Noise estimation Algorithms for Speech Enhancement in highly non-stationary Environments

Anuradha R. Fukane[1], Shashikant L. Sahare[2]

[1,2] Electronics and Telecommunication department
Cummins college of Engineering For Women, Pune 411052, Maharashtra, India

Abstract

A noise estimation algorithm plays an important role in speech enhancement. Speech enhancement for automatic speaker recognition system, Man–Machine communication, Voice recognition systems, speech coders, Hearing aids, Video conferencing and many applications are related to speech processing. All these systems are real world systems and input available for these systems is only the noisy speech signal, before applying to these systems we have to remove the noise component from noisy speech signal means enhanced speech signal can be applied to these systems. In most speech enhancement algorithms, it is assumed that an estimate of noise spectrum is available. Noise estimate is critical part and it is important for speech enhancement algorithms. If the noise estimate is too low then annoying residual noise will be available and if the noise estimate is too high then speech will get distorted and loss intelligibility. This paper focus on the different approaches of noise estimation. Section I introduction, Section II explains simple approach of Voice activity detector (VAD) for noise estimation, Section III explains different classes of noise estimation algorithms, Section IV explains performance evaluation of noise estimation algorithms, Section V conclusion.

Keywords: speech enhancement, Noise, VAD, FFT, Histogram.

1. Introduction

Speech enhancement plays an important role in numerous applications such as hearing aids, speech coding, cell phones, automatic recognition of speech signals by machines and many more. Speech signals from the uncontrolled environment may contain degradation components along with the required speech components. Degradation components include back ground noise, reverberation and speech from other speakers. Therefore the degraded speech components need to be processed for the enhancement. Speech enhancement algorithms improve the quality and intelligibility of speech by reducing or eliminating the noise component from the speech signals. Improving quality and intelligibility of speech signals

reduce listener's fatigue, improve the performance of hearing aids, cockpit communication, videoconferencing, speech coders and many other speech processing systems. In most speech enhancement algorithms it is assumed that an estimate of noise spectrum is available. Noise estimate is critical part and it is important for speech enhancement algorithms. Performance of speech enhancement algorithms depends on correct estimation of noise. Simple approach to estimate the noise spectrum of the signal using a Voice Activity Detector (VAD) another approach to estimate the noise using different noise estimation algorithms Noise estimation algorithms that continuously track the noise spectrum. It is challenging task to estimate the noise spectrum even during speech activity hence Researcher developed many noise estimation algorithms which are explained in next section.

2. Voice Activity Detection

Simple approach to estimate and update the noise spectrum during the silent segments of the signal using a Voice Activity Detector (VAD). The process of discriminating between the voice activity that is speech presence and silence that is speech absence is called voice activity detection. VAD algorithms typically extract some type of feature (e.g. short time energy, zero crossing etc.) from the input signal and compared against threshold value, usually determined during speech absent period. Generally output of VAD algorithms is binary decision on a frame-by-frame basis having frame duration 20-30 msec. A segment of speech is declared to contain voice activity (VAD = '1') if measured value exceed a predetermined threshold otherwise it is declared a noise (VAD = '0') figure 1shows VAD decisions. Several VAD algorithms

were proposed based on various types of features extracted from the signal. Noise estimation can have major impact on the quality and Intelligibility of speech signal. The early VAD Algorithms

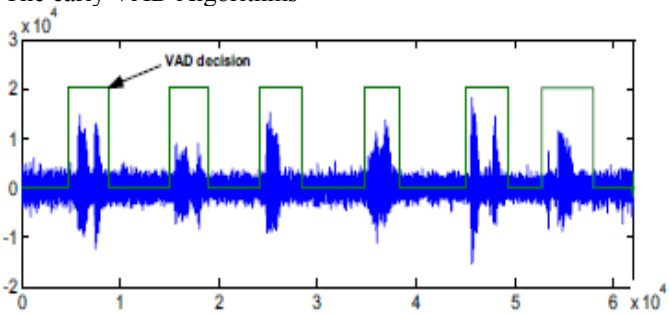

Figure -1 shows VAD decisions [3]

were based on energy levels and zero crossing [4], Ceptral features [4], the Itakura LPC spectral distance measures and the periodicity measures [2]. Some of VAD Algorithms are used in (GSM) System [3], cellular Networks [3], and digital cordless telephone systems [3]. VAD Algorithms are suitable for discontinues transmission in voice communication systems as they can be used to save the battery life of cellular phones. The majority of the VAD Algorithms encounter problems in low SNR conditions, particularly when the noise is non-stationary [1, 2]. Having an accurate VAD Algorithm in a non-stationary environment might not be sufficient in speech enhancement. Applications, as on accurate noise estimation is required at all times, even during speech activity. In case of Noise estimation algorithms they continuously track the noise spectrum therefore more suited for speech enhancement applications in non-stationary Scenarios.

3. Classes of Noise Estimation Algorithms

There are three classes of noise estimation algorithms. Minimal tracking Algorithms, Time Recursive Algorithms and Histogram based Algorithms. All algorithms operate in the following fashion. First the signal is analyzed using short time spectra computed from short overlapping frames, typically 20-30 msec. Windows with 50% overlap between adjacent frames. Then several consecutive frames called analysis segment are used in the computation of the noise spectrum. Typical time span of this segment may range from 400 msec. to 1 sec. The noise estimation algorithms are based on the assumptions that the analysis segment is too long enough to contain speech pauses and low energy signals segments and the noise present in the analysis segment is more stationary than speech, new assumption is that noise changes at a relatively slower rate than speech. The analysis segment has to be long enough to encompass speech pauses and low energy segments, but it also has to be short enough to track fast changes in the noise level, hence the chosen duration of the analysis segment will result from a track-off between these two restrictions. Now we will see different classes of noise estimation Algorithms.

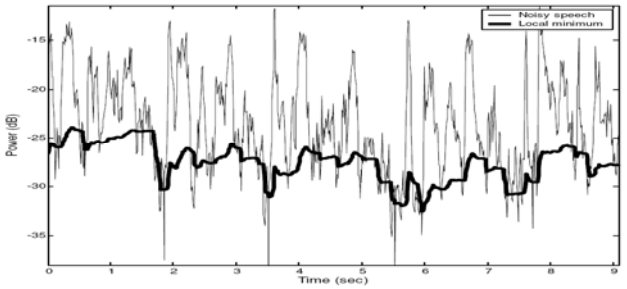

Figure-2 Plot of noisy speech power spectrum and local minimum [10]

3.1. Minimal – Tracking Algorithms

Minimal Tracking Algorithms are based on the assumption that the power of the noisy speech signal in individual frequency bands often decays to the power level of the noise, even during speech activity [12]. Hence by tracking the minimum of the noisy speech power in each frequency band, one can get a rough estimate of the noise level in that band. Two different algorithms were proposed for noise estimation first minimum statistics (MS) on noise estimation, which tracks the minimum of the noisy speech power spectrum within a finite window that is in analysis segment, and 2nd algorithm tracks the minimum continuously without requiring a window are explained in next section. Plot of noisy speech power spectrum and local minimum using (3) for a speech degraded by babble noise at 5dB SNR at frequency bin k=6 is shown in figure 2.

3.1.1. Minimum statistics (MS) Noise Estimation

The Minimum Statistics algorithm was originally proposed by Martin R. (1994) and later refined in [5] to include a bias compensation factor and better smoothing factor. Let y(n) = x(n) + d(n) denote the noise speech signal, where x(n) is the clean speech signal and d(n) is the noise signal, assume that x(n) and d(n) are statistically independent and zero mean. Noisy speech signal is transformed in the frequency domain by first applying a window w(n) to M samples of y(n) and then computing the M-point FFT of the windowed signal.

$$Y(\lambda, k) = y(\lambda M + m)\, w(m)\, e^{-j2\pi mk/M} \qquad (1)$$

Where λ indicates the frame index and k the frequency bin index varient from k = 0, 1, 2 ... M-1. Y(λ, k) is the short term Fourier Transform (STFT) of y(n). Periodogram of the noisy speech is approximately equal to the sum of periodogram of clean speech and noise given as

$$| Y (\lambda, k) |^2 \approx | X (\lambda, k)|^2 + |D(\lambda, k)|2 \quad (2)$$

Where $|Y(\lambda, k)|^2$ is the periodogram of noisy speed signal, $|X (\lambda, k)|^2$ is the periodogram of clean speed signal and $|D(\lambda, k)|^2$ is the periodogram of Noise signal. Because of this assumption, we can estimate the noise power spectrum by tracking the minimum of the periodogram $|Y(\lambda, k)|^2$ of the noisy speech over a fixed window length. The periodogram $|Y(\lambda, k)|^2$ fluctuates very rapidly over time, hence 1st under recursive version of periodogram can be used as

$$P (\lambda, k) = \alpha P (\lambda - 1, k) + (1-\alpha) |Y (\lambda, k)|^2 \quad (3)$$

Where α is the smoothing constant. The above recursive equation in recognized as an IIR Low pass filter, provides a smoothed version of periodogram $|Y(\lambda, k)|^2$. We can obtain an estimate of the power spectrum of the noise by tracking the minimum of P(λ,k).Our finite window smoothing constant α chosen experimentally not too low or too high. There are two main issues with the spectral minimal – tracking approach the existence of a bias in the noise estimate and the possible overestimate of the noise level because of inappropriate choice of the smoothing constant. More accurate noise estimation algorithm can be developed by deriving a bias factor to compensate for the lower noise values and by incorporating a smoothing constant that is not fixed but varies with time and frequency. The noise estimation algorithm using MS is summarized as below [12]. For each frame λ do following steps

1. Compute the short-term periodogram $|Y(\lambda, k)|^2$ of the noisy speech frame.
2. Compute the smoothing parameter α (λ, k) using equation.

$$\alpha(\lambda,k) = \frac{\alpha_{max} \cdot \alpha_c}{(1 + (P(\lambda - 1,k) / \sigma 2(\lambda - 1,k) - 1)}$$

3. Compute the smoothed power spectrum P (λ, k) using equation(3).

$$B_{min}(\lambda,k) \approx 1 + (D-1)\frac{2}{\widetilde{Q}_{eq}(\lambda,k)}$$

4. Compute Bias connection factor β_{min} (λ, k)
5. Search for the minimum psd Pmin (λ, k) over a D- Frame window. Update the minimum Whenever V (V < D) frames are processed
6. Compute α update the noise power spectral density (psd) according to equation

$$\sigma_d^2(\lambda, k) = B_{min} (\lambda, k). P_{min} (\lambda, k)$$

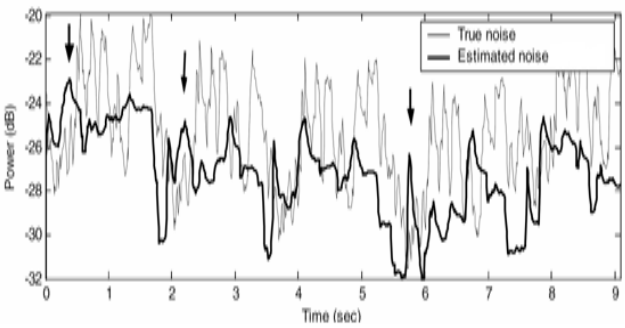

Figure-3 Plot of true noise spectrum and estimated noise spectrum using Continuous Spectral Minimum Tracking Arrows indicate regions where noise is overestimated [12]

3.1.2 Continuous Spectral Minimum

One of the drawbacks of the minimal tracking employed in the MS algorithm is its inability to respond to fast changes of the noise spectrum [12]. A different method for tracking the spectral minima was proposed in [10]. In contrast to using a fixed window for tracking the minimum of noisy speech is in [5] the noise estimate is updated continuously by smoothing the noisy speech power spectra in each frequency bin using a non-linear smoothing rule. For minimum tracking of the noisy speech power spectrum, a short time smoothed version of the periodogram of noisy speech is computed as before using the equation (3) and α is smoothing factor ($0.7 < \alpha < 0.9$). The non-linear rule used for estimating the noise spectrum based on tracking the minimum of the noisy speech power (P_{min} (λ, k)) in each frequency bin as follows

If P_{min} (λ-1, k) < P (λ, k) then

$$P_{min}(\lambda,k) = \gamma P_{min}(\lambda-1,k) + \frac{1-\gamma}{1-\beta}(P(\lambda,k) - \beta P(\lambda-1,k)) \quad (4)$$

else

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

end.

Where Pmin (λ, k) is the noise estimate and the parameter nocks set to α = 0.7, Po = 0.96 and γ = 0.998 [10], β is the look-ahead factor in minimum tracking, which can be adjusted if needed to vary the adaptation time of the algorithm. The typical adaption time using the values mentioned is 0.2–0.4msec figure 3 Shows exchange of continuous minimum tracking based in equation (4). The nonlinear tracking, maintains continuous psd smoothing without making any distinction between speech absent or present segments. Hence the noise estimation increases

whenever the noisy speech power spectrum increases, irrespective of the changes in the noise power level.

3.2 Time Recursive Averaging for Noise Estimation

The time–recursive averaging Algorithms exploit the observation that the noise signal typically has non uniform effect on the spectrum of speech [12], in that some regions of the spectrum will typically have a different effective signal to noise ratio (SNR). As a result, different from bands in the spectrum will have effectively different SNRs. More generally, for any type of noise we can estimate and update individual frequency bands of the noise spectrum whenever the probability of speech being absent at a particular frequency band is high or whenever the effective SNR at a particular frequency band is extremely low. This observation led to the recursive arranging type of algorithms in which noise spectrum is estimated as a weighted average of past noise estimates and the present noisy speech spectrum. The weights change adaptively depending either on the effective SNR of each frequency bin or on the speech present probability. All types of time recursive algorithms have following the general form as follows

$$\sigma_d^2(\lambda-1) = \alpha(\lambda,k)\,\sigma_d^2(\lambda-1,k) + (1-\alpha(\lambda,k))\,|Y(\lambda,k)|^2 \quad (5)$$

Where $|Y(\lambda,k)|^2$ is the periodogram of noisy speech, $\sigma_d^2(\lambda,k)$ denotes the estimate of the noise psd at frame λ, and frequency k and $\alpha(\lambda,k)$is the smoothing factor, which is timed and frequency dependent. Different algorithms were developed depending on the selection of the smoothing factor $\alpha(\lambda,k)$. Some chose to compute $\alpha(\lambda,k)$ based on the estimated SNR of each frequency bin [10] where as others chose to compute $\alpha(\lambda,k)$ based on the probability of speech being present or absent at frequency bin k [6]. Minima controlled Recursive Averaging (MCRA) Algorithm is based on this approach which is explained in next section. These two approaches are conceptually very similar. Other chose to use a fixed value for $\alpha(\lambda,k)$ only after a certain condition was met [6, 7].

3.2.1 Minima Controlled Recursive Averaging (MCRA) Algorithm

According to method explained in [6], the conditional speech presence probability $\hat{P}(\lambda,k)$ is computed by comparing the ratio of the noisy speech power spectrum to its local minimum against a threshold value. The probability estimate $\hat{P}(\lambda,k)$ and the time smoothing factor $\alpha(\lambda,k)$, is controlled by the estimate of spectral minimum and due to this reason this algorithm is called as Minima Controlled Recursive Averaging Algorithm (MCRA). This Algorithm is modified by researchers and some of them are MCRA-2 Algorithm explained in [7], improved MCRA Algorithm explained in [8]. The MCRA noise estimation algorithm proposed by Cohen in [6] is summarized [12] as

1. Smooth noisy psd $S(\lambda,k)$ as follows

$$S(\lambda,k) = \alpha_s S(\lambda-1,k) + (1-\alpha_s)|Y(\lambda,k)|^2 \quad (6)$$

Where α_s is smoothing constant.

2. Perform minimal tracking on $S(\lambda,k)$ using equation (4) to obtain $S_{min}(\lambda,k)$

3. Determine $P(\lambda,k)$ using equation(7)

If $S(\lambda,k) > \delta$ (threshold)
$\hat{P}(\lambda,k) = 1$ speech present
else
$\hat{P}(\lambda,k) = 0$ speech absent
end. (7)

4. Compute the time-frequency dependent smoothing factor $\alpha_d(\lambda,k)$ using equation (8) and the smoothed Conditional probability $\hat{P}(\lambda,k)$ from equation (9).

$$\alpha_d(\lambda,k) = \alpha + (1-\alpha)\,p(\lambda,k) \quad (8)$$

$$\hat{P}(\lambda,k) = \alpha\,\hat{p}(\lambda-1,k) + (1-\alpha p)\,\hat{p}(\lambda,k) \quad (9)$$

5. Update the noise psd $\sigma_d^2(\lambda,k)$ using equation (10)

$$\sigma_d^2(\lambda,k) = \alpha_d(\lambda,k)\,\sigma_d^2(\lambda-1,k) + [1-\alpha_d(\lambda,k)]|Y(\lambda,k)|^2 \quad (10)$$

3.3 Histogram – Based Noise estimation algorithms

In a general mathematical sense, a histogram is a function that counts the number of observations that fall into each of the disjoint categories known as bins, whereas the graph of a histogram is merely one way to represent a histogram as shown in figure-4. Histogram based noise estimation algorithms are motivated by the observation that the Most frequent value (that is the Histogram maximum) of energy values in individual frequency bands corresponds to the noise level of the specified frequency band, that is the noise level corresponds to the maximum of the histogram of energy values. In some cases, the histogram of spectral energy values may contain two modes 1st a low energy mode corresponding to the speech absent and low energy segments of speech and 2nd a high energy mode corresponding to the (noisy) voiced

segments of speech. The noise estimate is obtained based on the histogram of part power spectrum values [11] that is for each in coming frame, 1st construct the histogram of power spectrum

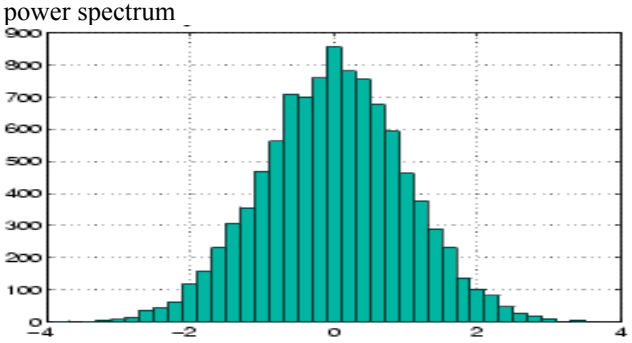

Figure-4 shows histogram of signal

values spanning a window of several hundreds of milliseconds and take as an estimate of the noise spectrum the value corresponding to the Maximum of the histogram values. This is done separately for each individual frequency bin. The histogram based noise estimation is summarized [12] as follows

1. Compute the noisy speech power spectrum
   $|Y(\lambda, k)|^2$.
2. Smooth the noisy psd using 1st order recursion.

   $$S(\lambda,k) = \alpha_S(\lambda-1, k) +(1 - \alpha)| Y(\lambda, k) |^2 \qquad (11)$$

   Where α is smoothing constant.

3. Compute the histogram of D part psd estimates

   S(λ, k) & S (λ-1, k) S (λ-2, k), --------- S (λ-D, k)
   using say 40 bins

4. Let C = [C1, C2, ---- C40] be the counts in each of the 40 bins in the histogram and

   S = [S1, S2, ------ S40] denote the corresponding centers of the histogram bins.

5. Let C_{max} be the index of the Maximum Count C_{max} = arg Max Ci and I varies < i < 40. Then take an estimate of the noise psd denoted by H_{max} (λ, k) the value corresponding to the maximum of the histogram

   $$H_{max} (\lambda, k) = S(C_{max}).$$

6. Smooth the noise estimate H_{max} (λ, k) using 1st order recursion

   $$\sigma_d^2(\lambda, k) = \alpha_m\sigma_d^2 (\lambda-1, k)+(1-\alpha_m) H_{max}(\lambda, k) \qquad (12)$$

   Where σ_d^2 (λ, k) is the smoothed estimate of the noise psd and α_m is a smoothing constant. Various histogram based methods are proposed in [12]

4. Performance Evaluation

The performance of the noise estimation algorithms was assessed using both objective and subjective measures [10]. In objective evaluation[5] the percentage relative estimation error variance were calculated between the true noise spectrum and estimated noise spectrum for white Gaussian noise, vehicular noise, and street noise, using sentences embedded in 15-dB SNR[8]. Results indicated mean errors on the order of a few percent for the white and vehicular noises and a larger error was noted for street noise, which is highly non-stationary in nature. The performance of MS Algorithm was compared with VAD algorithm [6]. Formal listening tests were conducted to evaluate the quality and intelligibility of the enhanced and coded speech. When compared with VAD algorithm, the MS algorithm approach yielded better quality and improved speech intelligibility scores [8]. The tracking of minimum in each frequency bin helped to preserve the weak voiced consonants, which might be classified as noise by most VAD algorithms on their energy is concentrated in a small number of frequency bins that is at low frequencies. Evaluation of the continuous Spectral Minimum Tracking algorithm with minimal tracking algorithm was reported in [7,8] when compared with Continuous Spectral Minimum Tracking algorithm was found to perform better in terms of both objective and subjective measures. Objective and subjective listening calculations of the MCRA–2 algorithm were reported in [10] when MCRA-2 algorithm was integrated in a speech-enhancement algorithm [10] and compared using subjective preference tests against other noise estimation algorithms including MS and MCRA. Subjective evaluation indicated that the speech quality of the MCRA–2 algorithm was better than MCRA and MS algorithm. Histogram based methods are not evaluated with MS, MCRA and other methods.

4. Conclusions

VAD Algorithms are not well suited for non-stationary environment. Noise estimation algorithms estimate and update the noise spectrum continuously, even during speech activity. Noise estimation algorithms are more suited for speech enhancement algorithms operating in highly non-stationary environments. Three different classes of noise estimation algorithms were presented.

Most of the noise estimation algorithms described provide an underestimate of noise the noise spectrum and are not able to respond fast enough to increasing noise levels. Objective evaluation and comparison between various noise estimation algorithms was also presented. Depending on the application one has to select noise estimation algorithm.

Acknowledgments

Anuradha R. Fukane thanks to Dr. S. D. Bhide, Dr. Madhuri Khambete and Prof. S. Kulkarni for their valuable guidance and support.

References

[1] Sohn J. and Kim N.(1999), "Statistical Model based voice activity detection", IEEE Signal Proc.Lett.6(1), 1-3.

[2] Tanyer S. and Ozer H. (2000), "Voice Activity Detection in Non stationary Noise", IEEE Speech Audio Procc.8 (4), pp. 478-482.

[3] Shrinivasan K. and Gersho A. (1993), "Voice Activity Detection for Cellular Network", Prec. IEEE Speech Coding Workshop pp. 85-86.

[4] Haigh J. and Mason J.(1993)"Robust Voice Activity Detection using Ceptral Features," Prec. IEEE TENCON 321-324A.

[5] Martin, R.(2001) "Noise power spectral density estimation based on optimal smoothing and minimum statistics". IEEE Trans. Speech Audio Process 9 (5), pp. 504–512

[6] Cohen, I., 2002. "Noise estimation by minima controlled recursive averaging for robust speech enhancement", IEEE Signal Proc. Letter 9 (1), pp.12–15

[7] Loizou, P, Sundarajan R. and Hu Y (2004) "Noise estimation Algorithm with rapid Adaption for highly non – stationary Environments", Prec. IEEE International Conference on Acoustic Speech signal Proc.

[8] Loizou P. , Sundarajan R. (2006) "A Noise estimation Algorithm for highly non–stationary Environments", speech Communication 48 (2006) Science direct pp. 220-231

[9] Sundarajan Rangachari (2004) "A Noise estimation Algorithm for highly non–stationary Environments", MS Thesis, Department of Electrical Engineering University of Texas-Dallas

[10] Doblinger, G., 1995. "Computationally efficient speech Enhancement by spectral minima tracking in subbands" Proc. Euro speech 2, pp. 1513–1516

[11] Hirsch, H., Ehrlicher, C. , (1995) "Noise estimation Techniques for robust speech recognition. Proc. IEEE Internat. Conf. on Acoust. Speech Signal Proc.pp 153–156.

[12] P. C. Loizou, "Speech Enhancement: Theory and Practice" 1st ed. Boca Raton, FL. CRC, 2007

[13] Saeed V. Vaseghi "Advanced Digital Signal Processing and noise Reduction" ,WILEY TK 2008

Anuradha R. Fukane received a Diploma in Electronics and Telecommunication Engg. in 1988, completed Diploma in Business Management in 1999, received bachelor's Degree in Electronics and Telecommunication Engg. in 2007 currently doing masters in Engg. (M.E.) in signal processing from Pune University ,will receive masters Degree in 2011. Associate with Cummins College of Engineering for Women Pune, Maharashtra since 1992. Author of "Theory and Solved Problems of Control system" Published by Satya Prakashan, New Delhi with ISBN NO. 81-7684-597-3, two papers are published in National Conferences and one paper Published in International Conference. Associate Member of IETE. Area of interest are Speech Signal processing, VLSI in Signal Processing. Currently working on Different Speech Enhancement Algorithms and their Performance Evaluation for Hearing Aids.

Shashikant L. Sahare received bachelor's Degree in Electronics Engg. in 2001, Master's Degree (M. Tech.) in Electronics Design and Tech. in 2004 from Center for Electronics Design Tech. Aurangabad . Associate with Cummins College of Engineering for Women Pune, Maharashtra, India since 2004. Three papers are published in National Conferences and one paper is published in International Conference. Areas of interest are Signal processing, Electronic Design. Currently working as Assistant Professor at Cummins College of Engineering for Women Pune, Maharashtra, India.