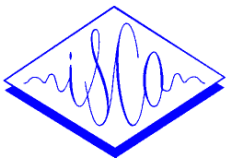
ISCA Archive
http://www.isca-speech.org/archive



3rd European Conference on
Speech Communication and Technology
EUROSPEECH'93
Berlin, Germany, September 19-23, 1993

# Noise-Adaptive Hidden Markov Models Based On Wiener Filters

S. V. Vaseghi B. P. Milner

School of Information Systems, University of East Anglia, Norwich, NR4 7TJ, UK,

10.21437/Eurospeech.1993-245

## Abstract

In noisy speech recognition, Wiener filters may be applied either directly to the noisy speech or, alternatively, the filters can be used to adapt the HMM mean cepstral vectors. In this paper we present experimental results which demonstrate that Wiener filters used for the adaption of HMM cepstral means perform better than the direct application of Wiener filters to the noisy signal.

Experiments indicate that the variance of the cepstral features decrease with increasing noise. A theoretical explanation of the decrease in variance is presented.

## 1- INTRODUCTION

Performance of hidden Markov model (HMM) speech recognition systems, trained in a noise-free environment, degrades rapidly with decreasing signal to noise ratio. HMM speech recognition systems achieve best performance when the models are trained and operated in matched environments. For most applications this is impractical because the operating environment varies with time and space, and it is necessary to employ a noise compensation scheme.

Optimal noise compensation is a probability maximisation problem. Consider the most common case of a signal x observed in additive independent random noise n. The noisy observation is y = x + n. For uncorrelated signal and noise, the probability of a signal estimate $\hat{x}$ given the noisy observation y is

$$p(\hat{x} \mid y) = p(x=\hat{x},\ n=y-\hat{x}) = p_x(\hat{x})\ p_n(y-\hat{x}) \quad (1)$$

Assuming that the signal and noise are normally distributed; $p_X(x) = N(x, \mu_X, \sigma_X)$ and $p_N(n) = N(n, \mu_n, \sigma_n)$, we have :

$$p(\hat{x}|y)=\frac{1}{2\pi\sigma_n\sigma_x}exp\left(-\frac{\sigma_n^2(\hat{x}-\mu_x)^2+\sigma_x^2(y-\hat{x}-\mu_n)^2}{2\sigma_x^2\sigma_n^2}\right) \quad (2)$$

The maximum log likelihood estimate is

$$\hat{x} = \frac{\sigma_x^2}{\sigma_x^2+\sigma_n^2}(y-\mu_n) + \frac{\sigma_n^2}{\sigma_x^2+\sigma_n^2}\mu_x \quad (3)$$

The estimate $\hat{x}$ is a weighted linear interpolation of the unconditional expected value of x, $\mu_X$, and the observed value (minus expected value of noise) y - $\mu_n$. The optimal estimator of eq(3) requires knowledge of the mean and variance of the signal and noise processes. When there is no prior knowledge of the signal statistics, the estimate of x is obtained using spectral subtraction as the term y-$\mu_n$ of eq(3). When in addition to the mean noise spectra, we also have the mean of signal spectra then the best linear filter is the Wiener filter.

In signal restoration, the classical methods for noise removal are the Wiener filter and spectral subtraction. In pattern classification, an alternative to filtering is to adapt the mean and variances of speech model prototypes [Roe 1987] [Nadas 1989] [Varga 1990] [Gales, Young 1992]. The Wiener filter is based on the least mean squared error optimisation and is the maximum likelihood filter for signals that are Gaussian distributed. Although the speech spectrum is log-normal, Wiener filters produce a substantial improvement in subjective quality and recognisability (by an automatic speech recognition system) of noisy speech.

## 2- WIENER FILTERS IN HMM SPEECH RECOGNITION SYSTEMS

For uncorrelated signal and noise processes the Wiener filters in time and frequency domains are given by the following equations :

$$w=[R_x+R_n]^{-1}P_x \Leftrightarrow W(\omega)=\frac{\mu_X(\omega)}{\mu_X(\omega)+\mu_N(\omega)} \quad (4)$$

Where R, P and $\mu(\omega)$ denote autocorrelation matrix, autocorrelation vector and power spectrum respectively, and the operator $\Leftrightarrow$denotes the Fourier transform relation. The Wiener filter is based on the mean spectra of

signal and noise and does not make use of information about the variance.

Application of the Wiener filter requires prior knowledge of signal and noise power spectra. The noise power spectra may be estimated and updated from speech inactive periods. The assumption is that between the estimation periods, noise characteristics do not change substantially. The power spectra of speech may be obtained from the mean cepstral vectors contained in HMMs. There are two different methods of using Wiener filters with HMMs : (a) Noise adaptive HMMs based on Wiener filters, and (b) direct application of Wiener filters derived from the most likely state sequence.

### 2.1 - Noise Adaptive HMMs Based on Wiener Filters

HMM based Wiener filters may be implemented by adaption of the mean cepstral vectors of HMM states. In the frequency domain Wiener filtering is a multiplication operation given as

$$\hat{X}(\omega)=Y(\omega)\cdot W(\omega)=Y(\omega)\cdot\frac{\mu_x(\omega)}{\mu_x(\omega)+\mu_n(\omega)} \tag{5}$$

and the equivalent Wiener filtering operation in the cepstral domain is

$$c_{\hat{x}}(m) = c_y(m) + c_w(m) \tag{6}$$

where $c_y(m)$ and $c_w(m)$ are the cepstra of the noisy signal and Wiener filter respectively. The cepstral transform of the Wiener filter equation gives

$$c_w(m)=c_{\mu_x}(m)-\underbrace{DCT\left(log\,[\mu_x(\omega)+\mu_n(\omega)]\right)}_{c_{wd}(m)} \tag{7}$$

where $c_{\mu x}(m)$ is the speech mean cepstral vector. In an HMM, $c_{\mu x}(m)$ is a mean vector of an HMM state. From eq(6) and (7) the filtered signal is given by

$$c_{\hat{x}}(m)=c_y(m)+c_{\mu_x}(m)-c_{wd}(m) \tag{8}$$

substituting the filtered signal of eq(8) into the HMM state observation Gaussian scoring function

$$exp\,[-\frac{(\overbrace{[c_y(m)+c_{\mu_x}(m)-c_{wd}(m)]}^{filtered\ signal}-c_{\mu_x}(m))^2}{2\sigma_x^2}]$$

$$= exp\,[-\frac{(c_y(m)-c_{wd}(m))^2}{2\sigma_x^2}] \tag{9}$$

From eq(9) it is evident that Wiener filtering is equivalent to replacement (or adaption) of the mean cepstral vector of each mixture component of the HMM with $c_{wd}(m)$. The block diagram of the adaption system is shown in figure(1).

### 2.2 - Wiener Filters Derived from Maximum Likelihood State Sequence

In this method, a maximum likelihood signal spectral sequence, obtained from the maximum likelihood state sequence of HMMs, is used to program a sequence of state-dependent Wiener filters [Vaseghi 1993]. The main disadvantage of this method is that at low signal to noise ratios, the maximum likelihood sequence is often incorrect. Experiments show that at below 20 dB SNR it is better to use the average state sequence instead of the Viterbi maximum likelihood sequence, for programming the Wiener filters. In general for signal classification, model adaption based on Wiener filters achieves better results than the direct application of Wiener filters to signal.

## 3 - Effects of Noise on the Variance of Cepstral Features

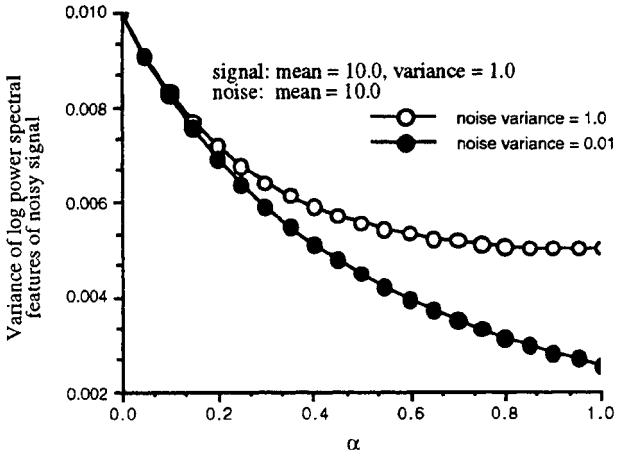
The effects of noise on linear power spectral features is to increase both the mean and the variance. The effects of noise on the variance of cepstral features are less well known. A large number of experiments indicate that the variance of cepstral features decreases with increasing noise. This may be explained by considering the nonlinear relation between the statistics of log-normal distributed power spectra of a noisy signal y=x+n, and those of normally distributed log power spectrum of y, logy :

$$\sigma^2_{logy}(ii) = \log\left(1+\frac{\sigma_x^2(ii)+\alpha^2\sigma_n^2(ii)}{[\mu_x(i)+\alpha\mu_n(i)]^2}\right) \tag{10}$$

Where $\mu(i)$ $\sigma^2(ii)$ and are the mean and variance, and the variable $\alpha$ is a measure of the amount of noise added to the signal, $\alpha$ increases for decreasing SNR. It can be seen that the variance of log spectral coefficients are a function of both the mean and the variance of spectral power variables. For a simple illustration of the decrease in variance as a function of signal to noise ratio consider the case when $\sigma_x^2(ii) = \sigma_n^2(ii)$ and $\mu_s(i) = \mu_n(i)$, therefore

$$\sigma^2_{logy}(ii) = \log\left(1+\frac{\sigma_x^2(ii)\,(1+\alpha^2)}{\mu_x(i)\,[1+\alpha]^2}\right) \tag{11}$$

The variance of noisy cepstrum is a function of the term $(1+\alpha^2)/(1+\alpha)^2$, and for $\alpha > 0$ this term decreases with increasing value of $\alpha$ (i.e decreasing SNR). The changes in variance of the cepstral variables is similar to that of log spectral variables as the two are related by a linear DCT transform. Figure(2) shows the decrease in log feature variance with the increasing noise power.




Figure( 2) - Illustration of decrease in variance of log features with increasing noise power.

## 4 - Experimental Results

The experiments are based on a data set of spoken English alphabet. For each of the 26 letters, the HMM was trained using 52 speakers with 3 utterances per speaker. The test data set consisted of a similar number of utterances from a different set of speakers. The feature vector consists of 10 cepstral coefficients including the zero$^{th}$ coefficients. To minimise the effects of variation in energy, all utterances were normalised to have the same power. The baseline HMM recogniser chosen is an 8-state left-right HMM without skip-state transition, and with multivariate Gaussian distribution and diagonal covariance matrices. Each state has 7 mixture densities.

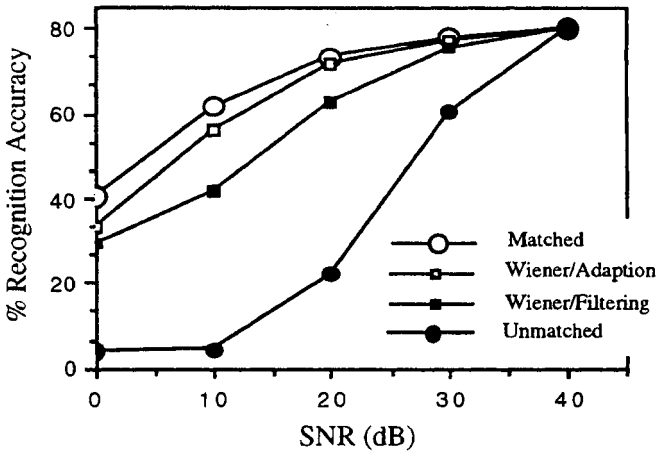
The recognition performance of HMMs for various signal to noise ratios were obtained under matched and unmatched conditions. These provide the upper and lower bounds on the performance of a noise compensation system.

In the direct application of Wiener filters, the signal spectral means from the HMM maximum likelihood state sequence, together with an estimate of noise from speech inactive periods, were used to program a sequence of state dependent Wiener filters. The filter sequence was applied to the noisy signal and a revised probability score for the filtered signal and each model was calculated. When the maximum likelihood state sequence was replaced with the mean state sequence the recognition performance improved significantly for low signal to noise ratios.

For noise adaption based on Wiener filters, the mean cepstral vector of each mixture component was converted to the power spectrum and adapted by adding the noise power spectrum to it. Table-1 and figure(3) show that noise adaption achieves better results.

Table-1

| SNR (dB) | Recognition Accuracy (%) | | | |
|---|---|---|---|---|
| | Matched | Unmatched | Wiener/ Adaptive | Wiener/ Filtering |
| 30.0 | 77.5 | 60.4 | 77.2 | 75.8 |
| 20.0 | 73.2 | 22.3 | 71.5 | 62.6 |
| 10.0 | 61.4 | 4.5 | 56.2 | 41.5 |
| 0.0 | 40.8 | 3.9 | 33.6 | 29.7 |




Figure(3) - Recognition accuracy of HMM Vs SNR for match, unmatched, Wiener-based adaption, and direct application of Wiener filters

## 5 - Conclusion

For HMM based speech recognition systems, Wiener filtering is equivalent to the adaption of state mean cepstral vectors. The adaption of the HMMs produce better result because : a) The adapted cepstral mean vectors approach those of the models trained in matched conditions and b) direct application of Wiener filters has the problem of unreliable maximum likelihood state sequence from which the Wiener filters are derived, this is particularly true in low SNR. Although the Wiener filters do not use information about the noise variance, the application of Wiener filters produces significant improvement as demonstrated by the experimental results.

**Acknowledgement**

The authors would like to acknowledge the support of Science and Engineering Research Council, and the British Telecom Research Laboratories, Martlesham, Ipswich, UK.

## REFERENCES


Boll, S. F., (1979) "Suppression of acoustic noise in speech using spectral subtraction", IEEE Trans. Acoust., Speech and Signal Proc., vol. ASSP-29, pages 113-120, April.

Carlson, B. A., Clements, M. A. (1992),"Speech Recognition in Noise Using a Projection-Based Likelihood Measure for Mixture Density HMM's", IEEE Proc. ICASSP-92, pages I-237-I240, San Francisco.

Gales,M.J.F., Young, S.,(1992),"An Improved Approach to the Hidden Markov Model Decomposition of Speech and Noise", IEEE Proc., ICASSP_92, pages I-223-I-226, San Francisco.

Lim, J. S., and Oppenheim, A. V., (1978),"All-pole modelling of degraded speech", IEEE Trans. Acoust., Speech and Signal Proc., vol. ASSP-26, pages. 197-210, June .

Nadas, A., Nahamoo, D., Pichney, A., (1989),"Speech Recognition Using Noise-Adaptive Prototypes", IEEE Trans. ASSP, vol. 37, No. 10, pages 1495-1503, October.

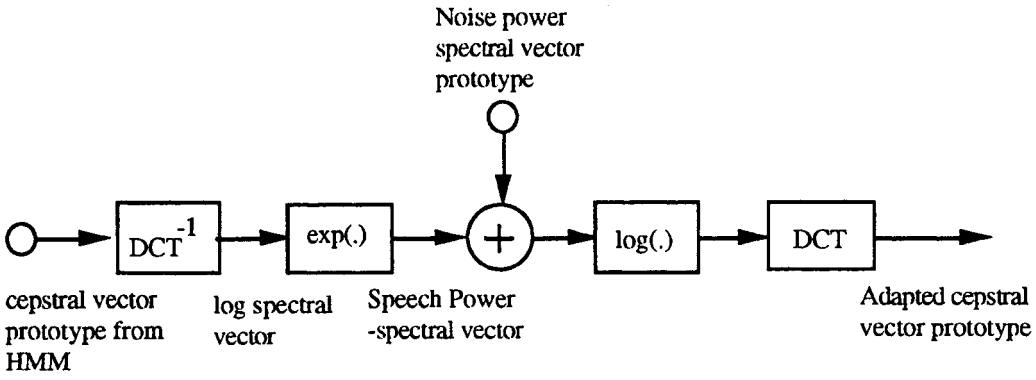
Porter, J. E. & Boll, S. F. (1984),"Optimal estimators for spectral restoration of noisy speech", IEEE Proc. ICASSP-84, San Diego, California, pages. 18A.2.1.-18A.2.4., March.

Roe, D. B., (1987)," Speech Recognition with a Noise-adapting Codebook", IEEE Proc. ICASSP-87, pages. 1139-1142, Dallas, Texas.

Van Compernolle, D., (1989), "Noise Adaptation in a Hidden Markov Model Speech Recognition System", Computer Speech and Language, pages. 151-167.

Varga, A.P., Moore, R.K.(1990),"Hidden Markov Model Decomposition of Speech and Noise",IEEE Proc. ICASSP-90, pages 845-848, NewMexico.

Vaseghi S. V. , Milner B. P. (1993) , " Noisy Speech Recognition Based on HMMs, Wiener Filters and Re-evaluation of Most Likely Candidates, ICAASP93- Vol. 2, pages 103-106.


Figure(1) - Noise-adaption of HMM cepstral prototypes. The noise spectra is estimated and updated from speech inactive perids.