Abolishing the Noise from Speech by Estimating the Variance using EM Algorithm

Gowder Praveena Hiriyan M.C.A., M.Phil.,
Department Of Computer Science,
KG College Of Arts and Science

ABSTRACT
A speech noise disturbs in almost all acoustic environments. The speech signal recorded by a microphone is generally infected by noise originating from various sources and it can corrupt or change the characteristics of the speech signals and degrade the speech quality and intelligibility in human-to-machine communication systems. To improve the quality of speech, here make use of the EM algorithm to estimate the clear variance and coupling of errors statistic.

1. INTRODUCTION
Speech enhancement as various techniques such as signal-subspace embedding, time-domain iterative, spectral subtraction etc. All these methods are frequently result in audible distortion of the signal, and are far from satisfactory in the real-world noisy environments. Recent neural network-based filtering methods utilize data sets where the clean speech is a target signal for training. These methods are often effective within the training set, but tend to generalize this for the actual speech with varying signal and noise levels (a review of neural-based approaches can be found in [16]). Furthermore, the neural network models in these methods do not fully take into account the non-stationary nature of speech. In the approach presented here, we assume the availability of only the noisy signal. Effectively, a order of neural networks is trained on the few noisy speech signals, resulting in a non-stationary model which can be used to remove noise from the given speech.

Nonlinear Speech Model
A noisy speech signal can be accurately modeled as a with both process and additive observation noise:

\[
x(k) = f(x(k-1),\ldots,x(k-M),W + V(K) \tag{1}
\]
\[
Y(k) = x(k)+n(k) \tag{2}
\]

where corresponds to the true underlying speech signal driven by process noise and is a nonlinear function of past values parameterized by . The speech is only assumed to be stationary over short part, with each part having a different model. The available observation is , which contains additive noise. If is linear, this reduces to the classic Linear Predictive Coding (LPC) model of speech. The optimal estimator given the noisy observations is .

The most direct way to approximate this conditional expectation would be to train on a set of clean data in which the true may be used as the target to a neural network. Our assumption, however, is that the clean speech is never available; the goal is to estimate itself from the noisy measurements alone. In order to solve this problem, we assume that is in the class of feed forward neural network models, and compute the dual estimation of both states and weights based on a Kalman filtering approach. In this paper we provide a basic description of the algorithm, followed by a discussion of experimental results.

2. DUAL EXTENDED KALMAN FILTERING
By formulating the dual estimation problem in a state-space framework, we can use Kalman filtering methods to perform the estimation in an efficient, recursive manner. At each time point, the Kalman filter provides an optimal estimation by combining a prior prediction with a new observation. Connor et al.[3] proposed using an extended Kalman filter with a neural network to perform state estimation alone. Puskorius and Feldkamp [13] and others have posed the weight estimation in a state-space framework to allow for efficient Kalman training of a neural network. In prior work, we extended these ideas to include the dual Kalman estimation of both states and weights for efficient maximum-likelihood optimization for robust nonlinear prediction, estimation, and smoothing [14]. The work presented here develops these ideas in the context of speech processing.

To apply the EKF, we first put the autoregression of Equation 1 and 2 in state-space form:

\[
X(k) = f(x(k-1)) + Bv(k) \tag{3}
\]
\[
Y(k) = Cx(k) + n(k) \tag{4}
\]

Where and . If the model is linear, then takes the form, and can be written as where is a matrix in controllable canonical form. We initially assume the noise terms and are white with known variances.

2.1. State Estimation
A linear model with known parameters, the Kalman filter (KF) algorithm can be readily used to estimate the states [8]. At each time step, the filter computes the linear least squares estimate.

In the linear case with Gaussian statistics, the estimates are the minimum mean square estimates. With no prior information on , they reduce to the maximum-likelihood estimates. When the model is nonlinear, the KF cannot be applied directly, they reduce to the maximum-likelihood estimates. When the model is nonlinear, the KF cannot be applied directly, but requires a linearization of the nonlinear model at each time step. The resulting algorithm is called the extended Kalman filter (EKF), and effectively approximates the nonlinear function with a time-varying linear one. The EKF algorithm is as follows.

\[
x = (k) = F[X^{k-1},W] \tag{6}
\]
\[
P_{x}(k) = A(k)P_{x}(k-1)A^{T}(k) + \sigma \tag{7}
\]
Where \( A(k) = F[x^a, w] \)
\[
    X^a(k-1)\quad (8)
\]
\[
    K(k) = P_x - (X(k)C^T(CP_x^T(K) + zn))^{-1} \quad (9)
\]
\[
    P_x(k) = I - K(k)C \quad (10)
\]
\[
    X^a(k) = X^a(k-1) + K(k)(y(k) - CX^a(X)) \quad (11)
\]

Note that the derivative in Equation 8 corresponds to the linearization of the neural network at the current operation point. This can be found by a single application of standard backpropagation. When the weights are not available, they must be replaced by an estimate.

### 2.2 Weight Estimation

Because the model for the speech is not known, the standard EKF algorithm cannot be applied directly. We approach this problem by constructing a state-space formulation for the underlying weights as follows:

\[
    W(k) = w(k-1) + v(k) + n(k) \quad (12)
\]

where the state transition is simply an identity matrix, and the neural network plays the role of a time-varying nonlinear observation on \( W(k) \). These state-space equations for the weights allow us to estimate them with a second EKF:

\[
    W^\wedge(k) = w(k-1) \quad (14)
\]

\[
    E_W(k) = P_w(k-1) + H(k)P_w(k-1)H^T(k) + Z^2n + Z^2n(k-1) \quad (15)
\]

\[
    E_W(k) = (1-Kw(k)H(k))P_w(k) \quad (16)
\]

where \( H(k) = Zx(k) \quad (17) \)

\[
    w^\wedge(k) = w^\wedge(k-1) + Kw(k)(y(k) - X(k)) \quad (19)
\]

The use of the EKF for weight estimation can be related to Recursive Least Squares (RLS), and thus represents an efficient second-order on-line optimization method. Note that when \( n(k) \) is not available, it must be replaced in the weight filter by an estimate. A maximum-likelihood interpretation of the EKF detailing the implications on the use of \( n(k) \) is given in [12].

### 3. EXPERIMENTS

#### 3.1. Nonstationary White Noise

To process noisy speech, the method is applied to successive 64ms windows of the signal (512 points at 8kHz sampling), with a new window starting every 8ms (64 points). A normalized Hamming window is used to emphasize data in the center of the window, and deemphasize data in the periphery. The standard EKF equations are also modified to reflect this windowing in the weight estimation. The result of applying the Dual EKF to a speech

#### 3.2. Colored Noise

For most real-world speech applications, we cannot assume the noise is white. For colored noise, the state-space equations 3 and 4 need to be adjusted before Kalman filtering techniques can be employed. Specifically, the measurement noise process is given its own state-space equations,

\[
    n(k) = A_n n(k-1) + B_n v(n)(k) \quad (20)
\]

\[
    n(k) = C_n n(k) \quad (21)
\]

where \( n(k) \) is a vector of lagged values of \( n(k) \), is white noise, is a simple state transition matrix in controllable canonical form, and \( A_n, B_n, C_n \) are of the same form as \( A, B, C \) and given in Equation 5. Note that this is equivalent to an autoregressive model of the colored noise, which may be fit from a small section of the noisy signal where speech
is not present. With this formulation for the colored noise, it is straightforward to augment both the state and the weight with , and write down combined state equations. Specifically, Equations 3 and 4 are replaced by:

\[
y(k)=\text{tr} \left [ \mathbf{C}_n \right ] \left [ \mathbf{x}(k)/n(k) \right ] \\
y(k)=f(x(k-1),w(k))+\text{Cnn}(k)+v(k)
\]

The noise processes in these state equations are now white, and the Dual EKF algorithm can be used to estimate the signal. Note that the colored noise explicitly affects not only the state estimation, but also the weight estimation.

An actual recording of highway noise through a cellular phone was added to a speech signal to produce the data shown in Figure 3 (3,500 points). Figure 4 shows a similar experiment with pink noise added (spectrograms shown in Figure 5). In both cases, the noise model and process noise variances were assumed known. Experiments were also run with estimated values using only the noisy speech, as described in the next section. Table 1 summarizes the results for several different initial SNR levels. Spectral subtraction results are included for comparison.

3.3. Estimating Noise Variances

In the implementation of the Dual EKF, it is assumed that the variances of and (or the SNR) are known quantities. Assuming stationarity of the additive noise, the noise variance (or its full autocorrelation for determining ) may be estimated from segments of the data that do not contain speech. Alternative methods for tracking nonstationary noise are given in [7, 2, 15].

To estimate the process noise variance (assuming an LPC model for the signal), Lim and Oppenheim [9] used an expression for the inverse Fourier transform of the signal power (which is a function of ). We have developed an alternative approach by noting that the process noise variance can be estimated directly by considering the relationship between the residual AR prediction error for clean and noisy speech [15].

All these approaches, however, are relatively “ad-hoc”, and estimating the noise variances remains a critical area for future work. Our current direction is to treat and as additional parameter values which may be optimized within the Kalman and maximum-likelihood framework.

Computational Statistics EM algorithm: Variance estimation

We consider again the problem of estimating the parameters in a mixture of a normal distribution \(N(\mu, \sigma)\) and a uniform distribution \(U([a, a])\) \(U([-a, a])\), where \(a\) is a known constant. The observed data are an iid sample \(w_1, \ldots, w_n\)

\[g(w_i) = (w_i; ) + (1 )c; \]

from \(W\) with pdf, \(c = (2a)^{-1}\), is the proportion of the Table 1: Comparison of methods on colored noise. Results for known noise statistics are indicated by (k), estimated statistics by (e). Spectral subtraction results were obtained using the Duke University Matlab Speech Processing Toolkit. While spectral subtraction was able to suppress noise, distortion of the signal resulted in poor SNR values.

Normal distribution in the mixture and \(a = ( ; ; )^T\) is the vector of pa-rameters. Typically, the uniform distribution corresponds to outliers in the data. The proportion of outliers in the population is then 1.\n
We have seen how to nd the MLE of using the EM algorithm. We know want to estimate the variance of .

1. \(S = (0; 1; 0.9)\) and \(a = 5\). Generate \(N = 1000\) samples of size \(n = 100\). For each sample, compute \(\hat{\sigma}\) using the EM algorithm. Estimate the variance of \(\sigma^2\).

2. We now consider one sample \(w_1; \ldots; w_n\) and we wish to estimate \(\sigma^2\) from that sample, without knowing the true value of . We will use two methods.

(a) Louis’ method: compute \(i_n(\sigma^2)\) and estimate \(i_{\hat{\sigma}}(\sigma^2)\) by Monte Carlo simulation; compute an estimate of \(i_n(\sigma^2)\) using the missing information principle equation, and its inverse \(i_{\hat{\sigma}}(\sigma^2)\).\n
(b) Bootstrapping: generate \(B = 1000\) bootstrap samples. Estimate \(\text{Var}(\sigma^2)\) by the sample variance of the bootstrap estimates of .

3. Compare the estimates of \(\text{Var}(\sigma^2)\) obtained by the different methods.

4. CONCLUSION AND FUTURE WORK

We have presented the Dual EKF algorithm with preliminary results on its application to speech enhancement in the presence of both nonstationary and colored noise. Initial results compare favorably to current state-of-the-art techniques. However, future work must involve more substantial evaluations based on both objective and subjective criteria. In addition, future algorithmic work will include alternative approaches to variance estimation, as well as the coupling of error statistics, windowing aspects, recurrent training implications, forward-backward methods for smoothing, and issues relating to maximum-likelihood estimation and the EM approach.

REFERENCES