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 avails a target signal for training. These methods are often effective within the training set, but tend to generalize this for the ac tual speech with varying signal and noise levels (a review of neural based approaches can be found in [16]). Furthermore, the net- work 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 signal, 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),…….x(k-M),W + V(K) (1)
Y(k)=x(k)+n(k), (2)

where corresponds to the true underlying speech signal driven by process noise and is a nonlinear function of past values of parameterized by . The speech is only assumed to be stationary over short part, with each part having a differ ent model. The available observation is , which contains addative 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 frame- work, 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. Puskorious 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) (3)
Y(k)=Cx(k)+n(k) (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 Kalmanfilter(KF) algorithm can be readily used to estimate the states[8].Aa 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 the 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.

$x \quad -(k) = F[X^{K-1}, W]$ (6)
Px-(k)=A(k)Px(k-1)A^T (k) + B€B^T (7)

Where $A(k)=\dfrac{F[x\hat{},w]}{X\hat{}(k-1)}$ (8)

$K(k)=px\text{-}(x)C^T(CPX\text{-}(K)C^T+zn)\text{-}1)$ (9)

$Px(k)=(I-K(k)C(PX\text{-}(K)$ (10)

$X\hat{}(k)=x\hat{}\text{-}(k)+K(k)(y(k)-CX\hat{}(X))$ (11)

Note that the derivative in Equation 8 corresponds to the lineariza- tion 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 separate state-space formulation for the underlying weights as follows:

$W(k)=w(k-1)$ (12)

$y(k)=f(x(k-1),w(k))+v(k)+n(k)$ (13)

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

$W\hat{}\text{-}(k)=w(k-1)$ (14)

$Pw\text{-}(k)=pw(k-1)$ (15)

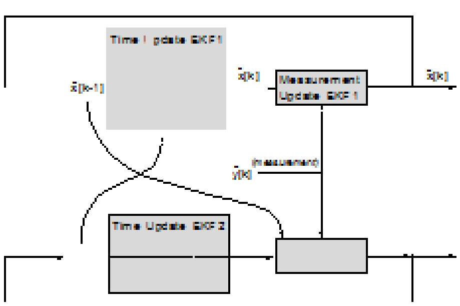
$Kw(k)=Pw\text{-}(K)[H^T(K)Pw\text{-}(k)\ H^T(k)+Z\hat{}2n+Z\hat{}2n]\text{-}1$ (16)

$Pw(k)=(I-Kw(k)H(k))Pw\text{-}(k))$ (17)

Where $H(k)=\dfrac{zx\text{-}(k)}{zw}$ (18)

$w\hat{}(k)=w\hat{}\text{-}(k)+Kw(k)(y(k)-x\text{-}(k))$ (19)

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



The linearization of the network in Equation 18 can be com- puted as a *dynamic derivative* (or full partial derivative) [17] to ac- count for the recurrent nature of the state-estimation filte r, includ- ing the dependence of the Kalman gain on the weights. Un- fortunately, the calculation of these derivatives is computationally expensive. Alternatively, this can be avoided completely by ignor- ing the dependence of on in Equation 6, resulting in a static linearization of the network. Early results on dynamic deriva- tives do not indicate performance advantages over the static deriva- tive. Thus, we report results only for static linearization in this pa- per.

2.3. Dual State and Weight Estimation

The essence of the Dual EKF algorithm is to run the state EKF and weight EKF in parallel (see Figure 1), simultaneously updating es- timates of and . At each time step, the current estimate of is used by the weight filter, and the current estimate of is used by the state filter. For finite data sets, the algorithm is run iteratively over the data until the weights converge.

This approach to dual estimation can be justified within a max- imum-likelihood framework and can also be related to the Expecta- tion Maximization (EM) algorithm. The approach is also related to work done by Nelson [11] in the linear case, and to Matthews' neu- ral approach [10] to the recursive prediction error algorit hm [6]. In the speech literature, the method is most closely related to Lim and Oppenheim's approach to fitting LPC models to degraded speech [9]. It also relates to Ephraim's model-based approach [5], but uses nonlinear estimation to fit the given data instead of using a fixed number of prespecified linear models.

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 estima- tion. 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,

$n(k)=A_n n(k-1)+Bnvn(k)(20)$ (20)

$n(k)=C_n n(k)$ (21)

where is a vector of lagged values of , is white noise, is a simple state transition matrix in controllable canoni- cal form, and and are of the same form as 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 straight for-ward 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)=[C Cn] [x(k)/n(k)] (22)

y(k)=f(x(k-1),w(k))+ Cnn(k)+ v(k) (23)

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 and 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 vari- ances of and (or the SNR) are known quantities. As- suming 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 not- ing that the process noise variance can be estimated directly by considering the relationship between the residual AR prediction er- ror for clean and noisy speech [15].

All these approaches, however, are relatively "ad-hoc", and es- timating the noise variances remains a critical area for future work. Our current direction is to treat and as additional parame- ters 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 mix-ture 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;) = (w; ;) + (1)c; (1)

from W with pdf.

where is the normal 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, dis- tortion of the speech signal resulted in poor SNR values.

Normal distribution in the mixture and = (; ;)^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 .

We have seen how to nd the MLE ^{b} of using the EM algorithm. We know want to estimate the variance of ^{b}.

1. Set = (0; 1; 0:9) and a = 5. Generate N = 1000 samples of size n = 100. For each sample, compute ^{b} using the EM algorithm. Estimate the variance of ^{b}.
2. We now consider one sample $w_1; : : : ; w_n$ and we wish to estimate Var(^{b}) from that sample, without knowing the true value of . We will use two methods.
   (a) Louis' method: compute $^b i_x(^b)$ and estimate $^b i_{zjy}(^b)$ by Monte Carlo simulation; compute an estimate of $^b i_y(^b)$ using the missing information principle equation, and its inverse $^b i_y(^b)$ ^{1}.
   (b) Bootstrapping: generate B = 1000 bootstrap samples. Estimate Var(b) by the sample variance of the bootstrap estimates of .
3. Compare the estimates of Var(b) 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 favor- ably 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 train- ing implications, forward-backward methods for *smoothing*, and issues relating to maximum-likelihood estimation and the EM approach.

REFERENCES

[1] S.F. Boll. Suppression of acoustic noise in speech using spec- tral subtraction. *IEEE ASSP-27*, pp. 113-120, April 1979.

[2] J. Cohen. Application of an auditory model to speech recog- nition. *Journal of the Acoustical Society of America*, vol. 85, no. 6, 1989.

[3] J. Connor, R. Martin, L. Atlas. Recurrent neural networks and robust time series prediction. *IEEE Tr. on Neural Networks*, March 1994.

[4] J. Deller, J. Praokis, J. Hansen. Discrete-Time Processing of

Speech Signals. Macmillan Publishing Company, NY, 1993.

[5] Y. Ephraim. Statistical model-based speech enhancement systems. *Proceedings of the IEEE,* Vol. 80, October 1992.

[6] G. Goodwin, K.S. Sin. *Adaptive Filtering Prediction and Control*. Prentice-Hall, Inc., Englewood Cliffs, NJ. 1994.

no. 2, 1994.

[7] H.G. Hirsch. Estimation of noise spectrum and its applica- tion to SNR-estimation and speech enhancement. *Technical Report TR-93-012, International Computer Science Institute*. 1993.

[8] F. Lewis. *Optimal Estimation*. John Wiley & Sons, Inc. New York, 1986.

[9] J. Lim, A. Oppenheim. All-pole modeling of degraded speech. *IEEE Trans. Acoust., Speech, Signal Process.,* vol. 26, June 1978.

[10] M. B. Matthews and G. S. Moschytz. The identification of nonlinear discrete-time fading-memory systems using neural network models. *IEEE Transactions on Circuits and Systems- II*, 41(11):740–51, November 1994.

[11] L. Nelson, E. Stear. The simultaneous on-line estimation of parameters and states in linear systems. *IEEE Tr. on Auto- matic Control*, February 1976.

[12] A. Nelson and E. Wan. A Two-observation Kalman frame- work for maximum-likelihood modeling of noisy time series, submitted to IJCNN'98.

[13] G. Puskorious, L. Feldkamp. Neural control of nonlinear dy- namic systems with Kalman filter trained recurrent networks. *IEEE Trn. on NN*, vol. 5,

[14] E. Wan, A. Nelson. Dual Kalman filtering methods for non- linear prediction, estimation, and smoothing. In *NIPS96 Pro- ceedings*, 1997.

[15] E. Wan and A. Nelson. Neural dual extended Kalman filter- ing: applications in speech enhancement and monaural blind signal separation. *Neural Networks for Signal Processing VII: Proceedings of the 1997 IEEE Workshop*, ed. Principe, Morgan, Giles, Wilson, 1997.

[16] E. Wan and A. Nelson. Networks for Speech Enhancement, in *Handbook of Neural Networks for Speech Processing*, Edited by Shigeru Katagiri, Artech House, Boston, 1998.

[17] P. Werbos. Neural networks, system identification, and con- trol in the chemical process industries. In *Handbook of In- telligent Control: Fuzzy, Neural, and Adaptive Approaches*, edit