Chapter 1

INTRODUCTION

1.1 Overview

Speech Signals are the primary source of direct transmitter-to-receiver human communication and falls in the category of acoustic signals. These signals are the mechanical waves represented in terms of analog signal and propagate as vibration in the channel. Only one thing that classifies the speech signals in acoustic signals is its origination by humans. The signals possess most diverse features as characteristics of human development and culture. Henceforth, the applications of speech signals range from music, medicine to security, authentication etc. The transmission of speech signals commence at generation, propagation and conclude by reception of signals (figure 1.1). The unwanted signals (figure 1.2) from different sources mix with these signals and generate noise. The noise in digital communication can surface due to various reasons for example: finite precision of equipments or use of local components, faulty lines, improper coding etc. This type of noise is a different subject as they manipulate the nature of speech signal. The voice signals in an open channel (air) are affected by surrounding noise that is ever present. These noise are the audible signals arising due to any operation, person or by an instrument. The spectral quality index [1] and signal to noise ratio [2][3] [4] of speech signals in this case are directly proportional to noise content in it. The first attempt for signal denoising was performed by Weiner et al. [5] who suppressed the noise from the source signal. However, for filtering the assumption is made that both noise and signal should be stationary and their statistic information is available priori. These conditions in present era communication are impossible to withstand. In past few decades the wireless channels have become too noisy in nature. The acoustic problems due to these have gained considerable amount of exponential growth with increase in technology and systems.

Figure 1.1: Enhancement of Speech Signals
1.1.1 History

In history, denoising of speech signals was always parallel to speech recognition researches. According to Suma Swamy [6] classification algorithms had gained 80% accuracy in starting of 21st century but the next decade witnessed a lapse of attention in this section. A major event “Google Voice Search app for iPhone” retained interest of researchers in this section as the voice of users were recognized and stored to generate speech models for better search. The initiative of this business demon elevated the interest of commercial research and more premium applications surfaced in past few years. Another application Siri draws knowledge about speaker for delivering the output [7].

1.1.2 Applications

Speech signals are the basic need for multimedia and entertainment purposes. A typical example of this is voice morphing with facial counterpart (the voice are recorded in different studio and face in different) and with change in character, the voice behind should attain respective pitch and amplitude respectively. Another application of voice is in the forensic laboratories. The original voice of victim and witness are tested for authentication purpose. The intelligence agencies hold a wide ‘sound bank’ for identification of suspects, as the communication can be in any language or with artificial modulation of voice.
1.2 Speech Signals Classification

1.2.1 Voiced/Unvoiced/Silence determination

Speech is composite mixture of two signals i.e. speech information and silent or noise selection existing in between utterances. The informative section of speech is further studied in two categories:

(a) Voiced Speech
(b) Unvoiced Speech

Voice speech generates due to forceful air pressure through larynx. The vocal cords produce the vowel sounds by stimulated tension through opening and closing of glottis and repeated production of periodic pulses of air. These pulses are nothing but the sound in common terms. Psychoacoustic experiments demonstrate that origination of most of speech occurs at this section and in unity characterize a speaker. Unvoiced speech originates due to constraints formed at vocal tract that produce turbulence at the intension of speakers. The speech signal analysis is distinguished from fellow signals by three main characteristics.

1.2.2 Characteristic features for v/un determination

1. Zero Crossing Rate: The rate of speech signal required to overcome zero provides information of source. Figure 1.3 depicts that mechanical sound signals possess higher ZCR than speech signals [8]. The energy in mechanical speech lies mainly at higher frequencies in comparison with voice and results high ZCR. The ZCR [8] is written in terms of components of signal.

\[ ZCR = \sum_{m=-\infty}^{\infty} |\text{sign}[x(m)] - \text{sign}[x(m-1)]| \cdot w(n-m) \quad (1.1) \]

\[ w(n) = \frac{1}{2N} \quad 0 \leq n \leq N - 1 \quad (1.2) \]
2. **Energy**: The mechanical voice has lower amplitude as in compared with voiced segments.

Equation 1.3 energy equation is show cast of amplitude information \[ E_n = \sum_{m=\infty}^{\infty} x^2(m) \cdot h(n - m) \] (1.3)

The \( E_n \) from equation 1.3 is passed through low pass filter for smoothing and reflection of amplitude variation in time. A second order hamming window \( h(n) \) sustains better outputs in terms of amplitude variation [8].

In voiced speech (Fig. 1.3) the short-time energy values are much higher than in unvoiced speech, which has a higher zero crossing rate.

3. **Cross-correlation**: 

The cross-correlation is measured among two adjacent pitch cycles of speech signals. If the cross-correlation is high (approximate to 1), the speech signal will be voice and vice versa.
Pitch detection:

If human voice is stated as quasi-periodic signal, the basic period would be defined as pitch period. The general frequency, pattern, time, gain and fluctuation vary from one person to another. For signal classification and analysis, it is necessary to determine pitch. A pitch detection method [9] states that high cross-correlation should be present in two consecutive pitch cycles.

The pitch detector’s algorithm can be given by equations 1.4 and 1.5.

\[
(x, y) = \int_{t_0}^{t_0+\tau} x(t) \cdot y(t) \, dt; \quad y(t) = x(t - \tau)
\]

(1.4)

\[
T_0 = \arg\max(p_\tau); \quad p_\tau = \frac{x y}{||x|| \cdot ||y||}; \quad ||x|| = (x, x)^{1/2}
\]

(1.5)

As the signals cannot transmit with 100% efficiency due to channel constraints, the signals require filtering. The channel in addition with system limitations introduce some unwanted signals that penetrate into speech and produce nuisance effects. These signals in collective term are referred as noise.
1.3 Noise

Many researchers have proposed different definitions of noise due to its broad category of existence. In general, the most common definition states- “Noise is random, undesirable electrical energy that enters the communications system via the communicating medium and interferes with the transmitted message. However, some noise is also produced in the receiver.”

Before initiating the study on noise reduction, it is recommended to overview noise as the starting point. Noise can be defined as the disturbance in the signal due to unwanted signal. The signal from any source that interfere the signal under consideration is termed as noise signal. Some common example of noise signals are the sound from mechanical devices, biological creatures, natural phenomenon (lightening, thunderstorm etc.). The noise from the surrounding mixes with the signal and disorients its information. This effect does not need noise signal to be noisy enough to be recognized by human ears. Ultrasound noise, radio waves (for example) too interferes with signal. In contrast with human intelligence machines cannot simply deduce noise and eliminate it. The noise prevention methods cannot be very effective since, two signals of different phase for two different receivers are noise for each other. In another words, the activity for one system produce signals that are noise for another system. Some examples of noise from day to day activities are mentioned in [13]. The effect of noise cannot be defined as the magnitude of noise is responsible for it. In this scenario, the study of noise elimination techniques becomes indispensible part of modern signal processing systems.

On the basis of the range of frequency present, noise can be broadly categorized into Broadband Noise and Narrowband Noise:

- **Broadband** - the energy is distributed equally across all frequency bands. E.g.low-frequency sound of a jet plane
- **Narrowband** - most energy is concentrated around specific frequencies. E.g.caused by rotating machinery.

Based on the origination of noise, noise classification is performed in two broad categories i.e. external noise and internal noise. The classification of noise is well documented and referenced from an online source [10].
External noise has random source of origination and generally cannot be controlled. In most of cases, the shifting of receiver is only solution else than passive noise control methods (section 1.4.1). Internal noise while on second note is computed by mathematical approaches for significant reduction. As the noise filtering of signals is possible, the knowledge of noise characteristics is essentials.

1.3.1 External Noise

1.3.1.1 Atmospheric Noise

Any natural entity occurrence such as lightening, thunderstorms has a supplementary sound wave with them. These waves travel in all directions in uncontrolled and explicit manner. The electrical disturbances are in order of radio waves and resonate with information signals present in channel. The strength of these signals is inversely proportional to their amplitudes in frequency spectrum. Hence, most of atmospheric noise is present in low and medium frequency ranges and little in UHF and VHF bands. Also the range of upper frequency bands lies approximately up to 80 KMS channel size and thus their existence become irrational in this domain (above 30 MHz).
1.3.1.2 Extra-terrestrial Noise

The high temperature of a body generates thermal noise. The sun and stars of near and far distance constantly radiate cosmic energy and thermal noise. This noise is roughly distributed in a uniform manner and inherits information during propagation. The galaxies and virtual point source also contribute to thermal noise.

1.3.1.3 Man-Made Noise

Man-made or industrial noise (for detailed analysis, refer 1.3.3) are originated due to human activities or the machines in operation. Automobiles, aircraft engines, motors, gearboxes, voltage leakage, light and other industrial operations lead to man-made noise. These noises are supplementary discharge of machine operations. The noise is intense and affects humans due to presence of industries near populated areas. The noise exceeds every other noise and falls in range 1 MHz to 600 MHz.

1.3.2 Internal Noise

1.3.2.1 Thermal Noise

Free electrons and ions due to vibration under molecular forces generate mechanical waves that are generally termed as noise. The vibration index of ions and electrons are subjected to thermal energy present in molecule. The energy from each ion unites and generates a significant value. For example, a resistance placed in circuit receives external forces. The thermal noise in small and medium circuits generally does not attain frequency susceptible to human ears, but in larger systems one can witness the noise.

1.3.2.2 Shot Noise

Shot noise can be best understood in terms of shock noise. The noise is significant and appears in sudden manner. This noise is not continuous in nature and possesses discreet waveform. The main prospect of this noise is uncertainty in projection of electrons and their cycle in a certain system. The resistance blocks the electrons to its maximum capacity and if this limit is high due to unknown reason, more electrons will collaborate at a place. The incoming of electrons from other end will create a push and after a certain resistance, the sudden discharge of electrons will create shot noise.
1.3.2.3 Transit Noise

The noise of transistors is termed as transit time noise. The transit time at higher frequencies are magnitudes and impact the system. At lower frequencies this phenomenon is irrelevant as the magnitudes do not affect the process. The random noise is generated at higher frequencies within device and directly proportional to operational frequency.

1.3.2.4 Miscellaneous Internal Noise

Miscellaneous noise is defined as one may or may not have an identification of internal noise. This noise generally exists in the electronic circuits and their magnitude is proportional to emitter current.

**1.3.3 Industrial Noise**

Industrial noise is the category that defines the worst scenario of noise. The name ‘industrial’ resembles the example as the heavy machinery produces significant sounds that affects the human health in direct or indirect manner. The industrial noise due to their volume of noise creates a noisy channel for speech signals. In other terms, if a speech filtering system is sustainable in industrial noise domain, it is assumed that same system have economical performances in any other environment. Also, due to this the industrial noises are commonly considered as the benchmark in performing the tests of any speech depression system.

1.3.3.1 Periodic Noise

Period indicates a repetition of pattern in architecture. A complex periodic signal is a repetition of non-sinusoidal pattern in complete wave or a wave with multiple frequency components (figure 1.6). Figure 1.6(a) represents the single frequency component and (b) is complex periodic waveform in amplitude spectrum.
Figure 1.6 (a): Periodic sinusoidal Waveform (b) complex periodic sinusoidal waveform

The periodic sounds have specific properties such as prediction via simple analysis of time domain signal. The complex periodic signals possess a harmonic spectrum with energy at whole number multiples of fundamental frequency.

1.3.3.2 Aperiodic Noise

Aperiodic sound does not reciprocate their pattern over the time (figure 1.7). For example in speech signals, the hiss and pop sound are produced by articulatory release of stop components. Aperiodic noises are classified in continuous and transient forms. Though the
signal possesses smooth edges, the duration of signal indicates the difference between transients and continuous aperiodic signals. The aperiodic signals lack periodicity in their waveform unlike the periodic noise that tends to repeat itself over the period of time.

![Figure 1.7: An aperiodic noise waveform](image)

The aperiodic noises are complex in nature with higher energy than the single frequency component. The signals are dense or continuous spectrum that contrasts the periodic signals of same category. The aperiodic signal does not follow ‘picket fence’ appearance and energy resonates freely in time space domain.

1.4 Noise Cancellation

To minimize the presence of noisy signals, active noise cancellation [11] method is well researched. This method introduces an ‘anti-noise’ signal through an independent source. These sources are powered by a signal processing algorithm required for cancellation of signals. In this research, algorithms to minimize the noise will be studied and results will be compared to locate their efficiency.

1.4.1 Noise Reduction

The consequences of exposing people to noise from various sources may vary from short term effects such as sleep disturbance to long term effects such as permanent hearing loss [12]. To reduce the noise reaching to our ear from source involves various methods which can be categorized into:
Passive Noise Control

Passive Noise control is a method in which the noise from the source is not allowed to reach the ear of the person. This is done by blocking the path of the noise using absorbing materials or by reflecting the noise in some other direction. Thermocol or polystyrene, clothes and wood are some examples of the materials which absorb the noise and reduce the adverse effects occurring from it.

Active Noise Control

Active Noise Control is a very effective electronic method to reduce the effect of the noise in an environment. In a general view point, active control is defined as, “A technique for suppressing unwanted disturbances by the introduction of controlled secondary sources such that their outputs interfere destructively with the incident primary disturbance”. It is basically generation of anti-noise, equal in magnitude and opposite in phase with the noise. The anti-noise and the noise are destructively interfered to remove the effects of noise from the path of the noise.

Though passive and active noise cancellation could be implemented independent of each other, the algorithms can be simultaneously processed for best results.

Speech and Noise Separation

Speech and noise separation is an efficient approach that for noise cancellation in case when sound to noise ratio is too poor to enforce active noise methods. Speech signal filtering follows classification approach and input signal is segmented into speech and noise signals. The speech signal then is input for noise cancellation methods. In frequency domain, the noise speckles have distinguished properties and appropriate algorithms are scheduled to eliminate them. The considerations of periodic and aperiodic type of noise are further estimated in following sections.

1.4.1 Adaptive Noise Cancellation

Adaptive Noise Cancellation (ANC) method creates the signal through electro-acoustic process that counters with noise signal and nullify its effect. The method of cancellation is termed as destructive interference. The signals are of opposite phase and on addition create a
plane wave of zero amplitude and phase. However in real testing, it is seen that accessing the knowledge of phase for noise signal is not ideal process. Hence, the overall result of technique leaves the traces of noise and the amount of trace is directly related to efficiency of ANC.

![Generated by ANC](image1.png)

![Unwanted noise](image2.png)

**Figure 1.8: Signal cancellation of two waves 180° out of phase**

1.4.2 Speech Signal Separation

The failure of adaptive methods in high MSE and low SNR were compensated by separating noise and speech signals. The statistical properties of signals (consider noise as signal) are exploited to differentiate noise and signal. The speech is processed to discard the traces of noise remaining in signal. The speech signal is decomposed into wavelets and soft thresholding filters the signal. The output of soft thresholding is subjected to number of wavelets and the value of thresholding selected as reference. These values are generally manually assigned but the performance can be enhanced, if automatic selection is done keeping the MSE low and high SNR.

1.4.3 Periodic Noise Separation

Periodic signals are the special case of noise and are generally overlooked. The periodic noise separation has limited literature and no system proposes complete denoising for periodic and aperiodic noises. The speech and periodic noise signals are separated through ICA filtering and further denoising mechanism as in aperiodic noise is followed.
1.5 Objective of Research Work

This research is the study of industrial noise of both periodic and aperiodic type. The industrial noise is most intense man made noise with severe impact on signals and human beings. Aperiodic noise term here has no individual significance, but used to derive the attention towards periodic noise. The researches till now had been intentionally or unintentionally concentrated on aperiodic noise. Following objectives are studied in this research:

- Creating a virtual platform to test mathematical algorithms for noise elimination. The virtual platform considers the hypothetical conditions of single noise and speech signal. In speech separation method, authenticated speech components are selected using the automated method. In practice, the components are optimized with objective function and components with minimum square error and high SNR are filtered. This method eradicated any possibility of vague components presence in speech signals even in heavy noise conditions.
- Testing the previous methods (one algorithm) of signal filtering in terms of standard evaluation parameters. Instead of acquiring the results of respective researchers, the models are created in this thesis to mitigate any possibility of functional difference in test scenarios. This approach provides limitations of existing methods and look into their flaws for better modeling.
- Combining one or more well demonstrated techniques of signal denoising to test their performance in speech signals. The thesis in all its experiments considers both noise types for evaluation.
- Optimization of techniques is then proposed as the overhead calculation in attempt to obtain the better response of evaluation techniques. The Particle Swarm Optimization technique has optimal performance in different fields and hence used for speech signal filtering.

In this research, the experiments are conducted to cope with both types of noises simultaneously, and possible solutions are entertained.
1.6 Methodology

The research starts with the introduction of various terms that would be required to progress the experimental work. The speech and noise signals are studied in the form of their nature and classification of noise is made as required. The literature survey of existing methods for signal modeling and noise estimation provides the necessary steps and architecture. By knowledge of system modeling and algorithms on virtual platform the adaptive algorithms are structured to mitigate aperiodic and periodic noises from speech signals. The results will be evaluated to measure the performance and ascertain their limitations. The limitations of adaptive methods are further considered as research objective and another model is prepared to overcome the flaws. The EMD and DWT methods are used as the solution of previous problem and both performed well. To gain the additional rise in PSNR and lower MSE, the model is optimized using PSO technique. Instead of good performance, the inability of EMD to differentiate the speech signals with periodic noise cannot be overlooked. To compensate this damage, EMD is replaced with ICA as the solution. The results sustain the performance of proposed method and optimization of PSO discards any possibility of further modification.

1.7 Outline of Thesis

Chapter 1- The chapter provides an introduction to the speech signals and noise that affects these signals. The research is commenced with a motivation and a problem domain to solve. The chapter also provides the classification for noise types and from the information about industrial noise, it is selected for computation.

Chapter 2- This chapter is the literature survey of previous methods on this topic. The chapter first classifies the types of noise cancellation techniques and approaches from various authors for noise cancellation in respective domains.

Chapter 3- This chapter is the study of simple techniques to cope with surrounding noise in speech signals. Three methods, LMS, NLMS and wiener filters are studied for their performance.

Chapter 4- This chapter studies the claims of EMD as the suitable technique for noise cancellation. In this chapter, additional supported methods are implemented to enhance the performance and obtain the best available results.
Chapter 5- The chapter studies the ICA as the solution for periodic noise problem domain. The ICA is supported with DWT and further optimized to gain the expected results. In addition the comparative analysis of all the techniques to ascertain the performance of all methods and evaluate each technique in comparison with others.

Chapter 6- The research ends with conclusion and future scope