A study on different linear and non-linear filtering techniques of speech and speech recognition

Minajul Haque[1] andKaustubh Bhattacharyya[2]

[1]School of Technology, Assam Don Bosco University
Airport Road, Azara, Guwahati - 781017, Assam. INDIA.
minajul.h.electronics@gmail.com
[2]School of Technology, Assam Don Bosco University
Airport Road, Azara, Guwahati - 781017, Assam. INDIA.
kaustubh.bhattacharyya@dbuniversity.ac.in

***Abstract:**In any signal noise is an undesired quantity, however most of thetime every signal get mixed with noise at different levels of theirprocessing and application, due to which the information containedby the signal gets distorted and makes the whole signal redundant.A speech signal is very prominent with acoustical noises like bubblenoise, car noise, street noise etc. So for removing the noises researchershave developed various techniques which are called filtering. Basicallyall the filtering techniques are not suitable for every application,hence based on the type of application some techniques are betterthan the others. Broadly, the filtering techniques can be classifiedinto two categories i.e. linear filtering and non-linear filtering.In this paper a study is presented on some of the filtering techniqueswhich are based on linear and nonlinear approaches. These techniquesincludes different adaptive filtering based on algorithm like LMS,NLMS and RLS etc., Kalman filter, ARMA and NARMA time series applicationfor filtering, neural networks combine with fuzzy i.e. ANFIS. Thispaper also includes the application of various features i.e. MFCC,LPC, PLP and gamma for filtering and recognition.*

Keywords:Adaptive filtering, Wiener filter, least mean square (LMS), normalizedleast mean square (NLMS), recursive least square (RLS), Kalman filter,ARMA, NARMA, Feature extraction (FE), MFCC, PLP, LPC, Gamma filter,Neural network and ANFIS.

I. INTRODUCTION

Speech is simple and reliable way of communication among the human beings. A speech signal has a frequency ranges from 300Hz to 3400 Hz, however the human beings have an audible frequency range of 20 Hz to 20 KHz [1, 2]. Different forms of information contained by the speech signals are basic information about the language being spoken and the emotion, gender and the identity of the speaker also could be the part of information. The individual characteristic like pitch, fundamental frequency, formant frequency can be distinguishing components of human speech. Sound generates by the three main processes, they are twisting of nerves, wire beating of membranes or blowing of air through holes but human voice mechanism is different as it comes out in different languages and feelings by control of brain [3, 4, 5]. Audio signal processing often suffering from noise trouble [6].

Let us consider the following 3 examples

- A person is moving in a vehicle at a high speed in the meantime he is talking on his mobile
- A political leader is delivering a speech in some place within a huge crowd and its being recorded
- A speech recognition system is running in an open area or inside a building

Now in these three different examples the speech/voice signal is the primary objective. How-ever, in all the three cases it is not possible to get only the speech signal of the speaker, because as the clean signal will get produce it will be corrupt by the signals which are generating from the surrounding environment of the speaker. These surrounding environmental signals are considered as noise because these are not desired signals [7].

Therefore it has cleared that some kind of technique is required which will extract the clean speech signal from the noise contaminated speech signal, called as the filtering of the speech signal. This filtering process also called the enhancement of the speech signal. Enhancement in terms of noise increase the signal to noise ratio (SNR) of the speech signal, which improves the quality and intelligibility of the speech signal. Different enhancement techniques are available like sub-band enhancement [8], blind signal separation [9], Enhancement in instantaneous amplitude and phase in noisy condition [10]. Jacob Benesty et al. have given a brief review of speech enhancement using linear filtering techniques [11]. In speech filtering the role of adaptive filters are very important. Adaptive filters are the filter which self-updates its weights based on the input data. Different adaptive filters are avail-able like wiener filter, LMS, NLMS, RLS filter [12, 13, 15, 16, 17, 18, 19]. Although Over the years engineers have developed a variety of theoretical and relatively effective techniques to comb at this filtering issue of speech. However, the problem of extracting a clean signal from noisy speech still poses challenges to the area of signal processing [20]. Therefore, the present research is going on nonlinear filtering different methods like Kalman filter, gamma filter NARMA process [21, 22, 23], Neural networks, adaptive neuro fuzzy system (ANFIS).

Feature extraction is also an integral part of any speech processing system. Feature could be used different application such as filtering, speech recognition, gender identification etc. A number of features could be extract from a speech signal, short time zeros crossing rate (ZCR), short time magnitude, MFCC, LPC, PLP and gamma coefficient are some of the popular feature. MFCC, PLP and LPC are the features which used very frequently in speech processing based applications.

This paper is organized as, section 2.1 describes the basic background of the natural human speech production system and digital signal processing (DSP) model of the human speech production system. Section 3 gives an idea regarding the types of noise that affect the speech signal. Section 3.1 explains enhancement and the different enhancement techniques which are used for speech filtering. Section 4 describes the different adaptive filters that are used for speech filtering, under which some of the popular filters are discussed in this section. Section 5 explains the how spectral subtraction can be used to enhance a noisy speech signal. Section 6 describes basics of Kalman filter with various application of Kalman filter to estimate clean speech from a noisy speech signal. Section 7 explains how features are used for different application of speech signal and this section also summarize some of the popular feature of the speech signal. Section 9 and 10 presents the non-linear filtering of speech signal with neural network and fuzzy logic. Section 11 gives a short list of application where the enhanced speech could be used for further processing. Lastly, section 12 give a conclusion of this paper.

II. BACKGROUND

Quantitative models of human speech production and perception provide important in-sights into our speech production and perception mechanisms and lead to high-quality computer synthesis of speech, robust automatic speech recognition (ASR), and efficient speech and audio coders. These issues are of importance in the development of effective human-computer communications through the medium of human language. In section 2.1 and 2.2 the natural production of human speech and the DSP representation for the same is described.

A. *Human speech production system*

Speech sounds are composed of sequence of sounds called phonemes produced as a result of acoustical excitation of vocal tract the vocal tract shape is determined from the position of the vocal organs, and speech is produced by controlling the speech production model using the vocal tract area [3]. Figure 1 show that the vocal tract is an acoustical tube which begins at the opening between the vocal cords and ends at the lips. Depending on the whether the vocal cord vibrates sound produced can be broadly classified as voiced and unvoiced sound. The classification of the speech signal into voiced, unvoiced, and silence provides a preliminary acoustic segmentation of speech, which is important for speech analysis. Many synthesis models assume constant voicing vocal-cord vibration, but it is quite clear that the binary distinction of vocal-cord vibration vs. no vocal-cord vibration is not accurate [4].

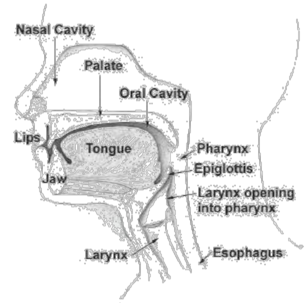

Fig. 1: Human speech production system [3]

B. *DSP model of human speech production system*

The basic assumption of speech processing system is that the source of excitation and the vocal tract system are independent [3]. To develop a model for speech production it would be necessary to accurately represent Excitation mechanism for both voiced and unvoiced sounds, vocal tract model, lip or nasal radiation. The block schematic in Figure 2 represents the model for the same [3, 4].

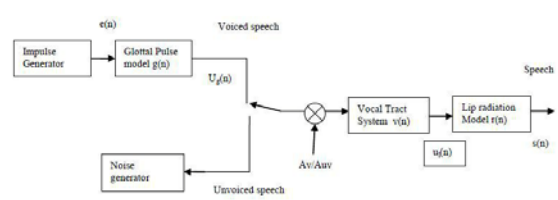

Fig.2: DSP model of human speech production [4]

III. DIFFERENT TYPES OF NOISES

The speech signal gets corrupt with noise very frequently, which degrades the quality and intelligibility of signal. Generally noise is a general term which is used for unwanted elements present in the desired signal. Noise is nothing but the changes that may occur in signals while storing, transmitting, processing and converting from one form to another. Noise in audio signals can be background noise due to spurious sounds during speech recording or 90 can be comfort noise added to fill silent gaps in the speech. The noise which mixed with the speech signal are broadly classifieds into two types [7]. They are:

- Additive noise: where noise gets added to the intended signal
- Multiplicative noise: where noise gets multiplied to the intended signal

A. *Different enhancement techniques*

To improve the quality of a noise corrupted speech signal, it is required to enhance the signal in terms of noise reduction. Enhancing of speech degraded by noise, or noise reduction is the most important field of speech enhancement, and used for many applications such as mobile phones, VoIP, teleconferencing systems , speech recognition, and hearing aids [20]. Therefore researchers have developed many methods to remove the noise and extract a clearspeech signal which contain the original information [15, 17, 24]. Among these spectral subtraction is one of the best possible way to remove the noise without much affecting the quality of the speech. So for getting a noise immune speech signal it has to go through some pre-processing technique. These pre-processing techniques are nothing but the filtering. Speech filtering process can be divided into two broad classes. They are linear filtering and nonlinear filtering. In linear filtering different linear adaptive filtering algorithms are applied to the noisy speech signal which attempt to enhance the signal by estimating the noise present in the signal. In linear filtering the noise which affect the speech signal considered to have a flat spectrum that is, affect the whole signal spectrum uniformly. However in real world the noise is not flat, thus it does not affect the signal uniformly i.e. some frequencies. 110 are affected more adversely than other's [20]. Therefore the nonlinear enhancement is more important. For nonlinear filtering various modified version of Kalman filters are present which gives good estimation of the noise [25, 26]. Normally the noise variation is from the silent regions of the signal. In speech signal this could be the first one or two seconds. Based on this silent periods the kalman filter estimate the noise overall the signal and subtract it from the noisy input to enhance the quality of the speech [25]. Recently some new speech enhancement technique has been develop which based on the IIR filter [27]. HwaSoo Kim et al. in [28] presented an IIR based gamma filter is presented for nonlinear filtering. Although this gamma filter has the same computational complexity as the conventional LMS algorithm, yet the gamma filter is more efficient. In speech enhancement applications, the input signal $x(n)$ with noise $n(n)$ can be expressed as equation 1:

$$y(n) = x(n) + n(n) \qquad (1)$$

The $x(n)$ is the pure speech signal, $n(n)$ is the noise signal, and $y(n)$ is the output noisy signal.

A simple block diagram for speech enhancement is shown in Fig. 3.

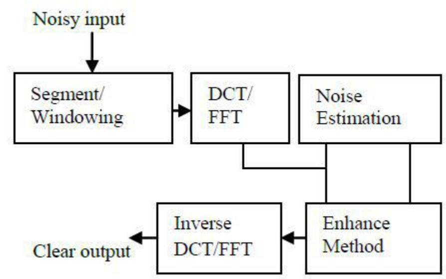

Fig.3: Block diagram of speech enhancement [15]

IV. DIFFERENT ADAPTIVE FILTERS

Digital signal processing systems are attractive due to their low cost, reliability, accuracy, small physical sizes, and flexibility. One such digital signal processing system is called filter. Filtering is a signal processing operation whose objective is to process a signal in order to manipulate the information contained in the signal. If the models are available for the signal and noise the linear filter can be used to enhance the quality of the signal. An adaptive filter is required when the models are unknown [12]. An adaptive filter is a filter that self-adjusts its transfer function according to an optimizing algorithm [13] [29]. Adaptive filtering can be considered as a process in which the parameters used for the processing of signals changes according to some criterion [30]. Usually the criterion is the estimated mean squared error or the correlation [15]. A general block diagram of an adaptive filter is shown in Fig. 4. There are different approaches used in adaptive filtering, which are shown in Fig. 5 an adaptive filter is a self-regulating system that take help of recursive algorithm for processing.

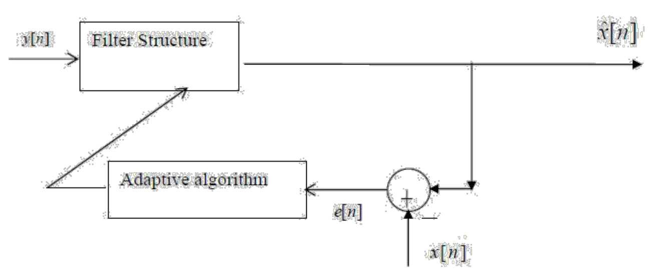

Fig.4: General block diagram of an adaptive filter [29]

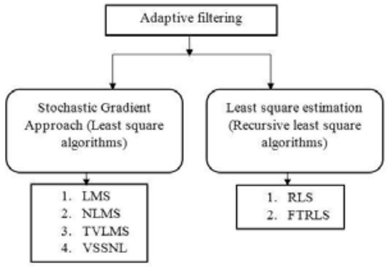

Fig.5: Hierarchy of adaptive filter [13]

First input and training is compared, accordingly error signal is generated and that is used to adjust some previously assumed filter parameters underffect of incoming signal. Filter parameter adjustment continue until steady state condition. As far as application of noise reduction from speech is concern, adaptive filters can give best performance. Reason for that noise is somewhat similar with the randomly generates signal and every time it's very fficult to measure its statistic. Design of fixed filter is completely failed phenomena for continuously changing noisy signal with the speech. Some of the signal changes with very fast rate in the context of information in the process of noise cancellation which requires the help of self-regularized algorithms with the characteristics to converge rapidly. Some application of adaptive filtering are system identification, channel equalization and signal enhancement by noise reduction [31].

A. Wiener filter

Among the numerous techniques that were developed, the optimal Wiener filter can be considered as one of the most fundamental noise reduction approaches. Wiener filter is the base of all adaptive filter. It is an optimum filter in the sense of producing the best estimate of the signal [13] [14]. Hwa Soo Kim et al. have proposed an improved wiener filter method for removing noise to provide a comfortable communication inside an automobile [28]. In this technique first a voice activity detector (VAD) is used to identify the input data as speech or noise at each frequency. The VAD used here is based on frequency instead of time domain, therefore, it becomes possible to detect within a frame also, which is not possible in timedomain VAD. In this way small musical noise and speech distortion problems can also be eliminate from the speech signal [28]. Another wiener filter based enhancement of the speech signal is given by BinWen Fan et al. which also utilizes the Mel-frequency domain processingwith wiener filter to improve the performance of noise reduction [32]. As in time domain the wiener filter estimates the noise based only on the silent period, so some noise might left in the speech period, causing the low SNR at the output. It is found that after joined the Mel treatment, the traditional Wiener filtering algorithm has a betterffect with speech signal estimating [32, 33]. The schematic of the proposed wiener filter is shown in Figure6.

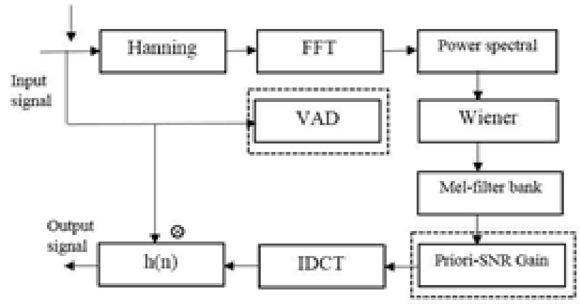

Fig. 6: Schematic diagram of the Mel-frequency based wiener filter for speech enhancement [28]

Jingdong Chen et al. presented a study on the quantitative performance behavior of the Wiener filter in context of noise reduction [14]. According to their method the quality of the filtered signal can be maintained by three ways i.e.

- If a priori knowledge of the clean speech signal
- If no priori knowledge is available then also a better result can be achieve by properly manipulating the Wiener filter parameter
- If multiple microphone sensor and multiple observation of the signal available

B. LMS and NLMS filter

LMS and NLMS both are adaptive filtering algorithm and stand for least mean squared algorithm and normalized least mean squared algorithm respectively. There are many digital signal processing application in which second order statistics cannot be specified. Such application includes channel equalization, echo cancellation, and noise cancellation. These application can be done with adaptive filter. In this paper the authors have given a com-parison between the adaptive filter algorithm like LMS, NLMS and Recursive least squares (RLS). Least mean squares (LMS) algorithms are class of adaptive filter used to mimic a desired filter by finding the filter coefficients that relate to producing the least mean squares of the error signal (dfference between the desired and the actual signal. The basic weight update equation is given by equation 2: [12, 16, 17].

$$\omega_{n+1} = \omega_n + \mu\Delta\varepsilon(n) \qquad (2)$$

Where, ω_n is the current weight, ω_{n+1} is the estimated weight, μ is the step size and $\varepsilon(n)$ is the current error.

The main drawback of the "pure" LMS algorithm is that it is sensitive to the scaling of its input, the NLMS algorithm solved this problem by normalizing with the power of the input. The Recursive least squares (RLS) adaptive filter is an algorithm which recursively finds the filter coefficients that minimize a weighted linear least squares cost function relating to the input signals. Here the comparison between different algorithms is described by SNR improvement table. It concludes that the best adaptive algorithm is Recursive Least Square according to the SNR improvement Table1 [17].

Table 1: SNR improvement of LMS, NLMS and RLS [17]

Noise variance	Sampling rate (kHz)	SNR Improvement (dB) LMS	SNR Improvement (dB) NLMS	SNR Improvement (dB) RLS
0.02	1.5	8.85	9.85	9.91
0.05	1.5	7.55	8.62	8.89
0.1	1.5	5.12	6.38	7.02

An optimal estimation of adaptive filtering using Unbiased and normalized adaptation noise reduction (UNANR) algorithm has been implemented for the noisy speech [15].

The Aim of this paper is to implement various adaptive noise canceler for speech enhancement based on gradient steepest descent approach. First the general LMS algorithm have used to filter the noisy input speech signal then the UNANR technique was used to filter the same signal and the results were compared. The adaptation process of the UNANR model is designed to modify the co-efficient that get convolved with the reference input in order to estimate the noise present in the given speech signal. A comparison based on the performance of the both technique is given in Table 2 and it can be conclude that the signal to noise improvement in the input signal after UNANR filtering is much higher up-to 10 dB (50%) than that LMS filter algorithm and 20dB than that of original signal [15, 16].

Table 2: The performance of LMS filter and UNANR algorithms for different speech signals [16]

Speech	Before filtering (dB)	After filtering using LMS	After filtering using UNR
S-I	SNR=28.4973	PSNR=42.9626 dB	PSNR=49.7023 dB
		RMSE=0.00011	RMSE=5.0529e-
		Time=0.2249 sec	Time=0.6411 sec
S-II	SNR=17.5265	PSNR=30.5199 dB	PSNR=37.1324 dB
		RMSE=0.000433	RMSE=0.000202
		Time=0.21238 sec	Time=0.62172 sec
S-III	SNR=23.2201	PSNR=33.8712 dB	PSNR=41.9642 dB
		RMSE=0.000257	RMSE=0.000101
		Time=0.22212 sec	Time=0.62613 sec

V.P.Patil gives the similar work in which LMS and UNANR filtering is used for recognition of human voice [34]. It was found that the signal to noise ratio of filtered speech is much higher than the noisy speech signal and the convergence rate of LMS algorithm compared to UNR algorithm is also high.

V. SPECTRAL SUBTRACTION

Spectral subtraction is historically one of the oldest and simple algorithm to implement easily and a minimal complexity of a speech enhancement. The spectral subtraction is based on the theory that the enhanced speech can be acquire by subtracting the estimated spectral factors from the continuous input noisy signal [1, 2]. In spectral subtraction basically the average magnitude of the noise spectrum is subtracted from the noisy speech spectrum. Clisnton Cloe et al. proposed an algorithm for speech noise removal based on the spectral subtraction, where they have varied the window length and data buffer overlapping factor to increase the SNR level of the enhanced speech [24]. A general block diagram of spectralsubtraction is shown in Figure 7

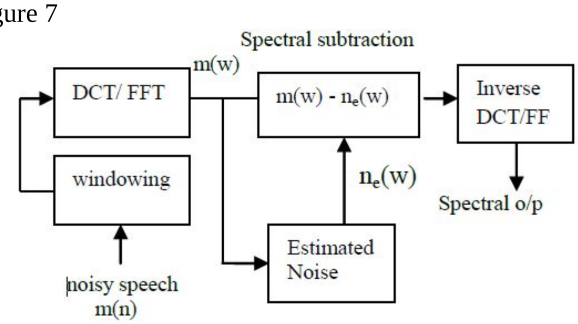

Fig. 7: Block diagram of spectral subtraction [1]

In the Figure 7 a complete step by step process of how to enhance a noisy signal by spectral subtraction method is presented. First the whole signal will divide into small segments then from these segments the noise spectrum will be calculated. Now these noise spectrum will be subtracted from each segments and finally inverse Fourier transform will be perform to get the enhanced signal.

Steven F.Boll implemented a pre-processing noise suppression algorithm using spectral subtraction [35]. It is assumed that the speech enhancement is possible by removing the noise from magnitude spectrum only. The algorithm estimated the noise spectrum during non-speech activity when only noise is there and during the speech activity it subtract noise spectrum from the noisy speech input.

VI. KALMAN FILTER

The Kalman filter is a mathematical procedure which operates through a prediction and correction mechanism. Kalman filter combines all the available data measured, plus the knowledge of the system and the measurement devices, to produce an estimation of the desired variables in such a manner that the error is statistically minimized [20]. Kalman filter is best suitable for reduction of white noise to comply Kalman assumption. However different methods developed to fit the Kalman approach to colored noises. M.Mathe et.al, in their paper Kalman filter has applied to three different noise namely white noise, random noise and color noise and the result of Kalman filter has compared with the spectral subtraction and wiener filter [20]. It is found that the Kalman filter gives better result than other two in terms of SNR as spectral subtraction only suitable for stationary signals and wiener filter suitable for stationary signal but it has the problem like musical noise. The author Rayan andsreenivasa described the application of Kalman filter in noise estimation for audio restoration in old analog media storage [25]. As application of Kalman filter requires a correct estimation of the measurement of noise variance, so here it is assumed that the first one or two second silent period of the recordings. An AR model (i.e. AR co-efficient) and driving noise variance and the measurement noise variance are assumed to be known priori. The standard Kalman filter with two different types of measurement noise variance estimations for broadband noise reduction of audio is studied in this paper. Author M.Gabrea and D.O. presented another application of kamlam filter where enhancement of speech signal in a cocktail party using a code book based approach is done [26].In some application, Kalman filter is based on the noise variance estimation from the silent region of the signal. It takes the same consideration like spectral subtraction where initial first one or two seconds of recording considered to be 250 noise [36]. Orchisama Das et.al. Proposed a Kalman filter technique to filter the noise by tuning its measurement noise covariance factor and distinct Kalman gain for silent and voiced frames [37].

VII. FEATURE EXTRACTION

The performance of a speech classifier which separate the voiced and non-voiced part of a speech signal in a noisy environment depend on the choice of the suitable features. So it is a crucial decision to choose the features of the speech in a noisy environment which will separate the voiced and non-voiced speech [33]. The vital part of the characterization of human speech is feature extraction. The human speech is mixed with the noises and it becomes difficult to identify whether the signal containing any speech or not. Therefore, an evaluation have performed on several time domain features for voiced/non-voiced classification of speech signal [33]. Multiple features can be done by using statistical models such as neural networks, Gaussian mixture model or hidden Markov model (HMM). The combination can significantly offer strong classifiers that depend on the number of features.

The combination of several features will improve the accuracy of the classification; however, the computation complexity is going to increase [33].

Therefore feature extraction is important and there are many techniques available for feature extraction methods are available. Some of them are listed below:

- MFCC (Mel frequency cepstral co-efficient)
- LPC (Linear predictive coding)
- PLP (Perceptual linear prediction)
- Gamma co-efficient

A. *Mel- frequency cepstral co-efficient (MFCC)*

The most significant method used to extract spectral features is calculating the Mel-Frequency Cepstral Coefficient (MFCC). MFCC are one of the most popular features extraction technique used in speech recognition based on frequency domain using the Mel scale which is based on the human ear scale. MFCC is an audio feature extraction technique which extract the parameter from the speech similar to the one which used by the human for hearing [38]. Wie han et al. proposed a novel and efficient way to calculate MFCC [39]. In conventional MFCC procedure first the speech signal is pre-emphasized with a pre-emphasis filter and then the pre-emphasis speech is separated into short segment called frame. Each frame of the conventional algorithm goes through a lot multiplication which required a huge amount of computational power but in the proposed algorithm only half of the computation required. There is no change in the algorithm just a little modification is done to eliminate the multiplication steps. Finally the new extraction algorithm reduces the number of multiplication from 1708 to 804 with only 1.5% drop in recognition accuracy. It is found that the new algorithm is more efficient for hardware implementation than the original algorithm [39]. Vibha Tiwari presented a work in which MFCC is used for designing text dependent speaker identification system. The paper gives the comparison between various features extraction technique used for speaker recognition like linear predictive coding (LPC), Local discriminant bases (LDB) etc [40]. Finally with the extracted features it design the speaker recognition system. Sunil Kumar and M. Laxminarayana shows the effects of re-sampling a speech signal on MFCC and also showed that, it is possible to extract the MFCC from a down sampled speech by constructing an appropriate Mel filter bank [41]. Figure 8 shows the complete process of calculating Mel frequency cepstrum. According to Utpal Bhattacharjee MFCC is more noise robust then other features like LPC i.e. it is less effected by the surrounding noise [42].

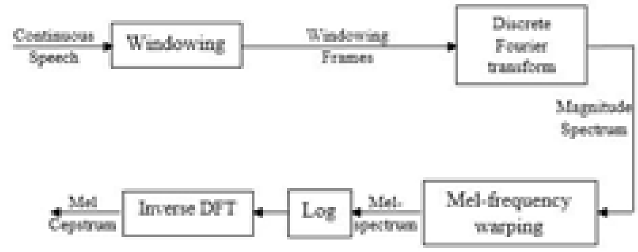

Fig. 8: Complete pipeline for MFCC [40]

The following equation 3 describes the mathematical relationship between the Mel scale and the linear frequency scale.

$$f_{Mel} = 1127.01\log(\frac{f}{700}+1) \quad (3)$$

SafdarTanweer et al. used the MFCC for recognizing the isolated words in artificial neural network. They have prepared a database of 10 classes containing isolated words i.e.0,1,2,3,4,5,6,7,8,9 for recognition in NN. The number of MFCC they took for their work was 13. They have found that the testing result shows accuracy of 59.2 percent since accuracy of classification among ten classes [0-9] above 50 percent is considered as the recognition of one class is as 100 percent the recognition can be enhanced further by using more sample vectors 305 of different groups of speaker [43]. Another similar work is found in [44] but here the authors have taken cubic-log energy at each of the Mel frequencies instead of logarithmic energy in standard MFCC process and they have also chosen the vector quantization (V Q) technique for recognition. A simple process of calculating MFCC in Matlab software is given by Siddhant C.Joshi et al. where the authors have described the complete mathematical process in a very simple way [45].

B. Linear predictive coding (LPC)

The Linear Predictive analysis is based on the assumption that the shape of the vocaltract governs the nature of the sound being produced. To study the property quantitatively,the vocal tract is modeled by a digital all-pole filter [46]. The transfer function in z-domain is given by equation 4

$$V(z) = \frac{G}{1-\sum_{k=1}^{p} a_k z^{-k}} \quad (4)$$

Where $V(Z)$ is the vocal tract transfer function, G is the gain of the filter and $\{a_k\}$ is a set of auto-correlation coefficients called Linear Prediction Coefficients (LPC). The upper limit of summation p, is the order of the all-pole filter. The set of LPC determines the characteristic of the vocal tract transfer function.

LPC analyzes the speech signal by estimating the formants, removing their effects from the speech signal, and estimating the intensity and frequency of the remaining buzz. The process of removing the formants is called inverse filtering, and the remaining signal is called the residue. In the past decades, several studies have been made that investigate how speech can be mathematically represented among these, linear predictive coding (LPC) coefficients is considered to be one of the most promising representation [47]. Bishnu S.Atal et al. describes a new approach to the excitation problem that does not require a priori knowledge of either the voiced-unvoiced decision or the pitch period. It uses the multi-pulse excitation model to synthesize the speech. The excitation for LPC speech synthesis usually consists of two separate signals - a delta-function pulse once every pitch period for voiced speech and white noise for unvoiced speech [48]. LPC is generally used for speech analysis and re-synthesis. In LPC system, each sample of the signal is expressed as a linear combination of the previous samples. This equation 3 is called a linear predictor and hence it is called as linear predictive coding [49]. It is used as a form of voice compression by phone companies, for example in the GSM standard. It is also used for secure wireless, where voice must be digitized, encrypted and sent over a narrow voice channel. In the LPC analysis one tries to predict x_n on the basis of the *p* previous samples [49].

$$x_n^{'} = \sum a_k x_{n-k} \quad (5)$$

Then $\{a1, a_2, a_3 \dots\dots\dots, a_p\}$ can be minimize the prediction powerQ_p, where

$$Q_p = E\left[\left|x - x_n^{'}\right|^2\right] \quad (6)$$

C. Perceptual linear prediction (PLP)

PLP is a model of feature extraction developed by Hermansky. PLP models the human speech based on the concept of psycho-physics of hearing [50].This model helps in removing irrelevant information of the speech and hence improve voice recognition rate. It might accept that PLP is an identical of LPC. PLP system is shown in the Figure 9

Figure9: Block diagram of PLP processing [50]

In PLP the power spectrum of windowed signal is calculated using the equation 7

$$P(w) = \mathrm{Re}(S(w))^2 + \mathrm{Im}(S(w))^2 \quad (7)$$

The PLP speech analysis method is more adapted to human hearing in comparison toLPC. In LPC model all poles transfer function of the vocal tract isassumed with the specifiednumber of resonance within the analysis band and it approximates power distribution equallywell at all frequencies of the analysis band which is in consistence with the human hearingbecause beyond 800Hz the spectral resolution of hearing decreases with frequency and hearingis more sensitive in the middle frequency range of the audible spectrum.

D. Gamma co-efficient

The quality and intelligibility of speech signal degrades under the noisy environment. The clean speech signal is corrupted due to different noises while it is in transmission or recording. Many approaches have been proposed for improving the quality of the speech among them the spectral subtraction and the Wiener filtering algorithm are widely used because of their low computational complexity and impressive performance. Kishor Odugu and B.M.S.Rao have presented a two stage noise reduction approach. First stage consists of Gamma tone filtering method for sub banding and synthesizing. To reduce the residual musical noise that remains after the first stage in the enhanced speech, perceptual weighting filter is developed based on simultaneous and temporal masking effects of human auditory system is used as a second stage in the proposed system. In this approach the noisy signal is applied to gamma tone analysis filters which will sub-band the noisy signal into M number of sub bands based on critical frequencies and then the noise reduction method is applied to each sub bands. Enhanced speech is reconstructed from the enhanced sub-bands. The system contains basically three stages [22]

- Gamma tone analysis filters bank (GA)
- Perceptual sub-banded speech enhancement system (H)
- Gamma tone synthesis filters bank (GS)

The authors have also is tested the proposed system with different types of real world noise like train, car and street noise and during the simulation the system leads to better improvements in high noise level. Lastly they compared enhanced speech with clean speech to know the effectiveness of the system [22]. Yang Shao et al. described that the new feature is based on the gammatone filtering and inspired by ASA (auditory scene analysis) studies which is the ability to segment the speech signal from a noisy background. The auditory feature is called as GFCC (Gammatone frequency cepstral coefficient). To get the GFCC, first the input signal pass through a bank of gammatone filter which are largely correlatedto each other, this bank of filter produce a GF feature vector then discrete cosine transformDCT is applied on the GF feature vector to produce the GFCC feature and those features are used for the robust speech recognition. The GFCC feature performs substantially better than conventional MFCC features. For testing the author have chosen some clean wordsthen those words mixed with speech shaped noise at different SNR level and The GFCC feature outperforms the MFCC and the PLP features across all SNR conditions. Author have also tested the various features under four other non-stationary noisy conditions. Theyare factory- noise, speech babble, destroyer operation room and F-16 cockpit from Noisex92 corpus [51]. Jun Qi et al. have implemented the GFCC feature on a grammar based speech recognition task in Mandarin and the comparison is taken out among MFCC, PLPand GFCC [52]. It is found that the time domain GFCC features improve the robust ASRperformance in noisy condition than frequency domain implementation. When compared to

MFCC and PLP the time domain GFCC provides better recognition performance under all noisy condition. Kanika Garg and Goonjan Jain have reviewed different filtering techniqueswhere they have found that with increasing SNR level the performance of filtering techniqueslike adaptive Wiener filter and spectral subtraction decreases but the performance of gammatone filter becomes better [7].

VIII. TIME SERIES ANALYSIS OF SPEECH

Auto regressive moving average (ARMA) model is a combination of both auto regressive (AR) and moving average (MA) model. An AR model is used to predict the current output based on the previous outputs but MA model use the previous outputs as well as the previous error to predict the current output [53].

AR(p) is an autoregressive model with *p* lags

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-1} + \varepsilon_t \qquad (8)$$

Where μ is a constant and γ_p is the coefficient for the lagged variable in time $t-p$ *MA(q)* is moving average model with q lags:

$$y_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_t \qquad (9)$$

Where θ_q is the coefficient for the lagged error term in time *t-q*

Therefore, *ARMA* models combine both *p* auto-regressive terms and *q* moving average terms, also called *ARMA* (*p, q*) [54, 55].

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \qquad (10)$$

SriramGanapathy in his paper presented a robust speech feature extraction technique for noisy speech and a degraded channel, based on ARMA modeling that emphasizes high energy region of the signal with data driven modulation filter. In this paper for feature extraction, Long segments of the input speech signal are transformed using DCT (discrete cosine transform). The full-band DCT signal is windowed into a set of overlapping sub-bands. The ARMA modeling is applied on the sub-band DCT components to estimate thesub-band envelope and it combines the peak estimation properties of AR approach along with the modulation filtering property of MA modeling. The new features tested for severalspeech recognition and language identification experiments in noisy and degraded channelconditions [54]. Sh. Oveisgharan et al., present an ARMA filtered fractionally difference

Gaussian noise (FdGn) model and a new AR filtered FdGn model is combine and applied onspeech signal to classify the voice, unvoiced, mixed and silence part using zero crossing rate(ZRC) feature. To estimate the parameter of the ARMA filter iterative maximum likelihoodapproach is used [56]. Speech signals have statistically non-stationary properties and cannotbe processed properly by means of classical linear parametric models (AR, MA and ARMA).

Therefore a nonlinear ARMA (NARMA) approach is given with neural networks to process any speech signal [53, 54, 56, 57]. The general form of an IIR nonlinear filter is given by equation:

$$y(k) = \phi\left(\sum^{p} a_i y(k-i) + \sum^{q} b_j e(k-j)\right) + e(k) \qquad (11)$$

Where y (k)is the output, e (k)is the input and - (:) is an unknown differentiable nonlinear function. This model is called as NARMA (p, q) (nonlinear ARMA) model. This NARMA model is implemented using recurrent perceptron neural network, which consist of three layer: input layer, processing layer and output layer. To find the parameter of the NARMA model Akaike information criterion is applied. To test the algorithm the TIMIT database is used which contains speech signals spoken by 630 persons in 8 different dialects of the English language. It is that the model provided by the Akaike criterion had an increasedgeneralization capacity i.e. the signal which were not used for training were well predicted but not for all signals [57].

IX. ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM

As the linear adaptive filtering has been successfully applied to many real world application, research has expanded to the non-linear adaptive filtering. The general approach for filtering any signal is to pass the noisy signal through a sort of filter, which will try to get rid of the noise while leaving unchanged the information signal. The filter might be fixed or adaptive. The design of fixed filters requires previous knowledge of both signal and noise, whereas adaptive filters require little or no knowledge in advance, since they have the ability to dynamically adjust their own parameters according to the characteristics of the noise being processed. To avoid the need of prior knowledge, the method requires a reference input which is supplied by one or more sensors located in the noise field where the information signal is weak or undetectable. One such nonlinear filtering system is adaptive neuro-fuzzy inference system (ANFIS). ANFIS is a neuro-fuzzy system that combines the learning capabilities of neural networks, with the functionality of fuzzy inference systems.A Neuro-fuzzy system is system that uses a learning algorithm derived from or inspired by neural network theories that determine rules created by fuzzy system by analyzing samples.In the ANFIS, ANN uses a back propagation gradient descent method for training fuzzysystem membership function parameters to emulate a given training data set. The ANFIS architecture can identify the near optimal membership functions of Fuzzy systems for achieving desired output. It is generated by the fuzzy toolbox available in MATLAB, which allows optimizing standard Sugeno Fuzzy model. The typical rule in a Sugeno model has the form-[58, 59, and 60].

If $X1 = x \; and \; X2 = y$ then output $z = ax + by + c$

The output z_i of each rule is weighted by the firing strength W_i of the rule. In the above case the firing strength is W_i and Method$(F1(x), F2(y))$.

The final output of the system is the weighted average of all the rule outputs computed as-

$$finaloutput = \sum_{i=1}^{n} W_i Z_i \Big/ \sum_{i=1}^{n} W_i \quad (12)$$

The advantages of ANFIS system are first, faster convergence is achieved during learning for a specific task and second, fuzzy network are capable of incorporating both numerical data (quantitative information) and expert's knowledge (qualitative information) and describe in the form of linguistic $IFTHEN$ rules [61]. A very good theoretical and mathematical explanation of NN, Fuzzy and ANFIS is given by Jasmin Thevaril et al. [62]. Kunjithapatham

Meena et al. have implemented the application ANFIS in solving the gender classification problem [63]. As the gender basically classified based on the feature pitch value but in some cases this features fails. Therefore the authors have used three new features i.e. energy entropy, short time energy, and zero crossing rates and trained them with fuzzy and neural networks to classify the gender of the speaker. The three features given input to the fuzzy system and then these variables are fuzzified into three various sets namely; large, medium and small and the output variable is fuzzified into three sets namely; male, female/male and female. In female /male the speech signal belongs to either male or female and the fuzzy rules generated. After training the fuzzy logic with fuzzy rules, the output of the fuzzy training are considered as features which belong to male or female. These features were used in neural network for classification. This method of gender classification is found to be better than other existing methods based on pitch. Anna Esposito et al. applied the ANFIS architecture to perform the nonlinear adaptive noise cancellation from speech signal [58]. A known noise signal is passes through a nonlinear model which makes the noise source nonlinear. If $n(t)$ is a known noise source and passes through a nonlinear passage, which transform it into a distorted and delayed noise $d(t)$ then the noisy signal $y(t)$ can be written as

$$y(t) = x(t) + d(t) = x(t) + f(n(t), n(t-1)) \quad (13)$$

Where $f(n(t), n(t-1))$ describes the dynamic nonlinear passage through which noise signal goes to corrupt the speech signal $x(t)$. Now to estimate $d(t)$ and to get a clean speech signal ANFIS system is used. Once $\hat{d}(t)$ is estimated the $\hat{x}(t)$ can be estimated by subtracting the$\hat{d}(t)$ from the noisy signal $y(t)$. T.Meera Devi et al. in [64] used the ANFIS technique for updating the weights of LMS, NLMS and RLS filter and found that with ANFIS, the results are better in terms of MMSE and SNR. The authors have stored the different types of noise such as babble noise, street noise car noise, airport noise and classified using fuzzy classification system. This pre-classified noise used as the reference signal with the noisy speech signal so the adaptive filter like LMS, NLMS and RLS can perform better. Table 3 gives the performance of LMS, NLMS and RLS with and without fuzzy system. Jay Kumar et al. also proposed a similar work with different types of noises i.e. song noise, clap noise, fan noise, and motor noise [59].

Table 3: SNR and Mean values of LSE algorithms before and after using ANFIS

Sl. No.	Algorithm	SNR (0 dB)		MMSE (0dB)	
		Without Fuzzy	With Fuzzy	Without Fuzzy	With Fuzzy
1	LMS	21	27	3.25	1.87
2	NLMS	24	27	2.10	1.86
3	RLS	27	30	1.97	1.49

A. Esposito et al. used the ANFIS to design a fast system for speech noise cancellation [58]. They have considered that the noise get its nonlinearity due to a dynamic passage which introduced the delays and distortion in the noise during the transmission. Thisdynamic channel is represented by a

$$d(t) = f\big(n(t), n(t-1)\big) = 4\,\frac{sinc(n(t).n(t-1))}{1+[n(t-1)]^2} \quad (14)$$

This $d(t)$ is estimated by the ANFIS and once $\hat{d}(t)$ is estimated then it is subtracted by the noisy speech signal to get a clean version of the noisy signal.

X. Post processing of enhanced speech

J.Bensety et al. have suggested a short list of speech communication application where a noise reduction algorithm is required are numerous [66]. Some of those application are listed below

- Hands free communication
- Voice over IP (VoIP)
- Hearing aids
- Answering machines
- Speech recognition
- Teleconferencing systems
- Car and mobile phones
- Cockpits and noisy manufacturing
- Multi-party conferencing

Also after the filtering some time the quality and intelligibility of signal get decreases, hence to regain the quality some techniques like harmonics regeneration cab be used [67].

XI. Conclusion

Filtering is the most important part of any speech based system. Because the noise degrades the quality and intelligibility of the signal so if the filtering is not done prior to any speech based application then the system doesn't work properly which misleads the output. 515 In this paper, based on the two varieties of noise i.e. linear and non-linear, different filtering techniques have been discussed. It is found that the filtering techniques are application based. If the effecting noise is linear i.e. noise who's PSD is constant, required a linear filtering approach like LMS and NLMS filtering and for non-linear noise i.e. PSD varies with time has to be dealt with the help of non-linear filtering techniques like modified versions of Kalman filter, Neural Networks, fuzzy based adaptive filter and ANFIS. Along with filtering techniques this paper comprises different feature extraction methods of speech. These features can be used for filtering as well as for any other speech based application like speaker recognition, gender recognition and automatic speech recognition system (ASR).

Acknowledgements

The authors would like to acknowledge Assam Don Bosco University, Guwahati for providing the resources necessary for compiling this paper.

References

[1] Kalandharan N, "Speech Enhancement by Spectral Subtraction Method", International Journal of Computer Applications (0975 – 8887) Volume 96– No.13, June 2014

[2] L. Wang, Z. Zhang, and A. Kai, "Hands-free speaker identification based on spectral subtraction using a multi-channel least mean square approach", IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, 2013, pp. 7224– 7228

[3] R.V.Pawar and R.M.Jalenkar, "Review on speech production model" International jour nal of engineering and innovation technology (IJEIT), Volume 3, Issue 9, March 2014

[4] Tohkura.Y, "Mathematical modeling of soeech production and its application to noise cancellation", IEEE transaction ASSP, Vol. ASSP-35

[5] Coker, C. H., "A model of articulatory dynamics and control", Proc. IEEE, pp. 452-460, 1989

[6] Zhe Chen, Dan Shi and Fuliang Yin, "Dynamic Noise Reduction Algorithm Based on Time-variety Filter", Awareness Science and Technology (iCAST), 2011 3rd Interna-tional Conference, ISBN Information: INSPEC Accession Number: 12577337 DOI: 10.1109/ICAwST.2011.6163085

[7] Kanika Garg and Goonjan Jain, "A Comparative Study of Noise Reduction Techniques for Automatic Speech Recognition Systems", International Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016, Jaipur, India

[8] G.Smith et al., "Subspace technique in speech enhancement", Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop

[9] Shuping Lv and Cheng Zhang, "Blind Signal Separation for Speech Signals With Noise", Proceedings of 2014 IEEE International Conference on Mechatronics and Automation August 3 - 6, Tianjin, China

[10] Yang Liu et al., "Speech enhancement in instantaneous amplitude and phase or application in

noisy reverberant environments", Speech communication, doi: 10.10.16/j.specom.2016.08.002
[11] Jacob Benesty et al., " A brief review of speech enhancement with linear filtering", EURASIP journal on advance in signal processing, 2014
[12] Vartika Anand et al., "Intelligent Adaptive Filtering For Noise Cancellation", Inter-national Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, ISSN ONLINE(2278-8875) PRINT (2320-3765)
[13] Monson H.hayes, "Statistical digital signal processing and modeling", chapter 4 and 9, Wiley edition
[14] Jingdong Chen et al., "New Insights Into the Noise Reduction Wiener Filter", IEEE transaction on audio, speech and language processing, Vol.14, No.4, July 2006
[15] Lalit P.patil et al., fficient algorithm for speech enhancement using adaptive fil-ter", International journal of electrical, electronics and computer engineering 3(1):98-103(2014)
[16] Sayed A. Hadei, M. lotfizad,"A Family of Adaptive Filter Algorithms in Noise Cancellation for Voice recognition" IJCEE, Vol.2, No.2, April 2010
[17] Jyoti Dhiman et al., "Comparison between adaptive filter algorithm (LMS, NLMS, RLS)", International journal of science, engineering and technology research (IJSETR), Volume 2, Issue 5, may 2013
[18] Sivaranjan Goswami et al., "A novel approach for design of a speech enhancement system using auto-trained NLMS adaptive filter", International journal of information and communication technology, Vol.6, Nos.3/4, 2014
[19] Anuja N. Untwale and Kishor S. Degaonkar, "Survey on noise cancellation technique of speech signal by adaptive filtering", Internation conference on pervasive computing (ICPC), 2015
[20] M.Mathe and S.P.Nadyala, " Speech enhancement using kalman filter for white, random and color noise", Signal processing conference, 2000
[21] M.S.Kavalekalam et al., "Kalman filter for speech enhancement in cocktail party sce-narios using a codebook-based approach", International conference on advance statistical signal processing, 2016
[22] Kishor Odugu and B.M.S.Rao, " A new speech enhancement using gammatone filters and perceptual wiener filtering based on sub banding", International conference on signal processing and communication, 2013, IEEE publisher
[23] Simon J. Godsill, "Robust modeling of noisy ARMA signals", IEEE international con-ference on acoustic, speech and signal processing 1997
[24] Cliston Cole et al. "Increasing Additive Noise Removal in Speech Processing Using Spectral Subtraction", Fifth International Conference on Information Technology: New Generations, April 2008, ISBN Information: INSPEC Accession Number: 9926122 DOI: 10.1109/ITNG.2008.86
[25] Rayan Kutty P.P and Sreenivasa Murthy A, "Kalman filter using quantile based noise estimation for audio restoration", Proceedings of ICETECT 2011
[26] M.Gabrea and D.O'shaughnessy, "Speech signal recovery in white noise using and adaptive kalman filter", Journal of mechanical science and technology, 23 march 2007, DOI:10.1007/BF02916349
[27] Zhe Chen, Dan Shi and Fuliang Yin, "Dynamic Noise Reduction Algorithm Based on Time-variety Filter", 3rd International conference on awareness science and tech nology (ICAST), ISBN Information: INSPEC Accession Number: 12577337 DOI: 10.1109/ICAwST.2011.6163085 Publisher: IEEE
[28] Hwa Soo Kim et al., "Speech enhancement via mel-scale wiener filtering with a frequency-wise voice activity detection", Signal processing conference, 2000, 10th European
[29] Harry Levitt, "Noise reduction in hearing aids: a review", Journal of Rehabilitation research and Development Vol. 38 No. 1, January/February 2001, Pages 111–121
[30] Premanada B S et al., "Speech Enhancement Algorithm to Reduce the Effect of Background Noise in Mobile Phones", International Journal of Wireless & Mobile Networks (IJWMN) Vol. 5, No. 1, February 2013
[31] Paulo S.R.Diniz, "Adaptive filtering algorithms and practical implementation", Third edition, ISBN: 978-0-387-31274-3 e-ISBN:, 978-0-387-68606-6, DOI: 10.1007/978-0-387-68606-6
[32] BinWen Fan et al., "The improvement and realization of speech enhancement algorithm based on wiener filtering", 8th Internation congress on image and signal processing (CISP 2015)
[33] F.Ykhlef and L.Bendaouia, "Evaluation of time domain features for voiced/non-voiced classification of speech", ICSES 2012-International conference on signal and electronic systems, WROCLAW @I 2012EEE
[34] V.P.Patil, "Voice signal recognition using LMS and UNANR filtering", IORD journal of science and technology, Volume 2, issue 2, pp 32-36
[35] Steven F.Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Transactions on acoustics, speech and signal processing, Vol. ASSP-27, No.2, April 1979
[36] Santos S.Pratapwar, "Reduction of Background Noise in Alaryngeal Speech using Spec-tral Subtraction with Quantile Based Noise Estimation"7th World Multiconference on Systemics, Cybernetics and Informatics (SCI 2003), Orlando, USA, July 27-30, 2003

[37] Orchisama Das, Bhaswati Goswami and Ratna Ghosh, "Application of the Tuned Kalman Filter in Speech Enhancement", 2016 IEEE First International Conference on Control, Measurement and Instrumentation (CMI)
[38] K.K. Paliwal and B.S. Atal, "Frequency related representation of speech", Proc. EU-ROSPEECH, p.p.65-68 Sep. (2003)
[39] Wei HAN et al., "An efficient MFCC extraction method in speech recognition", IEEE international symposium on circuits and system, 2006
[40] Vibha Tiwari, "MFCC and its application in speaker recognition", International journal on emerging technologies, ISSN:0975-8364
[41] Sunil Kumar and M.Laxminarayana, "Choice of mel filter bank in computing MFCC of a resampled speech", 10th international conference on information science, Signal processing and their applications (ISSPA 2010)
[42] Utpal Bhattacharjee, "A Comparative Study Of LPCC And MFCC Features For The Recognition Of Assamese Phonemes", International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue 1, January- 2013Vol. 2 Issue 1, January- 2013, ISSN: 2278-0181
[43] Safdar Tanweer, Abdul Mobin and Afshar Alam, "Analysis of Combined Use of NN and MFCC for Speech Recognition", World Academy of Science, Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:8, No:9, 2014 1736 International Scholarly and Scientific Research & Innovation 8(9) 2014 scholar.waset.org/1999.4/10000797 Inter-national Science Index, Computer and Information Engineering Vol:8, No:9, 2014 waset.org/Publication/10000797
[44] [44] Neeraj Kaberpanthi and Ashutosh Datar, "Speaker Independent Speech Recognition using MFCC with Cubic-Log Compression and VQ Analysis", International Journal of Computer Applications (0975 – 8887) Volume 95– No.26, June 2014
[45] Siddhant C. Joshi, Dr. A.N.Cheeran, "MATLAB Based Feature Extraction Using Mel-Frequency Cepstrum Coefficients for Automatic Speech Recognition", International Journal of Science, Engineering and Technology Research (IJSETR), Volume 3, Is-sue 6, June 2014
[46] Rabiner, L. and Schafer, R., "Digital Processing of Speech Signals". Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1978
[47] Lyne R.Palomar and Toshio Fukuda, "A comparative analysis of the topological struc-tres of different LPC feature-based speech models", International conference on Neural Networks, 1999. IJCNN '99
[48] Bishnu S. Atal and Joel R. Remde, "A new mpodel of LPC excitation for producing natural sounding speech at low bit rates", IEEE international conference on acoustic, speech and signal processing, june 2013
[49] R. L. K. Venkateswarlu et al., "Novelapproach for speech recognition by using self-organized maps", International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 4, August 2011
[50] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech", Acoustical society of America journal, vol. 87, pp.1738-1752, Apr.1990
[51] Yang Shao et al., "AN auditory-based feature for robust speech recognition", IEEE international conference on acoustical speecg signal processing, 2009
[52] Jun Qi et al., "Auditory feature based on gammatone filters for robust speech recogni-tion", IEEE international symposium oncircuits and systems (ISCAS), 2013
[53] Aurelio Uncini, "Audion signal processing by Neural networks", Elsevier, Volume55, Issue 3-4, pp.593-625
[54] Marina- Anca Cidota, "Choosing the parameter of the NARMA model implemented with recurrent perceptron for speech prediction", Cidotï¿œ, M. Neural Comput & Applic (2010) 19: 903. doi:10.1007/s00521-010-0375-7
[55] Brockwell PJ, Davis RA (1987) ,"Time series: theory and methods", Springer, New York
[56] Sh. Oveisgharan and M.B. Shamollahi, "Speech modeling and voiced/unvoiced/mixed/silence speech segmentation with fractionally Gaussian noise based model", International conference on acoustical speech signal processing, 2004
[57] Danilo P. Mandic et al., "Advance RNN based NARMA predictors", Journal of VLSI signal processing sytem for signal, image and video technology, Volume 26, pp 105-117, 2010
[58] Anna Esposito et al., "Designing a Fast Neuro-fuzzy System for Speech Noise Cancellation", Book Title MICAI 2000: Advances in Artificial Intelligence Pages pp 482-492 Copyright 2000 DOI 10.1007/10720076_44 Print ISBN 978-3-540-67354-5 Online ISBN 978-3-540-45562-2
[59] Jay Kumar et al., "environmental noise cancellation by using neuro fuzzy adaptive filtering", 2015 Fifth International Conference on Communication Systems and Network Technologies, IEEE publisher
[60] Sachin Lakra et al., "Selective noise filtering of speech signals using an adaptive neuro-fuzzy inference system as a frequency pre-classifier", Journal of Theoretical and Applied Information Technology 30th November 2015. Vol.81. No.3
[61] Mohammed Hussein Miry et al., "Adaptive Noise Cancellation for speech Employing Fuzzy and Neural

Network", Iraq J. Electrical and Electronic Engineering, Vol. 7, No. 2, 2011
[62] Jasmin Thevaril and H. K. Kwan, "Speech enhancement using Adaptive Neuro-Fuzzy Filtering", Proceedings of 2005 International Symposium on Intelligent Signal Process-ing and Communication Systems, December 13-16, 2005 Hong Kong
[63] Kunjithapatham Meena et al., "Gender Classification in Speech Recognition using Fuzzy Logic and Neural Network", The International Arab Journal of Information Tech-nology, Vol. 10, No. 5, September 2013
[64] T.Meera Devi et al. "Environmental Noise Classification and Cancellation using Fuzzy Classifier and Fuzzy Adaptive Filters", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 2, March 2012
[65] Radek Martinek et al., "A robust approach for acoustic noise suppression in speech using ANFIS ", Journal of electrical engineering, Vol. 66, No. 6, 2015, 301–310
[66] J.Benesty et al., "Speech enhancement", Springer series on signals and communication technology, ISBN 3-540-24039
[67] Yi Hu, "A simulation study of harmonics regeneration in noise reduction for electric and acoustic stimulation", The Journal of the Acoustical Society of America ï¿œ May 2010 DOI: 10.1121/1.3372718

AUTHOR PROFILE

Minajul Haque is currently serving as a lecturer in Guwahati Institute of technology. He has completed his Master of technology from Assam Don Bosco University in Electronics and Communication Engineering and Bachelor of technology from Biju Patnaik University of Technology in Electronics and Tele-communication in 2017 and 2014 respectively. His research interest is in Digital Signal Processing, Speech processing and Communication.

Kaustubh Bhattacharyya is currently serving as an Assistant Professor in the Electronics and Communication Engineering Department of Assam Don Bosco University, India. Prior to this he held research positions at Gauhati University, Assam, India. He completed his Bachelor of Science in Electronics, Master of Science in Electronics Science, Master of Philosophy in Electronics Science and Master of Technology in Electronics and communication Technology from Gauhati University, India in 2004, 2007, 2009 and 2012 respectively. His research interest is in high frequency communication and soft computing.