
where

$$G_{\bar{n}}^2 = R_{\bar{n}\bar{n}}(0)+\sum_{i=1}^{M} a_n(i)R_{\bar{n}\bar{n}}(i)$$

N = analysis window length (Nominally 20 ms)

Noise-Signal Correlation Filter

A block diagram indicating the noise cancellation procedure is shown in Figure (1).



s: speech
n: noise
$\bar{n}$: averaged noise
$H(z) = \sum_{i=0}^{L} h(i)z^{-i}$: correlation filter
x: noisy speech
u: filtered average noise
$\tilde{s}$: noise cancelled speech

Predictive Noise Cancellation Block Diagram
Figure 1

The purpose of $H(z)$ is to modify $\bar{n}(k)$ to approximate the noise $n(k)$ within the current analysis frame. The filter is estimated using a least square criterion. The tap parameters of $H(z)$ are estimated in order to minimize

$$\sum_k [\tilde{s}(k)]^2 = \sum_k [\hat{x}(k)-\sum_{i=0}^{L} h(i)\bar{n}(k-i)]^2 \tag{4}$$

Minimizing Equation (4) with respect to $h(i)$ results in a toeplitz system of linear equations:

$$\sum_{i=0}^{L} h(i)R_{\bar{n}\bar{n}}(i-j) = R_{\bar{n}\hat{x}}(j) \quad j=0,1,\ldots,L \tag{5}$$

where

$$R_{\bar{n}\hat{x}}(j) = \sum_{k=0}^{n-1} \bar{n}(k)\hat{x}(k+j)$$

$$\hat{x}(k) = -\sum_{i=1}^{M} a_x(i)\hat{x}(k-i)+G_x\delta_{k,0}$$

It was necessary to use the LPC minimum phase approximation $\hat{x}(k)$ to $x(k)$ since $\bar{n}(k)$ is an LPC minimum phase approximation. Note that $H(z)$ can be calculated using the two pass Levinson's recursion [2], [3].

After estimating $H(z)$ it is normalized to have a spectral average of unity by dividing each tap parameter $h(i)$ by $h(0)$. This normalization was included since the purpose of $H(z)$ is to shape the spectrum of $\bar{n}(k)$ but not to increase its total energy.

Autocorrelation Modification

Referring to Figure 2, the autocorrelation of the noise cancelled speech, $\tilde{s}(k)$ are given by

$$R_{\tilde{s}\tilde{s}}(m) = R_{\hat{x}\hat{x}}(m)-R_{\hat{x}u}(m)-R_{\hat{x}u}(-m)+R_{uu}(m) \tag{6}$$

$$m = 0,1,\ldots,M$$

It is not necessary to explicitly calculate $u(k)$ in order to obtain $R_{\hat{x}u}(m)$ and $R_{uu}(m)$. These correlation terms can be calculated from $R_{\hat{x}\bar{n}}(m)$ and $R_{\bar{n}\bar{n}}(m)$ as follows:

$$\text{Since} \quad u(k) = \sum_{i=0}^{L} h(i)u(k-i) \quad k=0,1,\ldots \tag{7}$$

then

$$R_{\hat{x}u}(m) = \sum_{k=0}^{N-1} \hat{x}(k)u(k+m)=\sum_{k=0}^{N-1}\hat{x}(k)\sum_{i=0}^{L} h(i)\bar{n}(k+m-i) \tag{8}$$

or

$$R_{\hat{x}u}(m) = \sum_{i=0}^{L} h(i)\sum_{k=0}^{N-1}\hat{x}(k)\bar{n}(k+m-i) \tag{9}$$

In terms of $R_{\hat{x}\bar{n}}(m)$ we have

$$R_{\hat{x}u}(m) = \sum_{i=0}^{L} h(i) R_{\hat{x}\tilde{n}}(m-i) \quad (10)$$

Likewise $R_{uu}(m)$ can be obtained from $R_{\tilde{n}\tilde{n}}(m)$ and h(i) as follows:

$$R_{uu}(m) = h(m)*h(-m)*R_{\tilde{n}\tilde{n}}(m) \quad (11)$$

let

$$R_{hh}(m) = h(m)*h(-m) = \sum_{i=0}^{L} h(i)h(i+m) \quad (12)$$

then

$$R_{uu}(m) = \sum_{i=-L}^{L} R_{hh}(i) R_{\tilde{n}\tilde{n}}(m-i) \quad m=0,1,\ldots,M \quad (13)$$

LPC Parameter Calculation

Having calculated $R_{\hat{x}u}(m)$ and $R_{uu}(m)$, the autocorrelations $R_{\tilde{s}\tilde{s}}(m)$ of the noise cancelled speech can be computed using Equation (6). From these the LPC coefficients can be calculated using the Levinson's recursion. A stable filter will result since $R_{\tilde{s}\tilde{s}}(m)$ is positive definite.

Implementation and Results

The algorithm was inserted into an LPC vocoder simulation and tested on a data base consisting of three types of noisy speech. Type one was clean speech plus known amounts of guassian noise digitized from an analog noise generator. Type two was clean speech plus known amounts of noise recorded in a helicopter cockpit. Type three was speech recorded in a helicopter. Specifications for the vocoder simulation were as follows:

Sampling Frequency = 6.667 kHz
Analysis Window Length, N = 19.2 ms
Predictor Order, M = 10
Correlation Filter Order, L = 10
Initial Averaging Period, N_c = 0.5 sec.

Results

An audio tape demonstrating the results will be played. A coarse measure of signal to noise improvement can be calculated by comparing the energy before cancellation $R_{xx}(0)$ with the energy after cancellation $R_{\tilde{s}\tilde{s}}(0)$. An improvement on the order of 10 to 20 db was observed for all types of noisy speech. Methods for measuring improvements in quality and intelligibility are currently being investigated.

Conclusion

An integrated system for noise cancellation coupled with LPC analysis has been presented. The method assumes that noise present during the current analysis frame can be estimated by filtering an all-pole average noise spectrum through an adaptively updated linear filter. The noisy speech autocorrelations are then modified to account for the noise estimate. The algorithm is currently being tested on a variety of noisy operating environments with preliminary results showing a signal to noise improvement of 10 to 20 db.

Acknowledgements

The author would like to thank William Done and Dennis Pulsipher for their assistance in this research.

This research is supported by the Information Processing Techniques Branch of the Advanced Research Projects Agency.

References

1. B. Gold, "Robust Speech Processing", Technical Note 1976-6, Lincoln Laboratory, M.I.T., January 1976.

2. J. Markel and A. Gray, "On Autocorrelation Equations as Applied to Speech Analysis", IEEE Trans. on Audio and Elect., Vol. 21, No. 2, April 1973, p 69-79.

3. E. A. Robinson, Multichannel Time Series Analysis with Digital Computer Programs, Holden Day, New York 1967, (Subroutine EUREKA).

4. M. Sambur and N. Jayant, "LPC Analysis/Synthesis from Speech Inputs Containing Quantizing Noise or Additive White Noise", IEEE Trans. on Acoust., Speech and Signal Processing, Vol. 24, No. 6, December 1976.