NOISE ESTIMATION ALGORITHMS FOR HIGHLY

NON-STATIONARY ENVIRONMENTS

APPROVED BY THE SUPERVISORY COMMITTEE:

_____

Dr. Philipos C. Loizou, Chair.

_____

Dr. Louis R. Hunt

_____

Dr. Issa Panahi

To my parents and my brother

NOISE ESTIMATION ALGORITHMS FOR HIGHLY NON-STATIONARY ENVIRONMENTS

by

SUNDARRAJAN RANGACHARI, B.E. in ECE

THESIS

Presented to the faculty of

The University of Texas at Dallas

in partial Fulfillment

of the Requirements

for the degree of

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS

August 2004

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Philipos C. Loizou, for his guidance in my studies and my work. He has offered me many helpful suggestions on conducting research and writing technical documents.

I would also like to thank Dr. Louis R. Hunt and Dr. Issa Panahi, for obliging to be on my defense committee and for their valuable feedback on this manuscript.

I also thank Arunvijay Mani for helping me format the thesis. And finally I take this chance to thank all of my friends for being my driving force and standing by me through thick and thin.

NOISE ESTIMATION ALGORITHMS FOR HIGHLY

NON-STATIONARY ENVIRONMENTS

Publication No. _____

Sundarrajan Rangachari, M.S.E.E.
The University of Texas at Dallas, 2004

Supervising Professor: Dr. Philipos C. Loizou

The quality and intelligibility of the speech in the presence of background noise can be improved by speech enhancement algorithms. This thesis addresses the issue of estimating the noise spectrum for speech enhancement applications. Two noise estimation algorithms are proposed for highly non-stationary noise environments. In method-1 a voice activity detector is first used to classify each frame of speech continuously into the speech present/absent frames, and the noise spectrum estimate is updated using a constant smoothing factor for speech absent frames and a frequency dependent smoothing factor for speech present frames. In method-2 the noise spectrum estimate is updated using a frequency dependent smoothing factor irrespective of speech present/absent frames. In both methods, the frequency dependent smoothing factor is calculated based on estimated speech presence probabilities in subbands. Speech presence is determined by computing the ratio of the noisy speech power spectrum to its local minimum, which is computed by averaging past values of the noisy speech power spectra with a look-ahead factor. The local minimum estimation algorithm adapts very quickly to highly non-stationary noise environments. This was confirmed with formal listening tests that indicated that the

proposed noise estimation algorithms when integrated in speech enhancement were preferred over other noise estimation algorithms.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Speech enhancement has found many applications particularly with the increase in automatic speech recognition and mobile communications. In automatic speech recognition systems, the performance degrades badly in the case of adverse environments with very low SNR. It was found that the recognition rate can be improved by applying a speech enhancement algorithm to the degraded speech to improve the intelligibility. Also in the case of mobile communication, the speech signal is degraded by different types of noise in the communication channel. Hence there is a need for speech enhancement system in the receiver. Many speech enhancement systems [6,7] have been developed based on spectral subtraction and Wiener filtering principles. The common features of all these methods are to estimate the power spectrum of clean speech using the power spectrum of noisy speech and noise. In single channel speech enhancement systems there will be access only to noisy speech and hence the noise statistics have to be estimated from the noisy speech itself.

Usually the noise spectrum estimate is obtained from the first few milli-seconds of noisy speech which are silence regions. This assumption is valid for the case of stationary noise in which the noise spectrum does not vary much over time. Traditional VADs [8,9] also track the noise only frames of the noisy speech to update the noise estimate. But the update of noise estimate in those methods is limited to speech absent frames. This is not enough for the case of non-stationary noise in which the power spectrum of noise varies even during speech activity. Hence there is a need to update the noise spectrum continuously over time and this is done by a

noise estimation algorithms [10,11]. This thesis proposes two new algorithms are proposed which are suitable for highly non-stationary noise environments.

This thesis is organized as follows. In chapter 2, some of the existing noise estimation algorithms are explained in detail and the drawbacks of these algorithms are discussed at the end of the chapter. In chapter 3, two proposed noise estimation algorithms are presented. Also the proposed methods are compared with some of the existing algorithms and the subjective and objective measures were derived. In chapter 4 the summary and conclusions are given along with the contributions of the thesis work.

CHAPTER 2

ANALYSIS OF EXISTING NOISE ESTIMATION ALGORITHMS

In this chapter some of the existing noise estimation algorithms which were based on tracking noise using the power spectrum of noisy speech are discussed. Most of these algorithms can be classified broadly into two classes. The first class is based on updating the noise estimate by tracking the silence regions of speech and the other class is based on updating noise estimate using the histogram of the noisy speech power spectrum.

2.1 Minimum Statistics (MS) noise estimation [1]

Martin's [1] method was based on minimum statistics and optimal smoothing of the noisy speech power spectral density. This method rested on two major observations. The first observation was independence of speech and noise which implied that the power spectrum of the noisy speech was equal to the sum of the power spectrum of clean speech and noise respectively. That is,

$$|Y(\lambda,k)|^2 = |X(\lambda,k)|^2 + |D(\lambda,k)|^2 \tag{2.1}$$

where $|Y(\lambda,k)|^2$, $|X(\lambda,k)|^2$ and $|D(\lambda,k)|^2$ were the power spectrum of noisy speech, clean speech and noise respectively and λ and k denoted the time index and frequency bin index respectively. The second observation was that the power spectrum of noisy speech often becomes equal to the power spectrum of noise. This happens during speech pauses and also between words and syllables. Hence the estimate of noise power spectral density was obtained by tracking the minimum of the noisy speech in each frequency bin separately. Also, since the

minimum was biased towards lower values, unbiased estimate was obtained by multiplying with a bias factor which was derived from the statistics of the local minimum.

2.1.1 Principles of the minimum statistics algorithm

Since the power spectral density of noisy speech was equal to the sum of noise power and speech power, noise variance was estimated by tracking the minimum of noisy speech power spectral density over a fixed window length. This window length was chosen wide enough to bridge the broadest peak in any speech signal. It was found out experimentally [12] that window lengths of approximately 0.8-1.4s gave good results.

For searching the minimum a first-order recursive version of the noisy speech power spectral density was used:

$$P(\lambda,k) = \alpha P(\lambda,k) + (1-\alpha)\left|Y(\lambda,k)\right|^2 \tag{2.2}$$

where α was the smoothing constant. To improve the performance of the minimum tracking procedure the following modifications were made.

1. Replacing the constant smoothing factor in Eq. (2.2) with the time-frequency dependent smoothing factor.
2. Deriving a bias factor for the noise estimate since the minimum tracking was biased towards lower values.
3. Improving tracking speed of the algorithm for increasing noise levels.

2.1.2 Deriving optimal time-frequency dependent smoothing factor

The smoothing parameter used in Eq. (2.2) had to be very low to follow the non-stationarity of the speech signal. On the other hand, it had to be close to one to keep the variance of the

minimum tracking as small as possible. Hence there was need for time and frequency dependent smoothing factors in place of a fixed smoothing factor.

This was derived for speech absent region. The requirement was that the smoothed power spectrum $P(\lambda,k)$ had to be equal to the noise variance $\sigma_D^2(\lambda,k)$ during speech pauses. Hence the smoothing parameter was derived by minimizing the mean squared error between $P(\lambda,k)$ and $\sigma_D^2(\lambda,k)$ as follows

$$E\left\{\left(P(\lambda,k)-\sigma_D^2(\lambda,k)\right)^2\middle|P(\lambda-1,k)\right\} \tag{2.3}$$

where

$$P(\lambda,k)=\alpha(\lambda,k)P(\lambda,k)+(1-\alpha(\lambda,k))\left|Y(\lambda,k)\right|^2 \tag{2.4}$$

Note that in Eq. (2.4) time-frequency dependent smoothing factor $\alpha(\lambda,k)$ was used instead of fixed α as defined in (2.2). Substituting (2.4) in (2.3) and setting the first derivative to zero gave the optimum value for $\alpha(\lambda,k)$:

$$\alpha_{opt}(\lambda,k)=\frac{1}{1+\left(P(\lambda-1,k)/\sigma_D^2(\lambda,k)-1\right)^2} \tag{2.5}$$

But in real time implementation, the value of estimated noise variance $\hat{\sigma}_D^2(\lambda,k)$ lags behind true noise variance $\sigma_D^2(\lambda,k)$. Hence some correction factor $\alpha_c(\lambda)$ was calculated using the ratio of averaged smoothed periodogram to estimated noise variance. The final smoothing parameter after the correction factor was given as

$$\alpha_{opt}(\lambda,k)=\frac{\alpha_{\max}\alpha_c(\lambda)}{1+\left(P(\lambda-1,k)/\hat{\sigma}_D^2(\lambda,k)-1\right)^2} \tag{2.6}$$

where $\alpha_{\max} = 0.96$. Figure 2.1 shows the smoothed power spectrum for a speech degraded by 5dB babble noise and the corresponding smoothing factor $\alpha_{opt}(\lambda, k)$ for a

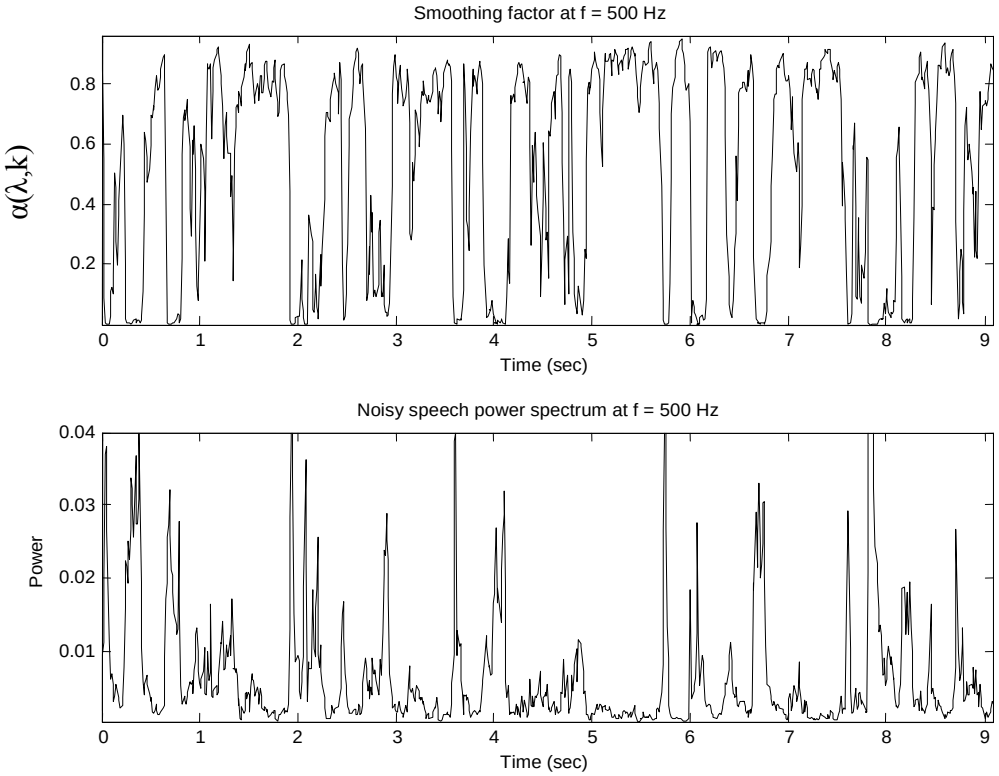

Fig 2.1. Top panel: Plot of time-frequency dependent smoothing constant derived for a noisy speech (5dB SNR) at f = 500 Hz. Bottom panel: Plot of noisy speech power spectrum for the same noisy speech at f = 500 Hz.

single frequency bin. It can be seen that the smoothing constant decreases to near zero when the speech power increases to follow the non-stationarity of speech and becomes close to one when speech was absent to reduce the variance of the minimum.

2.1.3 Bias factor

As the minimum was biased towards lower values, the bias factor for compensating the minimum of noisy speech power spectrum was derived using the statistics of minimum of the

correlated PSD estimates of noisy speech. It was stated that since pdf of $P(\lambda,k)$ was scaled by $\sigma_D^2(\lambda,k)$, the minimum statistics of the short term estimates $P_{\min}(\lambda,k)$ was also scaled by $\sigma_D^2(\lambda,k)$. Thus the bias term was derived by finding the mean of minimum PSD for some $\sigma_D^2(\lambda,k)$=1 which after simplification gave

$$B_{\min}(\lambda,k) \approx 1 + (L-1)\frac{2}{\tilde{Q}_{eq}(\lambda,k)} \tag{2.7}$$

where L was the window length over which the minimum was found and $\tilde{Q}_{eq}(\lambda,k)$ called "equivalent degrees of freedom", was a function of smoothed periodogram, and the previous noise variance. The unbiased noise estimate was then obtained as

$$\hat{\sigma}_D^2(\lambda,k) = B_{\min}(\lambda,k).P_{\min}(\lambda,k) \tag{2.8}$$

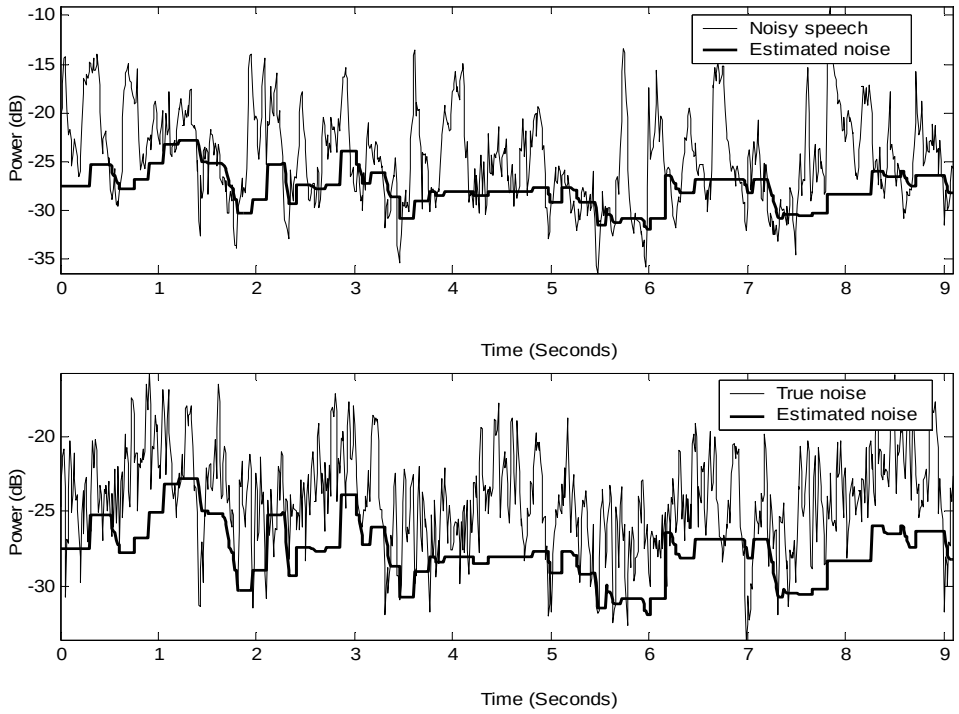

Fig 2.2. Top panel: Plot of noisy speech power spectrum and noise estimate using [1] for a noisy speech (5dB SNR) at f=500 Hz. Bottom panel: Plot of true and estimated noise power spectrum using [1] for the same noisy speech at f=500 Hz.

Figure 2.2 shows the power spectrum of noisy speech and the estimated noise power spectrum of a speech degraded by babble noise at 5dB SNR. The estimated noise spectrum is compared with the true noise spectrum for the same example.

2.1.4 Efficient minimum search algorithm

The minimum for the noisy speech was found over L consecutive frames. In the worst case of increasing noise levels, the minimum search lagged behind by 2L frames. To reduce this lag the whole window of L frames was divided into U subwindows of V samples each (L = UV). In this case, the maximum delay was reduced to L+V frames compared to 2L frames in the previous case.

2.1.5 Experimental results

The performance of the noise estimation algorithm was assessed using both objective and subjective measures. For objective measure, the percentage relative estimation error and the error variance were calculated between the true noise spectrum and estimated noise spectrum for white Gaussian noise, vehicular noise and street noise with a SNR of 15 dB using continuous English sentences and sentences with speech pauses. For subjective analysis, the noise estimation algorithm was combined with multiplicatively modified minimum mean square error log spectral amplitude (MM-MMSE-LSA) estimator [6,10] and the 2400bps MELP speech coder [13]. The diagnostic acceptability measure (DAM) and diagnostic rhyme test (DRT) scores were presented. All the results showed superior performance for minimum statistics approach compared to the VAD [1]. The tracking of minimum in each frequency bin separately helped in retaining the

weak voiced consonants which may be classified as noise only frames by most of the VAD's since their energy was concentrated in very small number of frequency bins.

2.2 Minima Controlled Recursive Averaging (MCRA) [2]

In [2] a new approach, called minima controlled recursive averaging (MCRA) was introduced for noise estimation. The noise estimate was updated by averaging the past spectral values of noisy speech which was controlled by a time and frequency dependent smoothing factors. These smoothing factors were calculated based on the signal presence probability in each frequency bin separately. This probability was in turn calculated using the ratio of the noisy speech power spectrum to its local minimum calculated over a fixed window time.

2.2.1 Noise spectrum estimation

The derivation for noise power spectrum from the signal presence probability was based on the following two hypotheses

$$\begin{aligned} H_0(\lambda,k): &\quad Y(\lambda,k) = D(\lambda,k) \\ H_1(\lambda,k): &\quad Y(\lambda,k) = X(\lambda,k) + D(\lambda,k) \end{aligned} \tag{2.9}$$

where $Y(\lambda,k)$, $X(\lambda,k)$ and $D(\lambda,k)$ represented the short time Fourier transform of the noisy speech, clean speech and noise respectively and $H_0(\lambda,k)$ and $H_1(\lambda,k)$ represented the speech absent and speech present hypotheses respectively. The noise variance was represented as $\sigma_D^2(\lambda,k) = E\left[\left|D(\lambda,k)\right|^2\right]$. The update of noise estimate for the above two hypotheses was written as follows.

$$\begin{aligned} &H_0^{'}(\lambda,k): \hat{\sigma}_D^2(\lambda+1,k) = \alpha_d \hat{\sigma}_D^2(\lambda,k) + (1-\alpha_d)\left|Y(\lambda,k)\right|^2 \\ &H_1^{'}(\lambda,k): \hat{\sigma}_D^2(\lambda+1,k) = \hat{\sigma}_D^2(\lambda,k) \end{aligned} \tag{2.10}$$

where $\hat{\sigma}_D^2(\lambda,k)$ was the estimate of noise variance and α_d (0<α_d<1) was the smoothing factor. Eq. (2.10) was based on the principle that noise estimate was updated whenever silence was detected, otherwise it was kept constant. The overall noise estimate was obtained based on speech presence probability as

$$\begin{aligned}\hat{\sigma}_D(\lambda+1,k) = [\hat{\sigma}_D(\lambda+1,k) \mid H_1^{'}(\lambda,k)]p'(\lambda,k) + \\ [\hat{\sigma}_D(\lambda+1,k) \mid H_0^{'}(\lambda,k)](1-p'(\lambda,k))]\end{aligned} \tag{2.11}$$

where $p'(\lambda,k) \triangleq P(H_1^{'}(\lambda,k) \mid Y(\lambda,k))$ was the speech presence probability. The noise variance for the two hypotheses defined in (2.10) was substituted and simplified as follows:

$$\hat{\sigma}_D^2(\lambda+1,k) = \tilde{\alpha}_d(\lambda,k)\hat{\sigma}_D(\lambda,k) + [1-\tilde{\alpha}_d(\lambda,k)]\left|Y(\lambda,k)\right|^2 \tag{2.12}$$

where

$$\tilde{\alpha}_d(\lambda,k) \triangleq \alpha_d + (1-\alpha_d)p'(\lambda,k) \tag{2.13}$$

.2.2.2 Computing speech presence probability

The speech presence probability $p'(\lambda,k)$ in each frequency bin was calculated using the ratio of the noisy speech power spectrum to its local minimum. For finding the local minimum, a smoothed power spectrum of noisy speech using first order recursive averaging was calculated as follows

$$P(\lambda,k) = \alpha_s P(\lambda-1,k) + (1-\alpha_s)\left|Y(\lambda,k)\right|^2 \tag{2.14}$$

where α_s was the smoothing factor. This helped reduce the variance of the local minimum of the noisy speech power spectrum. Then, the local minimum was found over a fixed window length of L time frames by sample wise comparison of smoothed power spectrum of noisy speech and the local minimum of noisy speech as follows

$$P_{\min}(\lambda,k) = \min\left\{P_{\min}(\lambda-1,k), P(\lambda,k)\right\}$$
$$P_{tmp}(\lambda,k) = \min\left\{P_{tmp}(\lambda-1,k), P(\lambda,k)\right\} \quad (2.15)$$

where $P_{\min}(\lambda,k)$ and $P_{tmp}(\lambda,k)$ were the local minimum and temporary variable respectively. For every L frames processed, the temporary variable was updated as follows to avoid $P_{\min}(\lambda,k)$ falling behind the global minimum:

$$P_{\min}(\lambda,k) = \min\left\{P_{tmp}(\lambda-1,k), P(\lambda,k)\right\}$$
$$P_{tmp}(\lambda,k) = P(\lambda,k) \quad (2.16)$$

This parameter L was the length of the window over which the minimum was updated. This L had to be chosen to cover the broadest peak of the speech spectrum. This parameter also controlled the update of the noise power spectrum particularly for increasing noise levels. Based on their experiments [2] with different speakers and noise types, window lengths of 0.5-1.5s were found to be optimum. Very small window length may result in overestimation of noise if the window length was less than the width of the speech peak itself. Also, very large window length will delay the update of noise variance estimate particularly for increasing noise levels.

The ratio of the noisy speech power spectra to its local minimum was defined as $S_r(\lambda,k) = P(\lambda,k)/P_{\min}(\lambda,k)$. A decision rule for finding speech present regions was derived by solving the Bayes minimum-cost function and was given as follows

If $S_r(\lambda,k) > \delta$ then

speech present $\rightarrow I(\lambda,k) = 1$ (2.17)

else

speech absent $\rightarrow I(\lambda,k) = 0$

Figure 2.3 shows the ratio $S_r(\lambda,k)$ for a noisy speech of SNR 5dB with babble noise at 500 Hz. The noisy speech spectrum at that frequency is also shown for reference. The dashed line in the

bottom panel of the figure shows the threshold δ. Hence, all bins for which the ratio $S_r(\lambda,k)$ was greater than the threshold were taken as speech present regions. Otherwise they were taken as speech absent region. The speech presence probability was obtained using the above decision rule as follows:

$$\hat{p}'(\lambda,k) = \alpha_p \hat{p}'(\lambda,k) + (1-\alpha_p) I(\lambda,k) \tag{2.18}$$

where α_p was a smoothing constant and $I(\lambda,k)$ was the indicator function for the speech presence decision defined in (2.17).

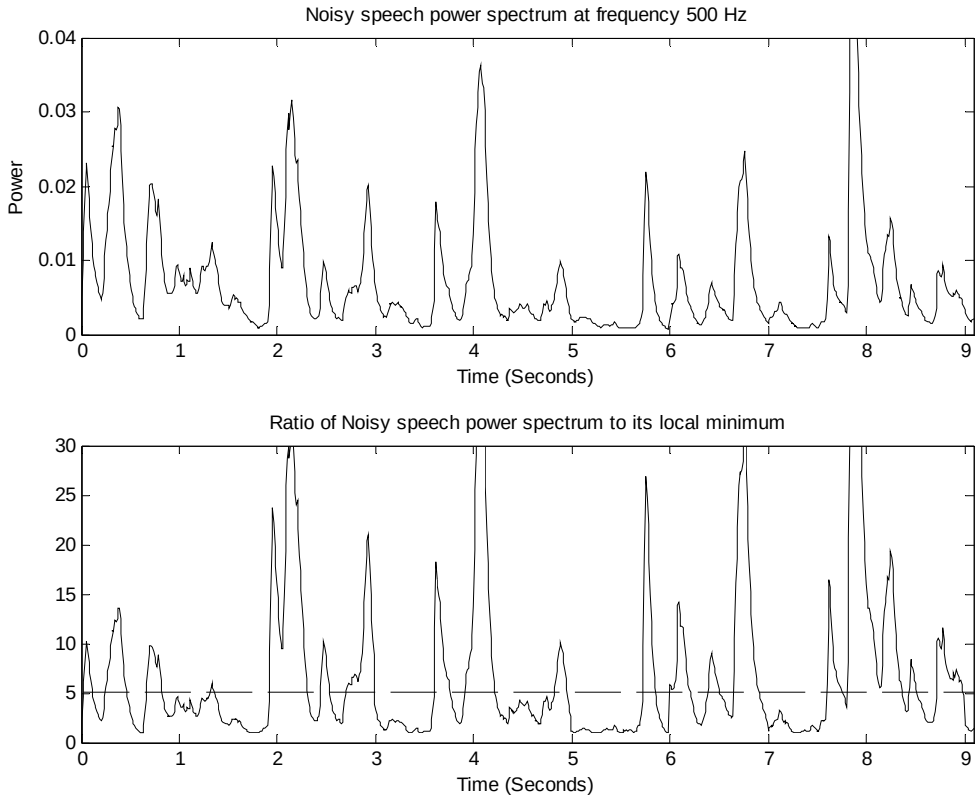

Fig 2.3. Top panel: Plot of noisy speech power spectrum for a noisy speech at SNR 5dB with babble noise at f=500 Hz. Bottom panel: Plot of $S_r(\lambda,k)$ for the same noisy speech at f =500 Hz. The dashed line shows the threshold δ.

After calculating the speech presence probability, the time-frequency dependent smoothing factor was calculated using (2.13) and then the noise variance was updated using (2.12).

The following are some of the merits of the algorithm.

- The parameter δ was not sensitive to types and intensity of the noise.
- Second, the ratio $S_r(\lambda,k)$ defined above was more correlated to the posteriori SNR, and hence the probability that the speech power was very high compared to noise when $S_r(\lambda,k) < \delta$ was very small. Hence, even a slight increase in noise estimate by deciding $H_0^{'}$ when $H_1^{'}$ was not very destructive.
- The correlation of speech in adjacent frames was exploited using α_p by first order recursion in (2.18).

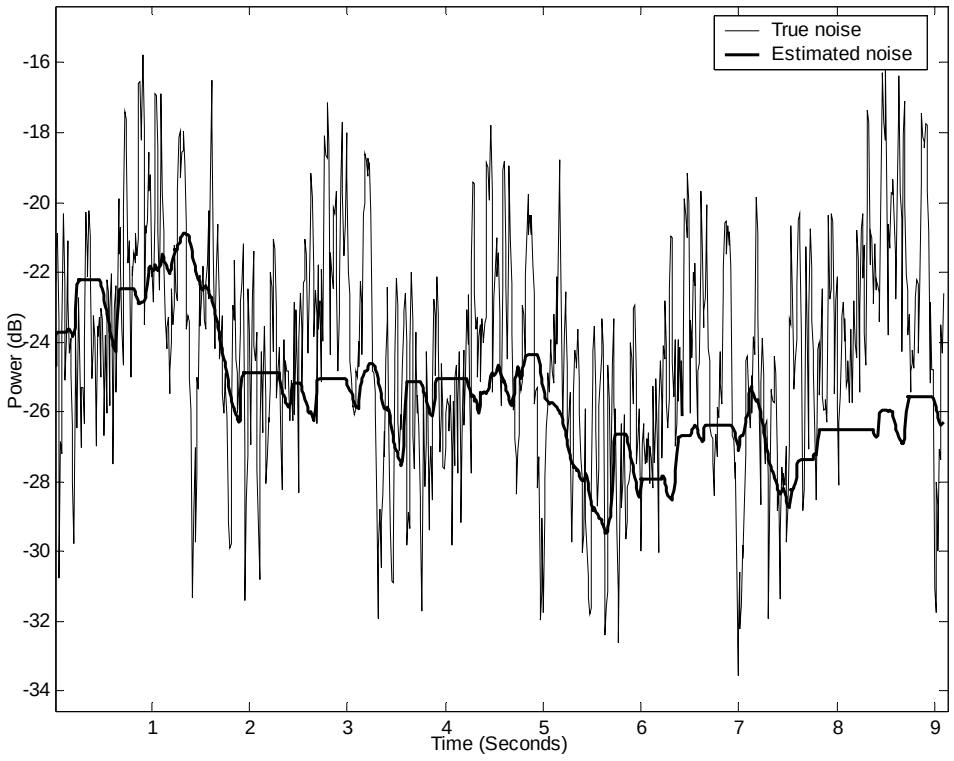

Fig 2.4. Plot of true noise power spectrum and estimated noise power spectrum using [2] for a noisy speech (5dB SNR) at f = 500 Hz.

Figure 2.4 shows the true noise power spectrum and the estimated noise power spectrum for the same example considered in Figure 2.1.

2.2.3 Experimental results

The performance of the MCRA approach was compared with weighted averaged method [5] after combining both the methods with optimally modified log-spectral amplitude estimator speech enhancement procedure [14]. For objective results, the improvement in segmental SNR was reported for white Gaussian noise, car interior noise and F16 cockpit noise for various noise levels from -5 to 10 dB. In all the cases, the MCRA approach showed a higher performance compared to weighted averaged method. Also, the methods were compared with a subjective study of spectrogram of enhanced speech and informal listening tests. The tracking ability of the algorithms was tested by comparing the spectrograms of enhanced speech for a signal recorded in a car by suddenly turning on the defroster in full. As expected, the weighted averaged method did not track the noise completely even after 12s whereas MCRA method tracked the noise floor in less than 3s.

2.3 Improved MCRA [3]

In [3], an improved MCRA noise variance estimator was proposed with improvements to [2] in the following aspects:

- Minimum tracking during speech activity.
- Speech presence probability estimation.
- Derivation of a bias compensation factor.

2.3.1 Noise spectrum estimation

As in [2] the derivation for noise estimate was based on the two hypotheses $H_0(\lambda,k)$ and $H_1(\lambda,k)$ representing speech absence and presence respectively. The PDF's of STFT of both speech and noise were assumed to be Gaussian. Then the conditional speech presence probability $p(\lambda,k) \triangleq P(H_1(\lambda,k) \mid \gamma(\lambda,k))$ was derived as [7]

$$p(\lambda,k) = \left\{1 + \frac{q(\lambda,k)}{1-q(\lambda,k)}(1+\xi(\lambda,k))\exp(-v(\lambda,k))\right\}^{-1} \tag{2.19}$$

where $q(\lambda,k) \triangleq p(H_0(\lambda,k))$ was a priori probability for speech absence and $v \triangleq \gamma\xi/(1+\xi)$, γ and ξ were a posteriori and a priori SNRs respectively. From the speech presence probability the noise estimate was updated using the time-frequency dependent smoothing factor as in [2].

$$\hat{\sigma}_D^2(\lambda+1,k) = \tilde{\alpha}_d(\lambda,k)\hat{\sigma}_D(\lambda,k) + [1-\tilde{\alpha}_d(\lambda,k)]\left|Y(\lambda,k)\right|^2 \tag{2.20}$$

where $\tilde{\alpha}_d(\lambda,k) \triangleq \alpha_d + (1-\alpha_d)p(\lambda,k)$. The speech presence probability was biased towards higher values to avoid speech distortion. Hence a bias compensation factor was introduced to the noise estimate as follows

$$\tilde{\sigma}_D^2(\lambda,k) = \beta\hat{\sigma}_D^2(\lambda,k) \tag{2.21}$$

and the value of β was determined as 1.47.

2.3.2 Computing speech absence probability

The minimum in each frequency bin was tracked individually using a procedure similar to [1] with the following variations:

- The number of past frames taken for minimum update depends on speech presence probability as compared to fixed window length in [1]. Very long window length was

used for frequent speech present regions and smaller window length for frequent absent regions.

- The smoothing was carried out in both time and frequency thereby exploiting the correlation of speech presence in adjacent frequency bins of consecutive frames.
- The smoothing of noisy speech was done in two iterations. The first iteration gave a rough VAD decision and the second iteration excluded high power speech regions to reduce the variance of minimum.

The first iteration of smoothing was given as follows

$$S_f(\lambda,k) = \sum_{i=-w}^{w} b(i)\left|Y(\lambda,k)\right|^2 \tag{2.22}$$

and

$$S(\lambda,k) = \alpha_s S(\lambda-1,k) + (1-\alpha_s)S_f(\lambda,k) \tag{2.23}$$

where $b(i)$ is the hamming window and α_s is the smoothing constant. Eq. (2.22) gives the frequency smoothing and Eq. (2.23) gives the time smoothing. The minimum of noisy speech was found over a window of L frames [1], as follows

$$S_{\min}(\lambda,k) \triangleq \min\left\{S(\lambda',k) \mid \lambda - L + 1 \le \lambda' \le \lambda\right\} \tag{2.24}$$

As stated in [1] the mean of the minimum was proportional to the noise variance and a constant bias term independent of noise spectrum was used as follows:

$$E\left\{S_{\min}(\lambda,k) \mid \xi(\lambda,k) = 0\right\} = B_{\min}^{-1}.\sigma_D^2(\lambda,k). \tag{2.25}$$

Then a rough speech presence decision was given as:

$$I(\lambda,k)=\begin{cases}1 & \begin{array}{l}\text{if } \gamma_{\min}(\lambda,k)<\gamma_0\\ \text{and } \zeta(\lambda,k)<\zeta_0\end{array} \quad (\text{speech is absent})\\ 0 & \text{otherwise} \qquad\qquad (\text{speech is present})\end{cases} \tag{2.26}$$

where

$$\gamma_{\min}(\lambda,k)\triangleq\frac{\left|Y(\lambda,k)\right|^2}{B_{\min}S_{\min}(\lambda,k)};\quad \zeta(\lambda,k)\triangleq\frac{S(\lambda,k)}{B_{\min}S_{\min}(\lambda,k)} \tag{2.27}$$

The threshold values γ_0 and ζ_0 were set so that the error margin in making a wrong decision was below a threshold. Eq. (2.26) was based on the same principle as in [2]. Whenever the ratio of noisy speech power spectrum to the local minimum was greater than a threshold, it was considered as speech present, otherwise, it was considered as speech absent. Instead of using only one ratio as in [2], two different ratio factors were used to make speech present decision with higher confidence.

The second iteration which excluded the high energy speech regions was as follows

$$\tilde{S}_f(\lambda,k)=\begin{cases}\dfrac{\sum_{i=-w}^{w}b(i)I(\lambda,k-i)\left|Y(\lambda,k-i)\right|^2}{\sum_{i=-w}^{w}b(i)I(\lambda,k-i)}, & \text{if } \sum_{i=-w}^{w}I(\lambda,k-i)\neq 0\\ \tilde{S}(\lambda-1,k) & \text{otherwise}\end{cases} \tag{2.28}$$

and

$$\tilde{S}(\lambda,k)=\alpha_s\tilde{S}(\lambda-1,k)+(1-\alpha_s)\tilde{S}_f(\lambda,k) \tag{2.29}$$

where Eq. (2.28) gives the smoothing in frequency and Eq. (2.29) gives the smoothing in time. Thus excluding high energy speech components as in Eq. (2.28) was useful in selecting smaller window length L, which reduced the delay in responding to the rising noise power compared to [1]. Let $\tilde{S}_{\min}(\lambda,k)$ be the minimum of the noisy speech $\tilde{S}(\lambda,k)$ in the second iteration. Then the actual speech absence probability was given by

$$\hat{q}(\lambda,k)=\begin{cases}1, & \text{if } \tilde{\gamma}_{\min}(\lambda,k)\le 1 \\ & \text{and } \tilde{\zeta}(\lambda,k)<\zeta_0 \\ \dfrac{(\gamma_1-\tilde{\gamma}_{\min}(\lambda,k))}{(\gamma_1-1)} & \text{if } 1<\tilde{\gamma}_{\min}(\lambda,k)<\gamma_1 \\ & \text{and } \tilde{\zeta}(\lambda,k)<\zeta_0 \\ 0, & \text{otherwise}\end{cases} \tag{2.30}$$

where

$$\tilde{\gamma}_{\min}(\lambda,k)\triangleq\frac{|Y(\lambda,k)|^2}{B_{\min}\tilde{S}_{\min}(\lambda,k)};\quad \tilde{\zeta}(\lambda,k)\triangleq\frac{\tilde{S}(\lambda,k)}{B_{\min}\tilde{S}_{\min}(\lambda,k)} \tag{2.31}$$

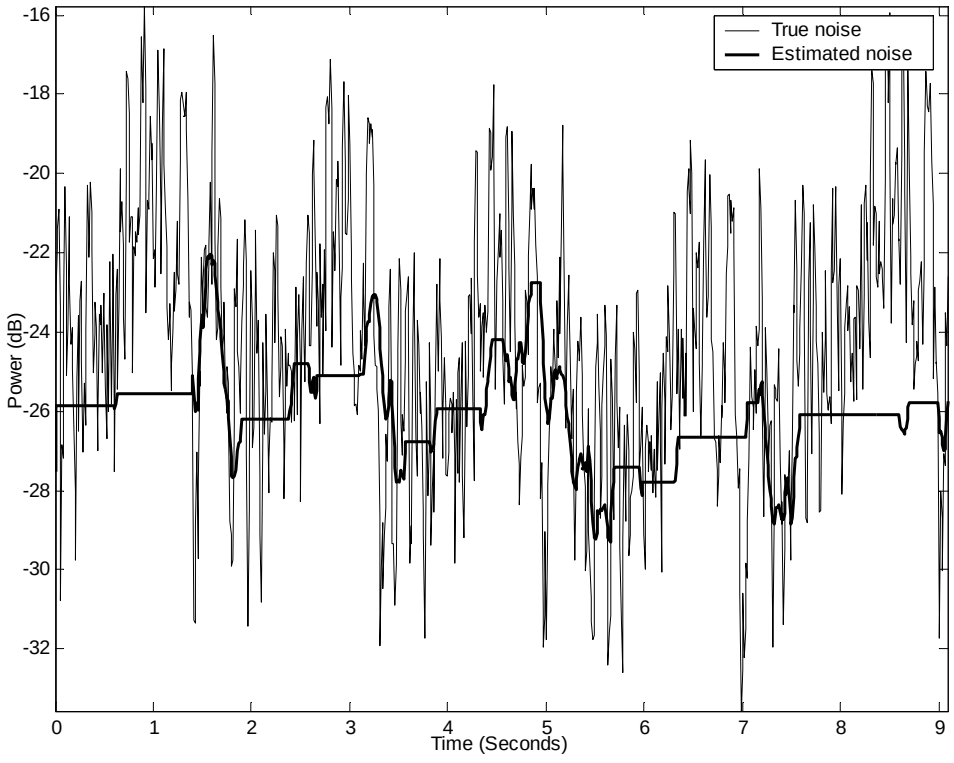

Fig 2.5. Plot of true noise power spectrum and estimated noise power spectrum using [3] at for a noisy speech (5dB SNR) at f=500 Hz.

The threshold γ_1 was set to satisfy a significant margin so that the probability of wrong decision was smaller than a threshold. Using the speech absence probability found from (2.30) the time-frequency smoothing factor was calculated and the noise estimate was updated. Figure 2.5 shows the power spectrum of true noise and the estimated noise for the same example considered above.

2.3.3 Performance evaluation

The performance of the algorithm was evaluated by both subjective and objective measures. For objective measure, the relative estimation error between true noise spectrum and estimated noise spectrum and the improvements in segmental SNR were calculated. Noise signals taken from Noisex92 database [15] include white noise, car noise and F16 cockpit noise. The speech signal was degraded at different SNR values in the range [-5,10] dB. The IMCRA method was compared with that of MS [1] for all the test cases and it showed improved performance in all cases. This was conformed with the subjective study of histogram and informal listening tests.

2.4 Continuous spectral minima tracking [4]

In [4] a computationally efficient noise estimation algorithm was proposed for speech enhancement. In contrast to using a specified window time for tracking the minimum of noisy speech as in [1-3], the noise estimate was updated continuously by smoothing noisy speech power spectra in each frequency bin separately using a non-linear smoothing rule.

For tracking the minimum of noisy speech power spectrum, a short-time smoothed version of the periodogram of noisy speech was computed as follows

$$P(\lambda,k) = \alpha P(\lambda-1,k) + (1-\alpha)\left|Y(\lambda,k)\right|^2 \tag{2.32}$$

with α as a forgetting factor between 0.7 and 0.9.

The non-linear rule used for estimating the noise spectrum by tracking the minimum of noisy speech power spectrum in each frequency bin separately was given as follows.

If $P_{\min}(\lambda,k) < P(\lambda,k)$ then

$$P_{\min}(\lambda,k) = \gamma P_{\min}(\lambda-1,k) + \frac{1-\gamma}{1-\beta}(P(\lambda,k) - \beta P(\lambda-1,k)) \tag{2.33}$$

else

$$P_{\min}(\lambda,k) = P(\lambda,k) \tag{2.34}$$

where $P_{\min}(\lambda,k)$ is the noise estimate and the values of the parameters were α=0.7, β=0.96 and γ=0.998.

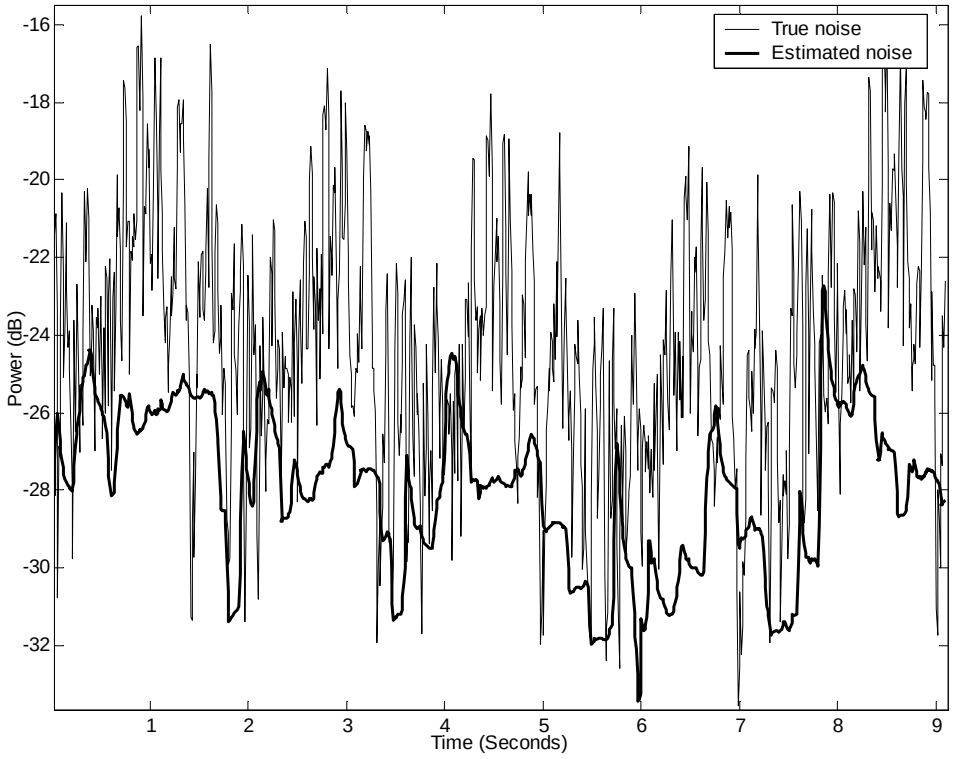

Fig 2.6. Plot of true noise spectrum and estimated noise spectrum using [4] for a noisy speech (5dB SNR) at f=500 Hz.

Also, the look-ahead factor β in the minimum tracking can be adjusted to vary the adaptation time of the algorithm. The typical adaptation time was 0.2 to 0.4s with the values mentioned above. Figure 2.6 shows an example of true noise power spectrum and the estimated noise power spectrum for the same example considered above.

2.5 Weighted averaging technique [5]

In [5] two techniques were presented for noise estimation using noisy speech power spectrum. Similar to the other methods discussed so far, these methods did not require explicit voice activity detection. In both methods the noise spectral estimate was updated by comparing the noisy speech power spectra to the current noise estimate.

The first method was called the weighted average technique, in which the noise estimate was updated continuously by smoothing the spectral values of noisy speech which were below the threshold. This smoothing was represented as follows

$$D(\lambda,k) = \alpha D(\lambda-1,k) + (1-\alpha)\left|Y(\lambda,k)\right|^2 \tag{2.35}$$

where α=0.85 was the smoothing factor.

The threshold for this method was taken as $\beta D(\lambda-1,k)$ where β takes a value in the range of 1.5 to 2.5. This threshold was adaptive in the sense that the threshold changes depending on the noise power level present in the noisy speech. Thus the threshold can follow the slow changes in noise power levels for slowly varying noise statistics. The overall algorithm can be summarized as follows

$$\begin{aligned}
&\text{If } \left|Y(\lambda,k)\right|^2 < \beta D(\lambda-1,k) \text{ then} \\
&\qquad D(\lambda,k) = \alpha D(\lambda-1,k) + (1-\alpha)\left|Y(\lambda,k)\right|^2 \\
&\text{else} \\
&\qquad D(\lambda,k) = D(\lambda-1,k)
\end{aligned} \tag{2.36}$$

Figure 2.7 shows the true noise spectrum and the estimated noise spectrum for the same example considered above.

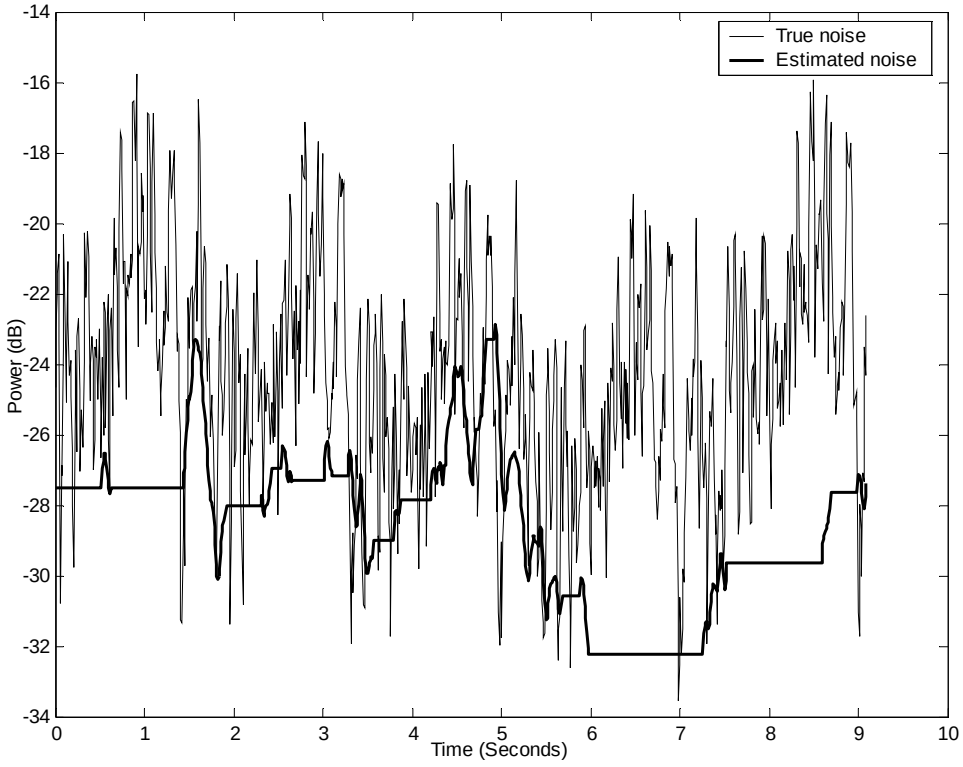

Fig 2.7. Plot of true noise spectrum and estimated noise spectrum using weighted averaged method [5] for a noisy speech (5dB SNR) at f=500 Hz.

An objective evaluation of the noise estimation algorithm was presented by calculating the relative mean squared error between the true noise spectrum and the estimated noise spectrum and the averaged value was around 3.4% for different SNR values of speech degraded by car noise. Also informal listening tests were conducted by combining the technique with nonlinear spectral subtraction and the result showed some appreciable suppression of the noise.

2.6 Histogram based technique [5]

In the second method in [5], the noise estimate was updated based on the histogram of the past noise segments. These noise segments were updated using the same threshold as mentioned in the first method in [5]. The histogram was found for the past 400ms of noise segments and the value which had the maximum occurrence was taken as the estimate of noise spectrum. This was done separately for each frequency bin. Also, a first order recursive smoothing was performed after finding the noise spectrum estimate from the histogram to avoid the presence of outliers in the estimate.

As explained in section 2.5 this method was also assessed by objective measure by finding the relative mean squared error between the true noise spectrum and the estimated noise spectrum. The results showed an improved performance for the histogram method over the weighted averaged method for all the different SNR values for the same example considered above. Also, this method showed better noise suppression in informal listening tests when combined with nonlinear spectral subtraction as compared to the weighted average technique.

2.7 Quantile based noise estimation [16]

Most of the algorithms presented so far track noisy only segments of noisy speech to update the noise estimate. This was done by tracking the minimum of the noisy speech as given by [1,2]. But since the minimum was sensitive to outliers, in [16] more reliable estimate was obtained by finding noise estimate from the q^{th}-quantile of the noisy speech spectrum. The method was based on the observation that speech is localized in both time and frequency. Even during speech present frames, speech is concentrated in a fraction of the entire frequency spectrum. Hence most of the regions of the noisy speech in the time-frequency plane are in the noise power level.

To find the quantile, first the noisy speech power spectrum for each frequency bin was sorted. The q[th] quantile noise estimate was defined as follows

$$D(k) = \bar{X}(t_{\lfloor qT \rfloor}, k) \tag{2.37}$$

where $\bar{X}(t,k)$ is the sorted noisy speech power spectrum such that $\bar{X}(t_0,k) \leq \bar{X}(t_1,k) \leq ..... \leq \bar{X}(t_T,k)$ and T is the duration of speech signal. For example q = 0 gives minimum, q=1 gives maximum and q=0.5 gives the median. Figure 2.8 shows $\bar{X}(t,k)$ for the same example considered above and for a single frequency bin (k = 10).

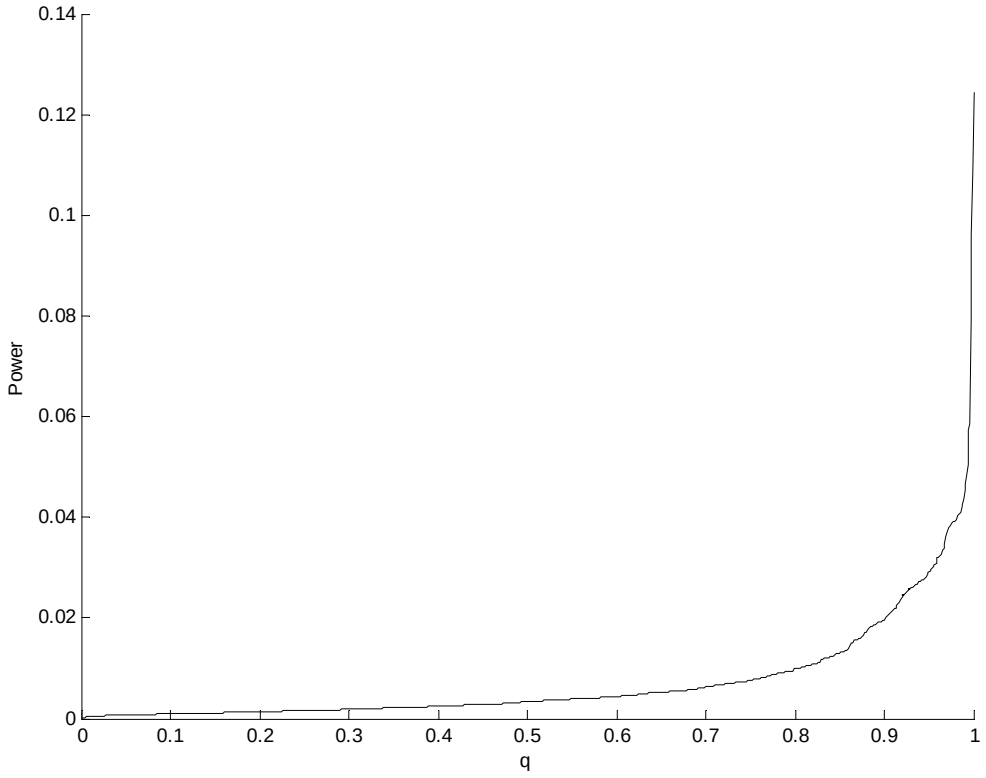

Fig 2.8. Plot of quantile of energy distribution in the noisy speech (5dB SNR) at f =500 Hz.

The value of q=0.5 was chosen to be optimum for this method. Also the implementation used the noisy speech spectrum over the entire time range in order to find the noise estimate at

time t. This was non-causal in the sense that it required noisy speech of future time to compute noise estimate of current time. To make the implementation causal, only the noisy speech till time t was used to find the noise estimate at time t. The error rate for causal case was slightly higher than the non-causal case since the initial noise estimates for the causal case had more error because of availability of only fewer noisy speech samples. Also in this method, the computational cost and memory requirements increased with time t. To improve the efficiency, finite length of noisy speech samples was used for noise estimation. Once the buffer was full, the smallest and largest values are removed from the buffer. As expected, the performance of the algorithm improved for increasing buffer lengths and asymptotically reached the performance of unlimited buffer case.

The performance of the noise estimation algorithm was evaluated using the word error rates on a speech recognition task by combining the algorithm with wiener filtering and spectral subtraction. When the word error rates of enhanced speech were compared with the word error rates without noise reduction, relative improvements of 25% was obtained for the choice of q = 0.55. As stated before, the word error rates for causal case were slightly higher compared to the non-causal case.

2.8 Chi-Square test based noise estimation [17]

In this method [17], the noise estimate was updated by finding the noise only frames of the noisy speech using a chi-square test. This test was based on the principle that if there was a good fit between the frequency distribution of the observed noisy speech frame and the estimated noise frame, the observed frame was taken as noise only frame. The noise estimate was also

updated whenever a noise only frame was found. It was also assumed that the distribution of noise was Gaussian if long frames were used.

First the noisy speech was bandlimited in 0.2-4KHz and then bandpassed into M (=8) subbands using a set of IIR bandpass filters. Then, each of these subband signals were divided into time frames of length L where L>>M. Since the noise pdf was assumed to be Gaussian for long frames, the current frame was combined with 7 previous frames giving a long frame size of 122ms. Then the chi-square test was applied to noisy speech data as follows.

Let $Y_{k,p}$ be the noisy signal in subband k and frame p. The samples in the noise vector $D_{k,p}$ was given by

$$D_{k,p} = [d_{k,p}(1), ..., d_{k,p}(L)] \tag{2.38}$$

where L was the frame length. The noise estimate was updated by finding noise-only frames based on the following two hypotheses

$$\begin{aligned} H_0 &: \ Y_{k,p} = \hat{D}_{k,p-1} \qquad \text{noise only frame} \\ H_1 &: \ Y_{k,p} = \hat{D}_{k,p-1} + X_{k,p} \ \text{noise-plus-speech frame} \end{aligned} \tag{2.39}$$

where $X_{k,p}$ was the speech signal present in frame p. The noise estimate in frame p-1 was grouped into 7 classes whose boundaries were chosen such that the number in each class was approximately equal. This was taken as the approximate noise pdf

$$\text{e} = [e_1, ..., e_i, ..., e_N] \text{ for N bins} \tag{2.40}$$

where e_i was the number of noise samples falling in bin *i*. Similarly the current frame of noisy speech was also grouped into same 7 classes as follows

$$\text{o} = [o_1, ...o_N] \tag{2.41}$$

The chi-square test was then applied to these bins where the chi-square statistic was given by

$$\aleph^2 = \sum_{i=1}^{N} \frac{(o_i - e_i)^2}{e_i} \tag{2.42}$$

The calculated value was then compared against a threshold value which depended on the allowed error probability and obtained from the standard chi-square tables. The hypotheses can thus be tested as follows

$$\aleph^2 \begin{matrix} > \\ < \end{matrix} \text{threshold} \Rightarrow \begin{cases} \mathrm{H}_1 \\ \mathrm{H}_0 \end{cases} \tag{2.43}$$

This hypothesis testing was done separately for each of the subbands. If H_0 was identified for all 8 subbands the frame was declared to be noise only. The noise estimate was then updated using a single-sided one pole recursive filter with a smoothing factor of 0.95 and the updated noise frame was used for the successive frames.

2.9 Drawbacks of existing algorithms

2.9.1 Drawbacks of MS [1]

As explained in section 2.1, the noise power spectrum estimate in [1] was obtained by tracking the minimum of noisy speech power spectrum over a specified window of length L frames. This length was based on the concept that it will encompass at least one silence period of the noisy speech and in turn means that it tracks at least one frame of noise only region. Also, since there was no way to adjust the length of this window based on the width of the speech peaks, this window length was chosen large enough to encompass the broadest peak possible in any speech waveform.

Considering the case of increasing noise levels, this minimum search algorithm had the delay of 2L frames which was approximately equal to 1.6-2.8s. Hence to reduce this delay, the whole window was divided into U subwindows of V frames each which resulted in L+V frames of

delay for increasing noise levels. Even this value was slightly greater than 1.5s [1]. This situation is illustrated by the Figure 2.9 where initially we have a high SNR of 20dB followed by a speech of low SNR of 5dB. Similar kind of situation was possible in cellular phone environments when the user moves from a quiet environment to a noisy environment.

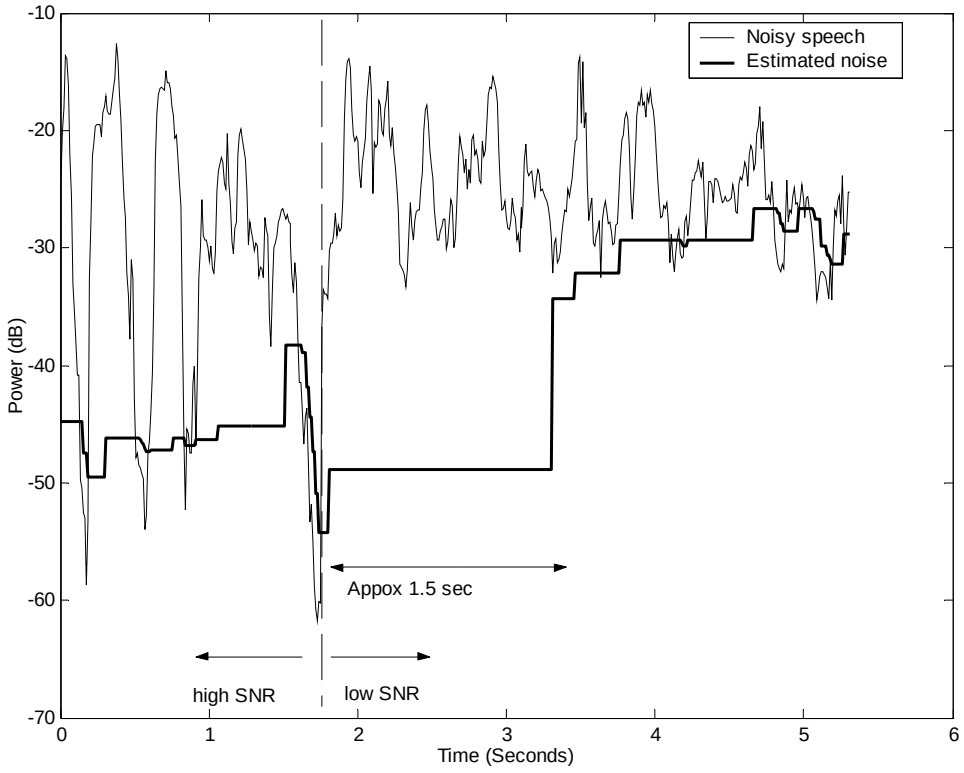

Fig 2.9. Plot of noisy speech power spectrum and noise estimate using [1] for a noisy speech at 20dB SNR (t<1.8s) followed by a noisy speech at 5dB SNR (t>1.8s) at f=500 Hz.

2.9.2 Drawbacks of MCRA [2]

Similar to [1] the major drawback with the noise estimation algorithm in [2] was the update of local minimum of noisy speech for increasing noise levels. According to the minimum tracking rule in [2], the minimum value was chosen as the minimum of previous local minimum estimate and the current noisy speech power as defined in (2.15). Also, to avoid falling down to

the global minimum, the temporary variable was updated for every L frames and set equal to the noisy speech power spectrum at that frame as in (2.16). The local minimum was updated using the temporary variable at the $2L^{th}$ frame. Hence this rule of minimum tracking takes at most 2L frames for updating the local minimum for increasing noise levels. Also, for typical window lengths of around 0.5-1.5s, the method takes around 1-3s for updating to higher noise levels. This is illustrated in Figure 2.10 using the similar example that was used in the previous section.

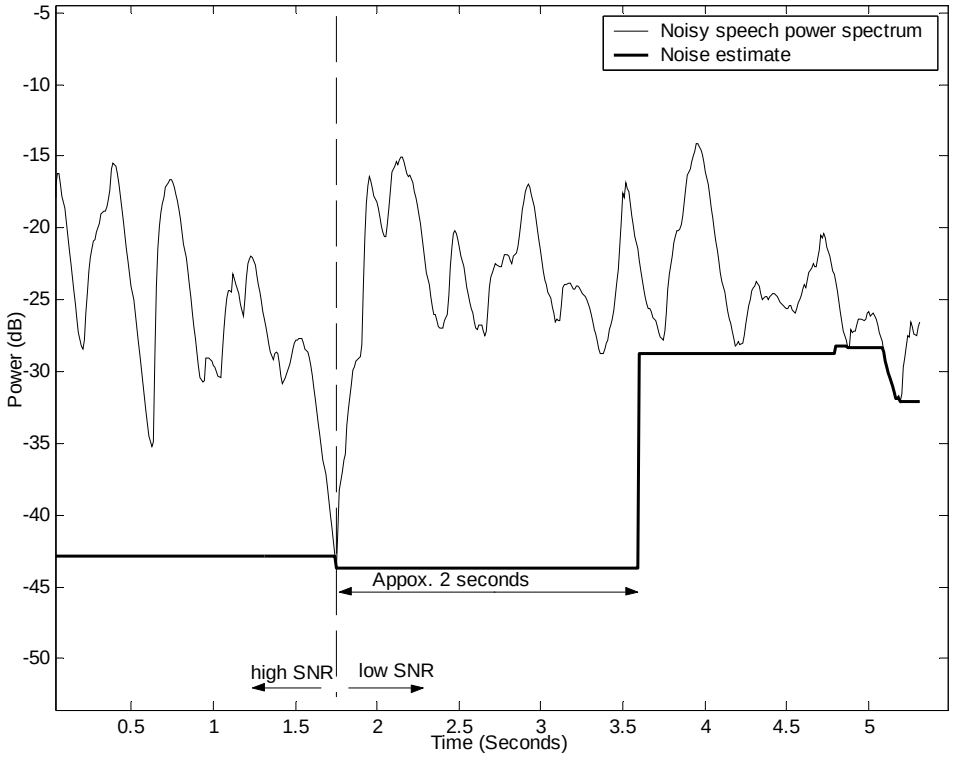

Fig 2.10. Plot of noisy speech power spectrum and noise estimate using [2] for a noisy speech at 20dB SNR (t<1.8s) followed by a noisy speech at 5dB SNR (t>1.8s) at f=500 Hz.

2.9.3 Drawbacks of IMCRA [3]

Even in the improved version of MCRA some variation of minimum statistics rule was used for minimum tracking. Even though the delay for this method was slightly less than that for

minimum statistics approach, this method takes slightly less than 1.5s to update noise estimate for increasing noise levels. Figure 2.11 shows the power spectrum of the true noise and estimated noise spectrum for the same example considered above.

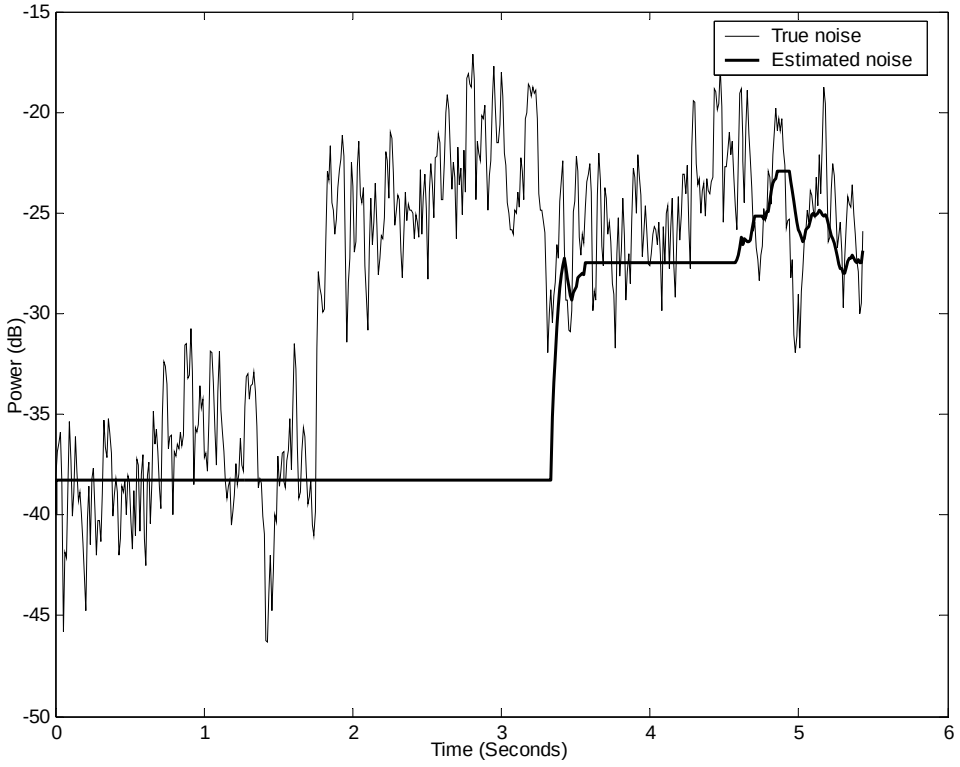

Fig 2.11. Plot of true noise and estimated noise spectrum using [3] for a noisy speech at SNR 20dB (t<1.8s) followed by a noisy speech at SNR 5dB (t>1.8s) at f=500 Hz.

2.9.4 Drawbacks of Continuous minima tracking [4]

The non-linear tracking used for noise estimation in [4] has continuous smoothing without any distinction between speech absent or present segments. Hence the noise estimate increases whenever the noisy speech power spectrum increases irrespective of the changes in noise power level.

This will result in overestimation of noise during speech activity and may result in clipping of speech power. Also, the increase in noise estimate during high speech region may result in clipping of low speech energy regions that immediately follows the high power speech region. Figure 2.12 shows the noise spectrum estimated using [4] and the true noise spectrum. The power spectrum of noisy speech is also shown for the same frequency bin to illustrate that the noise estimate increases whenever the noisy speech power increases. Some of those regions are indicated by arrows in the bottom panel of the figure.

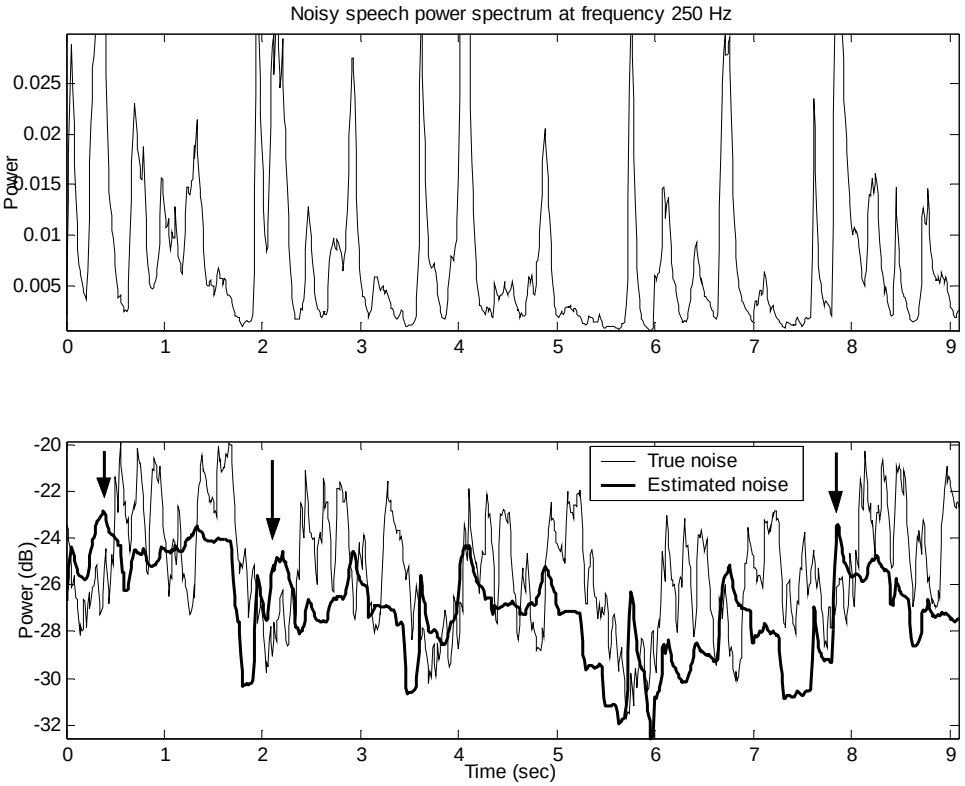

Fig 2.12. Top panel: Plot of noisy speech power spectrum at f = 250 Hz. Bottom panel: Plot of true noise spectrum and estimated noise spectrum using [4] for a noisy speech (5dB SNR) at f= 250 Hz .

2.9.5 Drawbacks of weighted averaging [5]

In [5] a very simple and computationally efficient procedure was used for noise estimation. This method however has a drawback when a high SNR speech segment was followed by a low SNR speech segment. In this case, the initial estimate of the noise power spectrum was very low during the high SNR regions. Hence the threshold which was based on the current noise estimate was also very low. During the low SNR regions the noise energy itself will be very high. But the threshold for finding noise only regions will be very low since it was based on the noise estimate at high SNR region. This may result in a situation where the noisy speech will never become smaller than the threshold. Thus the noise estimate will never be updated when the noise power stays at that higher level itself.

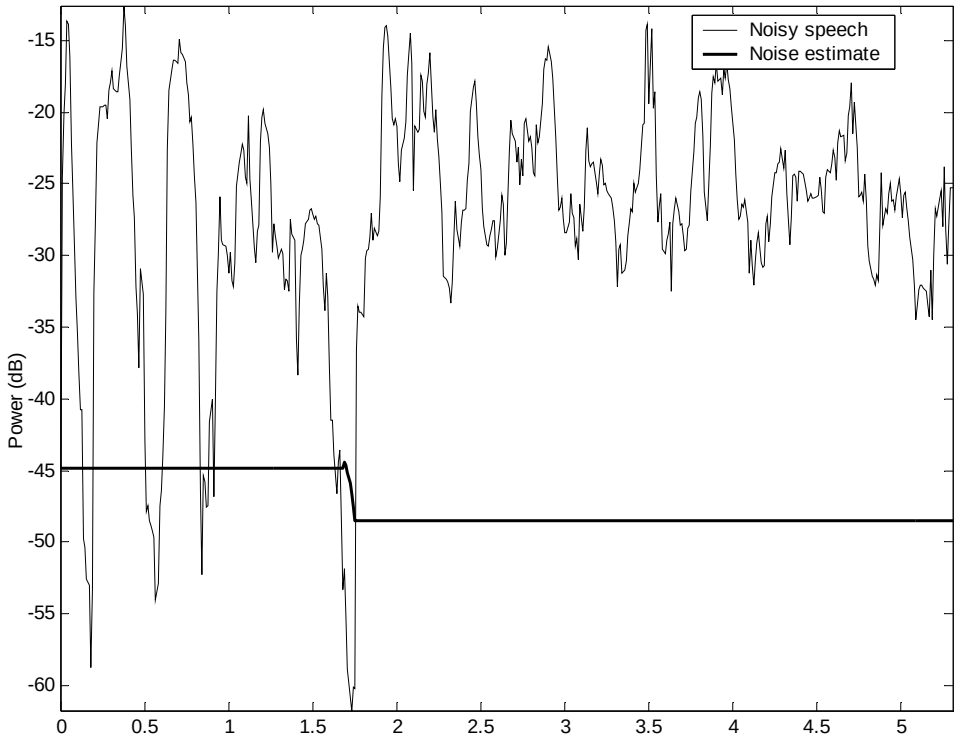

Fig 2.13. Plot of noisy speech power spectrum and noise estimate using [5] for a noisy speech at 20dB SNR (t<1.8s) followed by a noisy speech at 5dB SNR (t>1.8s) at f=500 Hz .

This situation was illustrated with the same example considered for the other algorithms above (a 20dB SNR speech followed by a 5dB SNR speech) in Figure 2.13. The figure shows the noisy speech power spectrum and the estimated noise power spectrum for a single frequency bin over time. Both weighted average technique and histogram based method given in [5] uses the same rule for updating the noise only segments. Hence both will fail to track the noise level completely for the case discussed above.

CHAPTER 3

PROPOSED NOISE ESTIMATION ALGORITHMS

3.1 Introduction

In this chapter two single channel noise estimation algorithms which are based on estimating the noise power spectrum using only the power spectrum of noisy speech were presented. As most of the other algorithms explained in the previous chapter, our methods are also based on tracking the minimum of the noisy speech power spectrum to track the noise only regions. However, our proposed algorithms do not wait for specific window time to update the noise estimate. Hence the update for varying noise power levels is much faster compared to most other algorithms and at the same time noise power spectrum is not overestimated. First the proposed noise estimation algorithm-1 is presented and then compared with some of the existing noise estimation algorithms. This algorithm was presented in [18]. Then, the proposed noise estimation algorithm-2 is presented which is a modified version of algorithm-1.

3.2 Proposed noise estimation algorithm-1

The first method uses two different algorithms to update the noise estimate. The first algorithm updates the noise power spectrum during speech-absent frames and the second algorithm updates the noise power spectrum during speech-present frames. Hence the noise estimate is updated continuously in every frame irrespective of speech present or absent frames. This is based on the concept that the power spectrum of speech was both localized in time and frequency, i.e. even in the speech present frames only a fraction of the entire frequency spectrum

has speech energy in it and all other regions have near zero speech power. This can be illustrated in Figure 3.1.

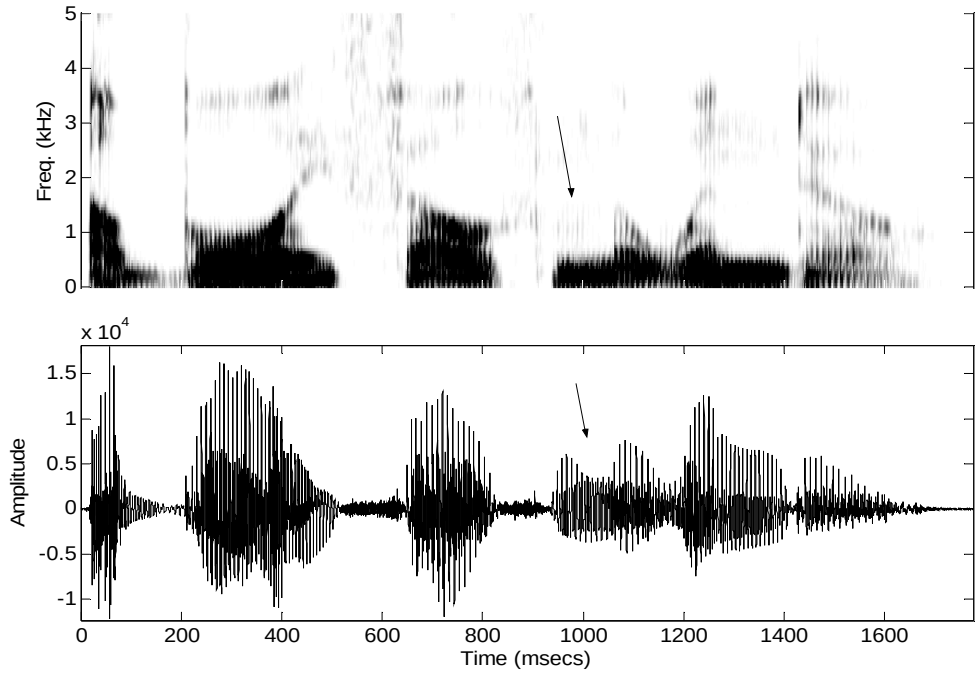

Fig 3.1. Plot of clean speech waveform and the corresponding spectrogram

In figure 3.1, consider the voiced segment in the speech waveform near 1000 msecs. Comparing the power spectrum of the corresponding frame in the spectrogram, it can be seen that it has primarily energy in the low frequency region. Energy in the high frequency for that particular frame is considerably low. On the other hand, unvoiced sounds have energy mainly in the high frequency regions. Hence by tracking the speech-absent frequency regions in each frame, the noise estimate can be updated even in the speech-present frames.

Let the noisy speech be denoted in the time domain as

$$y(n) = x(n) + d(n) \tag{3.1}$$

where $x(n)$ is the clean speech and $d(n)$ is the additive noise. The smoothed power spectrum of the noisy speech can be found using the first-order recursive formula as follows:

$$P(\lambda,k) = \eta P(\lambda,k) + (1-\eta)\left|Y(\lambda,k)\right|^2 \tag{3.2}$$

where $\left|Y(\lambda,k)\right|^2$ is the short time power spectrum of noisy speech and η is the smoothing constant. Now the overall algorithm can be summarized by the flow chart diagram shown in Fig 3.2.

3.2.1 Classifying noisy speech into speech present/absent frames

As explained in Chapter 2 the power spectrum of the noisy speech is equal to the sum of the speech power spectrum and noise power spectrum since speech and the background noise are assumed to be independent. Also in any speech sentence there are pauses between words which do not contain any speech. Those frames will contain only background noise. The noise estimate can be updated by tracking those noise-only frames. To identify those frames, a simple procedure is used which calculates the ratio of noisy speech power spectrum to the noise power spectrum at three different frequency bands:

$$\xi_L(\lambda) = \frac{\sum_{k=1}^{LF} P(\lambda,k)}{\sum_{k=1}^{LF} D(\lambda-1,k)}, \xi_M(\lambda) = \frac{\sum_{k=LF+1}^{MF} P(\lambda,k)}{\sum_{k=LF+1}^{MF} D(\lambda-1,k)}, \xi_H(\lambda) = \frac{\sum_{k=MF+1}^{Fs/2} P(\lambda,k)}{\sum_{k=MF+1}^{Fs/2} D(\lambda-1,k)} \tag{3.3}$$

where $D(\lambda,k)$ is the estimate of noise power spectrum for the current frame and LF, MF and Fs correspond to the frequency bins of 1 KHz, 3 KHz and sampling frequency respectively. If all the three ratios mentioned in (3.3) are smaller than the threshold σ, that frame is concluded as a noise-only frame. Otherwise, if any one or all the ratios are greater than threshold that frame is considered as speech present frame.

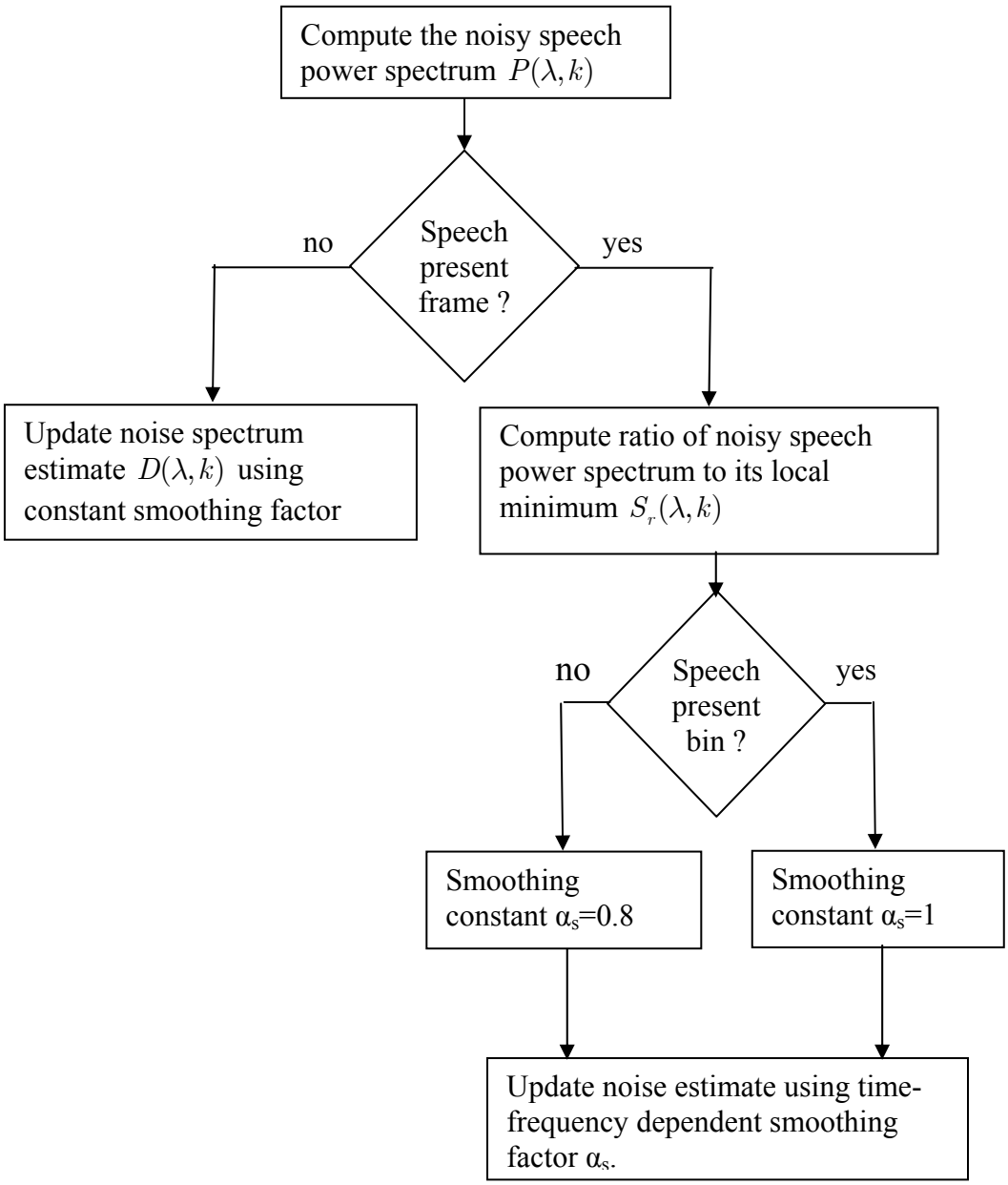

Fig 3.2. Flow diagram of the proposed algorithm-1

3.2.2 Update of noise estimate for speech absent frames

The noise estimate is then updated with a constant smoothing factor if the frame is classified as speech absent frame. This rule can be stated as follows

$$\begin{aligned} &\text{If } \xi_L(\lambda) < \sigma \text{ and } \xi_M(\lambda) < \sigma \text{ and } \xi_H(\lambda) < \sigma \text{ then} \\ &\quad D(\lambda,k) = \varepsilon D(\lambda-1,k) + (1-\varepsilon)\left|Y(\lambda,k)\right|^2 \\ &\text{end} \end{aligned} \tag{3.4}$$

where ε was the smoothing constant. Similar kind of rule was used in [16] in which the ratio was found for the whole frequency range as follows.

$$XNR(\lambda) = \frac{\sum_k P(\lambda,k)}{\sum_k D(\lambda-1,k)} \tag{3.5}$$

Eq. (3.5) is similar to Eq. (3.3) in the sense that both find the ratio of noisy speech power spectrum to the previous noise estimate. The drawback with Eq. (3.5) is that both the noisy speech and noise power spectrum were averaged over the entire frequency spectrum. This can mask the high frequency contents since the energy of low frequency bins are high compared to the energy of the high frequency bins. Hence some unvoiced frames may be classified as speech absent frames. To overcome this, the SNR was found in three different frequency bands separately and each of them was compared to a threshold to determine whether it was a speech present or absent frame. This also retained some weak consonants whose energy was concentrated in a very narrow frequency band.

3.2.3 Update of noise estimate for speech present frames

The proposed algorithm for updating noise spectrum in speech present frames was based on classifying speech present or absent frequency bins in each frame. This was done by tracking the local minimum of noisy speech and then deciding speech presence in each frequency bin

separately using the ratio of noisy speech power to its local minimum. Based on that decision a frequency-dependent smoothing parameter was calculated to update the noise power spectrum.

3.2.3.1 Tracking the minimum of noisy speech

Different methods [1,2] were proposed for tracking minimum of noisy speech based on finding the minimum over a fixed window length. These methods were more sensitive to outliers and also the noise update time was dependent on the length of the window used. In our method, a different non-linear rule was used for tracking the minimum of the noisy speech by continuously averaging the past spectral values [4].

$$
\begin{aligned}
&\text{If } P_{\min}(\lambda-1,k) < P(\lambda,k) \text{ then} \\
&\qquad P_{\min}(\lambda,k) = \gamma P_{\min}(\lambda-1,k) + \frac{1-\gamma}{1-\beta}(P(\lambda,k) - \beta P(\lambda-1,k)) \\
&\text{else} \\
&\qquad P_{\min}(\lambda,k) = P(\lambda,k) \\
&\text{end}
\end{aligned}
\tag{3.6}
$$

where $P_{\min}(\lambda,k)$ is the local minimum of the noisy speech power spectrum and β and γ are constants whose values are determined experimentally. The look-ahead factor β controlled the adaptation time of the local minimum. Figure 3.3 shows the power spectrum of noisy speech and the local minimum tracked with the above mentioned rule for a speech degraded by babble noise at 5 dB SNR. The adaptation time for the algorithm was approximately 0.5s for a general non-stationary noise environment.

3.2.3.2 Speech presence detection

The approach taken to determine the speech presence in each frequency bin was similar to the method used in [2]. Let the ratio of noisy speech power spectrum and its local minimum be defined as

$$S_r(\lambda,k) = P(\lambda,k) / P_{\min}(\lambda,k) \tag{3.7}$$

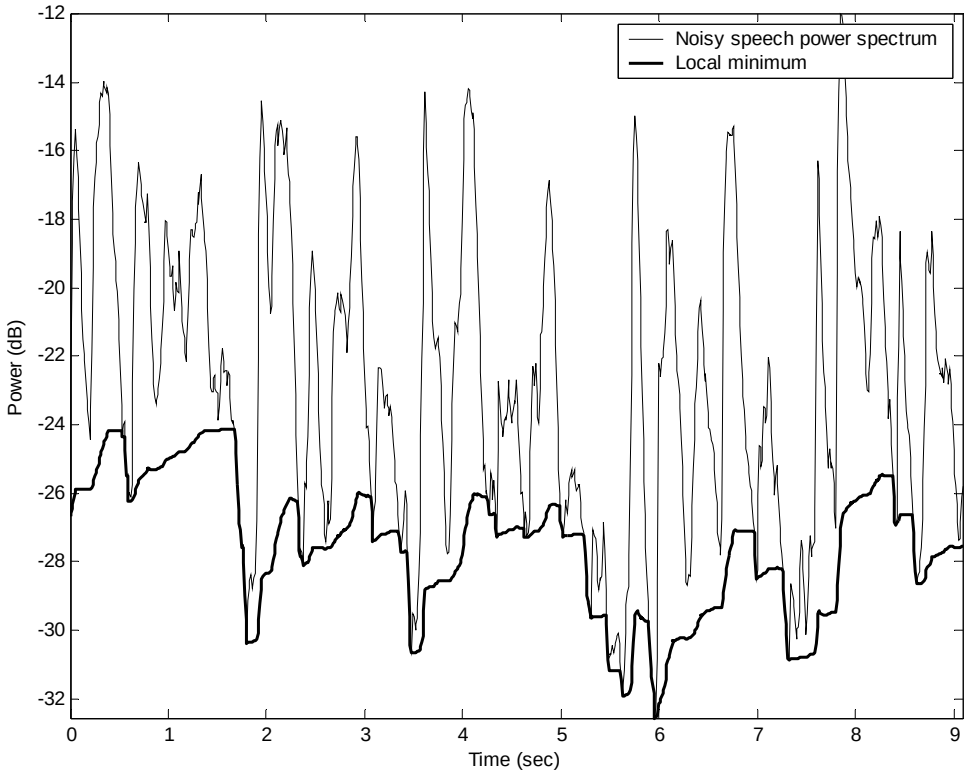

Fig 3.3. Plot of noisy speech power spectrum and local minimum using (3.6) for a noisy speech (5dB SNR) at f = 250 Hz.

This ratio is then compared with a frequency-dependent threshold, and if the ratio is greater than the threshold, it is taken as speech present frequency bin else it is taken as speech absent frequency bin. This is based on the principle that the power spectrum of noisy speech will be nearly equal to its local minimum when speech is absent. Hence smaller the ratio defined in (3.7)

the higher the possibility that it will be a noise-only region or vice versa. This can be summarized as follows:

If $S_r(\lambda,k) > \delta(k)$, then

speech present frequency bin *k*

else (3.8)

speech absent frequency bin *k*

where δ(k) is the frequency dependent threshold whose optimal value is determined experimentally.

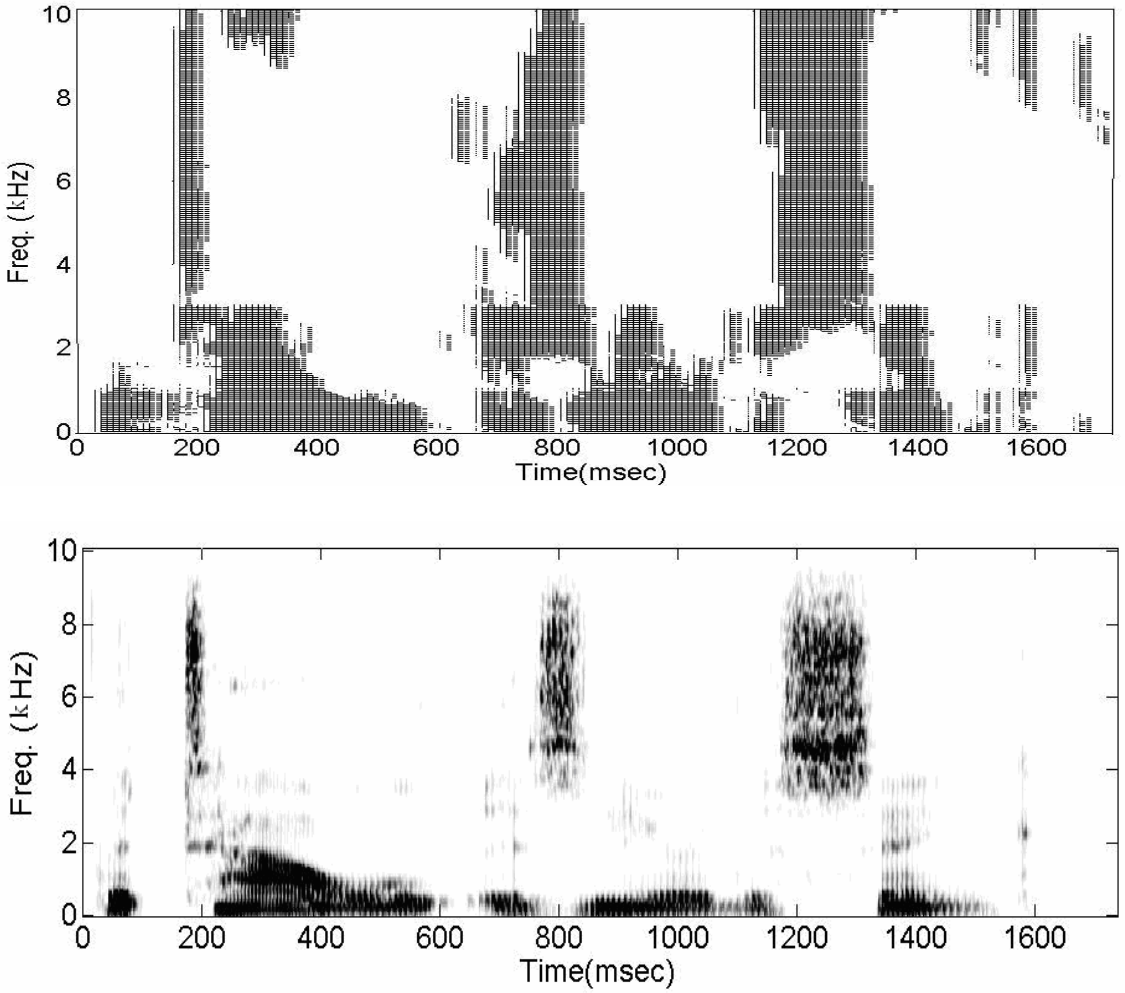

Fig 3.4. Top panel: Plot of speech presence detection from noisy speech based on the ratio $S_r(\lambda,k)$ using (3.8). Bottom panel: Spectrogram of the clean signal.

Figure 3.4 illustrates the speech presence or absence detection using the above rule. In the figure we compare the speech present/absent detection in a speech degraded by babble noise at 5 dB SNR with that of the spectrogram of the clean speech. In the top panel of Figure 3.4, the shaded regions are speech-present regions and the unshaded regions are speech-absent regions identified using the rule given in Eq. (3.8). It can be seen that our detection process had detected almost all the speech present regions correctly. Also, the threshold was chosen smaller than optimum so that we detect speech presence with higher confidence than speech absence to avoid any speech distortion. This can also be observed from the figure since some of the noise only regions were detected as speech. But only few of the low-energy speech regions were detected as noise-only regions. This may result in slight overestimate of the noise spectrum which will not have much effect on the overall enhanced speech.

3.2.3.3 Calculating frequency dependent smoothing constant

After classifying the frequency bins as either speech present or absent, the new frequency-dependent smoothing factor can be derived as follows

$$\alpha_s(\lambda,k) = \begin{cases} \alpha_1 & \text{speech present} \\ \alpha_2 & \text{speech absent} \end{cases} \tag{3.9}$$

where α_1 and α_2 are smoothing constants and $\alpha_1 < \alpha_2$. The value of α_1 was taken to be nearly equal to 1 so as to keep the noise estimate constant during high speech activity.

3.2.3.4 Update of noise spectrum estimate

After computing the frequency-dependent smoothing factor $\alpha_s(\lambda,k)$ using Eq. (3.9) , the noise spectrum estimate is updated as in Eq. (3.10).

$$D(\lambda,k) = \alpha_s(\lambda,k)D(\lambda-1,k) + (1-\alpha_s(\lambda,k))\left|Y(\lambda,k)\right|^2 \tag{3.10}$$

Hence the overall algorithm can be summarized as follows. If the ratios defined in Eq. (3.3) indicated that the current frame is a speech-absent frame, the noise estimate is updated with a fixed smoothing factor for all the frequencies using Eq. (3.4). Otherwise, only a fraction of the frequency bins of the noise spectrum is updated using Eq. (3.10) based on speech presence/absence detection.

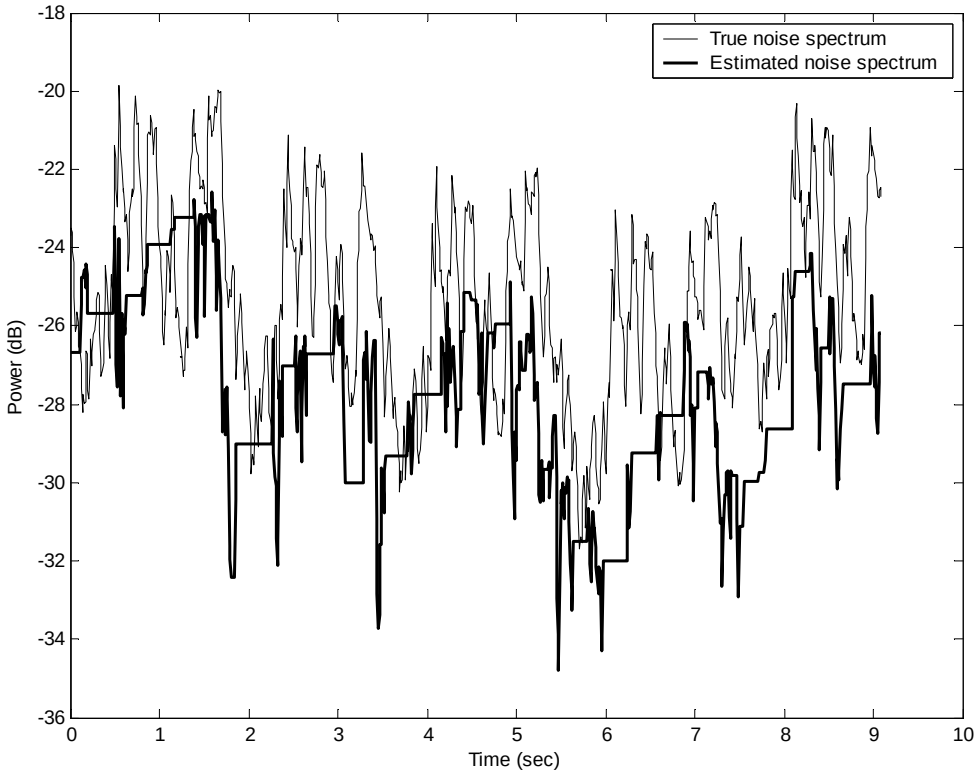

Fig 3.5. Plot of true noise spectrum and estimated noise spectrum using proposed method-1 for a noisy speech (5dB SNR) at f=250 Hz.

Figure 3.5 shows the true noise spectrum and the estimated noise spectrum calculated with proposed method-1 for speech degraded by babble noise at 5 dB SNR.

3.3 Comparison of proposed method-1 with existing algorithms

3.3.1 Comparison with MS [1]

Both the proposed method-1 and [1] track the minimum of the noisy speech in order to update the noise estimate. In both of the methods the adaptation time of the noise estimate depends on the adaptation time of the local minimum. For a non-stationary noise condition where the noise power varies slowly over time, both methods had same adaptation time. But for a sudden increase in noise level the adaptation time was slightly more than 1.5s for [1] whereas it is only 0.5s for our proposed method-1. Figure 3.6 shows a comparison between [1] and our proposed method-1 for the case where there was a sudden increase in noise power level. From the figure it can be seen that [1] takes more than 1.5s to update whereas our proposed method takes only 0.5s to update to the higher noise floor.

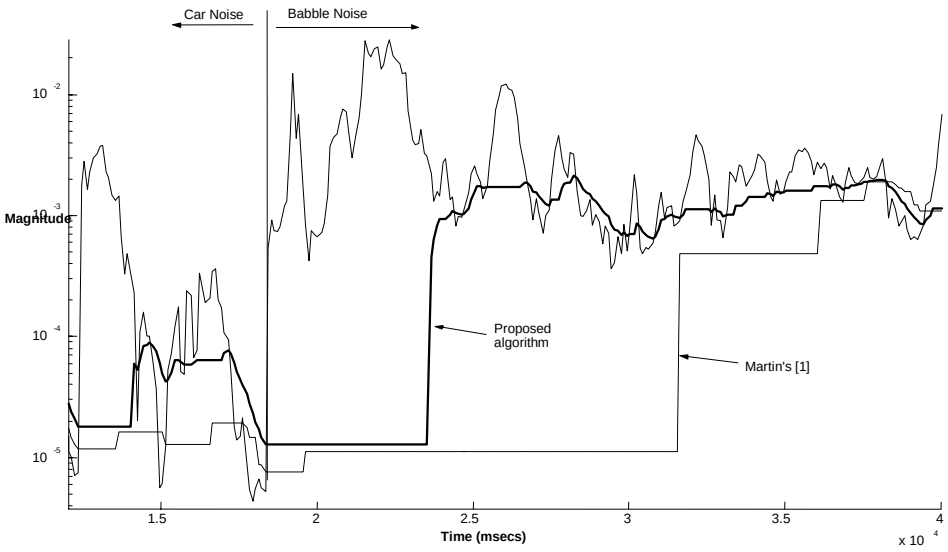

Fig 3.6. Comparison between the noise spectrum (for f=1.5 kHz) estimated using the proposed algorithm-1 (thick line) and Martin's [1] (dashed line) algorithm for a sentence corrupted by car noise (t <1.8 s) followed by a sentence corrupted by multi-talker babble (t>1.8 s).

3.3.2 Comparison with continuous minima tracking [4]

In [4] the noise estimate increases whenever the noisy speech power increases. This problem was avoided in our proposed method by using the ratio of the noisy speech to the local minimum to update the noise estimate. Whenever the noisy speech power increases the ratio between the noisy speech and local minimum also becomes greater than the threshold and the noise estimate is not updated. This can be seen from Figure 3.7 which compares the noise estimate from [4] and proposed method-1 with that of the true noise spectrum.

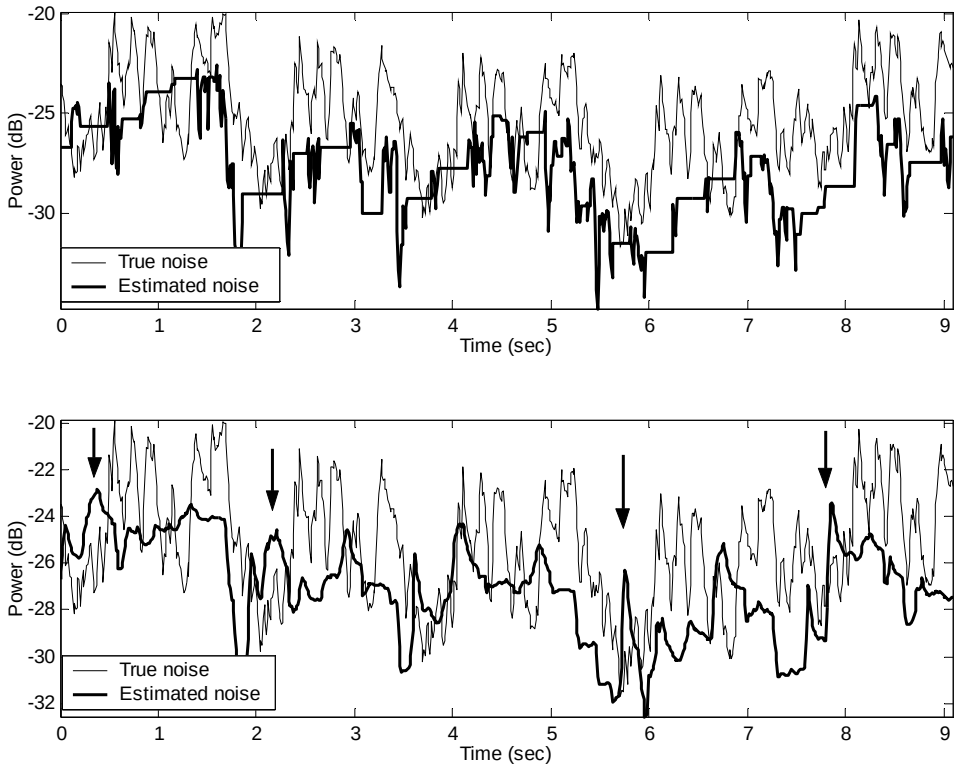

Fig 3.7. Top panel: Plot of true noise spectrum and estimated noise spectrum using proposed method-1 for a noisy speech (5dB SNR) at f=250 Hz. Bottom panel: Plot of true noise spectrum and estimated noise spectrum using [4] for a noisy speech (5dB SNR) at f=250 Hz. Arrows indicate regions where noise is overestimated.

3.3.3 Comparison with weighted average technique [5]

In [5], two methods were presented for noise estimation one based on weighted averaging and other based on the histogram of past noise segments. In both methods, the noise estimate was updated whenever the noisy speech became less than a threshold which was proportional to the previous noise estimate. Consider the same example as in section 3.3.1 where there was a sudden increase in noise level. This resulted in a situation in which the noisy speech spectrum is never smaller than the threshold since the threshold was based on the past noise estimates which was very low.

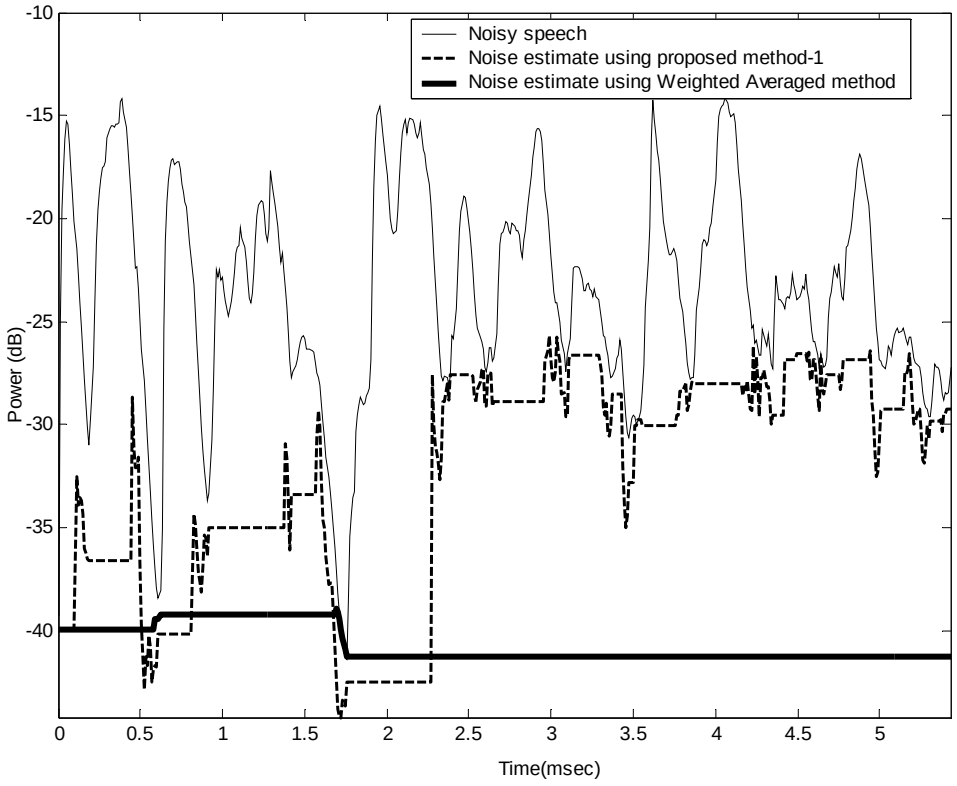

Fig 3.8. Comparison of estimated noise spectrum (f = 500 Hz) of proposed method-1(dashed line) with that of [5] (solid line) for a noisy speech of SNR 20dB (t<1.8s) followed by a noisy speech of SNR 5dB (t>1.8s).

Thus the noise estimate is never updated if the noise power remained at that higher level. In contrast, our proposed method-1 tracked the higher noise power in approximately 0.5s. Figure

3.8 shows the comparison of the noise estimate using our proposed method-1 and [5] with true noise spectrum for the same example as in Section 3.3.1.

3.3.4 Comparison with MCRA [2]

The local minimum in method [2] was found as a minimum of noisy speech over a window of length L frames. This had some drawbacks. Firstly, the minimum was sensitive to outliers. Secondly, the update of the minimum can take at most 2L frames for increasing noise levels. For typical window lengths of 1s, this delay is approximately 2s. But the local minimum tracked using our proposed method (Eq. (3.6)) takes only 0.5s to update to increasing noise levels. The threshold for comparing the ratio to identify the speech presence/absence detection was frequency dependent as opposed to a fixed threshold in [2].

In our proposed method the frequency-dependent smoothing factor has only 2 different levels for speech presence/absence detection. It does not exploit the property that adjacent frames have high correlation for speech presence.

3.4 Experimental results for proposed algorithm-1

The proposed noise estimation algorithm was combined with a Wiener-type speech enhancement algorithm [19], in which the spectral gain function was calculated as follows

$$G(\lambda,k) = \frac{C(\lambda,k)}{C(\lambda,k) + \mu_k D(\lambda,k)} \tag{3.11}$$

where $C(\lambda,k)$ was the estimated clean speech spectrum found from the noisy speech and noise estimate as follows

$$C(\lambda,k) = \max\left\{\left|Y(\lambda,k)\right|^2 - D(\lambda,k), \nu D(\lambda,k)\right\} \tag{3.12}$$

where ν=0.001 was a small positive number. The max(.) operation was used to avoid getting a negative value for the estimated clean speech spectrum if the noise estimate was greater than the noisy speech power spectrum. The over subtraction factor μ_k in Eq. (3.11) was determined from the a posteriori segmental SNR according to Fig. 3.9 [20].

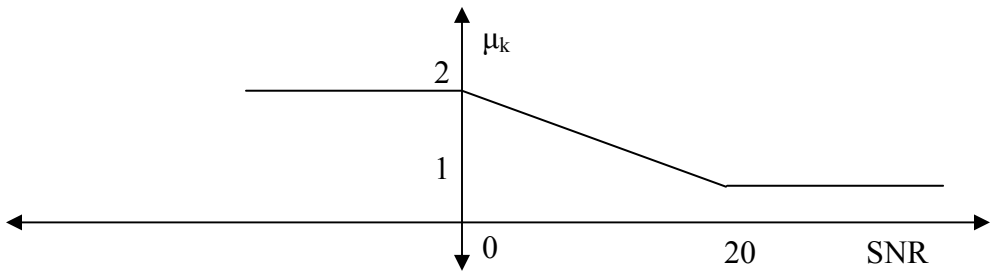

Fig 3.9. Plot of the multiplication factor μ_k in Eq. (3.11) for different values of a posteriori SNR of noisy speech.

The performance of the proposed method was evaluated using both subjective and objective measures. Speech was segmented into 20-ms frames using a Hamming window with 50% overlap. The following values were used in the implementation for Fs = 20.1 kHz: $\alpha_1 = 0.8$, $\alpha_2 = 1$, $\beta = 0.8$, $\gamma = 0.998$, $\eta = 0.7$, $\sigma = 1.3$, ε=0.8 and

$$\delta(k) = \begin{cases} 1.3 & 1 \le k \le LF \\ 3 & LF < k \le MF \\ 5 & MF < k \le Fs/2 \end{cases}$$

3.4.1 Objective measure

The relative mean squared error between the true noise spectrum and the estimated noise spectrum was calculated as follows

$$MSE = \frac{1}{L}\sum_{\lambda=0}^{L-1} \frac{\sum_k \left[\hat{\sigma}_D^2(\lambda,k) - \sigma_D^2(\lambda,k)\right]^2}{\sum_k \sigma_D^2(\lambda,k)} \tag{3.13}$$

where $\hat{\sigma}_D^2(\lambda,k)$ and $\sigma_D^2(\lambda,k)$ were the estimated and true noise spectrum respectively and L was the total number of frames in the noisy speech.

3.4.2 Subjective measure

The performance of the method was compared with that of the methods in [1,2,4,5] by formal listening tests which included two different noise types, namely a single noise source and three different sources, henceforth referred as triplet noise. The same speech enhancement algorithm [19] was used for all noise estimation algorithms and for all conditions. In the single-noise case, the speech sentence was degraded by either babble noise or factory noise. In the triplet-noise case, three different sentences are concatenated to evaluate the adaptation of the algorithm for different noise types. The three noise types include multi-talker babble (2 male and 2 female speakers), factory noise and white noise. Thus a noisy speech was composed of a sentence degraded by babble noise followed by a sentence degraded by factory noise and a sentence degraded by white noise without any pause between the sentences. The overall SNR of the noisy speech was 5dB for both cases. The speech sentences were taken from the HINT [21] database.

For the single noise case, 40 noisy speech sentences were used (20 sentences corrupted by babble noise and 20 sentences corrupted by factory noise) and for the triplet noise case, 20 sets of triplet sentences were used for each comparison. The listener was presented with a pair of sentences, one processed with our proposed method and the other processed with any one of the other [1,2,4,5] methods and asked to choose the sentence judged to be more natural, easier to listen and free of artifacts. The overall preference in percentage was found for the proposed method and compared to the other methods. The order of the sentences was randomized. Table

3.1 lists the mean percentage, that the proposed method-1 was preferred over other methods averaged over six listeners. Also the MSE was calculated using Eq. (3.13) for all methods in both noise types.

Table 3.1. Percentage of preference for the proposed method-1 compared to other methods for single and mixed type noise. The normalized mean squared error (MSE) between the estimated and true noise spectra is also given.

Method	Single Noise		Mixed Noise	
	Preference	**MSE**	**Preference**	**MSE**
Cohen and Berdugo [2]	60.8%	0.95	80.4%	1.12
Doblinger [4]	40.6%	0.52	82.2%	1.08
Hirsch and Ehrlicher [5]	47.8%	0.52	87.1%	0.87
Martin [1]	55.0%	0.53	58.8%	0.94
Proposed method-1	-	0.54	-	0.75

From the results it can be inferred that the listeners had equal preference for all the methods in single noise case since the percentage of preference was around 40-60% for all the methods. But for the triplet noise case, our algorithm had higher preference compared to the other methods. This was due to the fact that the adaptation time of our algorithm was smaller compared to the other methods for highly non-stationary environments and also the parameters were very robust for different types of noise.

3.5 Proposed noise estimation algorithm-2

As explained in Section 3.3.4, the proposed noise estimation algorithm-1 [18] did not exploit the property that adjacent frames of speech have high correlation for speech presence. Hence in the proposed method-2 this property was exploited. Also the classification of noisy speech frames into speech present/absent becomes more erroneous for low SNR noisy speech. Hence it

was preferable and more desirable that we avoid the use of a voice activity detector. This was done by computing the speech-presence probability for all frequency bins in every frame of noisy speech and then computing the time-frequency dependent smoothing factor based on the speech-presence probability. The noise estimate was then updated using the time-frequency dependent smoothing factor for every frame. The overall algorithm can be summarized by the flowchart diagram in Figure 3.10.

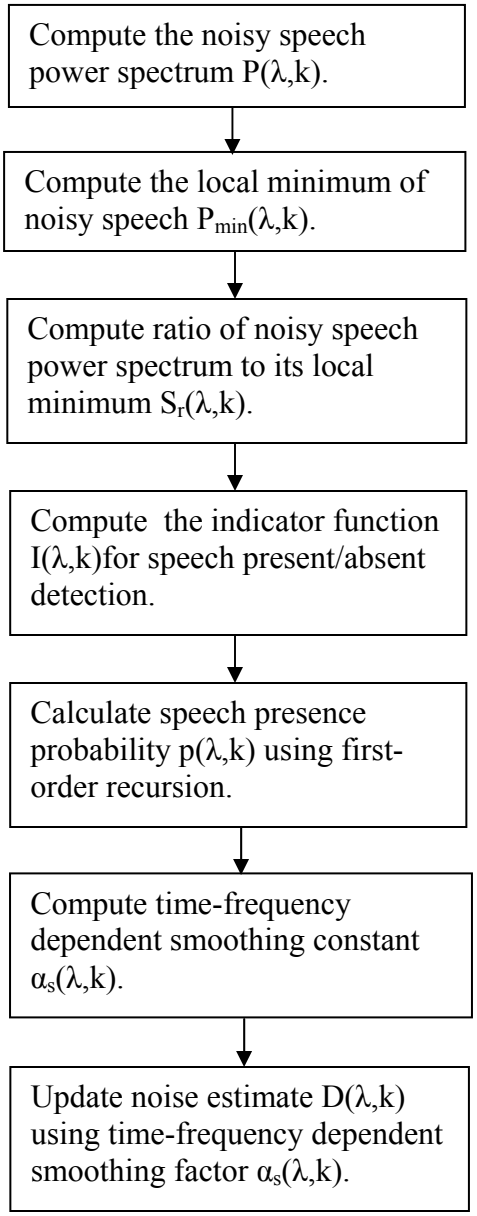

Fig 3.10. Flow diagram of proposed algorithm-2.

3.5.1 Tracking the local minimum of noisy speech

The local minimum $P_{\min}(\lambda,k)$ of smoothed power spectrum of noisy speech was found using the same non-linear rule as explained in Section 3.2.4.1 as follows

$$\begin{aligned}
&\text{If } P_{\min}(\lambda-1,k) < P(\lambda,k) \text{ then} \\
&\quad P_{\min}(\lambda,k) = \gamma P_{\min}(\lambda-1,k) + \frac{1-\gamma}{1-\beta}(P(\lambda,k) - \beta P(\lambda-1,k)) \\
&\text{else} \\
&\quad P_{\min}(\lambda,k) = P(\lambda,k) \\
&\text{end}
\end{aligned} \tag{3.14}$$

where β and γ were constants whose values were determined experimentally.

3.5.2 Computing speech presence probability

The speech presence/absence decision in each frequency bin was obtained by comparing the ratio of noisy speech power spectrum to the local minimum to a frequency dependent threshold as in Section 3.2.3.2 as follows.

$$\begin{aligned}
&\text{If } S_r(\lambda,k) > \delta(k) \\
&\quad \text{speech present} \rightarrow I(\lambda,k) = 1 \\
&\text{else} \\
&\quad \text{speech absent} \rightarrow I(\lambda,k) = 0
\end{aligned} \tag{3.15}$$

From this, the speech presence probability was updated using the first-order recursion as follows [2]

$$p(\lambda,k) = \alpha_p p(\lambda,k) + (1-\alpha_p) I(\lambda,k) \tag{3.16}$$

where α_p is the smoothing constant.

3.5.3 Calculating time-frequency dependent smoothing factor

Using the speech presence probability the time-frequency dependent smoothing factor can be obtained as follows [2]

$$\tilde{\alpha}_d(\lambda,k) \triangleq \alpha_d + (1-\alpha_d)p(\lambda,k) \tag{3.17}$$

Form the smoothing factor, the noise estimate is updated using Eq. (3.10). Figure 3.9 shows the true noise spectrum and the estimated noise spectrum for the same example used in the Section 3.2.3.4.

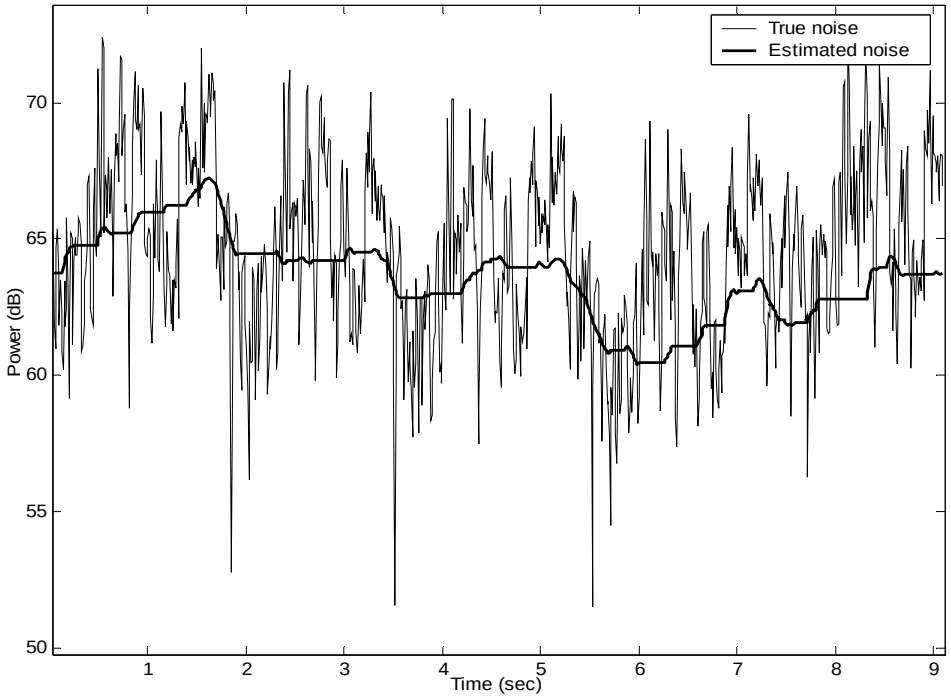

Fig 3.11. Plot of true noise spectrum and the estimated noise spectrum using proposed algorithm-2 for a noisy speech (5dB SNR) at f=500 Hz.

3.6 Experimental results for proposed algorithm-2

The proposed algorithm was combined with the same Wiener-type speech enhancement procedure [19] as explained in Section 3.4. Similar to the proposed method-1, this method was

also evaluated using both subjective and objective measures. The following values were used for the implementation for Fs = 20.1 KHz: $\alpha_d = 0.85$, $\alpha_p = 0.2$, $\beta = 0.8$, $\gamma = 0.998$, $\eta = 0.7$ and

$$\delta(k) = \begin{cases} 2 & 1 \le k \le LF \\ 2 & LF < k \le MF \\ 5 & MF < k \le Fs/2 \end{cases}$$

3.6.1 Objective measure

The relative mean squared error between the true noise spectrum and the estimated noise spectrum was calculated as follows

$$MSE = \frac{1}{L}\sum_{\lambda=0}^{L-1} \frac{\sum_k \left[\hat{\sigma}_D^2(\lambda,k) - \sigma_D^2(\lambda,k)\right]^2}{\sum_k \sigma_D^2(\lambda,k)} \tag{3.18}$$

where $\hat{\sigma}_D^2(\lambda,k)$ and $\sigma_D^2(\lambda,k)$ are the estimated and true noise spectrum respectively and L was the total number of frames in the noisy speech.

3.6.2 Subjective measure

Similar to the proposed method-1, this method was also tested using both single noise and triplet noise. Multi-talker babble noise was used for the single noise case and triplet noise condition was the same as in Section 3.4.2. For each set of comparison, 20 noisy speech sentences corrupted by babble noise were used in single noise condition and 20 sets of triplet sentences were used in the triplet noise condition. The percentage of preference for the proposed method over the other methods was computed for six listeners and tabulated in Table 3.2.

From the results, it can be seen that our proposed method had equal preference compared with the other methods in [1,3-5] for the single noise condition. But for the triplet noise

condition, the proposed method-2 had very high preference compared to all the other methods. This was due to the fact that our noise estimation algorithm adapts quickly to the highly non-stationary environments and also the parameters were very robust to different noise conditions.

Table 3.2. Percentage of preference for the proposed method-2 compared to other methods for single and mixed type noise. The normalized mean squared error (MSE) between the estimated and true noise spectra is also given.

Method	**Single Noise**		**Mixed Noise**	
	Preference	**MSE**	**Preference**	**MSE**
Cohen [3]	48.8%	0.40	81.67%	0.86
Doblinger [4]	53.8%	0.52	81.3%	1.08
Hirsch and Ehrlicher [5]	50%	0.52	78.8%	0.87
Martin [1]	50.83%	0.53	63.8%	0.94
Proposed method-2	-	0.43	-	0.87

CHAPTER 4

SUMMARY AND CONCLUSIONS

This thesis addressed the issue of noise estimation for enhancement of noisy speech. Two algorithms were proposed for noise estimation. The first method included voice activity detection in each frame. In speech absent frames, the noise estimate was updated with a constant smoothing factor. For speech present frames the noise estimate was updated based on speech present/absent detection in each frequency bin. In the second method, the noise estimate was updated continuously in every frame using time-frequency smoothing factors calculated based on speech-presence probability in each frequency bin of the noisy speech spectrum. The speech-presence probability was found using the ratio of noisy speech power spectrum to its local minimum. Unlike other methods, the update of local minimum was continuous over time and did not depend on some fixed window length. Hence the update of noise estimate was faster for very rapidly varying non-stationary noise environments. This was confirmed by formal listening test results which showed significantly higher preference for our methods compared to the other existing algorithms for estimating the noise spectrum.

The contributions of this thesis include.

- Development of a simple voice activity detection method based on estimation of the a posteriori SNR of noisy speech in three different frequency bands.
- Development of a noise estimation algorithm based on continuous tracking of the minimum of noisy speech power spectrum.

- Development of a frequency dependent threshold for comparing the ratio of noisy speech power spectrum to its local minimum for identifying speech present frequency bins.

The proposed noise estimation algorithms proved to be superior to the existing algorithms in many aspects. The algorithms were simple and computationally efficient. These algorithms did not suffer from the drawback of slow update for increasing noise power levels. At the same time, the noise spectrum was not overestimated.

BIBLIOGRAPHY

[1] Martin, R., "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. on speech and audio processing*, vol. 9, no. 5, pp. 504-512, July 2001.

[2] Cohen, I. and Berdugo, B., "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Signal Proc. Letters*, vol. 9, no. 1, pp. 12-15, Jan. 2002.

[3] Cohen, I., "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Trans. on speech and audio processing*, vol. 11, no. 5, pp. 466-475, Sept. 2003.

[4] Doblinger, G., "Computationally efficient speech enhancement by spectral minima tracking in subbands," in *Proc. 4th Eur. Conf. speech, Communication, and Technology, EUROSPEECH'95,* Madrid, Spain, pp. 1513-1516, Sept. 18-21, 1995.

[5] Hirsch, H. G. and Ehrlicher, C., "Noise estimation techniques for robust speech recognition," in *Proc. 20th IEEE Int. Conf. Acoustics, Speech, Signal Processing,* Detroit, MI, pp. 153-156, May 8-12, 1995.

[6] Ephraim, Y. and Malah, D., "Speech enhancement using a minimum mean-square error log spectral amplitude estimator," *IEEE Trans. Acoustic Speech and Signal Processing*, vol. ASSP-33, pp. 443-445, Apr. 1985.

[7] Ephraim, Y. and Malah, D., "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoustic Speech and Signal Processing*, vol. 32, pp. 1109-1121, Dec. 1984.

[8] Kim, N. S. and Chang, J.-H., "Spectral enhancement based on global soft decision," *IEEE Signal Processing Letters*, vol. 7, pp. 108-110, May 2000.

[9] Sohn, J., Kim, N.S. and Sung, W., "A statistical model-based voice activity detector," *IEEE Signal Process. Letters*, vol. 6, no. 1, Jan. 1–3 1999.

[10] Malah, D., Cox, R. V. and Accardi, A. J., "Tracking speech-presence uncertainty to improve speech enhancement in nonstationary noise environments," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing,* pp. 789-792, 1999.

[11] Sohn, J. and Sung, W., "A voice activity detector employing soft decision based noise spectrum adaptation," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing,* pp. 365-368, 1998.

[12] Martin, R., "Spectral subtraction based on minimum statistics," in *Proc. 7th Eur. Signal Processing Conf. (EUSIPCO'94),* Edinburgh, U.K., pp. 1182-1185, Sept. 13-16, 1994.

[13] McCree, A., Truong, K., George, E. B., Barnwell, T. P. and Viswanathan, v., "A 2.4 KBIT/S MELP coder candidate for the new U.S. federal standard," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing,* pp. 200–203-, 1996.

[14] Cohen, I. and Berdugo, B., "Speech enhancement for nonstationary noise environments," *Signal Processing*, vol. 81, pp. 2403-2418, Nov. 2001.

[15] Varga, A. and Steeneken, H. J. M., "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, vol. 12, no. 3, pp. 247-251, July 1993.

[16] Stahl, V., Fischer, A. and Bippus, R., "Quantile based noise estimation for spectral subtraction and wiener filtering," in *IEEE Conf. on Acoustics, Speech, Signal processing,* pp. 1875-1873, June 5-9 2000.

[17] Ahmed, B. and Holmes, H. H., "A voice activity detection using the Chi-Square test," in *Int. Conf. on Acoustics, Speech, Signal processing,* pp. I-625-I-628, May 17-21 2004.

[18] Rangachari, S., Loizou, P. C. and Hu, Y., "A noise estimation algorithm with rapid adaptation for highly non-stationary environments," in *IEEE Int. Conf. on Acoustics, Speech, signal processing,* pp. I-305-I-308, May 17-21 2004.

[19] Hu, Y., Loizou, P.C., "A subspace approach for enhancing speech corrupted by colored noise," *IEEE Signal Processing Letters*, vol. 9, no. 7, pp. 204-206, July 2002.

[20] Berouti, M., Schwartz, R. and Makhoul, J., "Enhancement of speech corrupted by acoustic noise," in *Proc. Int. Conf. on Acoustics, Speech, Signal Processing,* pp. 208-211, Apr. 1979.

[21] Nilsson, M., Soli, S. and Sullivan, J., "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," in *Journal of Acoustic Society of America,* pp. 1085-1099, 1994.

VITA

Sundarrajan Rangachari was born on May 16, 1980, the son of K. Rangachari and R. Padma. He finished his high school in “The higher secondary school for boys, Srirangam”, Trichy, India in 1997. In the same year he was admitted to the Bachelor program in the Electronics and Communication Engineering in Thiagarajar College of Engineering (under the Madurai kamaraj University), Madurai, India. He was awarded the Bachelor of Engineering in Electronics and Communication in April 2001. In November 2001, he joined Reliance Infocom Ltd., as a Project Trainee and worked there till February 2002. In August 2002, he was admitted to the Masters program in Electrical Engineering (specializing in Communications and Signal Processing), at the University of Texas at Dallas. From August 2002 he has been working in the Speech Processing Lab. Also, he was a Teaching Assistant in Electrical engineering department in UTD from August 2002 to May 2003. His research mainly concerns the noise estimation algorithms for speech enhancement.