Chapter 1

Introduction

Speech is the primary means of communication among human beings and is a result of complex interaction among vocal folds vibration at the larynx and voluntary articulators' movements (i.e. mouth, tongue, jaw, etc.). People use speech to communicate messages amongst themselves. When a speaker and a listener are nearer to each other in a quiet environment, communication is by and large easy and accurate. However, when people are separated by distance or if there is a noisy environment, they find it rather difficult to understand, that is to say, their ability to grasp receives a setback. Historically speaking the task of sophisticated speech signal enhancement in the field of communication engineering is said to have commenced just after the invention of telephone by Alexander Graham Bell in the year 1850. However, in the initial stages, the speech signal transmission, processing and reception was analog in nature and used only wired communication with a restricted number of users only. The meaningful work was started in this field after establishment of Bell Telephone Laboratories at New Jersey, USA. Since then the evolving discrete time signal processing techniques along with the development in digital hardware and software technologies have helped the rapid growth of purely digital speech signal processing applications like speech coding, speech synthesis, speech recognition, speaker verification and identification and speech enhancement [1]. At present the wireless communications industry is heavily dependent upon advanced speech coding techniques, while the integration of computers and voice technology (speech recognition, synthesis etc.) are poised for growth. Both the speech coding and recognition require some speech enhancement strategy to be embedded into them. An attempt has been made here to explain the requirements and scope in the field of speech enhancement and its real time implementation.

1.1 Requirements of Speech Enhancement

In electronic communication systems speech signal is transmitted electrically; the conversion media (microphone, loudspeaker, headphones, earphones), as well as the transmission media (wired or wireless), typically introduce distortions, yielding a noisy and distorted speech signal. Such degradation can lower the intelligibility (the likelihood of being correctly understood) and/or quality (naturalness and freedom from distortion, as well as ease for listening) of speech. The speech enhancement techniques are required to improve the speech signal quality and intelligibility in many applications either for the human listener or to improve the speech signal so that it may be better exploited by other speech processing algorithms. For human listeners speech enhancement technique should aim at high quality as well as intelligibility, while quality is largely irrelevant if the enhanced speech serves as input to a recognizer [2-4]. For coders or recognizers, speech could actually be "enhanced" in such a way as to sound worse, as long as the analysis process eventually yields a high-quality output i.e., if the "enhanced" input allows more efficient parameter estimation in a coder or higher accuracy in a recognizer, it serves its overall purpose. For example, pre-emphasizing the speech (to balance relative amplitudes across frequency) in anticipation of broadband channel noise (which may distort many frequencies) does not enhance the speech as such, but allows easier noise removal later (via de-emphasize). The present day speech enhancement techniques improve speech quality without increasing intelligibility; in fact in some cases it reduces intelligibility. Aspects of quality are worthwhile general objective. However, when there are distortions in speech, it is usually considered more important to make it intelligible rather than merely make it more pleasing.

In recent years with the increasing use of wireless communication in cellular and mobile phones with or without 'hands free' system, voice over internet protocol (VOIP) phones, voice messaging service (voice mail), call service centers, cord less hearing aids etc. require efficient real time speech enhancement strategies to combat with additive noise and convolutive distortion (e.g., reverberation and echo) that generally occurs in any communication system [6]. The other areas of application areas include aircraft and military communication, aids for hearing impaired persons, communication inside vehicles and telephone booths, enhancing emergency calls and black box recordings etc. Besides, speech enhancement is also required as a pre-processing block in other speech processing systems like speech recognition, speaker recognition, speaker identification and speech coding [4]. The speech codecs used in 3G cellular mobile phones require speech enhancement in a post-processing stage [41]. Most speech enhancement algorithms are needed to detect intervals in a noisy signal where speech is absent in order to estimate aspects of the noise alone. This is done by using voice activity detector (VAD) and hence the voice activity detection is an integral part of most speech enhancement techniques. The performance of most speech enhancement algorithms is highly dependent on VAD [4]. Hence speech enhancement and detection must be treated simultaneously. The VAD also finds application in mobile phones to detect speech/silence to reduce power consumption during non speech periods.

1.2 Scope of Research in Speech Enhancement and Detection

Speech enhancement algorithms have been made applicable to problems as diverse as background noise removal, cancellation of reverberation and multi-speech separation (speaker separation) in modern speech communication systems. This is outlined in figure 1.1. This figure depicts a speech signal being degraded by an additive background noise. In pursuance of this, a speech enhancement algorithm is used to restore the quality of the speech, before finally being presented to the listener. The listener here is taken to be either a human or a machine [7-8]. Besides, other sources of degradation also exist for speech signals, such as distortion from the microphone or reverberation from the surrounding environment. The approach to speech enhancement varies considerably depending upon the type of degradation.

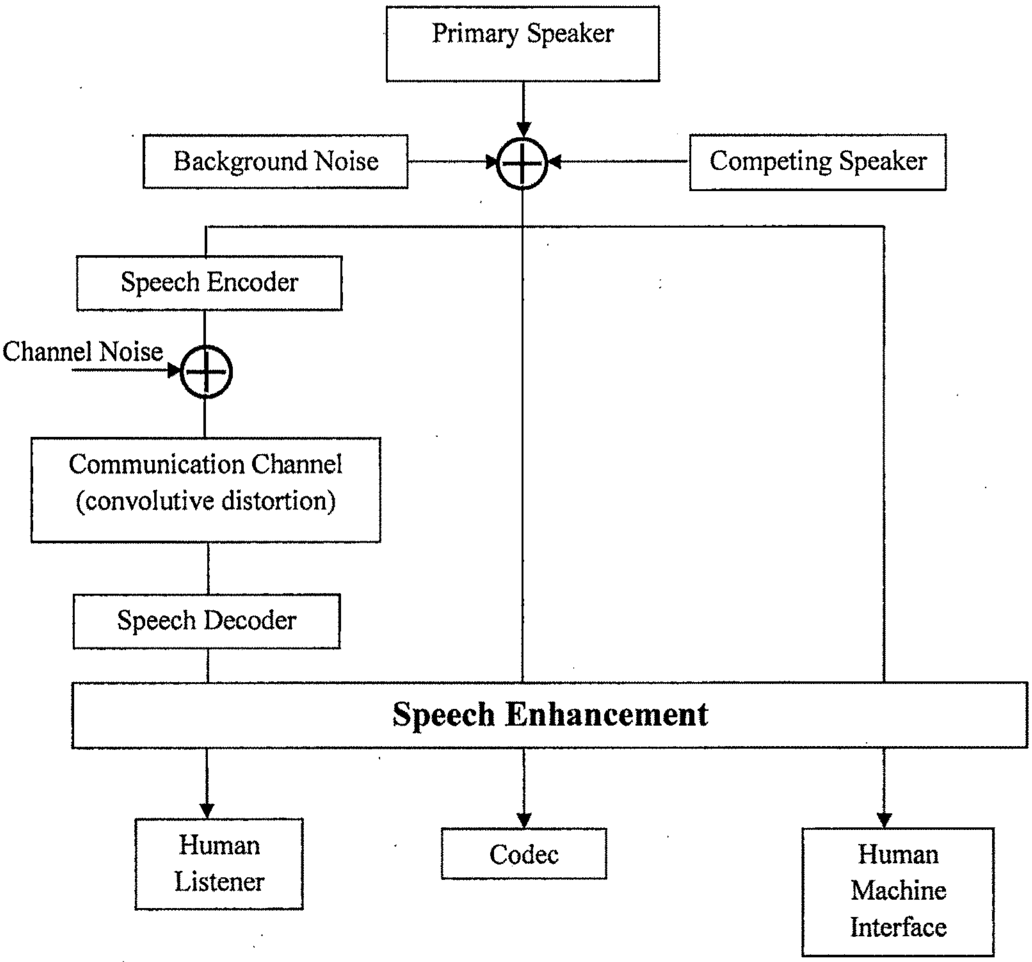

Fig. 1.1 Requirements of speech enhancement methods

The speech enhancement techniques can be classified into two basic categories: (i) Single channel and (ii) Multiple channels (array processing) based on speech acquired from single microphone or multiple microphone sources respectively [3]. However, single channel (one microphone) signal is available for measurement or pick up in real environments and hence the focus is here on single channel speech enhancement methods. That apart, the methods must also have other characteristics like real-time implementation, reasonable computational complexity while processing, low level of speech distortion, operation with low level SNR, separation as cleaned speech signal, adaptation to background noise, controlled level of noise suppression in speech, possibility of using a graphic equalizer for removing the stationary hindrances and easy integration with target applications etc. etc.

The pioneer work in the field has been done by Lim and Oppenheim [9] in 1979. Since then several methods have been evolved in the literature for single channel speech enhancement during last thirty years. The major contributors in this area are Boll and Berouti, (1979), Ephraim and Malah (1984), Sclarat (1986), Virag (1999), Kamath (2002) and so on [10-18]. The approach to speech enhancement varies considerably depending upon type of degradation. Various domains of speech enhancement are discussed throughout the thesis. Most of the methods assume the noise to be stationary and VAD estimates the noise characteristics during speech pauses or silent period [19-20]. However, some researchers have proposed the method to handle non-stationary noise [21].The limitations of these methods still pose a considerable challenge to researchers in this area. The objectives of speech enhancement vary widely, namely; reduction in noise level, increased intelligibility, reduction of auditory fatigue, etc. For communication systems, two general objectives depend on the nature of the noise, and often on the signal-to-noise ratio (SNR) of the distorted speech. With medium-to-high SNR (e.g., > 5dB), reducing the noise level can produce a subjectively natural speech signal at a receiver (e.g., over a telephone line) or can obtain reliable transmission (e.g., in a tandem vocoder application). For low SNR (e.g., $\leq$5dB), the objective could be to decrease the noise level, while retaining or increasing the intelligibility and reducing the fatigue caused by heavy noise (e.g., train or street noise). [1]

[1] A paper entitled "Requirements and Scope of Speech Enhancement Techniques in Present Speech Communication Systems" is presented in National Technical Paper Contest-2010 (NTPC-2010) for seniors at IETE Vadodara centre, Vadodara in March 2010 and won 3rd prize.

In the present work, the goal is to design a single channel speech enhancement algorithm having good noise suppression characteristic in the low SNR range (0-5dB) for various noise characteristics.

1.3 Objectives of the Research Work

The research topic is motivated by the fact that the speech is the most important signal transmitted using communication system and it is always subjected to background and surrounding noise and distortion. Accordingly the performance of speech communication system is greatly improved if speech enhancement is embedded into this system and if it works in real time. Several strategies have been suggested in the past for that and still however some of the challenges have remained unsolved. Hence it is essential to develop new strategies for real time embedded speech enhancement and detection considering the communication application [37-39]. The use of technical computing development support tools such as MATLAB, SIMULINK and related Toolboxes [42-48] make simulation study as well design of graphical user interface more simple and refined.

The work described in the thesis includes; amongst others the following:-

- Literature survey for existing techniques and modifications suggested by various researchers in present application scenario.
- Simulation of transform domain techniques using MATLAB/SIMULINK.
- Objective and subjective evaluation of simulated techniques.
- Limitations of existing techniques considering communication applications and suggesting new strategies to overcome it.
- Simulation, performance evaluation and comparison of suggested strategy using MATLAB/SIMULINK.
- Real time and hardware implementation of existing and modified techniques using SIMULINK on PC and using SIMULINK/ RTW/ Embedded Target for TI C6000 toolboxes of MATLAB and CCS V3.3 on DSK 6713 from Spectrum Digital Corporation.
- Hardware profiling of techniques considering it as embedded real time application.

1.4 Organization of the Thesis

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

Chapter: 1 **Introduction:** This chapter provides an overview and the context for the remainder of the thesis.

Chapter: 2 **Speech Enhancement Techniques: State of Art:** This chapter describes the survey of different speech enhancement techniques and existing algorithms which is the main part of the literature survey. The exhaustive search is done to find out the basic techniques and modifications suggested by various researchers. The classification is also presented in this chapter considering the type of degradation, processing domain and tools used. A case study of simulation and implementation work using Normalized Least Mean Square (NLMS) algorithm for noise and echo cancellation is described. MATLAB and SIMULINK are used for simulation and Real Time Workshop (RTW), Embedded Target for TI C6000 toolboxes from MATLAB, Code Composer Studio version 3.3 (CCS V3.3) and DSK 6713 hardware platform are used for implementation.

Chapter: 3 **Speech Enhancement and Detection Techniques: Transform Domain:** This chapter describes techniques for additive noise removal which are transform domain methods and based mostly on short time Fourier transform (STFT). The discrete Fourier transform is used as transformation tool in these techniques. These methods are based on the analysis-modify-synthesis approach. They use fixed analysis window length (usually 20-25ms) and frame based processing. They are based on the fact that human speech perception is not sensitive to spectral phase but the clean spectral amplitude must be properly extracted from the noisy speech to have acceptable quality speech at output and hence they are referred to as short time spectral amplitude or attenuation (STSA) based methods. The phase of noisy speech is preserved in the enhanced speech. The synthesis is mostly done using overlap-add method. They have been one of the well-known and well investigated techniques for additive noise reduction. Also they have less computation complexity and easy implementations. The detailed mathematical expression for the transfer gain function for each method is described along with the terms used in the function. The relative pros and cons of all available methods as well as applications are mentioned. However, they require the use of voice activity detector (VAD) and the performance depends on the accuracy of VAD. The magnitude spectral slope distance VAD is the simplest and reasonably accurate and its operation is described in this chapter. The other transformation tool used in speech enhancement is discrete wavelet transform (DWT) and the techniques based on DWT are also described in brief here. The performance evaluation of any algorithm is very important for comparisons. There are several objective measures are available to evaluate the speech enhancement algorithms. They are described in brief in this chapter.

Chapter: 4 **MATLAB Implementation and Performance Evaluation of Transform Domain Methods:** The simulation carried out to describe the functionality and behavior of STSA methods under various additive noise conditions is described in this chapter. The simulation work is concreted and converged by preparing a MATLAB GUI (Graphical User Interface). This GUI can be used to simulate any transform domain algorithm for different noise conditions. The IEEE standard database NOIZEUS (noisy corpus) is used to test algorithms. The database contains clean speech sample files as well as real world noisy speech files at different SNRs and noise conditions like airport, car, restaurant, train, station etc. The performance evaluation using different objective measures is also carried out and explained in this chapter. The GUI also includes evaluation of algorithms using objective measures. The limitations and present implementations of these methods are also mentioned.

Chapter: 5 **Relative Spectral Analysis-RASTA:** This chapter describes the Relative Spectral Analysis (RASTA) processing of speech which is originally proposed for automatic speech recognizers to work in reverberant environments. The original algorithm is modified later on for direct speech enhancement. In the present work this algorithm for speech enhancement is simulated and evaluated under different noise conditions. The original filter is redesigned to have better performance. The algorithm throws the challenge for real time implementation as it is non linear and non causal. However, it does not require the use of VAD and can be used to combat with additive and convolutive distortions.

Chapter: 6 **Hybrid Algorithm for Performance Improvement:** It is suggested here that for better performance the transform domain algorithm can be combined in some way with RASTA approach. The best performing transform domain algorithm is MMSE STSA85 (LSA) and it is combined with modified RASTA approach which is the modified and suggested approach for speech enhancement. This algorithm is also simulated and tested under different additive noise conditions using the NOIZEUS database and compared with the original algorithms. The results of performance evaluation using objective measures are described in this chapter. The comparison using alone objective measures is not sufficient as it will not ensure the quality of speech signal for human listeners and hence the subjective evaluation is also required to perform. The IEEE recommended and ITU-R BS.562-3 standard mean opinion score (MOS) listening test is carried out. The chapter describes the various guidelines followed to perform this test. The original and modified algorithms are compared based on this test and conclusion is made regarding quality of output of different algorithms. The algorithm is also tested under different reverberation condition using the Aachen impulse response (AIR) database developed by RWTH Aachen University, institute of communication systems and data processing (India). It is a set of impulse responses that were measured in a wide variety of rooms. This database allows realistic studies of signal processing algorithms in reverberant environments. The comments are made about performance of algorithms in the simulated reverberant conditions.

Chapter: 7 **Hardware Implementation Tools:** This chapter describes the suitable hardware implementation tools available for speech processing system. The SIMULINK, RTW, Target Support Package TC6 available with MATLAB package can be used for implementing algorithms on various DSP processors and microcontrollers. The 32 bit floating point TMS320C6713 DSP from Texas Instruments is suitable for embedding speech processing algorithms for real time applications. For rapid prototyping the DSP starter kit DSK 6713 is available from Spectrum Digital Incorporation. The code can be loaded on DSP using the compiler Code Composer Studio and DSK6713. All these tools together can be used to implement speech processing algorithm in real time. They are described in brief in this chapter.

Chapter:8 **Real Time and Embedded Implementation of Hybrid Algorithm:** The hybrid algorithm is first tried for real time implementation on PC. For dedicated hardware implementation the DSP implementation platform using TMS320C6713 processor from Texas Instruments is selected. Various profiling results are also obtained and compared in this chapter.

Chapter:9 **Conclusions and Future Scopes:** Final conclusions and future extension of the work and future scope in this field are elaborated in this chapter.

Chapter: 10 **References:** It contains Bibliography which includes the list of references used in each chapter.