A Family of Spectral Subtraction Algorithms for Tamil Speech Enhancement

Vimala.C, V.Radha

Abstract*— Speech enhancement aims to improve the speech quality by using various techniques and algorithms. Over the past several years there has been considerable attention focused on the enhancement of speech degraded by several types of noise. The degradation of speech due to the presence of noise causes severe difficulties in various communication environments. Noise suppression has numerous applications like Human Computer Interaction, hands-free communications, Voice over IP (VoIP), hearing aids, teleconferencing system etc. For this issue there is always a unique need for the technique which offers expected outcome with limited complexity in implementation. Hence, in this paper a family of spectral subtraction techniques is employed for Tamil speech noise cancellation due to its simplicity. The algorithms adopted for this research work are namely basic spectral subtraction, Non linear Spectral Subtraction, MultiBand Spectral Subtraction, Minimum Mean Square Error (MMSE), and Log Spectral MMSE. All these algorithms are analyzed and implemented for two types of noises namely white and babble noise. The performances of these algorithms are estimated based on SNR and MSE measures. Based on the experimental results, the Non linear spectral subtraction algorithm provides better results than any other adopted algorithms.

Index Terms*— Speech enhancement, Tamil Speech, Spectral Subtraction, Non linear Spectral Subtraction, MMSE, Log Spectral MMSE, SNR and MSE.

I. INTRODUCTION

The occurrence of background noise in speech significantly decreases the intelligibility and the quality of the signal. Degradation of speech severely affects the ability of a person to understand what is being said. Reducing or suppressing such background noise and improving the perceptual quality and intelligibility of a speech without disturbing the speech signal quality is a crucial task. Many applications like mobile communications, teleconferencing, speech and speaker recognition systems need more effective noise reduction algorithms for enhanced performance [1].

Speech enhancement systems can be classified in a number of ways based on the criteria such as number of input channels, time domain or frequency domain, adaptive or non adaptive and some additional constraints. During the last decades, various approaches such as spectral subtraction method, subspace methods, Hidden Markov Modeling, wavelet-based methods etc., have been proposed to solve this problem. Among these, spectral subtraction one of the earliest and widely used enhancement methods for all types of noise, has been chosen for its simplicity of implementation and low computational load. In this paper, the basic spectral subtraction technique and its modified versions are implemented for Tamil speech enhancement.

The paper is organized as follows. Section 2 presents the basic concept of spectral subtraction, section 3 explains the various spectral subtraction algorithms adopted for this research work, section 4 investigates the experimental results and section 5 deals with the performance evaluation. The summary and the conclusion are given in section 6.

II. SPEECH ENHANCEMENT USING SPECTRAL SUBTRACTION TECHNIQUE

The problem of cleaning noisy speech still poses a challenge in the area of signal processing. Eliminating various types of noise is difficult due to the random nature of the noise and the inbuilt complexities of speech. Due to the wide variety of noise characteristics, many speech enhancement routines fail to significantly improve the overall speech quality and can quite often introduce distortions to the voice portions of the signal. Noise which affects the signal may be in the form of white noise, colored noise, babble noise and many other types.

The spectral subtraction algorithm is historically one of the first algorithms proposed for additive background noise and it has gone through many modifications with time [1]. The greatest asset of spectral subtraction algorithm lies in its simplicity. The goal of spectral subtraction is to suppress the noise from the degraded signal. It is represented as (1)

$$y(n) = s(n) + d(n) \tag{1}$$

Where y(n) is the noisy speech which is composed of the clean speech signal s(n) and the additive noise signal d(n).

It operates by making an estimate of the spectral magnitude during periods of no speech and subtracting this spectral estimate of the noise from the subsequent speech spectral magnitude [4].

Revised Manuscript Received on March 2012.

Vimala.C, Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India, (e-mail: vimalac.au@gmail.com).

Dr.V.Radha, Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India, (e-mail: radhasrimail@gmail.com).

These algorithms attempt to be an omnipresent solution for all types of noise environments. However, the serious drawback of this method is that the enhanced speech is accompanied by unpleasant musical noise artifact which is characterized by tones with random frequencies. Although many solutions have been proposed to reduce this musical noise, results performed with these algorithms illustrate that there is a need for further improvement. The block diagram of the proposed work is shown in figure 1.

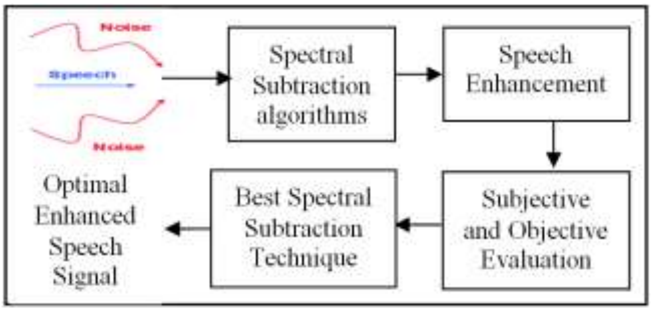

Fig 1. Block diagram of the proposed work

III. BASIC SPECTRAL SUBTRACTION ALGORITHM

The basic assumption of this algorithm is that the noise is additive and its spectrum does not change with time. This means noise is stationary or it is slowly time varying signal whose spectrum does not change significantly between the updating periods [1]. The noise spectrum can be estimated, and updated, during the periods when the signal is absent or when only noise is present [8]. The basic block diagram of spectral subtraction method [9] is shown below in figure 2.

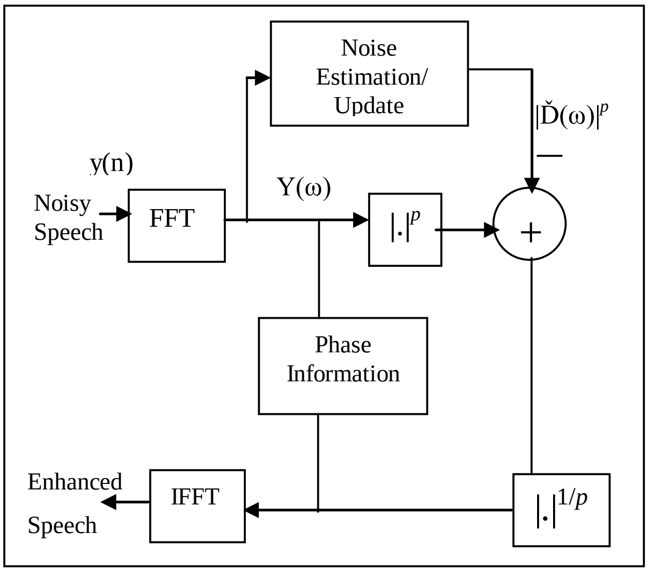

Fig 2. Basic block diagram of spectral subtraction method

The spectral subtraction techniques implemented for this work are explained below**.**

A. Boll's Spectral Subtraction

The first basic spectral subtraction technique was proposed by Boll which is popular due to its simple underlying concept and its effectiveness in enhancing speech degraded by additive noise [10]. The fundamental rule of this technique is to subtract the magnitude spectrum of noise from that of the noisy speech [2]. This method suppresses stationary noise from speech and an estimate of the noise signal is measured during silence or non-speech activity in the signal. It also concentrates on magnitude averaging and residual noise reduction [2].

B. Berouti's Spectral Subtraction

Berouti proposed a spectral subtraction method in 1979, for enhancing speech corrupted by broadband noise [3]. Original method entails subtracting an estimate of the noise power spectrum from the speech power spectrum, setting negative differences to zero, recombining the new power spectrum with the original phase, and then reconstructing the time waveform [1]. While this method reduces the broadband noise, it also usually introduces an annoying "musical noise". This method is based on the power spectral subtraction with adjusting subtraction factor. The adjustment is according to local a posteriori SNR [10].

This method differs from other conventional methods in two ways [1][10]:

- The method involves subtraction of a factor α times the noise spectrum, where α is a number greater than unity and varies from frame to frame.
- Then taking precautions to prevent the spectral components of the processed signal from going below a certain lower limit, designated as spectral floor.

Then express the spectral floor as a fraction β, of the original noise power spectrum $P_n(w)$. Berouti and Boll's subtraction methods form the basis of spectral subtraction to introduce new methods with further modifications [6].

C. Non–linear Spectral Subtraction (NSS)

In the case of NSS, proposed by Lockwood and Boudy the assumption is that the noise signal does not affect the speech signal uniformly over the whole spectrum. Certain types of noise may affect the low frequency region of the spectrum more than high frequency region [6]. Subtracting a constant ratio of the noise spectrum over the whole frequency spectrum may also remove parts of the speech signal. In order to prevent destructive subtraction of the speech while removing most of the residual noise, it is necessary to propose a nonlinear approach to improve the subtraction procedure [1].

Recent studies have focused on a non-linear approach to the subtraction procedure because it strongly suggests utilizing a frequency dependent subtraction factor for different types of noise [7]. Due to frequency dependent subtraction factor, subtraction process becomes nonlinear. To produce improved results, larger values are subtracted at frequencies with low SNR levels and smaller values are subtracted at frequencies with high SNR levels. It is estimated as follows [1].

$$|X_e(\omega)| = |Y(\omega)| - \alpha(\omega)N(\omega) \text{ if}$$
$$|Y(\omega)| > \alpha(\omega)N(\omega) + \beta|D_e(\omega)| \text{ else}$$
$$\beta = |Y(\omega)| \qquad (2)$$

Where β is the spectral floor, α (ω) is a frequency dependent subtraction factor and N(ω)is a non-linear function of the noise spectrum where $N(\omega) = Max(|D_e(\omega)|)$ (3)

$N(\omega)$ is the maximum of the noise magnitude spectra.

$$\alpha(\omega) = 1/r + P(\omega) \qquad (4)$$

Where r is a scaling factor and $P(\omega)$ is the square root of the posteriori SNR estimate given as

$$P(\omega) = |Y(\omega)| / |D_e(\omega)|. \qquad (5)$$

D. MultiBand Spectral Subtraction (MBSS)

Kamath & Loizou used the Berouti 's spectral subtraction method in several frequency bands for reduction of musical noise. In MBSS approach, the speech spectrum is divided into N overlapping bands and spectral subtraction is performed independently in each band [1][11]. The splitting of the speech signal into different bands can be performed either in the time domain by using band pass filters or in the frequency domain by using appropriate windows. Here in this algorithm the subtraction process is done with adjusting the subtraction factor and the adjustment is made according to local a posteriori SNR and the frequency band [11]. The estimate of the clean speech spectrum in the i^{th} band is obtained by[1]

$$|X_{ei}(\omega_k)|^2 = |Y_i(\omega_k)|^2 - \alpha_i \delta_i |D_i(\omega_k)|^2 \qquad (6)$$
$$b_i < \omega_k < e_i$$

Where $\omega_k = 2pi\ k/N,\ K=0,1\ldots..N-1$ are the discrete frequencies $|D_i(\omega_k)|^2$ is the estimated noise power spectrum obtained during speech absent segment, *αi* is the over subtraction factor of the *i*th band and *δi* is an additional band.

E. MMSE Spectral Subtraction Algorithm

Minimum Mean Square Error (MMSE) spectral subtraction algorithm was proposed by Ephraim and Malah in 1983. This is an effective method for optimally selecting the subtractive parameters in the mean error sense and also minimizes the Mean Square Error (MSE) [1][10]. Its structure is the same as that of spectral subtraction but it optimizes the estimate of the real rather than complex spectral amplitudes. Here, an estimate of the speech spectrum is available with known Power Spectral Density (PSD). In this approach, the spectral components are statistically independent and each follows Gaussian distribution [10]. The DFT phase follows uniform distribution and is independent of the amplitude. This method calculates a gain function based on the a priori and a posteriori SNRs[5].

F. Log Spectral MMSE STSA estimator

This algorithm is a modification to the MMSE STSA with the fact that a distortion measure based on the mean-square error of the log-spectra which is more suitable for speech processing [6]. It minimizes the mean square error between the logarithmic spectrum of enhanced speech and clean speech. This algorithm is very efficient for noise reduction with the knowledge of required parameters such as PSD and the absence of speech. These parameters are used in this estimation since the gain function should be modified by considering the uncertainty of speech presence in real environment [6]. It uses a soft-decision gain modification to indicate the narrowband speech presence probability for each frequency bin. It can be estimated as follows:

$$\hat{A}_k = \exp(E[\ln A_k \mid Y_k])$$
$$= \frac{\xi_k}{1+\xi_k} \exp\left(\frac{1}{2}\int_{v_k}^{\infty} \frac{e^{-t}}{t} dt\right) R_k \qquad (7)$$

Where

$v_k = \frac{\xi_k}{1+\xi_k}\gamma$, ξ_k and γ_k are a priori SNR and a posteriori SNR respectively. The gain curves of Log Spectral MMSE STSA are always lower than that of MMSE STSA, resulting in lower residual noise.

IV. EXPERIMENTAL RESULTS

For the experiments, the separate noise corpus from NOIZEUS were collected and added to the continuous Tamil Speech signal. Analysis is done on noisy speech signal corrupted by white noise and babble noise at 5dB, 10dB,-5dB and -10dB SNR levels for 10 datasets totally.

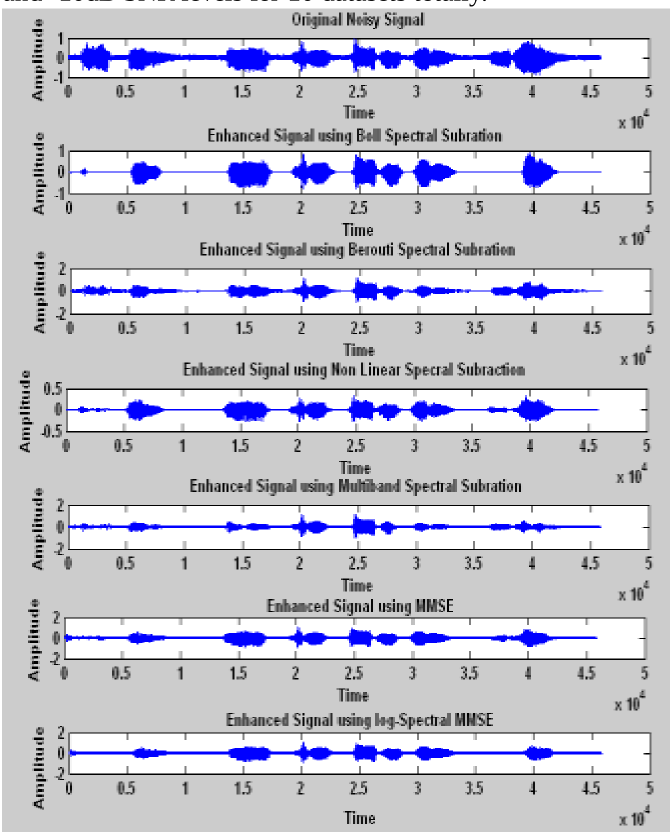

Fig 3. Results of noise cancellation for white noise at 10dB SNR using the above algorithms

The above figure 3 shows the speech enhancement for white noise at 10dB using the adopted spectral subtraction algorithms. Output SNR and MSE are estimated and compared to find the best technique for Tamil speech enhancement system and they are explained below.

V. PERFORMANCE EVALUATION

The above implemented algorithms are evaluated using both objective measures such as SNR and MSE and then subjective listening tests.

A. Mean Squared Error (MSE)

The Mean Squared Error (MSE) of an estimator is used to quantify the difference between values implied and the true values being estimated. It is calculated by the formula (8)

$$MSE = \frac{\sum\left(y_i - \hat{y}_i\right)}{n-p} \qquad (8)$$

B. Signal-to-Noise Ratio (SNR)

SNR is defined as the ratio of power between the signal and the unwanted noise. One of the most important goals of any speech enhancement technique is to achieve the highest possible SNR. SNR is calculated using the formula (9).

$$\frac{S}{N} = \frac{n_{signal}}{n_{noise}} \qquad (9)$$

Where n_{signal} is the original signal and n_{noise} is the noisy signal.

At 0 dB the two signals are of equal strength and negative values are associated with loss of intelligibility due to masking whereas positive values are usually associated with better intelligibility. It is observed from the experiments that these algorithms offer better speech quality but less speech intelligibility since it produces negative SNR values.

The following figures 4, 5, 6 and 7 show the performance evaluation of various Spectral Subtraction algorithms for white and babble noise based on MSE and SNR respectively.

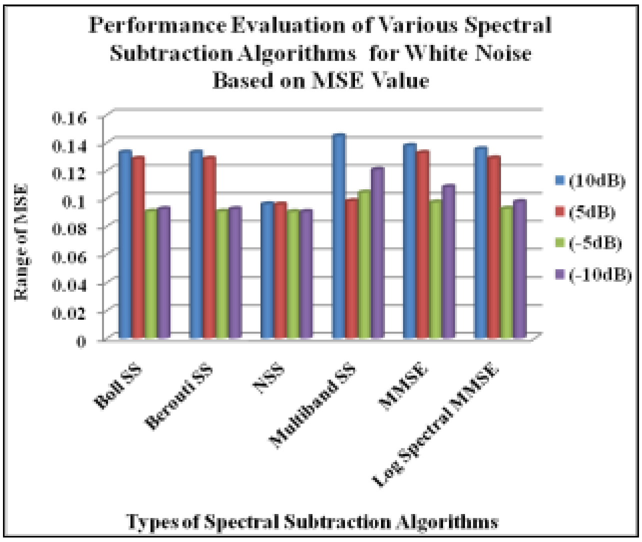

Fig 4. Performance Evaluation of Various Spectral Subtraction Algorithms for White noise based on MSE

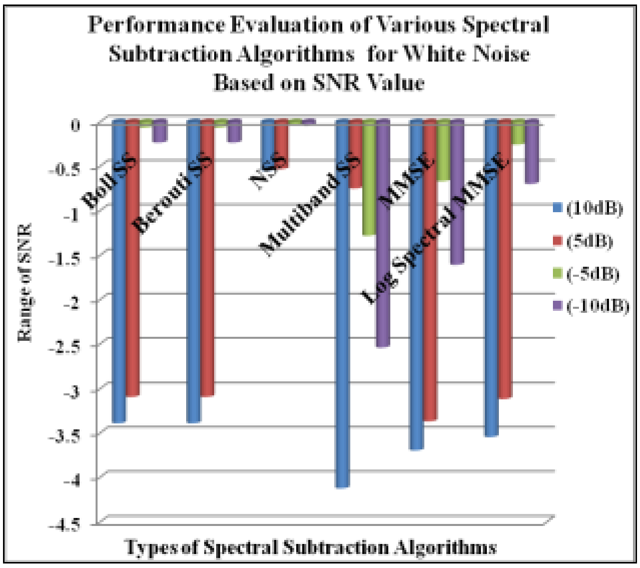

Fig 5. Performance Evaluation of Various Spectral Subtraction Algorithms for white noise based on SNR

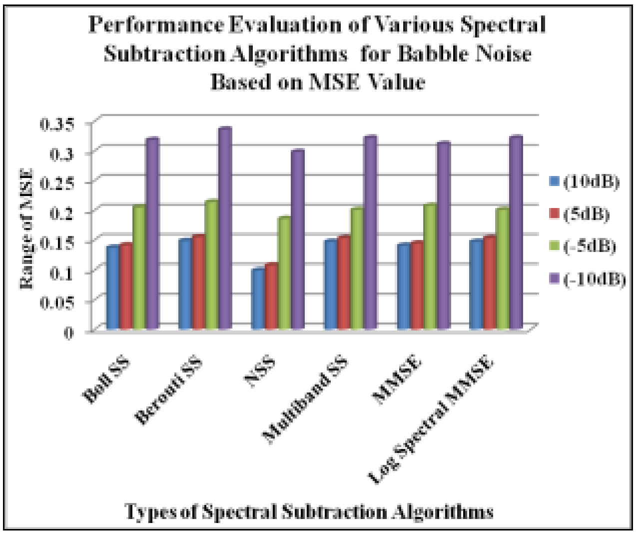

Fig 6. Performance Evaluation of Various Spectral Subtraction Algorithms for Babble noise based on MSE

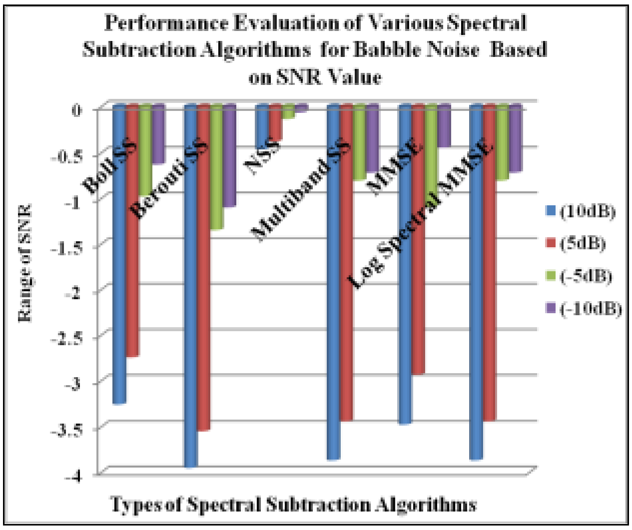

Fig 7. Performance Evaluation of Various Spectral Subtraction Algorithms for Babble noise based on SNR

The results indicated that the NSS algorithm performed the best constantly across the two types of noise conditions at various SNR values in terms of overall speech quality and minimum MSE value. The performance of this algorithm was found to be comparatively good with subjective listening tests also.

VI. CONCLUSION

The improved performance of a speech recognition system in noisy environment is an important challenge. To achieve the optimum clean signal which is affected by different types of noise, various spectral subtraction algorithms are analyzed and experienced for Tamil speech enhancement. These algorithms are computationally simple to implement as they involve a forward and an inverse Fourier transform. In this paper, it is brought out clearly that the Non Linear Spectral Subtraction algorithm provides better speech noise reduction and signal quality than the other spectral subtraction algorithms. It can be observed from the experiments as well as the overall studies from many people, that the spectral subtraction algorithm improves speech quality but not speech intelligibility.

REFERENCES

[1] Anuradha R. Fukane, Shashikant L. Sahare, "Different Approaches of Spectral Subtraction method for Enhancing the Speech Signal in Noisy Environments", International Journal of Scientific & Engineering Research, Volume 2, Issue 5, May-2011,ISSN 2229-5518.

[2] M. Berouti, R. Schwartz, and J. Makhoul, Enhancement of speech corrupted by acoustic noise,Proc. IEEE ICASSP , Washington DC, April 1979, 208-211.

[3] S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process., vol.27, pp. 113-120, Apr. 1979

[4] Ekaterina Verteletskaya, Boris Simak, "Noise Reduction Based on Modified Spectral Subtraction Method", IAENG International Journal of Computer Science, 38:1, IJCS_38_1_10.

[5] Gupta, V.K, Bhowmick, A, Chandra, M. and Sharan, S.N, "Speech Enhancement Using MMSE Estimation and Spectral Subtraction Methods",978-1-4244-9190-2/11/$26.00© 2011 IEEE.

[6] P.Krishnamoorthy and S.R. Mahadeva Prasanna, "Temporal and spectral processing Methods for Processing of Degraded Speech: A Review", IETE Technical Review,Vol 26,Issue 2,Mar-Apr 2009.

[7] P. Lockwood and J. Boudy, "Experiments with a nonlinear spectral subtractor (NSS), hidden markov models and the projection, for robust speech recognition in cars," Speech Communication, Vol. 11, Nos. 2-3, pp. 215-228, 1992.

[8]. Nicholas W. D. Evans, John S. Mason, Wei M. Liu and Benot Fauve, "On the fundamental limitations of spectral subtraction: An assessment by automatic speech recognition", School of Engineering, University of Wales Swansea,Singleton Park, Swansea, SA2 8PP, UK.

[9] Paurav Goel1, Anil Garg, "Developments in Spectral Subtraction for Speech Enhancement", International Journal of Engineering Research and Applications (IJERA), ISSN: 2248-9622, Vol. 2, Issue 1, Jan-Feb 2012, pp.055-063.

[10] R. Singaram, P. Guru Raghavendran, S. Shivaramakrishnan and R. Srinivasan*, "Real time speech enhancement using Blackfin processor BF533", J. Instrum. Soc. India 37(2) 67-79.

[11] Sunil D. Kamath and Philipos C. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise", Electrical Engineering (2002), Volume: 4, Issue: 2, Publisher: IEEE, Pages: 2-5, ISBN: 0780374029, DOI: 10.1109/ICASSP.2002.1004852.

AUTHORS PROFILE

Ms. Vimala.C, currently doing Ph.D and working as a Project Fellow for the UGC Major Research Project in the Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women. She has more than 2 years of teaching experience and 2 years of research experience. Her area of interest includes Automatic Speech Recognition and Synthesis. She has 10 publications at National and International level conferences and journals. Email id: vimalac.au@gmail.com

Dr. V. Radha, is the Associate Professor of Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, TamilNadu, India. She has more than 21 years of teaching experience and 7 years of Research Experience. Her Area of Specialization includes Image Processing, Optimization Techniques, Voice Recognition and Synthesis, speech signal processing and RDBMS. She has more than 63 Publications at national and International level journals and conferences. She has one Major Research Project funded by UGC. Email id: radhasrimail@gmail.com.