SPEECH ENHANCEMENT
FOR LINEAR-PREDICTIVE-ANALYSIS-BY-SYNTHESIS CODERS
Marcin Kuropatwinski∗, Dieter Leckschat†, Kristian Kroschel‡, Andrzej Czyzewski§, Chaz Hales¶

1SIEMENS AG, Information and Communication Products, Dept. ICP CD TI24, Bocholt, Germany
2Institut f. Nachrichtentechnik, Universität Karlsruhe, Karlsruhe, Germany
3Sound Engineering Dept., Politechnika Gdanska, Gdansk, Poland

ABSTRACT

Speech coding techniques commonly used in low bit rate analysis-by-synthesis linear predictive coders (LPAS coders) create a model that emphasizes the important features of a speech signal. The utilization of these coding methods for speech enhancement is shown. Specifically, the speech signal will be modeled as the output of a cascade of an adaptive formant filter and an adaptive pitch filter, driven by a white Gaussian process with a time changing variance. The parameter and signal estimation method, which is based on the Expectation Maximization (EM) algorithm and implements this speech signal model, is investigated. The proposed approach yields better performance both in SNR and subjective impression than do speech enhancement methods which use only AR speech signal parameters.

1. INTRODUCTION

Research on speech coding has resulted in a number of solutions to the problem of accurately representing a speech signal digitally at the lowest bit rate possible. These solutions involve achievements in source coding theory, like vector quantization, as well as methods which came from the experimental trials aimed at finding the most appropriate speech signal representation. Knowledge about the nature of the speech signal gained through this research can be utilized in other areas of speech signal processing such as speech enhancement and noise reduction in mobile telephony.

It is a known fact that for noisy speech inputs, the performance of low bit rate coders deteriorates considerably. This is illustrated in Table 1, which shows the mean opinion scores (MOS), estimated using the method standardized by the ITU [1], for a GSM Enhanced Full Rate (GSM EFR) Coder and a GSM Half Rate (GSM HR) Coder. It can be seen that the reduction of the signal quality, i.e. the reduction of the MOS score of the noisy signal compared to that of the clean signal, is greater as the bit rate decreases.

<table>
<thead>
<tr>
<th></th>
<th>Reference signal</th>
<th>GSM EFR 12.2kb/s</th>
<th>GSM HR 5.6kb/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>clean speech</td>
<td>4.03</td>
<td>3.88</td>
<td>3.31</td>
</tr>
<tr>
<td>noisy speech</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5dB SNR</td>
<td>4.03</td>
<td>3.50</td>
<td>2.62</td>
</tr>
</tbody>
</table>

Table 1.: Estimated MOS scores for two GSM Coders with clean and noisy speech inputs

The quality of the transcoded noisy signal vs. bit rate is a particularly important consideration in adaptive mode rate coders. Using speech enhancement with these coders allows the coder to operate at a higher channel rate (lower source rate) in the noisy environment without compromising the transcoded signal quality.

To make the speech coders robust against noise, the vector quantizers of the short term parameters are trained with a training set including noisy data [2]. This implies that more bits must be used to code the STP parameters.

The information theory point of view (communication through an acoustic channel in the presence of noise) adds interesting insights into the speech enhancement problem. The possibility of fully recovering a clean signal, i.e. a signal whose entropy is bounded above by the bit rate of the currently available coders, about 2.4kb/sec for the telephone band speech, exists only if we assume a white Gaussian channel (a general, physically reasonable assumption) with SNR greater than −3dB. Assuming such a channel, the problem is how to make use of this possibility.

The method proposed in this paper of creating a noise reduction system that attempts to achieve this theoretical upper bound is to use the speech signal representation given by the synthesis model of an LPAS coder [3] to model the joint probability density of the speech samples in a given frame. The resulting probability density function is used within a statistical framework to estimate the parameters of the coder in the presence of additive noise. The method of estimating the formant and pitch filter parameters of an LPAS coder, based on the Expectation Maximization (EM) algorithm, is provided. This method allows us to incorporate in the estimation procedure the a priori distributions of the parameters in the form of a vector quantizer. The proposed method is compared to the method of modeling the speech signal as an AR process.

2. SIGNAL MODEL FORMULATION

The sum of statistically independent speech and noise signals is observed:

\[ r(t) = s(t) + n(t) \]

where \( r(t) \) is the time index. The noise is assumed to be white Gaussian noise with standard deviation \( \sigma_n \).

The probability density function of the speech signal is derived from the Code Excited Linear Prediction coder system [4] shown in Fig. 1. For our purposes, we model the

* Corresponding author: marcin.kuropatwinski@bch.siemens.de
excitation of the CELP system as a zero mean, variance one, white Gaussian noise process. The validity of this model is shown by observing the output of an analysis filter system excited by clean speech signals.

\[ a = [a_p, a_{p-1}, ..., a_1]^T \] - speech signal AR parameters

\[ b \] - pitch predictor tap

\[ L \] - pitch predictor lag

\[ \sigma^2_{\theta} \] - variance of the driving term (excitation signal)

\[ \rho \] (subscript) - STP synthesis filter order

\[ s_n(t) = [s(t), ..., s(t-n+1)]^T \] - vector of the recent \( n \) signal samples

\[ s = [s(1), ..., s(N)]^T \] - vector of the speech samples in the current frame.

The probability distribution of one frame of Gaussian noise with time changing variance.

\[ \text{Gaussian excitation of the CELP system as a zero mean, variance one, white Gaussian noise process.} \]

The probability distribution of one frame of Gaussian noise with time changing variance.

\[ \text{Gaussian excitation of the CELP system as a zero mean, variance one, white Gaussian noise process.} \]

The probability distribution of one frame of Gaussian noise with time changing variance.

\[ \text{Gaussian excitation of the CELP system as a zero mean, variance one, white Gaussian noise process.} \]

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.

The probability distribution of one frame of Gaussian noise with time changing variance.
E-step:

Since s and n are statistically independent, the log-likelihood function of the pdf of the complete data is equal to:

\[ \ln p(s,n | \theta, \sigma_n) = \ln p_s(s | \theta) + \ln p_n(n | \sigma_n) \quad (2) \]

with:

\[ \ln p_s(s | \theta) = C - N \ln(\sigma_s) - \frac{1}{2\sigma_s^2} \sum_{i=1}^{N} s(t) + a_i^2 s_i(t-1) + b s_i(t-L)^2 \]

and:

\[ \ln p_n(n | \sigma_n) = C - N \ln(\sigma_n) - \frac{1}{2\sigma_n^2} \sum_{i=1}^{N} n_i^2(t) \]

To compute the function \( Q(\theta, \theta_{(k)}) \), we take the conditional expectation of the log-likelihood function, given the observed vector r and the parameter values in the k-th iteration:

\[ E[\ln p_s(s | \theta)] = C - N \ln(\sigma_s) - \frac{1}{2\sigma_s^2} \sum_{i=1}^{N} E[s(t) + a_i^2 s_i(t-1) + b s_i(t-L)^2 | r, \theta_{(k)}] \quad (3) \]

Taking the conditional expectation of the log-likelihood function of the noise signal is analogous to the above operation for the speech signal.

The a posteriori second order statistics, which are needed to compute (3), are defined as:

\[ R(t | t) = E[x(t)x^T(t) | r, \theta_{(k)}] \quad (4) \]

and can be obtained using a Kalman filter [6]. From Kalman filter theory, the value \( R(t | t) \) can be expressed in terms of the error covariance matrix \( K(t | t) \) and the system state \( x(t) \):

\[ R(t | t) = K(t | t) + x(t)x^T(t) \quad (5) \]

To use the Kalman filter, we put the introduced signal model into the state space form given by:

\[ x(t) = F(t | t-1)x(t-1) + v(t) \]

(6)  (process equation)

\[ r(t) = Cx(t) + v(t) \]

(7)  (measurement equation)

The corresponding Kalman filter matrices can be found as follows. Samples of the speech signal in terms of the short and long term residuals can be expressed as:

\[ s(t) = \sum_{k=1}^{N} a_k s(t-k) + s_n(t) \quad (8) \]

where the short term residual signal given by:

\[ s_n(t) = b s_n(t-L) + s_n(t) \quad (9) \]

As shown above, the long term residual signal can be represented as white Gaussian noise with time changing variance.

For the case where the state vector of the Kalman filter is defined as:

\[ x(t-1) = \begin{bmatrix} s_n(t) \\ s_{p+1}(t-1) \end{bmatrix} \quad (10) \]

the state transition matrix based on (8) and (9) is given by:

\[ F(t | t-1) = \begin{bmatrix} B & 0_{N \times (N-1)} \\ 0_{(N-1) \times (N-1)} & A \end{bmatrix} \quad (11) \]

where:

\[ B = \begin{bmatrix} 0_{N \times (N-1)} \[ I_{N-1} & 0_{N \times (N-1)} \] \end{bmatrix} \quad (12) \]

and:

\[ A = \begin{bmatrix} 0_{N \times (N-1)} \[ f_{pp} & 0_{N \times (N-1)} \] \end{bmatrix} \quad (13) \]

In these equations, \( 0_{m \times n} \) is the \( m \times n \) zero matrix, and \( I_{m \times n} \) is the \( m \times n \) identity matrix.

The covariance matrix of the process noise vector, \( v(t) \) in (6), is given by:

\[ Q(t | t) = \sigma_v^2 g g^T \]

(14)

The additive noise \( n(t) \) in the state space equations as the measurement noise, \( v(t) \) in (7), and the corresponding covariance matrix is:

\[ Q_n = \sigma_n^2 \quad (15) \]

The measurement matrix is given by \( C = [0_{(N-1) \times p} \ [1] \] .

The Kalman filtering operation is accomplished for \( t = 1...N \) by the following steps:

1. Compute the predicted state error covariance matrix:

\[ K(t | t-1) = F(t | t-1) F(t | t-1)^T + Q(t) \]

2. Compute the Kalman gain matrix:

\[ G = K(t | t-1) C^T (C K(t | t-1) C^T + Q_2)^{-1} \]

3. Compute the state estimate:

\[ \hat{x}(t) = F(t | t-1) x(t-1) + G(r(t) - C F(t | t-1) x(t-1)) \]

4. Compute the filtered state error covariance matrix:

\[ K(t | t) = K(t | t-1) - G C K(t | t-1) \]

The initial values for the given frame are taken from the frame processed previously.

M-step:

Using the results from the E-step of the EM algorithm the parameters maximizing the \( Q(\theta, \theta_{(k)}) \) function are computed according to the following set of formulas:

\[ a_{(k+1)i} = \sum_{i=1}^{N} E[s_i(t-1) s_i(t-1) | \theta_{(k)}, \theta_{(k)}]^{-1} \sum_{i=1}^{N} E[s_i(t-1) s(t-1) | \theta_{(k)}, \theta_{(k)}] \]

The computation of the formant filter parameter is derived under the assumption that the interaction between the speech
samples at the time instant \( t \) and the past short residual signal is negligible. As shown in [7], this is a reasonable assumption which considerably simplifies the M-step, which is otherwise an iterative procedure.

The remaining speech signal parameters are computed according to:

\[
L_{k+1} = \arg \max_{L} \left( \sum_{j=1}^{N} \mathbb{E} [s_{p+1}(t) s_{n}(t-L)] \right) \tag{17}
\]

\[
h_{k+1} = - \frac{1}{N} \sum_{j=1}^{N} \mathbb{E} [s_{p}(t) s_{n}(t-L_{k+1})] \tag{18}
\]

\[
\sigma_{n}^{2} = \frac{1}{N} \left[ \mathbb{E} [h_{k+1} : a_{k+1}^{T}] \right] \times \left( \frac{1}{N} \sum_{j=1}^{N} \mathbb{E} [s_{p}(t-L_{k+1}) : s_{n}(t-L_{k+1})] \right) \times \left[ \mathbb{E} [h_{k+1} : a_{k+1}^{T}] \right] \tag{19}
\]

The formant filter parameters for the first iteration of the EM algorithm were derived from a correlation sequence computed in the frequency domain, which was found using the spectral subtraction algorithm. The starting parameter of the long term predictor was computed using the standard open-loop LTP parameter estimation procedure running on the noisy observation.

This straightforward implementation of the proposed algorithm is computationally very complex, however by exploiting the fact that the Kalman filter matrices are sparse, further research can be done to find a more efficient way to program the algorithm.

4. EXPERIMENTS

Computer simulations were carried out to compare the performance of the proposed method with the established speech signal estimation method using only STP parameters [8]. The utterance of duration ca. 3 seconds was contaminated with synthetic noise at three different signal to noise ratios (SNR) to assess the proposed method. In Tables 2 and 3, the SNR results for the conventional Kalman filtering method, STP only, and the new method, both STP and LTP, are shown.

The spectrograms, in Figures 3a and b, show the effect of the new noise reduction method. These spectrograms correspond respectively to the speech signal contaminated with 0dB SNR noise, and the processed signal resulting from our noise reduction method. The periodicity of the speech signal can be seen in the harmonic structures of the spectrograms.

5. DISCUSSION

It has been shown that the application of speech coding techniques for speech enhancement can result in significantly better quality of the enhanced speech. A new method of estimating the speech signal and the parameters of the typical LPAS coder in the presence of noise has been proposed. Further research will be focused on STP- and LTP-parameter estimation in the presence of noise and also vector quantization of the STP parameters.

6. REFERENCES