Speech is the fundamental and common medium, hence important for us, to communicate and most effective and reliable means for expressing oneself for personal communication. With advancement in hardware technologies, there are so many electronic and mobile personal communication based devices available, today in market and that too in cheaper cost and with easy availability. Fig. 1.1 shows some typical speech communication applications [1] and most effective and reliable means for expressing oneself for personal communication.

Fig. 1.1 Speech communication applications

The applications like speech recognition, mobile and personal communication, public address system are few of the applications from long list of speech based systems. However, undesired noises in environment like sound from heavy machines, vehicles are also present in one or other form everywhere. These noises cause undesired effects in speech transmission and acquiring systems. Recently, restricted or usable vicinity of applications is moving from one place and close room to more open and multiple locations, leading to several types of undesired signals of mixing with desired speech signal making speech more corrupt with noise. Not only human communications but intelligent machines which trying to automate the things and sometimes also takes decision based on what it receives as a speech, also suffers from the degraded performance.

Since last five decades, various approaches for noise reduction and speech enhancements have been investigated and developed. Among, very early and fundamental approach of noise reduction was introduced to use the theory of the optimum Wiener filter. Given a desired signal and an input signal, the Wiener filter produces an estimate of the desired signal that is optimal, i.e. the squared mean error or difference between the signals is minimized. The Wiener filter can also be adaptively estimated used in an environment where the surrounding noise has time-varying characteristics. Adaptive algorithms such as Least Mean Square (LMS) and Recursive Least Squares (RLS) are well known examples and also widely used. There are so many applications of speech still to be far from reality just because of lack of efficient and reliable noise removal mechanism and
preserving or improving the intelligibility for the speech signals [1-2]. In audio signal processing applications auditory system parameters\(^2\) like echo, multiple echo, reverberation, flanging and equalizer reduces the overall computational complexity and memory requirement of the system. Before doing speech enhancement we need to generate and analyze these parameters [3]. An acoustic \textit{echo} is one of the simplest acoustic modeling problems. Echoes occur when a sound arrives via more than one acoustic propagation path. Reverberation is the persistence of sound in a particular space after the original sound is removed. A reverberation, or reverb, is created when a sound is produced in an enclosed space causing a large number of echoes to build up and then slowly decay as the sound is absorbed by the walls and air [4]. The "flange" effect originated when an engineer would literally put a finger on the flange, or rim of one of the tape reels so that the machine was slowed down, slipping out of sync by tiny degrees. Equalization is the process of adjusting the strength of certain frequencies within a signal. The most well known use of equalization is in sound recording and reproduction [5]. Various audio effects were simulated in MATLAB. The implementation of effects performed using digital signal processing components. To get the simulated results, we have taken a sample wav file. Using this input sample file various audio effects were simulated. Figure 1.2 shows the sample input wav file, response of echo, reverberation, flanger and equalizer effects. Figure1.3 shows responses of LP, HP and BP filters for equalizer.

Recent advances in CPU and multi-core hardware has provided ample amount of computational power and thus, need for today is to design the complex but yet efficient and realistic approach for noise reduction to achieve speech enhancement. The speech enhancement is not only useful for storage and transmission of speech data but it can play vital role in improving much need system based speech recognition where accurate identification of words and sentences can provide automation in most of the human-machine based interface and also be useful in machine-machine interaction based automation. Robotics is a familiar example where speech recognition systems can become boon for today’s advanced society at social level in addition to during natural calamities and on war fields.

It is obvious that speech enhancement can boost up the performance of speech recognition systems by keeping low word error rate (WER). Types and sources of noise that can be considered in speech enhancements are also discussed in further section.

Fig. 1.2 The sample input wav file, response of echo, reverberation, flanger and equalizer effects

Fig. 1.3 Responses of LP, HP and BP filters
There are various types of advanced speech enhancement algorithms in literature and they can be classified in main three categories, namely: filtering/estimation based noise reduction, beam forming and active noise cancellation (ANC) techniques. Detailed knowhow of these techniques can aid the research in speech enhancements. In this thesis, we have attempted towards surveying the methodologies for speech improvement. It is also investigates, how these techniques affect the performance of various application systems like speech recognition and speech communication.

Speech is the fundamental way for “we humans” to communicate. This way of expressing oneself is probably one of the most effective and reliable means for personal communication. For centuries, efforts have been made to enable individuals to communicate over great distances, distances that render normal face-to-face speech communication impossible. The invention of the radio telegraph and the telephone in the nineteenth century was a great leap forward in the direction of seamless personal communication between persons on the geological locations.

At the same time the industrial revolution introduced new difficulties for personal communication in the form of high sound pressure levels from vehicles and other kinds of machine. Today, we live in a world where silence is a rarity and noise is almost constantly present. This noise sometimes impairs our ability to communicate reliably regardless of what communication means we choose. Not only human communications but intelligence machines which trying to automate the things and sometimes also takes decision based on what it receives as a speech, also suffers from the degraded performance. During the last decades, since the 1940’s, different approaches to noise reduction and speech enhancements have been developed. One early and fundamental method of noise reduction was to use the theory of the optimum Wiener filter. Given a desired signal and an input signal, the Wiener filter produces an estimate of the desired signal that is optimal, i.e. the squared mean error or difference between the signals is minimized. The Wiener filter can also be adaptively estimated used in an environment where the surrounding noise has time-varying characteristics. Adaptive algorithms such as Least Mean Square (LMS) and Recursive Least Squares (RLS) are well known examples and also widely used.

We can today retrospect notice a development from the earlier analogue techniques into digital techniques used since 1980’s. Recent advances in computers have rendered it possible to implement rather complex signal processing algorithm in real-time on digital processors. These processors are capable of performing rapid additions and multiplication which are two fundamental arithmetic operations used when filtering a digital signal. There are some well researched techniques and well studied situations. Like, frequency domain techniques, voice activity detection (VAD), multi-microphone techniques (Beam-forming), noise level estimation (SNR estimation) and active noise cancellations (ANC) techniques are few of many popular terms in speech enhancements.
Detailed knowhow of these techniques can aid the research in speech enhancements. Our research work aims at developing effective speech enhancement techniques particularly to improve the performance of applications like speech recognition or personal communication via mobile or Bluetooth.

1.1 Scope of Research

Speech is medium of communication to express message of the speaker. Along with the message of the speaker other information like language, dialect, gender and age of the speaker are embedded in the speech signal. Listener can perceive this information along with the message in the speech. In fact, human ears are capable of decoding this supplementary information in order to be more informed about speaker. Due to automation and law enforcement needs, there are so many applications derived from the field of speech processing like speech recognition system, gender classification based on the speech of speaker etc. However, most of the times, the speech signal is affected by interference from various sources of noise and consequently, speech processing in various application takes place with degraded speech, which may result in poor performance of the developed system.

In general, there exists a need for digital voice communications, human-machine interfaces, and automatic speech recognition systems to perform reliably in noisy environments. For example, in hands-free operation of cellular phones in vehicles, the speech signal to be transmitted may be contaminated by reverberation and background noise. In many cases, these systems work well in nearly noise-free conditions, but their performance deteriorates rapidly in noisy conditions. Therefore, development of pre-processing algorithms for speech enhancement is always of interest. The goal of speech enhancement varies according to specific applications, such as to boost the overall speech quality, to increase intelligibility, and to improve the performance of voice communication devices.

Enhancement of speech is useful in many applications like aircraft, mobile, military and commercial communications. An enhancement of speech is also useful to improve perceived speech for hearing impaired persons or to improve the speech of speaker with defective speech production process. In all these applications the end users are human beings. Apart from these, there are other applications which involves enhancement as a pre-processing step for other speech processing tasks such as speaker recognition.
1.2 Problem Formulation

Speech enhancements involve processing of speech signals in temporal and/or spectral domains. Any such processing introduces distortion into speech signal. Trade-off between the reduction of noise and introduction of new distortion depends on the perception by the human auditory system. Enhancement in the processes signal is measured in terms of quality and intelligibility. Quality refers to naturalness and case of listening to speech, whereas intelligibility refers to ease of understanding speech. In general, high quality speech signal can be considered as highly intelligible, but highly intelligible speech signal need not be of high quality.

In order to have effective speech enhancement techniques applicable for wide range of applications, following are the research objectives of our work:

- Literature survey for existing techniques and modifications suggested by various researchers in present application scenario.
- To study the effect of various auditory parameters, this reduces computational complexity and memory requirement of digital processor system.
- To understand the effect of noise on various applications based on speech processing like speech recognition, speaker identification, gender classification, personal and mobile communication etc.
- To identify the specific noise based issues and their severity for different speech processing applications.
- To explore and implement the various noise removal and speech enhancement techniques.
- To analyze the effectiveness and usefulness of speech enhancement techniques in one or more than one speech processing applications.
- To study the effect of various noise levels on the speech processing applications and formulating the new SNR estimation strategy in order to improve the performance of particular speech enhancement technique.
- To conduct the experiment based on dataset available publicly or self created dataset to evaluate the performance of speech enhancement techniques.
- To embed the speech enhancement technique in one or more applications of speech processing and observe the change in performance of the system using dataset available publicly or self created dataset.
• Real time and hardware implementation of speech recognition technique using MATLAB and CCS V3.1 on DSK 6713 from Spectrum Digital Corporation.
• Hardware profiling of technique considering it as real time speech processing for multimedia applications.

1.3 Thesis Contribution

In this research work, the problem of speech enhancement to be used in multimedia application like speech recognition has been considered. In order to improve the performance of speech recognition using specialized speech enhancement technique. In the part of our work, we proceed with two-fold objectives. First is to improve the speech recognition performance in multi-microphone environment. Second, we attempted to analyze the performance of speech recognition against the filter-bank parameters; filter length and number of sub bands. In the remaining part of the research work, we have improved the performance of beamforming based speech recognition system using evolutionary computational algorithms (Genetic algorithm, GA). Additionally, the system is made to be working in real-time as time required for classifier has been reduced dramatically. This is particularly achieved by including the zeros at random places and in random amount in initial population chromosomes, which were generated randomly in the range of 0 to 1. This results in the reduction of feature elements in feature descriptor and have feature vector length.

1.4 Limitations and remedial action during research work

This research work involves development of new strategies based on soft computing for real time speech processing in multimedia application. We will be using MATLAB software for speech processing. The speech signals used in our work will be either dataset available at reputed research groups working in same field or new dataset designed by us. Validation of results will be with either annotated dataset or manual annotation using ground truth. Despite of the pre-structured work planned, there could be limitations as described below. In case of some limitation occurred, the corrective action is taken as indicated in table 1.1.
Table 1.1 Limitations and remedial action

<table>
<thead>
<tr>
<th>Limitation</th>
<th>Remedial Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>The speech signal dataset available is not capable of serving the purpose of our research work.</td>
<td>Construction of new data set or changing the objective.</td>
</tr>
<tr>
<td>Non-availability of methods to annotate the data.</td>
<td>Finding method to know approximated ground truth subjectively.</td>
</tr>
<tr>
<td>The new methodology to be explored is not giving satisfactory performance.</td>
<td>Change the methodology or changing the objective.</td>
</tr>
<tr>
<td>Software constraints or unavailability.</td>
<td>Choosing the new software or developing the methodology with basic software such as C/C++ or JAVA or using speech processing libraries.</td>
</tr>
<tr>
<td>Lack of validation measures.</td>
<td>Coming up with new measures or taking experts views.</td>
</tr>
<tr>
<td>Non-availability of real speech data from real environment.</td>
<td>Simulation of speech signals or changing the objective.</td>
</tr>
<tr>
<td>Non-availability of annotated data.</td>
<td>Manual annotation or changing the objective.</td>
</tr>
<tr>
<td>Impossible to implement mathematical process.</td>
<td>Approximating the mathematical process.</td>
</tr>
</tbody>
</table>
Chapter-1 Introduction

1.5 Outline of Thesis

The thesis is organized in the form of eleven chapters as follows:

Chapter-1 Introduction

The “speech signal” is an integral part of most of the multimedia applications apart from the personal communication. Here we wish to produce and analyze a variety of effects, viz. echo, multiple echo, reverberation, flanging and equalizer on pre-recorded speech, which will be used in sound processor to generate various musical effects. The broad objective of this thesis is to devise new strategies using soft-computing techniques so that performance of speech based application can be improved. In this chapter, introduction of speech recognition and enhancement is presented. How speech is influencing lifestyle of citizens and what kind of research is required for speech related applications are briefly discusses here. The brief note on scope of this topic is placed in this section. Apart from scope and applications, the specific problems dealt in this thesis work are mentioned. More precise contributions of this thesis are also highlighted here. It is interesting to foresee the effect of this research in future speech related applications especially in multimedia area.

Chapter-2 Basic Concepts of Speech Enhancement and Recognition

The effect of speech purity is visible at the performance of any speech based applications. The speech recognition, being a multimedia application under consideration here, has been a very important system in almost every area of life. Here an attempt has been made towards studying and implementation of speech enhancement techniques like Spectral Subtraction, Minimum Mean Square Error (MMSE), Kalman, Wavelet Transform, Wiener and Adaptive Wiener filter. The intelligent speech enhancement techniques can raise the outcome of speech recognition and hence it is very important to know the basics involved in it. This chapter is devoted to know the fundamental concepts behind the various classes of speech image enhancements techniques. The short description of speech recognition is also included in this chapter. Finally, brief note on various performance measures in speech enhancements and recognition has been incorporated.
Chapter-3 Literature Review

The speech, being a fundamental way of communication for the humans, has been embedded in various essential applications like speech recognition, voice-distance-talk and other forms of personal communications. There are so many applications of speech still to be far from reality just because of lack of efficient and reliable noise removal mechanism and preserving or improving the intelligibility for the speech signals. The broad categories of speech enhancement techniques can be listed as speech filtering techniques, beam forming techniques and active noise cancellation methods. In this chapter, an attempt has been stepped towards surveying the methodologies for speech improvement. It was also interesting to discuss, how these techniques affect the performance of various application systems like speech recognition and speech communication. Essentially, we also discussed here about the types and sources of noise that can be considered in speech enhancements.

Chapter-4 Soft-computing: Evolutionary Computations

The continuing advances of computational technology such as availability of large memories in small space, parallel GPUs, have changed the paradigm of the way the problem used to be solved. The soft-computing is the new paradigm for the computationally solvable complex problems and has been heavily relied on the computational power of devices. It includes the probabilistic theory in addition to the elements covered by computational intelligence. In this chapter, we have given brief description of main elements of the soft-computing, such as neural network, fuzzy logic and genetic algorithm.

Chapter-5 Speech Enhancement and Beamforming

A new generation of speech acquisition applications is emerging as a result of advances in technology and the prevalence of mobile and broadband communication. Thus, it becomes essential to have reliable speech processing based applications. The speech is corrupted with so many different types of noises and by cross voices. This presents the need of cleaning out the speech so that applications can perform without any flaws. In
this chapter, we have described in detail about the speech enhancement theory and especially with beamforming techniques.

Chapter-6 Experimental Setup and Dataset

While experimenting with speech enhancement using beamforming technique and later for the speech recognition experiments, there is a need of dataset with ground truth. There are not many datasets available in public across the world. It was necessary to construct the dataset using beamforming parameters. In doing so, we have attempted to do the simulation of speech database to be used for speech recognition experiments with beamforming parameters. In this chapter, the detail of this simulation has been provided about this dataset.

Chapter: 7 Speech Recognition using Beamforming technique

In this chapter, our work has two-fold objective. First is to improve the speech recognition performance in multi-microphone environment. Second, we attempted to analyse the performance of speech recognition against the filter-bank parameters; filter length and number of subbands. The experiments were performed for 20 words including numbers and commands, 10 words of numbers only and 10 words of commands only for different values of filter bank parameters. The results obtained have proved the speech enhancing capability of the beamforming technique in multi-microphone network where noise and echo-interference can degrade the original speech signal.

Chapter-8 Evolutionary Computation based Real Time Speech Beamforming for Multimedia Applications

Here we have presented the approach of evolutionary computation in form of genetic algorithm to select the features that are responsible for discriminating the different words. The system is made to be working in real-time as time required for classifier has been reduced dramatically. This is especially an important requirement in the mobile devices where power, memory and processing power are available with large constraints. The experiments were performed for 20 words including numbers and
commands, 10 words of numbers only and 10 words of commands only for different values of filter bank parameters. The results show the effectiveness of the GA optimization in all the subsets of experiments with different parameters of beamforming.

**Chapter-9 Real-time implementation of speech recognition**

In this chapter we depicted the real time hardware implementation of speech recognition using DSP processor software development kit, DSK-TMS320C6713 with Code Composer Studio (CCS). MFCC algorithm calculates cepstral coefficients of Mel frequency scale. After feature extraction from recorded speech, each Euclidian Distance (ED) from all training vectors is calculated using Gaussian Mixture Model (GMM). The command/voice having minimum ED is applied as similarity criteria. The timing analysis is done for various individual blocks of algorithm. The time required for processing in DSP and PC processors are compared.

**Chapter-10 Conclusions and future scopes**

In this chapter final conclusions, future extension of the work and future scope in this field are elaborated.

**Chapter-11 References**

Thesis ends with Bibliography which includes the list of references used in each chapter, research project details, list of short term program attended, list of publications and presentations, additional resources used and list of MATLAB programs simulated for research work.

**1.6 Conclusion**

The “speech signal” is an integral part of most for the multimedia applications apart from the personal communication. The broad objective of this thesis is to devise new strategies using soft-computing techniques so that performance of speech based application can be improved. In this chapter, introduction of speech recognition and enhancement is presented. How speech is influencing
lifestyle of citizens and what kind of research is required for speech related applications are briefly discusses here. The brief note on scope of this topic is placed in this section. It is interesting to foresee the effect of this research in future speech related applications especially in multimedia area. Apart from scope and applications, the specific problems dealt in this thesis work are mentioned. More precise contributions and outline of this thesis are also highlighted here.