BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY

Acoustic Echo and Noise Cancellation Schemes Using Time and Frequency Domain Adaptive Techniques.

by

Upal Mahbub

MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING

Department of Electrical and Electronic Engineering
BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY

November 2011

The thesis entitled **"Acoustic Echo and Noise Cancellation Schemes Using Time and Frequency Domain Adaptive Techniques"** submitted by Upal Mahbub, Student No.: 1009062047, Session: October, 2009 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING on November 28, 2011.

BOARD OF EXAMINERS

1. ____

(Dr. Shaikh Anowarul Fattah) **Chairman**
Associate Professor (Supervisor)
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka - 1000, Bangladesh.

2. ____

(Dr. Md. Saifur Rahman) **Member**
Professor and Head (Ex-officio)
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka - 1000, Bangladesh.

3. ____

(Dr. Mohammed Imamul Hassan Bhuiyan) **Member**
Associate Professor
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka - 1000, Bangladesh.

4. ____

(Dr. Farruk Ahmed) **Member**
Professor (External)
School of Engineering and Computer Science (SECS)
Independent University, Bangladesh (IUB)
Block-B, Bashundhara R/A, Dhaka-1229.

CANDIDATE'S DECLARATION

I, do, hereby declare that neither this thesis nor any part of it has been submitted elsewhere for the award of any degree or diploma.

Signature of the Candidate

Upal Mahbub

To my parents.

Contents

Acknowledgments

I would like to express my sincere gratitude and profound indebtedness to my supervisor Dr. Shaikh Anowarul Fattah for his guidance, encouragement, constructive suggestions, and support during the span of this research. I also want to thank him for spending so many hours with me in exploring new areas of research and new ideas and improving the writing of this dissertation.

I would also like to thank the head of the department of Electrical and Electronic Engineering for allowing me to use the lab facilities, which contributed greatly in completing the work in time. I wish to express note of thanks to Dr. Celia Shahnaz, for providing inspiration and thoughtful comments. I express appreciation to my colleague Mr. Hafiz Imtiaz and my friend Mr. Partha Pratim Acharjee for their suggestions, support and friendship.

Special note of thanks goes to my sister Ms. Upoma Mahbub and my wife Ms. Tasnuva Chowdhury for their continuous moral support and friendly cooperation. Finally, and most importantly, I wish to thank my parents, who raised me, supported me and taught me with utmost patient and love.

Abstract

Acoustic echo occurs in real life environment when speech signal coming out from a loudspeaker is delayed, attenuated and reflected back to the source microphone. Most communication systems are prone to acoustic echo which can severely degrade the quality and intelligibility of the signals transferred through the communication channels. In conventional acoustic echo cancellation (AEC) methods gradient based adaptive filter algorithms, such as least mean squares (LMS) and normalized LMS are employed where an error function is minimized to obtain the optimal filter coefficients corresponding to the acoustic echo path. The main problem of these methods is the necessity of the dual channels, one for the reference signal and the other for the echo corrupted signal. However in many practical applications only one channel is available, such as a conference hall environment with single microphone and a loudspeaker. Due to the unavailability of separate reference signal in single channel scenario, the task of echo cancellation becomes extremely difficult and is attempted by a few researchers. In this thesis, first a single channel echo cancellation scheme is developed based on the gradient based LMS adaptive filter algorithm, where, unlike conventional dual channel schemes, a delayed version of an estimated echo cancelled signal is utilized as a reference signal. In the proposed formulation, the effect of flat delay, i.e. the time required to produce an echo, is incorporated with a view to reduce the number of unknown parameters of the acoustic echo path, which offers a faster convergence. Moreover, based on energy and cross-correlation coefficients of the reference and current frames, a multi-step stopping criteria is developed, which can efficiently control the updating procedure of the proposed LMS adaptive filter. Extensive experimentation is carried out on real life speech signals corrupted by echoes using the proposed single channel LMS algorithm with and without the multi-step update constraints. It is found that the performance of former one, the controlled LMS algorithm, is far better than that of the later one in terms of (a) the average echo return loss enhancement (ERLE) in dB and (b) the difference between input- and output-signal to distortion ratio (SDR) in dB. In real life applications, inclusion of noise with the speech signals is obvious in most of the cases, which makes the task of single channel echo cancellation even more difficult. In view of handling the challenging task of cancelling the echo in the presence of noise, a two step algorithm is developed where a spectral subtraction based noise reduction scheme is introduced after the single channel echo cancellation. It is shown that even under severe noisy conditions in different acoustic environments the proposed two-step single channel acoustic echo and noise cancellation (AENC) method can significantly reduce the effect of both echo and noise.

As an alternate to the gradient based approaches, the problem of dual channel echo and/or noise cancellation can also be realized based on some optimization algorithm driven adaptive filters. However, undoubtedly the problem would be very difficult for the single channel scenario which is the case under consideration. Thus, the single channel AEC problem is formulated as an optimization problem introducing the particle swarm optimization (PSO) algorithm, which offers a quick convergence to the desired solution. For proper operation of the PSO algorithm, a frame by frame processing is required for which the overlap-add method is adopted. In order to estimate the unknown coefficients of the acoustic echo path, the PSO based algorithm is formulated both in the time and frequency domain separately and it is found that the frequency domain approach performs better in comparison to the time domain approach. The performance of the proposed PSO algorithms are also investigated for different controlling parameters, namely number of particles, maximum particle velocity etc. The PSO based algorithm is also extended for the complicated case of adaptive echo and noise cancellation. From detailed simulations it is found that the performance of the proposed PSO based AENC algorithm outperforms that of the proposed gradient based algorithm under different noisy conditions at various acoustic environments.

List of Tables

List of Figures

List of abbreviations

AEC	Adaptive Echo Cancellation
ANC	Adaptive Noise Cancellation
LMS	Least Mean Square
DFT	Discrete Fourier Transform
NLMS	Normalized Least Mean Square
SDR	Signal to Distortion Ratio
SNR	Signal to Noise Ratio
INTEL	INTelligibility Enhancement by Liftering
RLS	Recursive Least Mean Square
PSO	Particle Swarm Optimization
PSO-TD	Proposed PSO based method in Time Domain
PSO-FD	Proposed PSO based method in Frequency Domain
SPSO	Standard Particle Swarm Optimization
ERLE	Echo Return Loss Enhancement
AENC	Acoustic Echo and Noise Cancellation
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
PA	Public Address
SS	Spectral Subtraction
STSA	Short-Time Spectral Amplitude
ML	Maximum-Likelihood
MAP	Maximum *A Posteriori*

Chapter 1

Introduction

Echo is the repetition of a waveform due to reflection from points where the characteristics of the medium through which the wave propagates change. Acoustic echo results from a feedback path set up between the speaker and the microphone in a mobile phone, hands-free phone, teleconference or hearing aid system. Acoustic echo can result from a combination of direct acoustic coupling and multi-path effect, where the sound wave is reflected from various surfaces, such as wall, ceiling and floor of a room and then picked up by the microphone. If the time delay is not too long, the acoustic echo may be perceived as a soft reverberation, which up to a certain level, in some applications like concert halls and church halls, may be acceptable. However, in its worst case, acoustic feedback can result in howling if a significant proportion of the sound energy transmitted by the loudspeaker is received back at the microphone and circulated in the feedback loop. Incorporation of an acoustic echo canceller is thus a necessity for designing a communication channel or establishing a conference room environment. Moreover, in real life scenarios, different types of noise, for example machine noise, fan noise etc., are present everywhere making the problem of acoustic echo cancellation more difficult. Therefore, designing a generalized integrated acoustic echo and noise canceller for the enhancement of speech signals has become challenging task for sound engineers for the past few decades.

1.1 Need for Single-Channel Echo Cancellation

With the rapid growth of modern technology in recent decades, the whole dimension of communications has been changed. Today people are more interested in hands-free communication which allows a person to have both hands free and to move freely in the room during a conversation. However, the presence of a large acoustic coupling between the loudspeaker and microphone may produce a loud echo resulting disturbance in conversation. If the echo is produced from multiple surfaces it may cause reverberation, which is, in effect, a multiplicity of echoes whose speed of repetition is too quick for them to be perceived as separate from one another. In real life scenarios, such as a talk in a large conference hall or in the public address system of mosques, churches and trade fair, reverberation is a very common phenomenon, which may degrade the quality of the speech signal to a great extent leading to complete loss of intelligibility. An echo corrupted public address system or conference room acoustic system could arise public annoyance and produce severe sound pollution. Furthermore, the acoustic systems in these scenarios could become instable, which would produce a loud howling noise.

Apart from these, in this new age of global communications, wireless phones are regarded as essential communication tools and have a direct impact on people's day-to-day personal and business communications. As new network infrastructures are implemented and competition among wireless carriers increases, digital wireless subscribers are becoming more cautious about the service and voice quality they receive from network providers. Subscriber's demand for enhanced voice quality over wireless networks has driven a new and key technology termed echo cancellation, which can provide a better voice quality across a wireless network. Now-a-days, speech quality over the phone is treated as one of the most important as criteria for assessing the overall quality of a wireless network. Regardless of whether the subscriber's opinion is subjective or not, it is the key to maintain subscriber's loyalty. For this reason, the effective removal of hybrid and acoustic echoes, which are inherent within the telecommunication network infrastructure, is the key to maintaining and improving the perceived voice quality of a call. The search for improved voice quality has led to intensive research in the area of echo cancellation. Such research is conducted with the aim of providing solutions that can reduce background noise and remove hybrid and acoustic echoes before any transcoder processing occurs. By employing echo cancellation technology, the quality of speech can be improved significantly.

Though the problem of automatic removal of acoustic echo in dual telephone line or wireless communication links has been investigated by several researchers, the more critical problem of single channel echo cancellation in large room environment has rarely been attempted. Nevertheless, the acoustics engineers generally employ different manual measures to suppress the room echoes, such as implementing sound absorbers in room walls, reducing reflecting planes in a room or using manually controlled filters in the path of sound propagation. Hence, development of an automatic environment adaptive acoustic echo canceller for the purpose of single channel echo cancellation us still in great demand.

1.2 Acoustic Echo and the Room

Delayed and attenuated version of an original sound being reflected back to the source is termed as echo. In an acoustic room environment, sounds may get reflected from the walls, ceiling, floor and other neighboring objects and result in multi-path echo as well as multiple harmonics of echoes, which are transmitted back to the listeners unless otherwise eliminated. If a direct coupling exists between the source and the listener, the listener will listen the direct sound first, followed by reflections of the nearby surfaces, the later being called early reflections. The acoustic echo phenomenon is illustrated in Fig. 1.1 by a schematic diagram. If a reflected sound arrives after a very short duration of the direct sound, it is considered as a spectral distortion or reverberation. However, when the leading edge of the reflected sound arrives a few tens of milliseconds after the direct sound, it is heard as a distinct echo. In communication systems, acoustic echo arises when sound from a loudspeaker is picked up by the microphone in the same room, for example, sound from the earpiece of a telephone handset may be picked up by the microphone in the very same handset. Examples of acoustic echo are found in everyday surroundings such as: a. Hands-free car phone systems b. A standard telephone or cell-phone in speaker-phone or hands-free mode c. Dedicated stand-alone conference phones d.

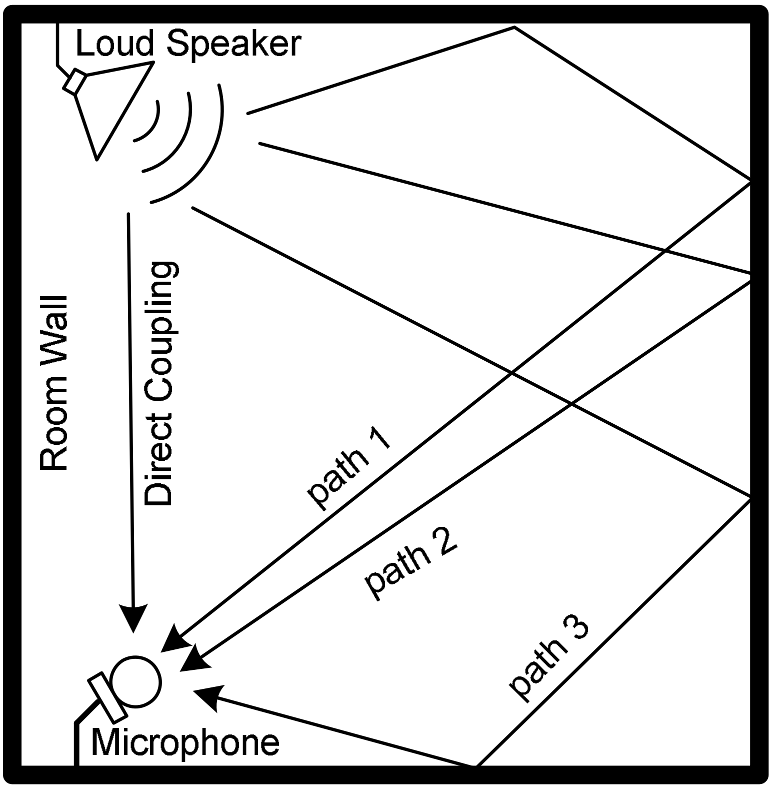

Fig. 1.1: Sources of Acoustic Echo in a Room

Installed room systems which use ceiling speakers and microphones on the table e. Physical coupling (vibrations of the loudspeaker transfer to the microphone via the handset casing).

The difficulties in cancelling acoustic echo stem from the alteration of the original sound by the ambient space. Acoustic echo cancellers greatly enhance the audio quality of a multi-point hands-free communications system. They allow conferences to progress more smoothly and naturally, keep the participants more comfortable, and prevent listener fatigue. Acoustic echo is most noticeable (and annoying) when delay is present in the transmission path. This would happen primarily in long distance circuits, or systems utilizing speech compression (such as in video-conferencing or digital cellular phones). Even though this echo might not be as annoying when there is no delay (such as with short links between conference rooms in the same building or distance learning over fiber-optic cable), room acoustics will still affect the sound and may hamper communication. Also, howling can occur if the microphone is positioned too close to the speaker whether or not there is transmission delay, and is eliminated by most acoustic echo cancellers [1] [2].

The primary beneficiaries of an echo canceller are the people at the far (or remote) end of the transmission path. The near (or local) echo canceller prevents the echo of the remote peoples voices from being returned (i.e. echoed) to them through the audio system. People speaking on the same (local) end as the AEC should not notice the AEC if it is doing its job properly. While the people on the far end receive the benefit of better audio quality, it also enables the conversation to flow more smoothly, benefiting both parties.

An AEC solution that was designed to operate in an office may not work properly in a conference room. If an echo canceller were compliant in one room and not another, it would most likely be due to a tail length that was too short for the second room. The tail length of an AEC is the length of time over which it can cancel echoes. The tail length of the echo canceller should meet the requirements of the room it is to be operated in. The flat delay, which is the time taken by the direct sound to propagate from the loudspeaker to the microphone, is included in the tail length. This is directly related to the reverberation time of the room. As the room reverberation time increases, a longer tail length will be needed in that room. If the reverberation time is much longer than the tail length, a significant amount of the echo will remain audible. However, excess tail length will not improve or degrade the performance of the canceller.

There are two main factors that affect the reverberation time of a room. They are room size, and the materials used to construct the walls and objects in the room. Most sound is absorbed when it strikes walls or other surfaces. If materials are used that absorb sound well (such as carpet, curtains, or acoustic tile), the reverberation will die out more quickly than if the room contains mostly reflective materials (hard wood, glass, or plaster). If a room is small, the sound waves will bounce off the walls more frequently, and will be absorbed more quickly [3].

Howling rejection is important in cases where both parties are using hands-free communications systems. In these types of systems, it is very easy for the open microphones and loudspeakers to produce acoustic feedback, resulting in squealing tones (much like the feedback from a microphone in an auditorium). This obviously prevents any useful conversation from taking place. The most common way to avoid this problem is to implement howling rejection, typically done by shifting the

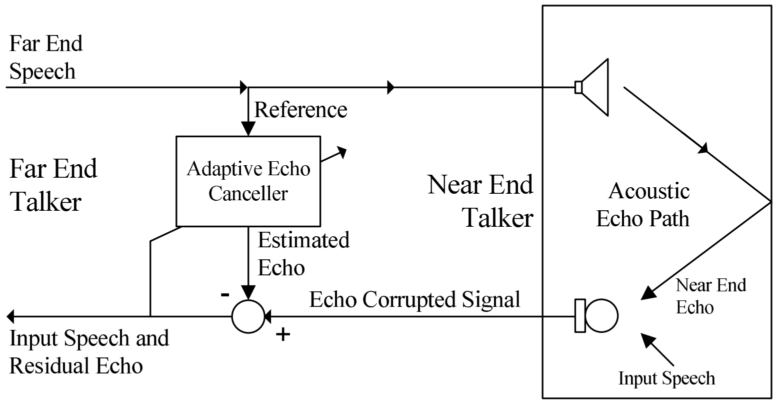

Fig. 1.2: Acoustic echo cancellation in a communication channel using adaptive filters

frequency of the signal as it goes through the canceller.

Acoustic echo cancellation (AEC) has widespread applications in many real-life situations, such as cellular phone communication, hands-free phones, teleconferences, hearing aid systems, and in sound system for large conference halls, churches etc..

1.3 Traditional Echo Cancellers

Echo suppressors, earphones and directional microphones have been conventionally used in order to combat the problem of acoustic echo. However, these instruments generally place physical restrictions on the talkers [4]. In order to mitigate this problem, now-a-days, adaptive filter algorithms are commonly employed in the acoustic system for echo cancellation.

In communication system, an echo canceller is basically a device that detects and removes the echo of the signal from the far end after it has echoed on the near ends equipment. Despite the fact that adaptive echo cancellation was conceived by the mid-1960s, practical implementation had to wait for the very large scale integration (VLSI) systems.

Most of the AEC methods are based on estimating room impulse response using different adaptive filter algorithms. In Fig. 1.2 a general scenario of adaptive acoustic echo cancellation is depicted. The adaptive filter at the near end speaker saves some recent samples of the far end speech as reference and tries to produce an exact replica of the near end room response which is the acoustic echo path for the far end signal. The difference between the echo corrupted near end speech signal and the estimated echo produced by the adaptive filter is the error which consists of the near end input speech along with some residual echo. The task of the adaptive filter is to minimize this error signal by iterative estimation of the echo path, i.e. by adapting the characteristic of the room response at near end. The situation is different in case of closed room echoes or reverberations. There is only one channel that has a single microphone in one end and a loud speaker in the other. Example of such systems are conference hall sound systems, PA systems etc.. Modeling of these single channel AEC problems with adaptive filters is difficult as there is no channel for obtaining reference signals.

The most popular adaptive filter algorithms are the Least Mean Square (LMS) algorithm, the Normalized Least Mean Square (NLMS) algorithm and Recursive Least Square (RLS)algorithm [5]. Among these algorithms, RLS offers fastest convergence. However, LMS based algorithms are more commonly used because of their less computational burden [5]. Recently, proportionate NLMS and its variants are employed for echo cancellation in sparse channels [6] [7].

1.4 Acoustic Echo and Noise

Usually, acoustic environments are surrounded with environmental noise. Therefore, most effective echo cancellers should take into account the effects of noise too. For adaptive filters, noise cancellation is a different task then echo cancellation, thus only a single adaptive filter cannot cancel both. Adaptive signal processing methods are also used for adaptive noise cancellation (ANC) [8]. However, in real life applications, when both echo and noise are present, an AEC or an ANC alone will not be able to enhance the speech signal. Thus, an adaptive integrated echo and noise canceller should be used. The echo cancellation part is similar to the basic AEC, however the noise reduction task is not easily solved since a ”noise only” reference signal which is sufficiently correlated to the noise within the microphone signal in a typical acoustic environment can not be obtained. Besides this principal restriction, the solutions using single microphone are favorable in most commercial mobile products. Thus, although the multi-microphone systems might yield better performance, the single microphone spectral weighting techniques (e.g., Wiener filtering, spectral subtraction, minimum mean square error (MMSE), etc.) are often preferred [9]. These methods, however, have well-known disadvantages such as limited performance at low signal-to-noise ratio (SNR) environments and artificial sounding residual called musical noise. The latter problem can be somewhat alleviated by applying spectral floor [10].

An adaptive echo and noise cancellation (AENC) scheme is proposed in [11] where a sub-band noise cancellation is incorporated.

1.5 Single Channel AENC

A major problem of all the adaptive filter algorithms, as stated previously, is that, they require two channels, one for receiving echo corrupted signal and the other for the reference signal. For example, in the AEC systems employed in communication channels or multi-path systems, the near end signal, which is available at hand, is fed to the adaptive filter as a reference. It is to be noted that the problem of AEC would be much difficult if, instead of two channels, only a single channel is available, which may arise in many other applications, such as acoustic echo in conference room environment. Presence of environmental noise along with the echo causes degradation of performance of the adaptive echo cancellation algorithms, making the problem very much challenging and rarely addressed by researchers. Single channel noise suppression problem, although difficult, has been dealt by many researchers over the last decades [12]. However, the problem of single channel echo and noise cancellation (AENC) is yet to be addressed.

1.6 Optimization methods for echo cancellation

LMS based echo cancellers lack the flexibility of controlling the convergence rate, number of iterations, and tolerance consistency. Moreover, adaptive algorithms generally do not allow modifying the possible ranges of the filter coefficients. In this regard, use of an optimization algorithm could provide much more flexibility.

Conventional optimization algorithms, for example genetic algorithm, Tabu search, and simulated annealing are difficult to implement and they exhibit slow convergence rate [13]. However, the particle swarm optimization algorithm (PSO), proposed by Kennedy and Eberhart in 1995 [14], provides ease of implementation and faster convergence rate [15]. The PSO is a stochastic search algorithm and unlike the gradient-based algorithms, it can converge to solutions even for non-differentiable systems if properly parameterized and constrained [16].

The implicit rules adhered to by the members of bird flocks and fish schools, that enable them to move synchronized, without colliding, resulting in an amazing choreography, was studied and simulated by several scientists [17] [18]. In simulations, the movement of the flock was an outcome of the individuals (birds, fishes etc.) efforts to maintain an optimum distance from their neighboring individuals [16]. The social behavior of animals, and in some cases of humans, is governed by similar rules [19]. However, human social behavior is more complex than a flocks movement. Besides physical motion, humans adjust their beliefs, moving, thus, in a belief space. Although two persons cannot occupy the same space of their physical environment, they can have the same beliefs, occupying the same position in the belief space, without collision. This abstractness in human social behavior is intriguing and has constituted the motivation for developing simulations of it. There is a general belief, and numerous examples coming from nature enforce the view, that social sharing of information among the individuals of a population, may provide an evolutionary advantage. This was the core idea behind the development of PSO [16].

PSO is similar to EC techniques in that, a population of potential solutions to the problem under consideration, is used to probe the search space [20]. However, in PSO, each individual of the population has an adaptable velocity (position change), according to which it moves in the search space. Moreover, each individual has a memory, remembering the best position of the search space it has ever visited [16]. Thus, its movement is an aggregated acceleration towards its best previously visited position and towards the best individual of a topological neighborhood. Since the *acceleration* term was mainly used for particle systems in Particle Physics [21], the pioneers of this technique decided to use the term particle for each individual, and the name swarm for the population, thus, coming up with the name Particle Swarm for their algorithm [14].

Recently, the PSO is recieving much attention in areas of power system, computer architecture, and control system [22], [23]. PSO was originally introduced by Eberhent and Kennedy in 1995 [16]. This method rooted on the notion of swarm intelligence of insects, birds, etc. Standard PSO (SPSO) has already been applied successfully in many applications such as training of an artificial neural networks, power flow scheduling [24] , generating interactive and improvised music [25], and assigning tasks in distributed computing systems [26].

It can be expected that an integrated echo and noise canceller by utilizing an optimization algorithm driven adaptive filter and a single channel noise suppressor may certainly suggest an effective solution to the age-old problem of AENC.

1.7 Literature Review

Acoustic echo may lead to total unintelligibility of the near-end speaker in hands free communication systems. Acoustic echo cancellation is the most important and well-known technique to cancel the acoustic echo [27]. This technique enables one to conveniently use a hands-free device while maintaining high user satisfaction in terms of low speech distortion, high speech intelligibility, and acoustic echo attenuation. The acoustic echo cancellation problem is usually solved by using an adaptive filter in parallel to the acoustic echo path [27] - [30]. The adaptive filter is used to generate a signal that is a replica of the acoustic echo signal. An estimate of the near-end speech signal is then obtained by subtracting the estimated acoustic echo signal, i.e., the output of the adaptive filter, from the microphone signal. Sophisticated control mechanisms have been proposed for fast and robust adaptation of the adaptive filter coefficients in realistic acoustic environments [30], [31]. In practice, there is always residual echo, i.e., echo that is not suppressed by the echo cancellation system. The residual echo results from 1) the deficient length of the adaptive filter, 2) the mismatch between the true and the estimated echo path, and 3) nonlinear signal components.

It is widely accepted that echo cancellers alone do not provide sufficient echo attenuation [29] - [32]. Approaches of combining acoustic echo cancellation and residual echo reduction have been considered to achieve sufficient quality of the transmitted speech [33] [34]. The realization of such a combined system is, however, a challenging task. The difficulties in canceling acoustic echo are caused by the high computational complexity and some influences such as background noise, near-end speech, and variations of the acoustic environment which disturb the adaptation of the canceler [35]. In practice, especially in mobile hands-free applications where all these factors play a significant role, the residual echo remains at the output of an adaptive echo canceler due to the misadjustment of the adaptive algorithm and the constraint of finite filter length [36]. Turbin et al. compared three postfiltering techniques to reduce the residual echo and concluded that the spectral subtraction technique, which is commonly used for noise suppression, was the most efficient [9]. This single microphone spectral weighting technique, however, has well-known disadvantages such as limited performance at low signal-to-noise ratio (SNR) environments and artificial sounding residual called musical noise. The latter problem can be somewhat alleviated by applying spectral floor [10]. In a reverberant environment, there can be a large amount of so-called late residual echo due the deficient length of the adaptive filter. In [32], Enzner proposed a recursive estimator for the short-term power spectral density (PSD) of the late residual echo signal using an estimate of the reverberation time of the room. The reverberation time was estimated directly from the estimated echo path. The late residual echo was suppressed by a spectral enhancement technique using the estimated short-term PSD of the late residual echo signal.

In some applications, like hands-free terminal devices, environmental noise reduction becomes necessary due to the relatively large distance between the microphone and the speaker. The first attempts to develop a combined echo and noise reduction system can be attributed to Grenier et al. [37], [38] and to Yasukawa [11]. Both employ more than one microphone. A survey of these systems can be found in [30] and [39]. Beaugeant et al. [40] used a single Wiener filter to simultaneously suppress the echo and noise. In addition, psychoacoustic properties were considered in order to improve the quality of the near-end speech signal. They concluded that such an approach is only suitable if the noise power is sufficiently low. In [41], Gustafsson et al. proposed two postfilters for residual echo and noise reduction. The first postfilter was based on the log spectral amplitude estimator [42] and was extended to attenuate multiple interferences. The second postfilter was psychoacoustically motivated. When the hands-free device is used in a noisy reverberant environment, the acoustic path becomes longer and the microphone signal contains reflections of the near-end speech signal as well as noise. Martin and Vary proposed a system for joint acoustic echo cancellation, dereverberation, and noise reduction using two microphones [43]. A similar system was developed by Dörbecker and Ernst in [44]. In both papers, dereverberation was performed by exploiting the coherence between the two microphones as proposed by Allen et al. in [45]. Bloom [46] found that this dereverberation approach had no statistically significant effect on intelligibility, even though the measured average reverberation time and the perceived reverberation time were considerably reduced by the processing.

It should however be noted that most hands-free devices are equipped with a single microphone. A single-microphone approach for dereverberation is the application of complex cepstral filtering of the received signal [47]. Bees et al. [48] demonstrated that this technique is not useful to dereverberate continues reverberant speech due to so-called segmentation errors. They proposed a novel segmentation and weighting technique to improve the accuracy of the cepstrum. Cepstral averaging then allows to identify the acoustic impulse response (AIR). Yegnanarayana and Murthy [49] proposed another single microphone dereverberation technique in which a time-varying weighting function was applied to the linear prediction (LP) residual signal. The weighing function depends on the signal-to-reverberation ratio (SRR) of the reverberant speech signal and was calculated using the characteristics of the reverberant speech in different SRR regions. Unfortunately, these techniques are not accurate enough in a practical situation and do not fit in the framework of the postfilter which is commonly formulated in the frequency domain. Recently, practically feasible single microphone speech dereverberation techniques have emerged. Lebart proposed a single microphone dereverberation method based on spectral subtraction of the spectral variance of the late reverberant signal [50]. The late reverberant spectral variance is estimated using a statistical model of the AIR. This method was extended to multiple microphones by Habets [51]. Recently, Wen et al. presented results obtained from a listening test using the algorithm developed by Habets [52].

These results showed that the algorithm in [51] can significantly increase the subjective speech quality. The methods in [50] and [51] do not require an estimate of the AIR. However, they do require an estimate of the reverberation time of the room which might be difficult to estimate blindly. Furthermore, both methods do not consider any interferences and implicitly assume that the sourcereceiver distance is larger than the so-called critical distance, which is the distance at which the direct path energy is equal to the energy of all reflections. When the sourcereceiver distance is smaller than the critical distance the contribution of the direct path results in overestimation of the late reverberant spectral variance. Since this is the case in many hands-free applications, the latter problems need to be addressed.

Global Optimization (GO) methods can be classified into two main categories: deterministic and probabilistic methods. Most of the deterministic methods involve the application of heuristics, such as modifying the trajectory (trajectory methods) or adding penalties (penalty-based methods), to escape from local minima. On the other hand, probabilistic methods rely on probabilistic judgements to determine whether or not search should depart from the neighborhood of a local minimum [53] - [60]. In contrast with different adaptive stochastic search algorithms, Evolutionary Computation (EC) techniques [61] exploit a set of potential solutions, named population, and detect the optimal problem solution through cooperation and competition among the individuals of the population. These techniques often find optima in complicated optimization problems faster than traditional optimization methods. The most commonly met population-based EC techniques, such as Evolution Strategies (ES) [62] - [70], Genetic Algorithms (GA) [71] [72], Genetic Programming [73] [74], Evolutionary Programming [75] and Artificial Life methods, are inspired from the evolution of nature. The Particle Swarm Optimization (PSO) method is a member of the wide category of Swarm Intelligence methods [76], for solving GO problems. It was originally proposed by J. Kennedy as a simulation of social behavior, and it was initially introduced as an optimization method in 1995 [16] [14]. PSO is related with Artificial Life, and specifically to swarming theories, and also with EC, especially ES and GA. PSO can be easily implemented and it is computationally inexpensive, since its memory and CPU speed requirements are low [16]. Moreover, it does not require gradient information of the objective function under consideration, but only its values, and it uses only primitive mathematical operators. PSO has been proved to be an efficient method for many GO problems and in some cases it does not suffer the difficulties encountered by other EC techniques [16].

1.8 Objective

The objectives of this research are:

1. To develop single channel acoustic echo cancellation scheme using gradient based adaptive filter algorithm.

2. To formulate the problem of single channel acoustic echo cancellation as an optimization task and thereby use the Particle Swarm Optimization (PSO) algorithm in time and frequency domain.

3. To incorporate a spectral subtraction based noise reduction method for single channel acoustic echo and noise suppression using gradient based adaptive filter algorithms.

4. To develop a scheme for single channel acoustic echo and noise cancellation using the PSO algorithm in time and frequency domain.

The outcome of this thesis is a single channel combined echo plus noise cancellation scheme based on the time and frequency domain PSO algorithm, which can be used to enhance speech signals obtained from conference room environments, PA systems or data communication channels with high degree of accuracy and efficiency yet low computational burden.

1.9 Organization of the Thesis

First, the process of single channel acoustic echo generation in a large conference hall environment is explained and then the problem of acoustic echo cancellation (AEC) is formulated. A comparison between the single channel and a dual channel echo cancellation scheme is also discussed. Afterwards, a novel AEC scheme based on the conventional least mean squares (LMS) algorithm is proposed for single channel echo suppression. A detailed analysis on the iterative update procedure of the proposed algorithm is carried out in order to develop a better adaptation criterion. A set of update constraints is introduced resulting a modified scheme based on the LMS algorithm for obtaining better echo cancellation performance. It is to be noted that extensive experimentations is performed to evaluate the performance of the proposed algorithms throughout the thesis on several speech frames taken from the most commonly used standard TIMIT speech dataset [77].

In chapter 3, considering the problem of single channel AEC as an optimization problem, the particle swarm optimization (PSO) algorithm is introduced. The PSO algorithm is employed for both time and frequency domain echo cancellation. In the simulation result section of this chapter a comparison of echo cancellation performance between the proposed time domain PSO (PSO-TD), frequency domain PSO (PSO-FD) and the previously proposed modified LMS method is compared in terms of the signal to distortion ratio (SDR) difference between input and output speech and the average echo return loss enhancement (ERLE). Also, variation in echo cancellation performance by varying some important parameters of the PSO algorithm is also observed and analyzed.

The problem of single channel AEC becomes more challenging in chapter 4 when noise is introduced at the input speech signal. A modified LMS algorithm is proposed in this chapter for developing a single channel integrated acoustic echo and noise cancellation (AENC) scheme. A spectral subtraction based noise cancellation method is employed for single channel noise cancellation.

The single channel AENC problem is again addressed in chapter 5 and this time the proposed solutions are based on time and frequency domain representations of the PSO algorithm, denoted as PSO-TD-AENC and PSO-FD-AENC, respectively. A comparison of the performance of the PSO-TD-AENC, PSO-FD-AENC and modified LMS algorithms in cancelling echo at noisy environment is presented at the simulation section.

In the final chapter, chapter 6, the whole scenario of this literature is summarized with some concluding remarks and some ideas of future improvements.

Chapter 2

Single Channel Acoustic Echo Cancellation Based on Adaptive LMS Algorithm

The phenomenon of acoustic echo occurs when the output speech signals from a loudspeaker gets reflected from different surfaces, like ceilings, walls, and floors and then fed back to the microphone. Due to these feedback paths, acoustic echo may originate in several real-life applications, such as cellular phones, hands-free phones, tele-conferences, hearing aid systems, and large conference halls [2]. In its worst case, acoustic echo can cause howling of a significant portion of sound energy [2] [4].

Echo suppressors, earphones and directional microphones have been conventionally used in order to combat the problem of acoustic echo. However, these instruments generally place restrictions on the talkers movement [4]. As an alternate of such hardware based solutions, adaptive filter algorithms are widely being applied for echo cancellation. Among different adaptive filter algorithms, the gradient based least mean square (LMS) algorithm and its modifications, such as normalized least mean square (NLMS) algorithm are well-known for their satisfactory performances and less computational burden [78]. LMS/NLMS are popular for their ease of implementation [5]. Besides these algorithms, the recursive least mean square (RLS) algorithm is well-known for its fast convergence at the expense of computational complexity [5]. A common problem of all these adaptive filter algorithms is that, they require two channels, one for receiving echo corrupted signal and the other for the reference signal. For example, in the adaptive echo cancellation (AEC) systems employed in communication channels or multi-path systems, the near-end signal, which is available at hand, is fed to the adaptive filter as a reference to cancel the far-end echoed signal. It is to be noted that the problem of AEC would be much difficult in some applications where, instead of two channels, only a single channel is available, such as acoustic echo in conference room environment [2]. In this case, a desired reference signal is not available.

A single channel echo cancellation scheme is developed in this chapter, using gradient based adaptive LMS algorithm. In the proposed formulation, unlike conventional adaptive filter algorithms, the effect of flat delay is incorporated by pre-calculating the number of filter coefficients corresponding to the flat delay based on the distance between the speaker and the microphone. This offers advantage of a huge reduction of unknown parameters of the room response. Utilizing the prior information on flat delay, we propose necessary modification in the gradient based adaptive algorithm to achieve a faster convergence [79].

Moreover, based on energy and cross-correlation coefficients of the reference and current frames, we propose to impose a multi-step stopping criteria, which can efficiently control the update sequence of the adaptive filter. In the proposed updating algorithm, nine different critical scenario are taken into consideration depending on the speech properties of the reference frame and the current frame. For example, performance of the AEC when the reference and current frame are voiced-unvoiced, voiced-voiced, voiced-pause, unvoiced-unvoiced, unvoiced-voiced, unvoiced-pause etc. It is shown that the proposed algorithm can successfully handle all these difficult conditions resulting in a high signal to distortion ratio (SDR) and also a high average echo return loss enhancement (ERLE) in dB.

2.1 Dual Channel Vs. Single Channel Acoustic Echo Cancellation

In a dual channel communication, generally two channels carry speech signals of two different speakers. Along with the original speech signal of a particular speaker, an echo signal generated from a reflected speech signal of the other speaker may be mixed up, resulting in a echo-corrupted signal. The task of an AEC scheme, is to reduce the effect of echo from the echo-corrupted signal by using some adaptive algorithm, thereby transmitting a signal which is more close to the original signal.

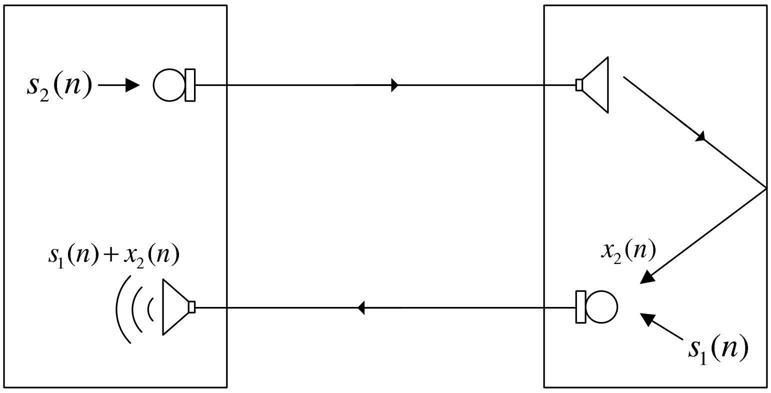

Fig. 2.1: Acoustic echo in dual channel communication systems.

In Fig. 2.1, a typical dual channel communication system is shown, where, $s_1(n)$ and $s_2(n)$ are speech signals generated by two different persons talking at the two ends of a communication link. The signal $s_2(n)$ from person 2, traveling through one of the communication channels, reaches the other end to person 1 when played through a loudspeaker. The output of the loudspeaker, if reflected from the walls, ceiling and floor of a room and fed back to the microphone of person 1, could be transmitted along with the speech signal $s_1(n)$ of person 1. As a result, instead of listening only $s_1(n)$, person 2 will also hear a slightly attenuated form of his own voice $x_2(n)$, known as an echo signal for person 2. In Fig. 2.1, for simplicity, echo generation process in a dual channel communication system is demonstrated only for person 2. A similar scenario is also applicable for the case of person 1.

In Fig. 2.2 a conventional echo cancellation block is introduced to remove the effect of echo from the echo corrupted signal,

$$y_1(n) = s_1(n) + x_2(n). \tag{2.1}$$

For this purpose, most commonly the LMS adaptive filter algorithm is used, where given a reference signal an estimate of the echo part $x_2(n)$ of $y_1(n)$ is generated based on the minimization of an error function $e_1(n)$ defined as,

$$\begin{aligned} e_1(n) &= y_1(n) - \widehat{x}_2(n) && (2.2)\\ &= s_1(n) + x_2(n) - \widehat{x}_2(n), && (2.3) \end{aligned}$$

As mentioned earlier, $x_2(n)$ is an attenuated and delayed version of $s_2(n)$ which,

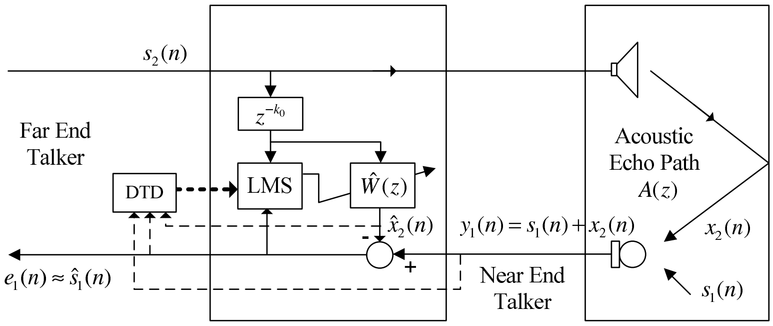

Fig. 2.2: LMS adaptive algorithm for dual channel acoustic echo cancellation in communication systems.

based on linear prediction theory can be expressed as,

$$\begin{aligned} x_2(n) &= \mathbf{a}_n^T \mathbf{s}_2(n-k_0) \qquad (2.4) \\ &= \sum_{k=1}^{p} a_n(k) s_2(n-k_0-k), \qquad (2.5) \end{aligned}$$

where, $\mathbf{s}_2(n-k_0) = [s_2(n-k_0-1), s_2(n-k_0-2), \ldots, s_2(n-k_0-p)]^T$ is a vector of p previous values of s_2 with predefined flat delay k_0 and $\mathbf{a}_n = [a_n(1), a_n(2), \ldots, a_n(p)]^T$ is the vector of the unknown room response coefficients. The number p of unknown attenuation coefficients $a_n(k)$ depends on the characteristics of the room.

Here, the task of an adaptive filter is to produce optimum values of unknown filter coefficients $\widehat{\mathbf{w}}_n$ from given $s_2(n-k_0)$ such that the resulting signal $\widehat{x}_2(n)$ closely matches $x_2(n)$, i.e,

$$\begin{aligned} \widehat{x}_2(n) &= \widehat{\mathbf{w}}_n^T \mathbf{s}_2(n-k_0) \qquad (2.6) \\ &= \sum_{k=1}^{p} \widehat{w}_n(k) s_2(n-k_0-k), \qquad (2.7) \end{aligned}$$

Here, $\widehat{\mathbf{w}}_n = [\widehat{w}_n(1), \widehat{w}_n(2) \ldots \widehat{w}_n(p)]^T$ is the estimated attenuation vector. The value of p also signifies the number of unknown parameters to be estimated from the system.

Under optimum condition, $\widehat{\mathbf{w}}_n = \mathbf{a}_n$. For LMS adaptive filter algorithm, the desired values of $\widehat{\mathbf{w}}_n$ are estimated adaptively by using the following updated equation [5],

$$\widehat{\mathbf{w}}_{n+1} = \widehat{\mathbf{w}}_n + 2\mu e_1(n) \mathbf{s}_2(n-k_0) \qquad (2.8)$$

An extremely important issue of designing adaptive echo cancelers for dual channel is to handle double talk, which occurs when the far-end and near-end talkers

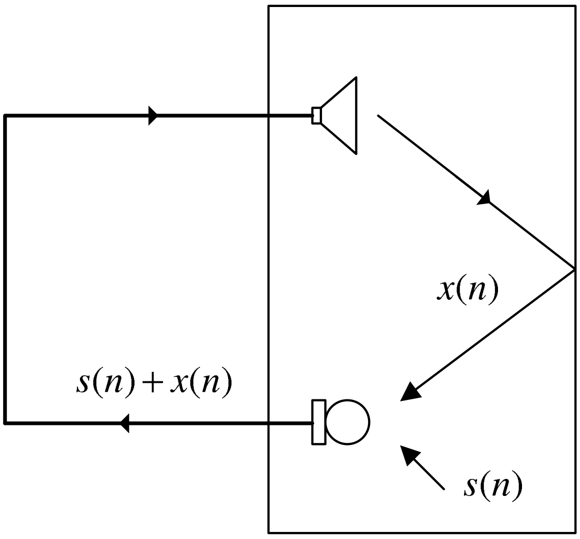

Fig. 2.3: Single channel adaptive acoustic echo cancellation in room environment.

are speaking simultaneously. In this case, the far end signal consists of both echo $x_1(n)$ and far-end speech $s_2(n)$. During the double-talk periods, the error signal $e(n)$ described in (2.3) contains the residual echo and the near-end speech $s_1(n)$. To correctly identify the characteristics of $A(z)$, the near-end signal must originate solely from its input signal from the far end. An effective solution, as shown in figure 2.2, is to detect the occurrence of double talk using a double talk detector (DTD) and then to disable the adaptation of $\widehat{W}(z)$ during the double-talk periods. If the echo path does not change during the double-talk periods, the echo can be canceled by the previously estimated $\widehat{W}(z)$, whose coefficients are fixed during double-talk periods.

The above scenario will be drastically changed for single channel echo environments such as, conference room environment and hearing aid systems. As shown in Fig. 2.3, unlike the dual channel scenario, the speech signal $s(n)$ itself is reflected and fed back to the sole microphone as echo $x(n)$, producing an echo corrupted signal $y(n) = s(n) + x(n)$. Echo cancellation in single channel environment would be extremely difficult in comparison to that in two channel case because of the following reasons,

(1) In dual channel AEC, two channels are dedicated to receive inputs from two different speakers and generally, a dual talk detector (DTD) is used. In this case, one channel carries the speech signals from person 1, namely $s_1(n)$, along with the echo signal corresponding to person 2, namely $x_2(n)$. Because of the presence of the

DTD, the echo canceller can exploit the advantage of having a reference of echo-free signal from channel 2, namely $s_2(n)$, to cancel the echo portion of the input signal of channel 1, namely $x_s(n)$ and vice versa. On the other hand, single channel AEC deals with a one speaker and the echo itself is originated by the same speaker speaking in the microphone. Both speech and echo propagation is carried out by a single channel. The most difficulty here, unlike the dual channel case, is to obtain a separate reference signal for the AEC block to cancel out the echo portion from the input echo-corrupted signal. There is no scope of receiving reference signals for echo estimation from another channel.

(2) As a result, in the proposed single channel AEC scheme, a cleaned speech sample is used as reference for the next samples. When suppressing echo in a certain sample, there may be some residual echo present in the cleaned speech (that is why it is denoted by $\widehat{s}(n)$ rather that $s(n)$ itself). Thus, if the currently cleaned speech sample is used as reference for cancellation of echo of a future sample, it would obviously generate some error. So, getting a very high degree of echo cancellation performance using only the traditional adaptive filter algorithms may not be expected in case of single channel AEC.

(3) In case of single channel AEC, the speech from a speaker is contaminated by attenuated previous samples of speech of the same speaker, which increases the probability of the speech and echo to be correlated to some extent. Whereas, in the case of two channel communication, since echo and speech signals are coming from two different speakers, the degree of correlation would be much lower.

(4) In a real conference room environment, the flat delay k_0 should be very large for human perception to distinguish an echo from the original signal because when dealing with audible frequencies, the human ear cannot distinguish an echo from the original sound if the delay is less than 100 millisecond. Thus, the echo estimation task has to deal with a large filter on the order of thousand coefficients. However, the value of $\widehat{w}_n(p)$ is generally considered to be zero for lower values of p. That is why, the variable k_0 is introduced. The value of k_0 can be thousand or more depending on the room acoustic and it symbolizes the amount of flat delay (for which the value of $\widehat{w}_n(p)$ is zero). On the other hand, in case of two channel communication, the value of k_0 may be as small as a single sample and is not significant at all. The goal

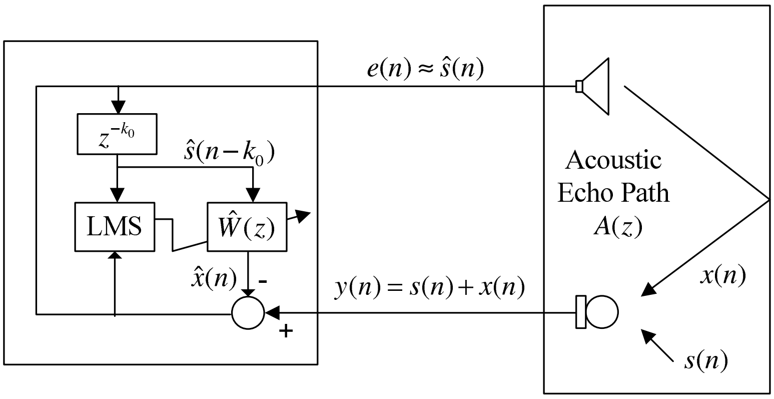

Fig. 2.4: LMS based single channel acoustic echo cancellation in room environment.

of two channel echo cancellation is to cancel the echo from the other channel so that the person speaking in one channel could not hear his/her own voice through the loudspeaker while talking. It is not customary in this case for the room environment of the other end, where the signal is being fed back to the microphone from the loudspeaker, to be like a large conference hall which will produce large delay or long echo trail. Dual channel echo occurs simply when, the loudspeaker output is coupled to the microphone input in any end of the communication link in any possible way.

Hence, there is no doubt that simple use of a conventional adaptive filter algorithm would not be sufficient to effectively suppress echo in single channel environment. In what follows, our objective is to develop an echo cancellation scheme to reduce $x(n)$ from $y(n)$ based on adaptive filter algorithm. Since there is no direct reference available for the adaptive filter, in this case, the most crucial part would be to generate an appropriate reference signal given only the echo corrupted signal. In the proposed scheme, we have considered a delayed versions of an estimated echo cancelled signal as the reference signal to the adaptive filter and thereby established the necessary update relations for the desired filter coefficients.

2.2 Analysis of the proposed single channel AEC based on LMS algorithm

2.2.1 Formulation of LMS Update Equation

In order to reduce the effect of echo, in the proposed scheme, the echo-corrupted signal $y(n)$ is passed through an adaptive filter block, as shown in Fig. 2.4.

As mentioned in section 2.1, unlike dual channel AEC problem, the echo signal $x(n)$ corrupting the speech signal $s(n)$ is generated from the delayed and attenuated version of the same signal $s(n)$ and similar to equation (2.5) can be expressed as,

$$\begin{aligned} x(n) &= \mathbf{a}_n^T \mathbf{s}(n-k_0) && (2.9)\\ &= \sum_{k=1}^{p} a_n(k) s(n-k_0-k), && (2.10) \end{aligned}$$

where, $\mathbf{s}(n-k_0) = [s(n-k_0-1), s(n-k_0-2), \ldots, s(n-k_0-p)]^T$ is a vector of p previous values of $s(n)$ with predefined flat delay k_0. The number p of unknown attenuation coefficients $a_n(k)$ depends on the characteristics of the room.

The task of the adaptive filter block is to produce an estimate of $x(n)$ given $y(n)$ and a reference signal. Since, there is no scope to provide a separate reference in single channel AEC problem, we intend to utilize some delayed versions of the adaptive filter output as the reference signal. The error signal which the adaptive filter tries to minimize can be defined as,

$$\begin{aligned} e(n) &= y(n) - \widehat{x}(n) && (2.11)\\ &= s(n) + x(n) - \widehat{x}(n) && (2.12) \end{aligned}$$

where, $\widehat{x}(n)$ is an estimate of the echo signal generated by the adaptive filter utilizing its coefficient vector $\widehat{\mathbf{w}}_n$ and the echo suppressed input signal $\widehat{s}(n)$ and can be expressed as

$$\begin{aligned} \widehat{x}(n) &= \widehat{\mathbf{w}}_n^T \widehat{\mathbf{s}}(n-k_0) && (2.13)\\ &= \sum_{k=1}^{k=p} \widehat{w}_n(k) \widehat{s}(n-k_0-k). && (2.14) \end{aligned}$$

Here,the vector $\widehat{\mathbf{w}}_n = [\widehat{w}_n(1), \widehat{w}_n(1), \ldots, \widehat{w}_n(p)]^T$ consists of the current estimate of the room attenuation parameters. It is assumed that $\widehat{\mathbf{w}}_n$ has p parameters corresponding to p number of reflection paths and the values of the reflection parameters vary from 0 to 1. It is to be mentioned that the reason behind considering $\widehat{\mathbf{s}}(n-k_0)$ is that the task of AEC starts after the flat delay period of k_0 samples. In the adaptive filter algorithm, the effect of echo in $y(n)$ is iteratively minimized utilizing a certain number of previous samples of $\widehat{s}(n)$ as reference. In this iterative process an estimate of $s(n)$ can be written as

$$\widehat{s}(n) = s(n) + \varsigma(n), \tag{2.15}$$

where $\varsigma(n)$ is an estimation error, which predominantly exhibits noise-like behavior. Although, it may also posses speech-like characteristics, because of its low intensity in comparison to the original signal $s(n)$, such behavior can be considered insignificant. With the increasing iterations towards an optimum solution, $\varsigma(n)$ tends to vanish gradually resulting $\widehat{s}(n) = s(n)$.

Thus the objective function in this case can be defined as the mean square estimation of the error function, namely,

$$\begin{aligned} J_n &= E\{e^2(n)\} = E\{[y(n) - \widehat{x}(n)]^2\} && (2.16) \\ &= E\{[s(n) + x(n) - \widehat{x}(n)]^2\} && (2.17) \\ &= E\{s^2(n)\} + E\{[x(n) - \widehat{x}(n)]^2\} \\ &\quad +2E\{[s(n)(x(n) - \widehat{x}(n))]\}, && (2.18) \end{aligned}$$

where, the last term of right hand side of the objective function can be expressed as

$$\begin{aligned} & 2E\{[s(n)(x(n) - \widehat{x}(n))]\} \\ = \; & 2\sum_{k=1}^{k=p}\{(a_n(k) - \widehat{w}_n(k))r_{ss}(k_0+k) - r_{s\varsigma}(k_0+k)\} && (2.19) \end{aligned}$$

where, $r_{ss}(n)$ corresponds to the cross-correlation between $s(n)$ and $s(n-k_0-k)$ and $r_{s\varsigma}(n)$ is the cross-correlation between $s(n)$ and the noise-like term $\varsigma(n)$. The magnitude of $r_{ss}(n)$ strongly depends on speech characteristics and the amount of flat delay k_0. Optimal performance of the filter occurs when $r_{ss}(n)$ is minimum, i.e. the least possible correlation between $s(n-k_0-k)$ and $s(n)$ is desired. In that case, the correlation between reverberant and non-reverberant part of the input signal will also be minimum making the single channel echo cancellation problem easier. On the other hand, as explained before, because of the noise-like characteristics of $\varsigma(n)$, the term $r_{s\varsigma}(n)$ can also be neglected. As a result, the objective function in equation (2.18) reduces to,

$$E\{e^2(n)\} = E\{s^2(n)\} + E\{[x(n) - \widehat{x}(n)]^2\} \tag{2.20}$$

which may further be expanded as

$$\begin{aligned} E\{e^2(n)\} &= E\{s^2(n)\} + E\{[\sum_{k=1}^{k=p} a_n(k)s(n-k_0-k) \\ &\quad -\sum_{k=1}^{k=p} \widehat{w}_n(k)\widehat{s}(n-k_0-k)]^2\} \qquad (2.21) \\ &= E\{s^2(n)\} \\ &\quad +E\{[\sum_{k=1}^{k=p}(a_n(k)-\widehat{w}_n(k))s(n-k_0-k) \\ &\quad +\sum_{k=1}^{k=p} \widehat{w}_n(k)\varsigma(n-k_0-k)]^2\} \qquad (2.22) \end{aligned}$$

Note that, equation (2.22) reveals a more critical difference between dual channel and single channel AEC. In dual channel AEC, the task of the adaptive filter is to minimize the MSE by producing an estimate $\widehat{\mathbf{w}}_n$ of the room response, such that $\widehat{\mathbf{w}}_n = \mathbf{a}_n$. However, this is not true for the single channel case, which can be clearly observed from equation (2.22), where even in the case when $\widehat{\mathbf{w}}_n = \mathbf{a}_n$ an additional term $\sum_{k=1}^{k=p} \widehat{w}_n(k)\varsigma(n-k_0-k)$ remains. Thus, an additional task of the adaptive filter in this case is to diminish the estimation error term to zero while estimating the optimum filter coefficient.

Minimization of the objective function (2.20) results in,

$$\frac{\delta J_n}{\delta \widehat{\mathbf{w}}_n^T} = 0 \tag{2.23}$$

$$E\{[x(n) - \widehat{x}(n)] \sum_{k=1}^{k=p} \widehat{s}(n-k_0-k)\} = 0, \tag{2.24}$$

Now, using (2.10), (2.14) and (2.15) and also employing the assumptions that

$r_{ss}(n) = 0$ and $r_{s\varsigma}(n) = 0$ we obtain,

$$\begin{aligned} & E\{x(n)s(n-k_0-k)\} = \\ & \sum_{l=1}^{p} \widehat{w}_n(l) E\{s(n-k_0-l)s(n-k_0-k)\}. \end{aligned} \tag{2.25}$$

The above equation is similar to Wiener-Hopf equation and its solution can be written as

$$\widehat{\mathbf{w}}_n = \mathbf{R}_{ss}(n-k_0)^{-1}\mathbf{r}_{xs}(n-k_0), \tag{2.26}$$

where, $\mathbf{r}_{xs}$ is the cross-correlation matrix between $x(n)$ and $s(n)$, while $\mathbf{R}_{ss}$ is the auto-correlation matrix of $s(n)$. There is no doubt that $\widehat{\mathbf{w}}_n$ is the most optimum solution possible. Hence it is shown that even for a single channel AEC problem, the most optimum solution $\widehat{\mathbf{w}}_n$ can be achieved under the assumptions stated earlier.

Note that, in real time implementation, the echo signal described in (2.10) will be depending on every sample values of echo-reduced signal and can be expressed as

$$x(n) = \mathbf{a}_n^T \widehat{\mathbf{s}}(n-k_0) \tag{2.27}$$

$$= \sum_{k=1}^{p} a_n(k)\widehat{s}(n-k_0-k), \tag{2.28}$$

As a result, the objective function defined in equation (2.21) will reduce to a simpler form defined by

$$\begin{aligned} E\{e^2(n)\} &= E\{s^2(n)\} + E\{[\sum_{k=1}^{k=p} a_n(k)\widehat{s}(n-k_0-k) \\ &\quad - \sum_{k=1}^{k=p} \widehat{w}_n(l)\widehat{s}(n-k_0-k)]^2\} \end{aligned} \tag{2.29}$$

Proceeding in a similar fashion, similar optimum solution $\widehat{\mathbf{w}}_n$ can be achieved.

However, the common problem in obtaining the Wiener-Hopf solution is the inversion of the autocorrelation matrix. As an alternative adaptive filter algorithms are very popular for iterative estimation of optimal filter coefficients, which does not require any correlation measurements or matrix inversion. The update equation of the weight vector is generally expressed as

$$\widehat{\mathbf{w}}_{n+1} = \widehat{\mathbf{w}}_n - \mu\nabla\xi(n) \tag{2.30}$$

where, μ is the step factor controlling the stability and rate of convergence, $\xi(n)$ is the cost function and ∇ is the gradient operator. The LMS algorithm simply approximates the mean square error by the square of the instantaneous error, i.e. $\xi(n) = e^2(n)$. Using (2.12), the gradient of $\xi(n)$ can be written as

$$\nabla\xi(n) \quad = \quad \frac{\delta\xi(n)}{\delta\widehat{\mathbf{w}}_n^T} = -2e(n)\widehat{\mathbf{s}}(n-k_0). \tag{2.31}$$

Thus, the update equation for LMS from equation (4.29) is,

$$\widehat{\mathbf{w}}_{n+1} = \widehat{\mathbf{w}}_n + 2\mu e(n)\widehat{\mathbf{s}}(n-k_0) \tag{2.32}$$

For the k-th unknown filter parameter at the n-th iteration,

$$\widehat{w}_{n+1}(k) = \widehat{w}_n(k) + 2\mu e(n)\widehat{s}(n-k_0-k), \tag{2.33}$$

where, $k = 1, 2, \ldots, p$.

2.2.2 Convergence Analysis of the LMS Update

In this section, our objective is to show that the proposed LMS update equation (2.32) for the single channel AEC converges to the optimum solution. In what follows, starting from the proposed update equation (2.32) we show that the average value of the weight vector $\underline{\widehat{\mathbf{w}}}_n$ converges to the Wiener-Hopf solution given by (2.26).

Considering expectation operation on both sides of equation (2.32) we obtain,

$$\underline{\widehat{\mathbf{w}}}_{n+1} = \underline{\widehat{\mathbf{w}}}_n + 2\mu E\{e(n)\widehat{\mathbf{s}}(n-k_0)\} \tag{2.34}$$

where, $\underline{\widehat{\mathbf{w}}}_n = E\{\widehat{\mathbf{w}}_n\}$. Now, for the k-th unknown weight vector(where $k = 1, 2, \ldots, p$), using (2.12) and considering $r_{ss}(n) = 0$ the term $E\{e(n)\widehat{\mathbf{s}}(n-k_0)\}$ of (2.34) can be written as,

$$E\{e(n)\widehat{\mathbf{s}}(n-k_0)\} = E\{[x(n) - \widehat{x}(n)]\widehat{s}(n-k_0-k)\}. \tag{2.35}$$

Similar to the procedure followed in the previous section, using (2.10), (2.14) and (2.15), and also employing the assumptions that $r_{ss}(n) = 0$ and $r_{s\varsigma}(n) = 0$ we obtain,

$$E\{[x(n) - \widehat{x}(n)]\widehat{s}(n-k_0-k)\} = \mathbf{r}_{xs}(n-k_0) - \mathbf{R}_{ss}(n-k_0)\widehat{\mathbf{w}}_n \tag{2.36}$$

Now, using (2.36), (2.34) can be written as

$$\underline{\widehat{\mathbf{w}}}_{n+1} = \underline{\widehat{\mathbf{w}}}_n - 2\mu\mathbf{R}_{ss}(n-k_0)\underline{\widehat{\mathbf{w}}}_n + 2\mu\mathbf{r}_{xs}(n-k_0) \tag{2.37}$$

In order to obtaion a homogeneous solution of equation (2.37), we consider,

$$\underline{\widehat{\mathbf{w}}}_{n+1} = \underline{\widehat{\mathbf{w}}}_n - 2\mu\mathbf{R}_{ss}(n-k_0)\underline{\widehat{\mathbf{w}}}_n \tag{2.38}$$

For correlation matrix $\mathbf{R}_{ss}$, using eigenvalue decomposition we obtaion,

$$\mathbf{R}_{ss} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T \tag{2.39}$$

where, each column of the matrix $\mathbf{U}$ consists of eigenvectors corresponding to eigenvalues constituting the diagonal elements of the matrix $\boldsymbol{\Lambda}$ and $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. Now, multiplying both sides of (2.38) by $\mathbf{U}^T$ we get,

$$\underline{\widehat{\mathbf{w}}}_{n+1}^U = \underline{\widehat{\mathbf{w}}}_n^U - 2\mu\boldsymbol{\Lambda}\underline{\widehat{\mathbf{w}}}_n^U \tag{2.40}$$

where, $\mathbf{U}^T\underline{\widehat{\mathbf{w}}}_n = \underline{\widehat{\mathbf{w}}}_n^U$. The k-th coefficient of the weight vector can be expressed as,

$$\underline{\widehat{w}}_{n+1}^U(k) \quad = \quad (1-2\mu\lambda(k))\underline{\widehat{w}}_n^U(k). \tag{2.41}$$

Hence, the homogeneous solution can be obtained as

$$\widehat{w}_{h.s} \quad = \quad C_k(1-2\mu\lambda(k))^n, \tag{2.42}$$

where, C_k is a constant. Next, in order to obtain the particular solution for the k-th coefficient, based on (2.37) one can get,

$$\widehat{w}_{p.s} \quad = \quad \widehat{w}_{p.s} - 2\mu\lambda(k)\widehat{w}_{p.s} + 2\mu r^U(k) \tag{2.43}$$

For a particular solution $\widehat{w}_{p.s} = K_p r^U(k)$ (2.43) can be written as,

$$K_p r^U(k) \quad = \quad K_p r^U(k) - 2\mu\lambda(k)K_p r^U(k) + 2\mu r^U(k) \tag{2.44}$$

which leads to $K_p = \frac{1}{\lambda(k)}$ and the particular solution,

$$\widehat{w}_{p.s} \quad = \quad \frac{1}{\lambda(k)} r^U(k) \tag{2.45}$$

Hence, the total solution of (2.40) becomes

$$\underline{\widehat{w}}_{n+1}^U(k) \quad = \quad C_k(1-2\mu\lambda(k))^n + \frac{1}{\lambda(k)} r^U(k). \tag{2.46}$$

In the iterative update procedure, obviously the homogeneous part $(1 - 2\mu\lambda(k))^n$ decays to zero with iterations. From the rest of the terms, it can be shown that,

$$\underline{\widehat{\mathbf{w}}} = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^T\mathbf{r}_{xs} = \mathbf{R}_{ss}^{-1}\mathbf{r}_{xs}. \tag{2.47}$$

Thus, it is found that the average value of the weight vector converges to the wiener-hopf, which is the optimum solution with increasing number of iteration.

2.3 Development of Adaptive Characteristics

In general, the common problems of an adaptive LMS algorithm are: (i) very slow convergence (ii) fluctuation around a desired value (iii) in some cases, tendency of not converging to an optimum solution or even diverging to a wrong solution. In the proposed LMS based algorithm, one or more of these problems may occur, since in some practical cases, the assumptions on the negligibility of the cross-correlation terms $r_{ss}(n)$ and $r_{s\varsigma}(n)$ may not strictly hold. One possible solution to overcome these problems is to exploit some adaptive characteristics, which along with the proposed LMS update algorithm can guarantee a better convergence performance. In view of developing such adaptive characteristics, following three factors are considered in the proposed algorithm: (i) the degrees of the cross-correlation terms $r_{ss}(n)$ and $r_{s\varsigma}(n)$, (ii) the amount of signal power for the two signals under consideration: the reference signal $s(n-k_0)$ and the current signal $s(n)$, (iii) the mean square error between consecutive estimates of the unknown filter coefficients.

In order to demonstrate the performance of the proposed LMS update algorithm, speech samples of different characteristics, such as voiced, unvoiced and pause are taken into consideration. It is found that the negligibility of the cross-correlation terms $r_{ss}(n)$ and $r_{s\varsigma}(n)$ strongly depends on the characteristics of the speech samples. For example, because of the inherent periodicity of the voiced speech, the degree of cross-correlation between two voiced speech frames of a person becomes higher in comparison to that between two unvoiced speech frames which are random in nature. In this case, ratio of power of two different speech frames may also carry some significant information. For example, if we consider a voiced frame and an unvoiced frame, their power ratio is generally higher in comparison to that of two voiced speech frames.

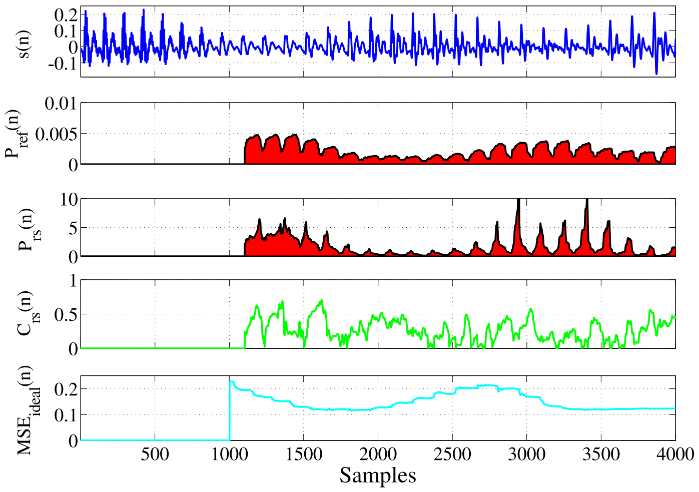

Fig. 2.5: A voiced frame followed by another voiced frame (a) Original Signal $s(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$ (e) MSE of coefficient updated from ideal values $MSE_{ideal}(n)$.

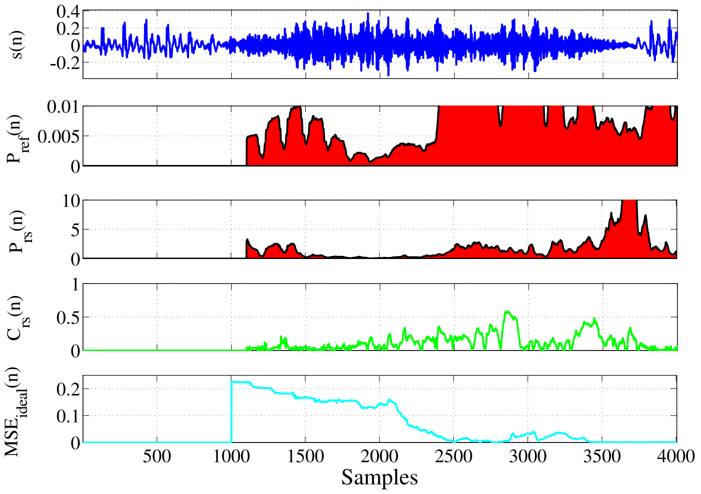

Fig. 2.6: A voiced frame followed by an unvoiced frame (a) Original Signal $s(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$ (e) MSE of coefficient updated from ideal values $MSE_{ideal}(n)$

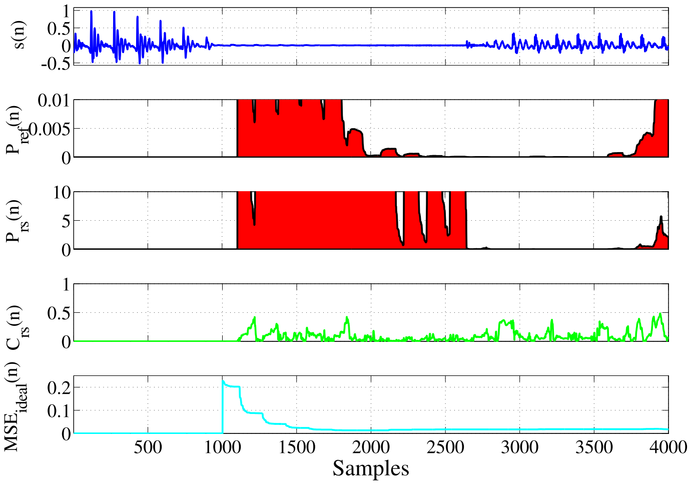

Fig. 2.7: A voiced frame followed by a pause (a) Original Signal $s(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$ (e) MSE of coefficient updated from ideal values $MSE_{ideal}(n)$

In Fig. 2.5(a), a male utterance $/iy/-/r/$ of a duration of 250 ms with a sampling frequency of 16 kHz is shown. in this figure, a few samples of voiced phoneme are followed by another few samples of voiced phoneme. The strong periodicity of the utterance $s(n)$ clearly indicates its voiced characteristics. Considering the flat delay of $k_0 = 1000$ samples, from the starting point of $s(n)$, this utterance will act as a reference signal for the generation of echo that corrupts the current samples at or after k_0 samples. Employing the proposed LMS algorithm on the echo-corrupted signal $y(n)$, an echo reduced signal $\widehat{s}(n)$ is obtained. In Fig. 2.5(b), power of the reference signal $\widehat{s}(n-k_0)$, namely $P_{ref}(n)$ is depicted, which is computed at every input instances considering a window of M samples and is defined as

$$P_{ref}(n) = \frac{\sum_{i=-\frac{M}{2}}^{\frac{M}{2}-1} [\widehat{s}(n-k_0+i)]^2}{M} \tag{2.48}$$

. Here we consider $k_0 >> M$ and $M = 100$. In this connection, we also consider the average power $P_{sup}(n)$ of the last M samples of the echo suppressed speech signal

$\widehat{s}(n)$, which is defined as

$$P_{sup}(n) \quad = \quad \frac{\sum_{j=0}^{M-1} [\widehat{s}(n-j)]^2}{M}. \tag{2.49}$$

The ratio of $P_{ref}(n)$ and $P_{sup}(n)$ is denoted as the power ratio $P_{rs}(n)$, which is shown in Fig. 2.5(c). In Fig. 2.5(d) the cross correlation coefficient $C_{rs}(n)$ between the reference signal $\widehat{s}(n-k_0)$ and the current signal $\widehat{s}(n)$ is shown. A coefficient of correlation, $C_{rs}(n)$, is a mathematical measure of how much one number can expected to be influenced by change in another. It is defined as,

$$C_{rs}(n) = \frac{cov(\widehat{s}(n-k_0+i)\widehat{s}(n-j))}{\sigma_{\widehat{s}(n-k_0+i)}\sigma_{\widehat{s}(n-j)}} \tag{2.50}$$

Here, $-M/2 \leq i \leq M/2-1$ and $0 \leq j \leq (M-1)$. If $C_{rs}(n) = \pm 1$ then there is a strong positive/negative correlation between two signals. If it is zero then there is no correlation among the matrices. In order to demonstrate the performance of the proposed LMS update algorithm, in terms of convergence rate and parameter estimation accuracy, in Fig 2.5(e), the mean square error $MSE_{ideal}(n)$ between the estimated coefficients w_n and the true coefficients a_n is depicted.

In a similar fashion, in Fig. 2.6 and 2.7, first a voiced phoneme /*ih*/ followed by an strong unvoiced phoneme /*sh*/ and then a voiced phoneme /*ih*/ followed by pause are considered, respectively. It is to be mentioned that in these figures Fig. 2.5-Fig. 2.7, the reference signal is always a voiced frame and the current frame is voiced, unvoiced or pause respectively. It is found that the power of the reference voiced frame is always quite high in comparison to unvoiced or pause frames. However, the power ratio not only depends on the power of the reference voiced frame but also on the power of the echo suppressed signal. If the current frame is a pause or weakly unvoiced frame then the power ratio is very high, otherwise, for voiced and strong unvoiced frames the power ratio is lower. The correlation coefficient is very small when measured between a voiced and a unvoiced frame, but is quite large for two voiced frames.

The presence of voiced frame as a reference strongly governs the rate of convergence and the estimation error of the proposed LMS algorithm. For example in Fig. 2.5, because of althrough presence of the voiced frame as reference, the convergence performance becomes very poor and even in some cases the algorithm diverges and in all cases, the estimation error was higher. On the contrary, in Fig. 2.7 it is observed that, when the current frame is pause, even in the presence of voiced reference frame a very fast convergence is obtained with a small estimation error. Moreover, in Fig. 2.7, as the current frame is unvoiced instead of pause, a slower convergence is observed with a high estimation error.

It is quite interesting that the performance characteristics of the proposed LMS update algorithm drastically changes when the reference frame is considered unvoiced, as shown in Fig. 2.8, 2.9 and 2.10. In this case a very fast convergence is obtained with a high level of estimation accuracy.

The reason behind this drastic change in characteristics can be explained based on the cross-correlation that may exist between the reference frame and the current frame. In case of voiced reference frame, a strong correlation persists between each samples of the voiced frame, which makes it difficult for the LMS to estimate the room response as the assumption of the negligibility of the cross-correlation terms $r_{ss}(n)$ and $r_{s\varsigma}(n)$ does not hold anymore. Moreover, when the current frame has a high energy speech along with the echo, i.e. when the power ratio is lower, the convergence performance of the LMS algorithm may degrade because of the chances of suppression of the input speech. In the case when the current frame is pause, no matter whether the reference frame is voiced or unvoiced, a fast convergence with high estimation accuracy is achieved using the proposed LMS algorithm. The reasons behind are, (i) negligible cross-correlation between reference frame and current speech frame and (ii) a comparatively higher power ratio. In case of unvoiced reference frame, because of existence of a little correlation between the input and the reference frame the convergence performance of the proposed LMS algorithm is found quite satisfactory irrespective of the power of the reference signal(strong unvoiced or weakly unvoiced).

Finally, in Fig. 2.11, 2.12 and 2.13 the outcomes of three different cases when the references are always from a pause or stop frame are shown. As can be seen from the figures, the presence of pause or stop as reference, no significant update occurs in the proposed LMS algorithm and in some cases, as expected, the convergence performance degrades. This is because of the lack of reference data as well as signal energy, which are required for LMS updates.

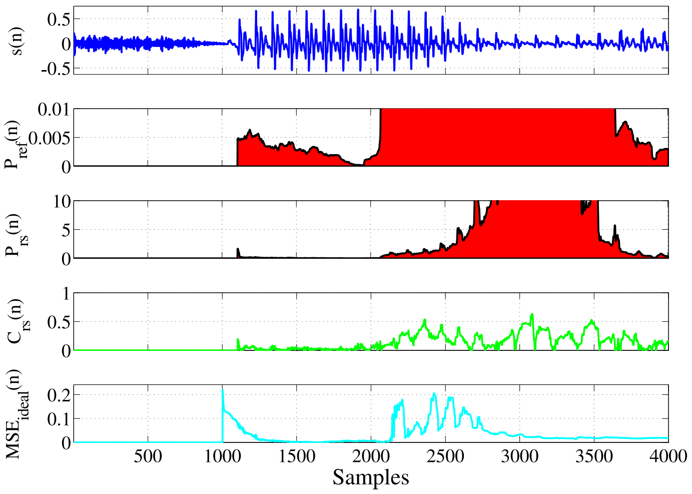

Fig. 2.8: An unvoiced frame followed by a voiced frame (a) Original Signal $s(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$ (e) MSE of coefficient updated from ideal values $MSE_{ideal}(n)$

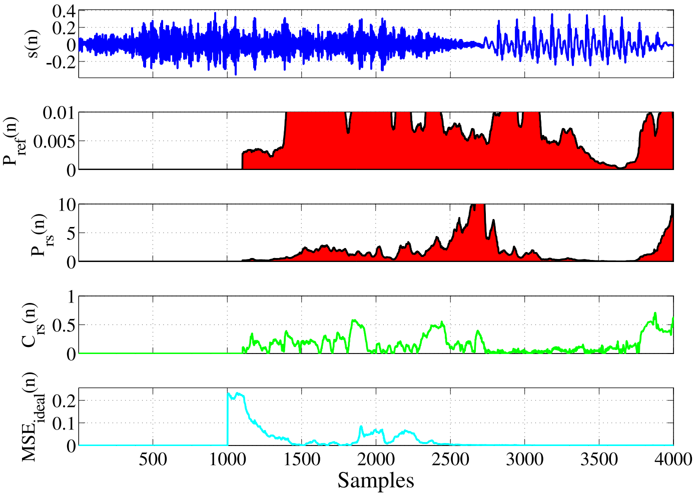

Fig. 2.9: An unvoiced frame followed by another unvoiced frame (a) Original Signal $s(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$ (e) MSE of coefficient updated from ideal values $MSE_{ideal}(n)$

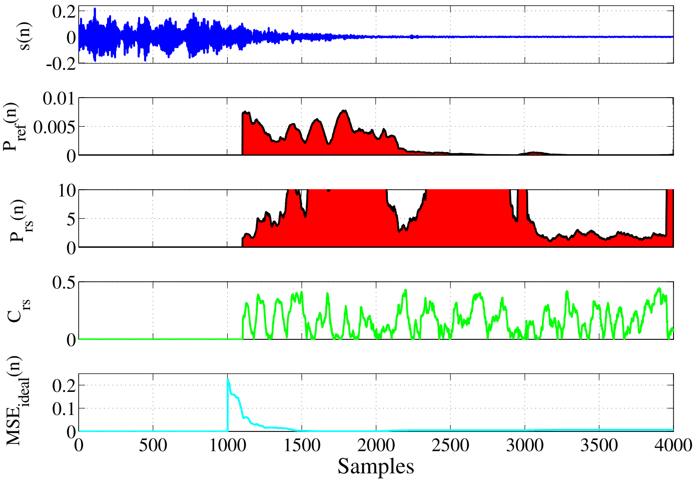

Fig. 2.10: An unvoiced frame followed by a pause (a) Original Signal $s(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$ (e) MSE of coefficient updated from ideal values $MSE_{ideal}(n)$

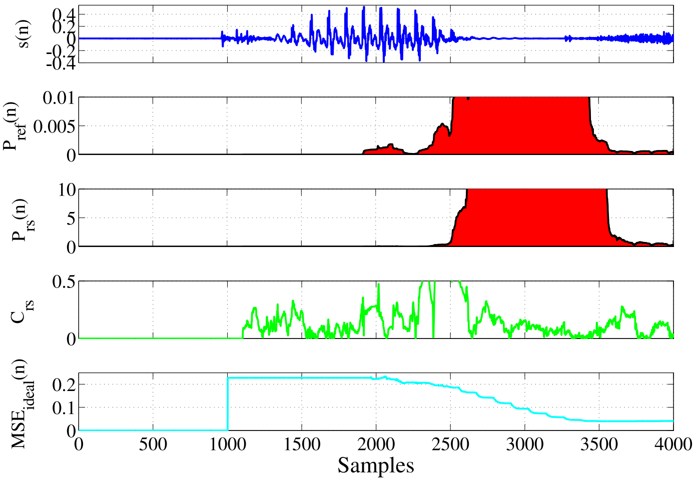

Fig. 2.11: A pause followed by a voiced frame (a) Original Signal $s(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$ (e) MSE of coefficient updated from ideal values $MSE_{ideal}(n)$

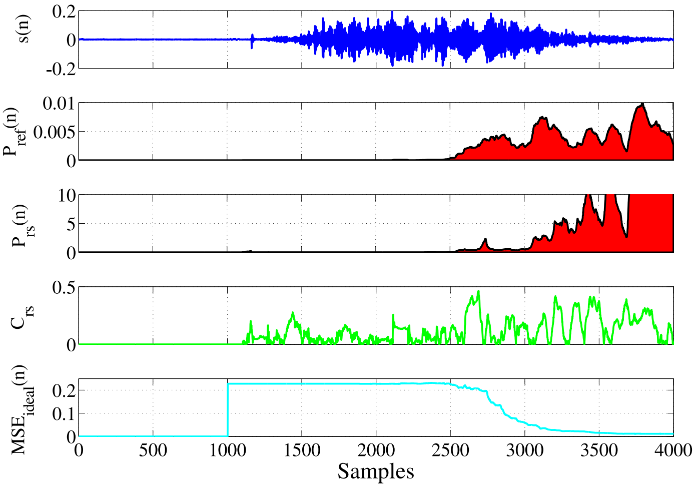

Fig. 2.12: A pause followed by an unvoiced frame (a) Original Signal $s(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$ (e) MSE of coefficient updated from ideal values $MSE_{ideal}(n)$

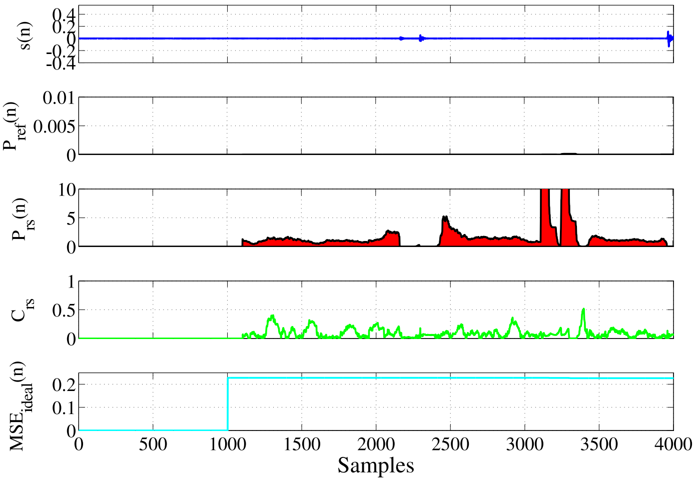

Fig. 2.13: A pause followed by another pause (a) Original Signal $s(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$ (e) MSE of coefficient updated from ideal values $MSE_{ideal}(n)$

2.4 Proposed Update Constraints

The insight obtained from extensive experimentation on several such case as presented in Fig. 2.5 - Fig. 2.13 are summarized in table 2.1. It is clearly observed from the table that in many cases the performance of the proposed algorithm is not satisfactory or even poor, e.g. when the reference and the input signal both are voiced frames, when the reference signal is pause etc.. In view of overcoming these cases we are going to propose some conditions which will guarantee a fast convergence with a low estimation error. It is obvious that if the proposed algorithm is used, there is always a possibility to obtain poor convergence or even in some cases divergence with a high estimation error. Based on the results obtained from Table 2.1 and some more experimentation we hereby propose three conditions on LMS update, which are designed to indicate whether the updating should be carried out or halted. Implementation of these conditions in the proposed LMS update will provide assurance of fast convergence with a high estimation accuracy.

Following conditions are proposed for constraining the LMS update,

Condition 1: $P_{rs}(n) \geq \zeta$ and $P_{ref}(n) \geq \beta$

Condition 2: $C_{rs}(n) \leq \Upsilon$ and $P_{ref}(n) \geq \beta$

Condition 3: $e_{coeff}(n) \leq \aleph$

where, ζ, β, Υ and $\aleph$ are threshold values and $e_{coeff}(n)$ is the mean square error of the estimations of successive iterations defined as,

$$e_{coeff}(n) = \sum_{K=1}^{p} (\widehat{w}_n(k) - \widehat{w}_{n-1}(k))^2 / p. \tag{2.51}$$

The update of the proposed LMS algorithm will be carried out if any one of the first two conditions is true. The third condition will be checked after each estimation and if it is true the new estimation will take effect. Now, let us discuss the facts behind choosing these three conditions for our proposed method.

2.4.1 Condition 1: The Power Ratio and Reference Power Constraint

From Table 2.1 we know that the performance of the LMS is best when the reference sample $\widehat{s}(n-k_0)$ is from a voiced or unvoiced segment with high energy while the

Table 2.1: Dependence of LMS update on the acoustic characteristics of the reference and current speech frame

Reference Speech	**Speech at the Echo Corrupted Sample**	**LMS Update performance**
Voiced	Voiced	unsatisfactory
Voiced	Unvoiced	unsatisfactory
Voiced	Pause	satisfactory/excellent
Unvoiced	Voiced	excellent
Unvoiced	Unvoiced	excellent
Unvoiced	Pause	excellent
Pause	Voiced	poor
Pause	Unvoiced	poor
Pause	Pause	poor

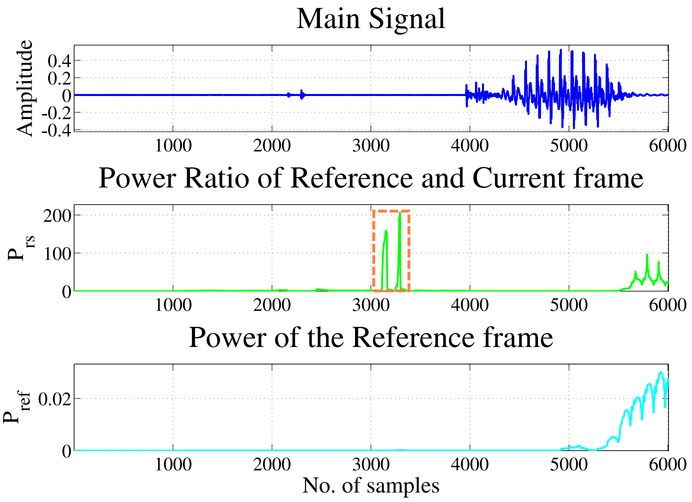

Fig. 2.14: Power ratio when reference and current sample both are from pause/stop.

current sample $s(n) + x(n)$ is from a speech pause or stop. This fact leads us to decision that updating the LMS at high power ratios only, will improve the performance of the adaptive algorithm as it will ensure upto a certain limit that the reference frame is a high energy frame (whether voiced or unvoiced) and the current frame may be a stop or a pause. Thus, condition 1 is aimed to ensure that,

$$P_{rs}(n) \text{ or}, \frac{P_{ref}(n)}{P_{sup}(n)} \text{ or}, \frac{P_{voiced}(n)}{P_{pause}(n)} \text{ or}, \frac{P_{unvoiced}(n)}{P_{pause}(n)} \geq \zeta \tag{2.52}$$

However, considering a high power ratio may not always be the sufficient condition to ensure that the reference has a high energy. For example consider Fig. 2.14. The marked area in the figure , though shows a high power ratio, comes from an initial silence where only a very little amount of noise was present. Obviously, the reference was also from a speech pause as the person has not started speaking yet and coincidentally the power of the reference (though very small) was very large compared to the current frame which is also a silence. To prevent the update of LMS at these situations, we also employ a threshold on the reference frame power, which is, LMS would update if, $P_{ref}(n) \geq \beta$.

2.4.2 Condition 2: Effect of Cross-Correlation Coefficient

In case of single channel AEC, the reference and the echo corrupted signal may both be speech signals of the same person. Thus, it is very much probable that the two signals may be correlated. If the reference is also correlated with the speech input, then it is certain that the adaptive algorithm would try to suppress speech also and its update equation would show unusual degradation. The first part of condition 2 is employed to ensure that LMS will only update its estimation when the cross-correlation between the reference and the current frame is smaller than a certain threshold Υ.

Now, there is another critical scenario which needs to be addressed. In Fig. 2.8, it can be seen that though the power ratio is very small, the update is quite good for LMS, where as, in Fig. 2.6, the power ratio is small and the update is poor. In both cases the correlation coefficient is very small. The reference frame power is, however, quite high. If we set condition 1 for these scenarios, LMS update would be halted for low power ratio, though correlation coefficient and reference power are high. To prevent this, we set a combined condition. If the reference power $P_{ref}(n) \geq \beta$ and the correlation coefficient $C_{rs}(n) >\geq \Upsilon$, LMS would update itself irrespective of condition 1.

2.4.3 Condition 3: Effect of change in estimated coefficients

It has been observed that, in case of single channel adaptive echo cancellation, the previous two conditions could not always prevent the degradation of performance. As the cleaned signal at present time is stored as reference for the future, there is always an estimation error (however small that is) present in echo cancellation. Thus, using this cleaned signal as reference may degrade the update of LMS even though the reference has a high energy and the signal currently being processed is a pause. To limit these degradations, a condition on the variation of updated estimation is employed which dictates the LMS update in a way, that any abrupt and huge change in the estimated coefficients is not allowed. Thus, if there is a huge difference between the currently estimated coefficients of the LMS algorithm and the immediate previous estimation, LMS would discard the current estimation and hold on to the previous estimation.

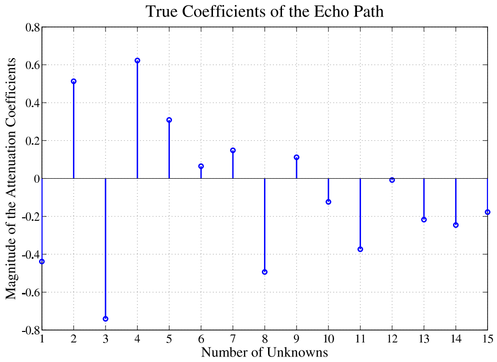

Fig. 2.15: True coefficients of the impulse response of echo path

2.5 Simulation Results

2.5.1 Room response and flat delay

For simulation purpose an FIR filter a_n of length 1016 taps was chosen. The coefficients of $a_n(k)$ are shown in Fig. 2.15. The coefficients were obtained as [80],

$$\begin{aligned} a_n(k) &= (\frac{R}{B})e^{-20Ak}, 1001 \leq k \leq Length \\ &= 0,\ otherwise \end{aligned} \tag{2.53}$$

where R being a random number between -1 and 1, $Length = 1016$, $A = 0.004$ and $B = 1$. With 16kHz sampling frequency, the time span of the entire filter length (1016 samples) is about 63.5ms. However, the response differs significantly from zero for about 1ms. Thus there is a lengthy leading zero part in the response (not shown in Fig. 2.15, known as the flat delay. The length of the flat delay is equal to the round-trip delay between the echo canceller and the point of echo reflection. Usually the flat delay is many times the length of the significant part [81].

The conventional echo canceller which uses the FIR structure just depends on increasing the numbers of filter tap to cover the whole echo path impulse response region, these greatly degrade the cancellation performance, resulting a long convergence time and large residual echo error. The most effective technique to handle flat delay is to introduce the idea of a delay buffer, where the time delay is pre-calculated from the measured distance between the microphone and the speaker. In this way, the computational time is greatly reduced [2]. In Fig.2.16, the parameters involved in the pre-calculation is shown. Considering that, there is a distance d

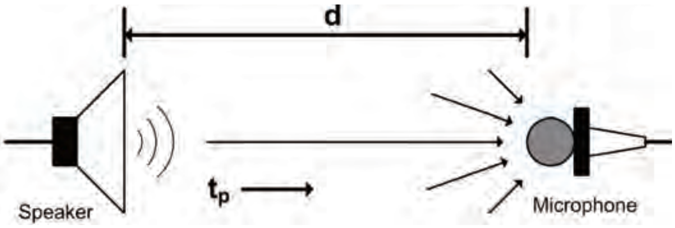

Fig. 2.16: Pre-calculation of delay coefficients.

meters between microphone and speaker, sampling frequency is F_s Hz, the sound propagation time to traverse this distance is $t_p = d/S$ with the speed of sound through air $S = 332\ m/s$, the required number of zeros k_0 (delays) in the filter can be computed as

$$k_0 = t_p \times F_s. \tag{2.54}$$

In our simulation, an acoustic room environment is simulated using a tap-delay-line filter, where it is assumed that the speaker to microphone distance is $d = 20.75$m, which according to (2.54) for a sampling frequency of 16 KHz corresponds to a delay of 1000 taps. Obviously, the coefficients corresponding to these taps are all zero.

Thus, although the overall filter length is usually very large for acoustic echoes, because of the implicit zeros which correspond to a specified delay, it is evident that a few number of unknown coefficients has to be determined.

2.5.2 Speech Sample and Performance Measure

Simulations were performed on two different speech signals uttering (1) "Good service should be rewarded by big tips" by a male voice and (2) "She had your dark suit in greasy wash water all year" another male voice. Both of the speech were taken from the TIMIT database [77]. The step size for the LMS adaptive filter was varied from $1/p$ to 0.02 where p is the number of unknown coefficients of the room response.

The echo return loss enhancement (ERLE) in dB is computed as a performance measure. ERLE is a smoothed measure of the amount (in dB) that the echo has been attenuated. It is defined as the ratio of the power of the residual echo signal and the input echo signal power [2],

$$ERLE = -10 \log \frac{E((echo_{residual}(n))^2)}{E((echo_{original}(n))^2)}, \tag{2.55}$$

The ERLE indicates the amount of loss introduced in the echo cancellation process by the adaptive filter alone. The average value of ERLE over time is used as a criteria of performance evaluation in this experiment.

Another criteria for performance evaluation termed Signal to Distortion Ratio (SDR) expressed in dB is defined as,

$$SDR = 10 \log \frac{P_s(n)}{P_d(n)}, \tag{2.56}$$

where $P_s(n)$ is the original signal power and $P_d(n)$ is the amount of distortion (noise and echo) present in a distorted signal. The difference of SDR in system output and input is an indicator of the system performance. The higher the SDR difference the better is the improvement. In case of single channel echo cancellation, even the reference signal is corrupted with noise-like residual echo, which effects the performance of the adaptive algorithm to a great extent. Thus, SDR provides a better view of the echo cancellation performance over the traditional ERLE in this particular case.

2.5.3 Results and Comments

Based on the limitations in update imposed by characteristics of speech frame, three conditions are employed in the proposed LMS based adaptive echo cancellation method for controlling the update of the adaptive algorithm, as discussed previously.

In Fig. 2.17.a and Fig. 2.17.b a plot of the original echoless speech signal 1 $s(n)$ and the power ratio P_{rs} for a room response with 2 unknown coefficients and 1000 sample flat delay is shown. It can easily be seen that the power ratio is very high in speech pauses and stops. From the figure, we can assume the value of ζ to be around 2.0 for equation (2.52). However, if the current frame power is too small than the reference frame the ratio would be very high, irrelative of the order of magnitude. This fact led to another condition based on the power of reference frame, which would ensure that the reference frame is a high energy voiced or unvoiced frame at high power ratios. From Figure 2.17.c, the value of β can be considered to be near 0.003 for the condition $P_{ref} \geq \beta$ for LMS update.

In Fig. 2.18.a, the original speech signal 1 $s(n)$ is shown, while, in Fig. 2.18.b and Fig. 2.18.c the values of C_{rs} for $M = 100$ and the values of $MSE_{ideal}(n)$ are

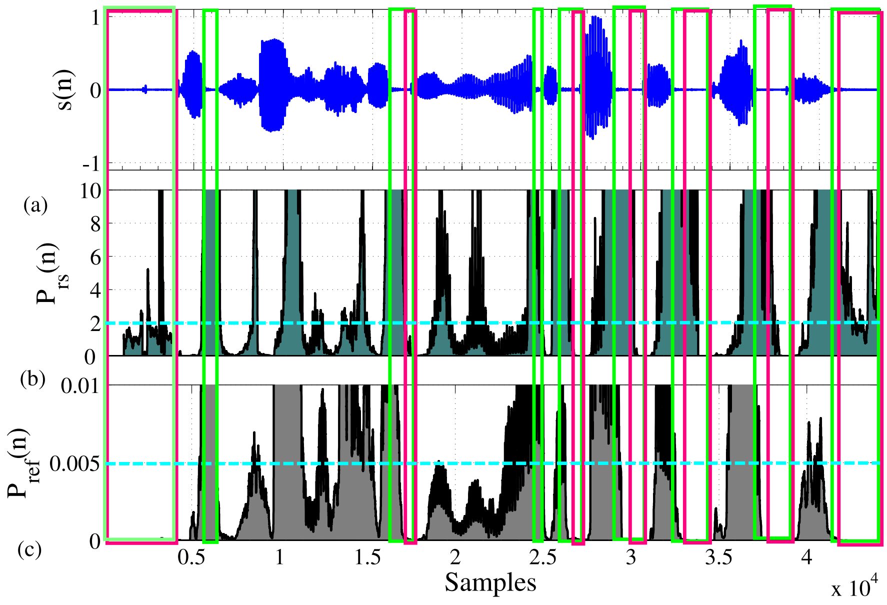

Fig. 2.17: (a) Echoless signal s_1 (b) Power ratio P_{rs}(M=100) (c) Power of the reference frame P_{ref}.

shown. In case of the second condition, related to the correlation coefficient C_{rs}, we considered the value of Υ to be 0.25 which ensures that no speech is being suppressed by confusing it with the echo.

Now, the two conditions, condition 1 and 2 have been simultaneously applied to the nine different cases of voiced, unvoiced and pause frames described in section 2.3, and the results are shown in Fig. 2.19- Fig.2.27.

As can be seen from these figures, the application of condition 1 and condition 2 in the proposed method improved the convergence performance of the LMS algorithm to a large extent. The $MSE_{ideal}(n)$ curve is now totally non-diverging and the estimation error is minimized in every cases.

For continuous speech signals, it has been observed that extreme changes in coefficient estimation resulted in degradation of SDR improvement and ERLE in the retrieved signal. Large and abrupt fluctuation of the filter estimate increases estimation error and makes the system unstable. Thus, condition 3 is introduced to suppress abrupt change in coefficient estimation of the LMS algorithm. If the change in updated coefficients is smaller then a threshold $(0.7e^{-4})$ then LMS would update, otherwise the algorithm would hold the previous coefficients.

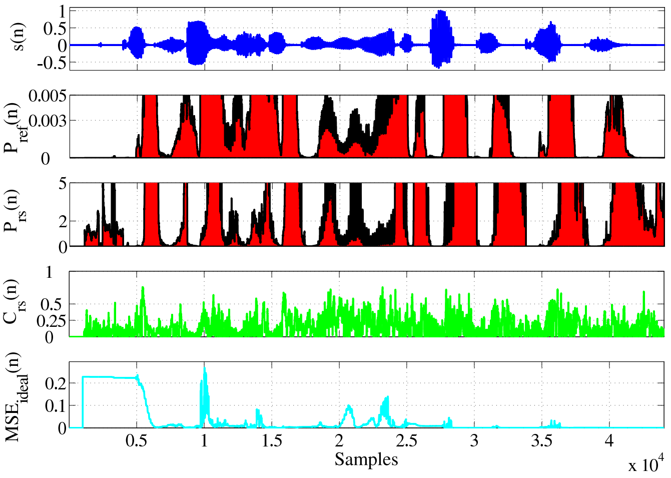

Fig. 2.18: LMS update on speech signal 1(2 unknown coefficients, no conditions applied) (a) Original Signal $s(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame C_{rs}(e) MSE of coefficient updated from ideal values $MSE_{ideal}(n)$

.

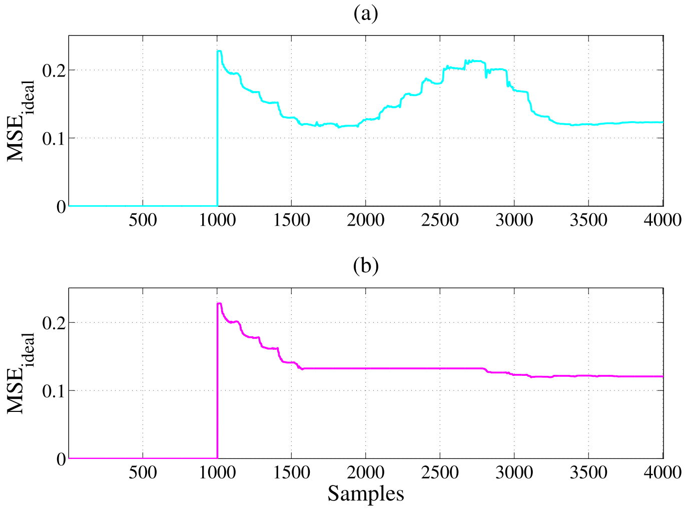

Fig. 2.19: MSE of estimation coefficients from ideal values for a voiced frame in reference and a voiced frame in current samples (a) without applying any condition (b) applying condition 1 and 2.

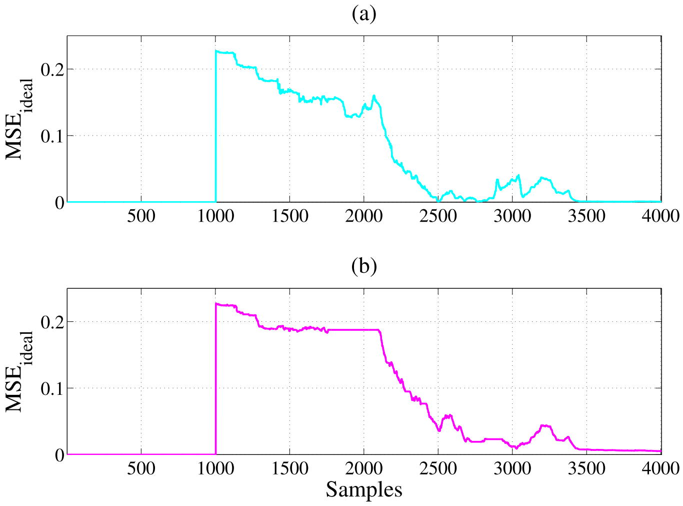

Fig. 2.20: MSE of estimation coefficients from ideal values for a voiced frame in reference and a unvoiced frame in current samples (a) without applying any condition (b) applying condition 1 and 2.

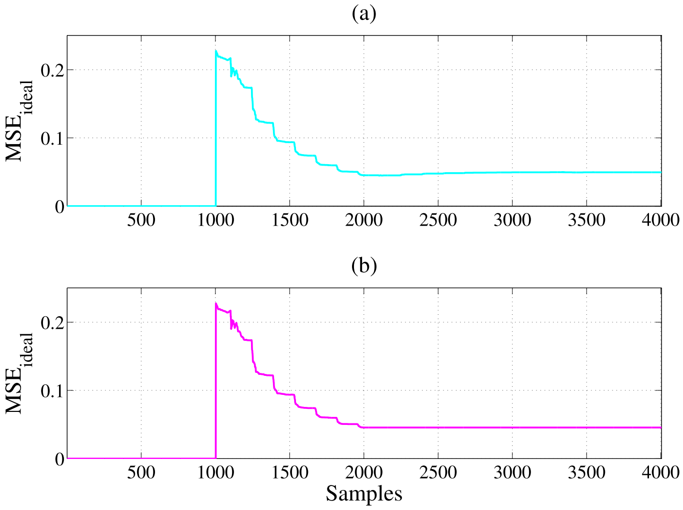

Fig. 2.21: MSE of estimation coefficients from ideal values for a voiced frame in reference and a pause frame in current samples (a) without applying any condition (b) applying condition 1 and 2.

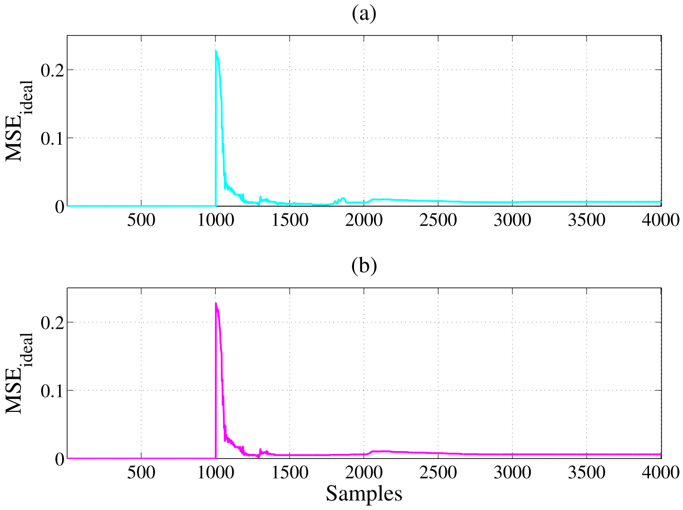

Fig. 2.22: MSE of estimation coefficients from ideal values for a unvoiced frame in reference and a voiced frame in current samples (a) without applying any condition (b) applying condition 1 and 2.

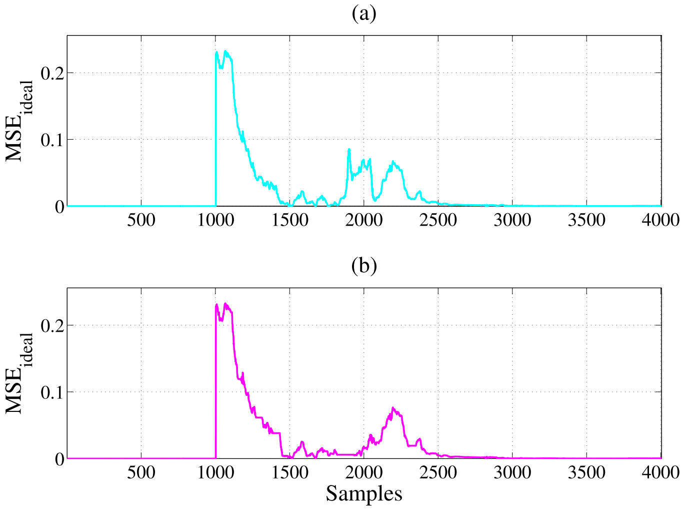

Fig. 2.23: MSE of estimation coefficients from ideal values for a unvoiced frame in reference and a unvoiced frame in current samples (a) without applying any condition (b) applying condition 1 and 2.

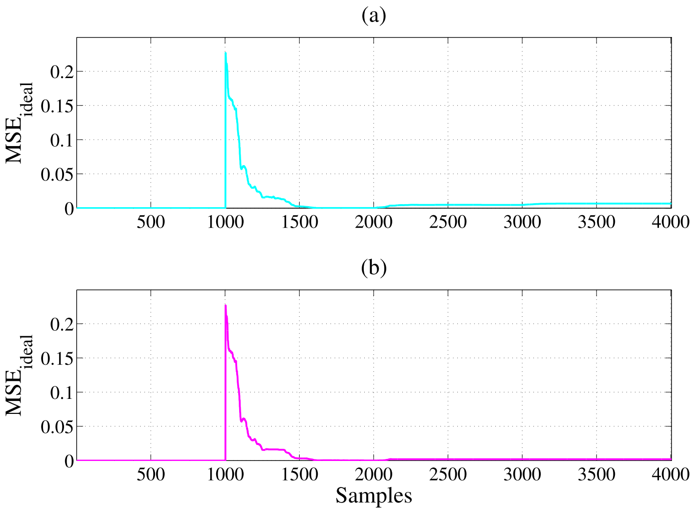

Fig. 2.24: MSE of estimation coefficients from ideal values for a unvoiced frame in reference and a pause frame in current samples (a) without applying any condition (b) applying condition 1 and 2.

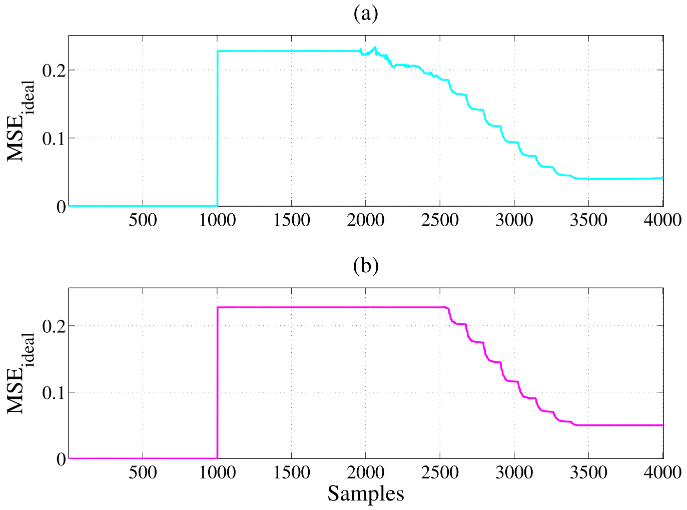

Fig. 2.25: MSE of estimation coefficients from ideal values for a pause frame in reference and a voiced frame in current samples (a) without applying any condition (b) applying condition 1 and 2.

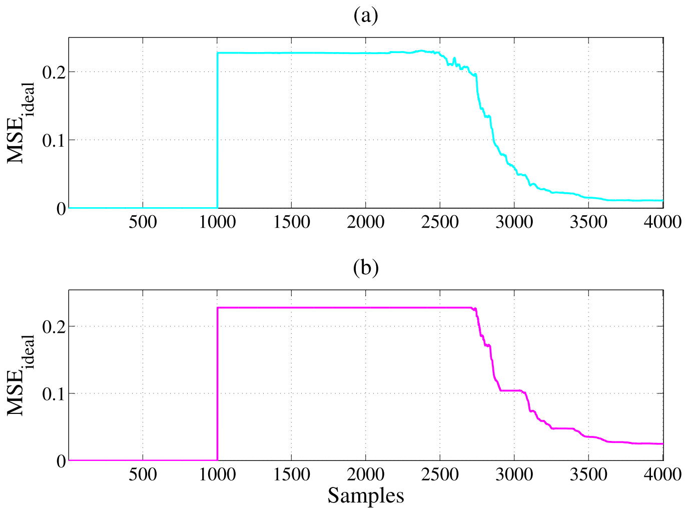

Fig. 2.26: MSE of estimation coefficients from ideal values for a pause frame in reference and a unvoiced frame in current samples (a) without applying any condition (b) applying condition 1 and 2.

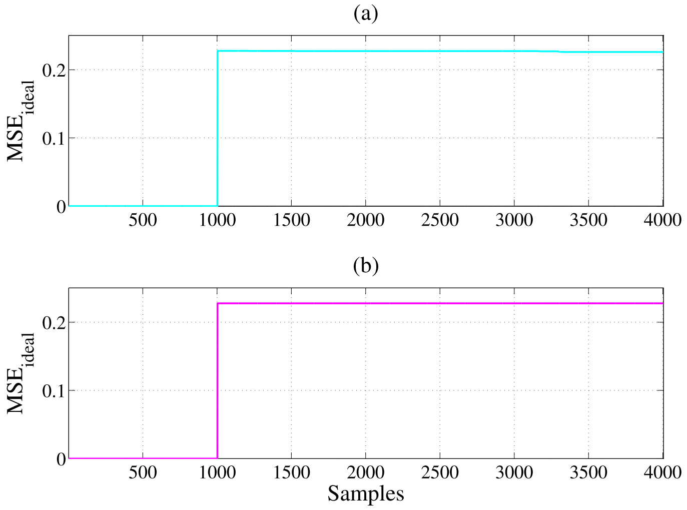

Fig. 2.27: MSE of estimation coefficients from ideal values for a pause frame in reference and a pause frame in current samples (a) without applying any condition (b) applying condition 1 and 2.

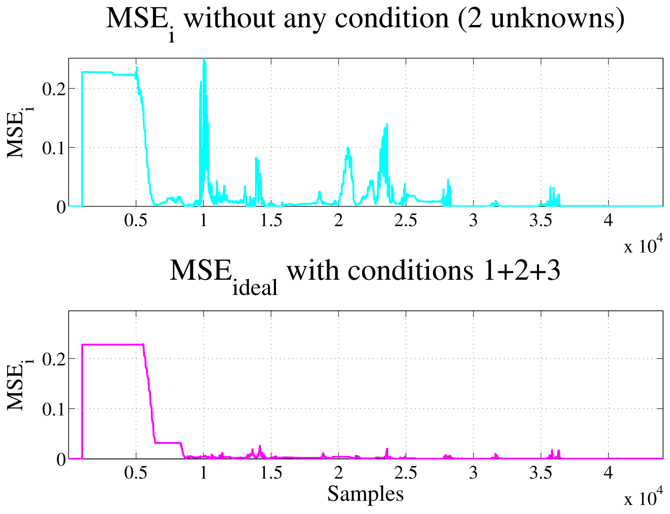

Fig. 2.28: MSE of LMS estimations from ideals on speech signal 1 (2 unknown coefficients) (a) Without any condition (b) With condition 1, 2 and 3 simultaneously applied

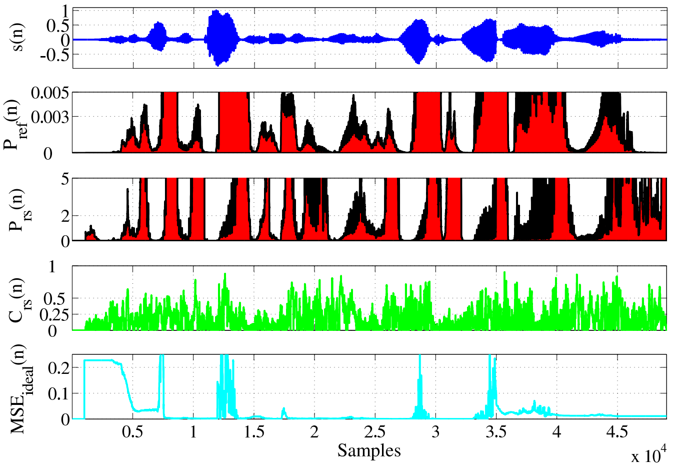

Fig. 2.29: LMS update on speech signal 2(2 unknown coefficients, no conditions applied) (a) Original Signal s_1 (b) Power of the reference frame P_{ref} (c) Power ratio P_{rs}(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame (e) MSE of coefficient updated from ideal values .

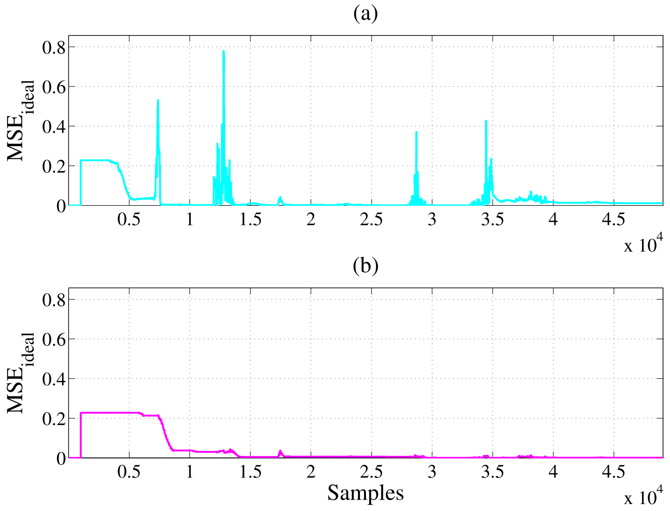

Fig. 2.30: MSE of LMS estimations from ideals on speech signal 2(2 unknown coefficients) (a) Without any condition (b) With condition 1, 2 and 3 simultaneously applied

In Fig. 2.28.a the MSE of coefficient estimates of LMS with respect to ideal values of unknown coefficients without applying any conditions is shown, while in Fig. 2.28.b the MSE with conditions 1, 2 and 3 simultaneously applied is depicted. It can easily be seen that the update of LMS has been stabilized to better results robustly when the conditions are applied and degradation of estimation is prevented. In Fig. 2.29 and 2.30 the reference power, power ratio, correlation coefficients and LMS update with and without conditions for speech signal 2 is illustrated. These figures also demonstrates the efficiency of the applied conditions in controlling the update of LMS algorithm for producing better results.

In Table 2.2 and Table 2.3 the performance of the proposed LMS algorithm in terms of ERLE and SDR with application of the three conditions in different combinations is shown for speech sample 1 at 2 and 15 unknown parameters of room response respectively. As can be seen from tables, applying all the three conditions (condition 1, 2 and 3) gives the consistent better results in terms of SDR improvement and ERLE in both cases. A combination of any two conditions or applying any one of the three conditions may also give good results in some unusual cases.

In Table 2.4 and Table 2.5 the performance of the proposed LMS update method

Table 2.2: Performance for two unknown coefficient on speech singal 1(μ is varied from 0.5 to 0.02)

Method	SDR improvement	ERLE (dB)
No Condition	4.2383	4.9455
Condition 1	8.1188	7.7073
Condition 2	4.0429	6.0975
Condition 3	7.3819	6.6764
Condition 1,2	4.0584	5.4551
Condition 1,3	11.0503	9.0704
Condition 2,3	9.0494	7.1331
Condition 1,2,3	9.4747	7.1320

Table 2.3: Performance for fifteen unknown coefficient on speech signal 1(μ is varied from 0.5 to 0.02)

Method	SDR improvement	ERLE (dB)
No Condition	6.8771	1.9387.
Condition 1	6.9705	1.4418
Condition 2	4.6360	0.4678
Condition 3	7.1388	2.0022
Condition 1,2	4.9232	0.8749
Condition 1,3	6.9379	1.4370
Condition 2,3	6.3789	0.9437
Condition 1,2,3	7.0163	1.2941

Table 2.4: Performance for increasing number of unknown coefficients for the input speech signal 1(μ is varied from $1/p$ to 0.02)

	No Conditions		With Conditions 1+2+3	
No. of coefficients	SDR Improvement (dB)	ERLE (dB)	SDR Improvement (dB)	ERLE (dB)
2	4.2383	4.9455	9.4747	7.132
4	3.2086	0.7677	4.2036	1.7477
6	7.3659	3.4404	8.4879	4.5733
8	4.7824	-0.6398	3.6733	-1.5853
10	4.836	-0.1011	4.2362	-0.8793
12	5.4874	0.8732	5.5778	0.1671
14	6.806	2.0228	6.9683	1.5306

Table 2.5: Performance for increasing number of unknown coefficients for the input speech signal 2(μ is varied from $1/p$ to 0.02)

	No Conditions		With Conditions 1+2+3	
No. of coefficients	SDR Improvement (dB)	ERLE (dB)	SDR Improvement (dB)	ERLE (dB)
2	-1.2179	5.7513	6.1706	9.5454
4	1.7307	3.4837	4.4491	4.6295
6	6.0433	5.2986	10.1695	7.7976
8	2.9494	1.2432	5.0501	1.1656
10	3.202	1.6676	5.2221	1.5762
12	5.1768	2.503	6.9767	2.4338
14	6.6419	3.667	7.685	3.4796

with and without applying the three proposed conditions is compared in terms of SDR difference (dB) and ERLE (dB). Performance is evaluated for different number of unknown coefficients ranging from 2 to 14 for speech signal 1 and 2 respectively. The tables clearly demonstrate the superiority of the proposed LMS algorithm with update constraints over that without any constraints.

2.6 Conclusion

In this chapter, a novel approach of single channel acoustic echo cancellation scheme using gradient based adaptive LMS algorithm is proposed. The proposed scheme differs from dual channel adaptive algorithm in several critical issues which are highlighted through comparison between these two schemes. Afterwards, the validity of the proposed scheme was proved mathematically by showing that the estimated coefficients obtained by the proposed scheme may reach wiener-hopf solution in the long run based on two critical assumptions. Later, the LMS update equation for the proposed scheme was derived and validated mathematically. The two assumptions on signal correlation that degrades the performance of the proposed scheme in reality were handled next by setting some constraints in the update procedure of the LMS algorithm. The constraints were obtained by analyzing the properties of speech frames and also by following the mean square change in consecutive estimations of the LMS filter. In the simulation section, performance of the proposed scheme is evaluated based on improvement of the Signal to distortion ratio (SDR) in dB and also based on the traditional Echo Return Loss Enhancement (ERLE) parameter measured in dB. It is shown in the result section that the performance of the single channel echo cancellation scheme is enhanced to a great extent if the proposed conditions are applied.

Chapter 3

Single Channel Acoustic Echo Cancellation Based on Particle Swarm Optimization Algorithm

The traditional gradient-based adaptive algorithms, such as LMS, Normalized LMS and Recursive-Least-Square (RLS) that have been used in speech enhancement are not suitable for multi-modal error surface, as they are likely to stick in local optima. Moreover, they may sometimes lack the flexibility of controlling some major characteristic parameters: i. the convergence rate, ii. number of iterations, iii. range of variation of filter coefficients, and iv. tolerance consistency. An alternative to gradient-based techniques is the class of stochastic optimization algorithms which are popular in a wide variety of applications. In these algorithms, the probability of encountering the global optimum instead of the local ones is increased and much more flexibility of controlling the characteristic parameters can be achieved.

Conventional optimization algorithms, for example genetic algorithm, Tabu search, and simulated annealing are difficult to implement and they exhibit slow convergence rate [82]. Recently, the particle swarm optimization algorithm (PSO), proposed by Kennedy and Eberhart in 1995 [16], is receiving much attention in areas of power system, computer architecture, and control system [22], [23]. The reason behind choosing PSO as an optimization tool is that it can provide an ease of implementation and a faster convergence rate in comparison to many other optimization algorithms [15].

The main idea in this chapter is to model the task of single channel echo cancellation as an optimization problem instead of employing the conventional gradient-based adaptive filter algorithms. For this purpose, we propose to employ the PSO

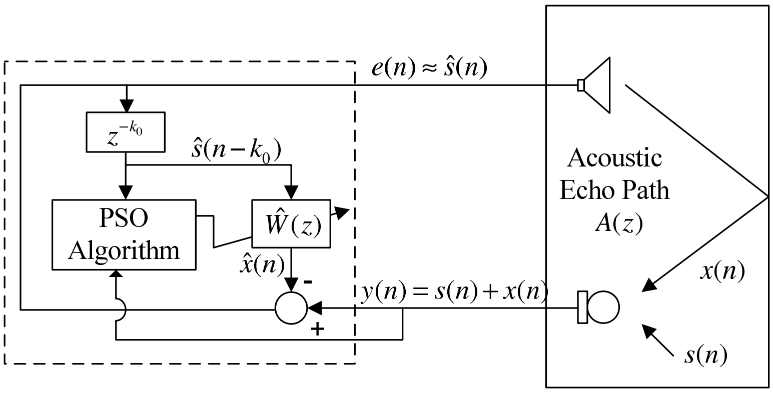

Fig. 3.1: Implementation of the Particle Swarm Optimization (PSO) Algorithm for acoustic echo cancellation in a conference room environment.

algorithm that can efficiently handle the task of multi-variable coefficient optimization. In the proposed PSO based echo cancellation scheme, the filter parameters corresponding to the room response are obtained via an iterative coefficient adaptation while minimizing an error function. The performance of the echo cancellation scheme with respect to the variation of some PSO parameters, such as number of particles and maximum particle velocity is investigated. In order to reduce the computational cost, the flat delay is pre-calculated based on the distance between the speaker and the microphone [2], [4]. Both time domain and frequency domain PSO based echo cancellation schemes have been tested under various acoustic environments [83] [84]. Moreover, the performance of the proposed technique is compared with that of traditional adaptive filter algorithms, namely LMS.

3.1 Proposed Scheme of Adaptive Echo Cancellation

In the previous chapter, it is shown that, in order to solve the single channel acoustic echo cancellation problem gradient based adaptive filter algorithms can be used. However, in this section our objective is to develop a scheme based on the PSO algorithm to handle the echo cancellation problem. In Fig. 3.1, a schematic diagram of the proposed PSO based echo cancellation scheme is shown. Here the microphone input signal $y(n)$ consists of the input speech signal $s(n)$ and the corresponding echo signal $x(n)$. As stated in the previous section, the echo signal $x(n)$ corrupting the speech signal $s(n)$ is generated from the delayed and attenuated version of the same signal $s(n)$ and can be expressed as

$$\begin{aligned} x(n) &= \mathbf{a}_n^T \mathbf{s}(n-k_0) \qquad (3.1) \\ &= \sum_{k=1}^{p} a_n(k) s(n-k_0-k), \qquad (3.2) \end{aligned}$$

where $\mathbf{s}(n-k_0) = [s(n-k_0-1), s(n-k_0-2), \ldots, s(n-k_0-p)]^T$ is a vector of p previous values of $s(n)$ with predefined flat delay k_0. The number p and the values of unknown attenuation coefficients $a_n(k)$ depend on the characteristics of the room.

The task of the adaptive filter block is to produce an estimate of $x(n)$ given $y(n)$ and a reference signal. Since there is no scope to provide a separate reference signal in case of single channel AEC problem, we propose to utilize some delayed versions of the adaptive filter output as the reference signal. However, in case of optimization algorithm based processing, a frame by frame based operation is required. Given a flat delay of k_0 samples, the optimization process starts from k_0 samples and continues frame by frame with a certain percentage of overlap between successive frames. For a frame of N samples, the sum square error $E_{st}(l)$ between the input $(l+1)$-th frame and the corresponding reference frame that the adaptive filter tries to minimize can be defined as

$$\begin{aligned} E_{st}(n) &= \sum_{r=0}^{r=N-1} [y(n-lN-r) - \widehat{x}(n-lN-r)]^2 \qquad (3.3) \\ &= \sum_{r=0}^{r=N-1} [s(n-lN-r) + x(n-lN-r) - \widehat{x}(n-lN-r)]^2 \qquad (3.4) \end{aligned}$$

where, $l = 0, 1, 2 \ldots$ corresponds to the frame number. Here the reference signal $\widehat{x}(n)$ is an estimate of the echo signal generated by the adaptive filter utilizing its estimated coefficient vector $\widehat{\mathbf{w}}_n$ and the echo suppressed input signal $\widehat{s}(n)$ and can be expressed as

$$\begin{aligned} \widehat{x}(n) &= \widehat{\mathbf{w}}_n^T \widehat{\mathbf{s}}(n-k_0) \qquad (3.5) \\ &= \sum_{k=1}^{k=p} \widehat{w}_n(k) \widehat{s}(n-k_0-k). \qquad (3.6) \end{aligned}$$

In the proposed method, unlike conventional approaches, we propose to optimize the objective function stated in equation (3.4) using the particle swarm optimization algorithm (PSO).

Note that unlike sample by sample operation involved in gradient based adaptive filter algorithm, in the proposed PSO based AEC the speech is processed frame by frame with a certain overlap. In order to further demonstrate this frame by frame operation, in Fig. 3.2, the process of estimating the third frame of original input speech from the first and second frames is shown. It can be seen that the first frame has direct effect on the third frame because of the flat delay present in the room response. In this figure, a time shifting window is taken where 25% overlap between successive frames are considered. For frame 2, only frame 1 is available at hand, so we could convolve frame 1 with the estimated room response w_n and get some output. Then again frame 1 is shifted by 25% of its length and the output is obtained. This way the frames are shifted and after shifting it for certain times the output of the whole frame 2. An overlap and add scheme is employed on the resulting outputs to generate a proper estimation of the echo at frame 2. Similar procedure is followed for successive frames as shown in Fig. 3.2. The overlapped segments are obtained by considering time shifted windows from frame 1 and frame 2 as inputs of the estimated filter and thus the echo at frame 3 is obtained.

In the following section two new echo cancellation schemes using PSO in time and frequency domain are proposed.

3.1.1 PSO Approach in Time Domain for Solving AEC

PSO is a population based search procedure in which the individuals, called particles, adjust their position to search through the solution space. Particle swarm adaptation has been shown to successfully optimize a wide range of continuous functions (Angeline, 1998; Kennedy and Eberhart, 1995; Kennedy, 1997; Kennedy, 1998; Shi and Eberhart, 1998). The algorithm, which is based on a metaphor of social interaction, searches a space by adjusting the trajectories of individual particles which can be considered as moving points in multidimensional space. The individual particles are drawn stochastically toward the positions of their own previous best performance and the best previous performance of their neighbors. PSO is a computational intelligence-based technique that is not largely affected by the size and nonlinearity of the problem, and can converge to the optimal solution in many problems where most analytical methods fail to converge [85].

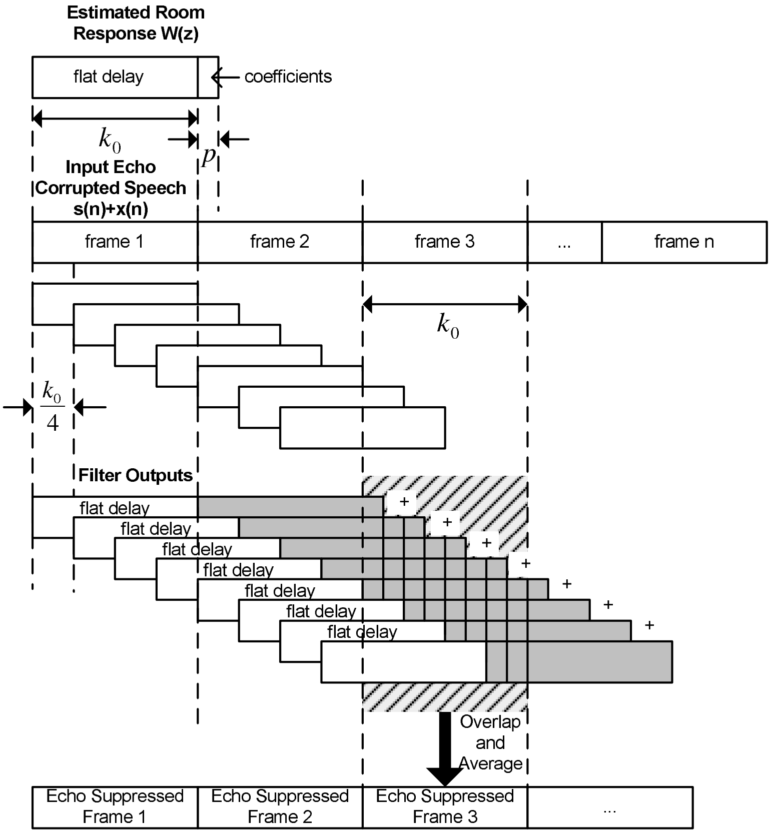

Fig. 3.2: Overlapping and averaging procedure for obtaining proper estimation of echo suppressed signal for frame by frame processing approach

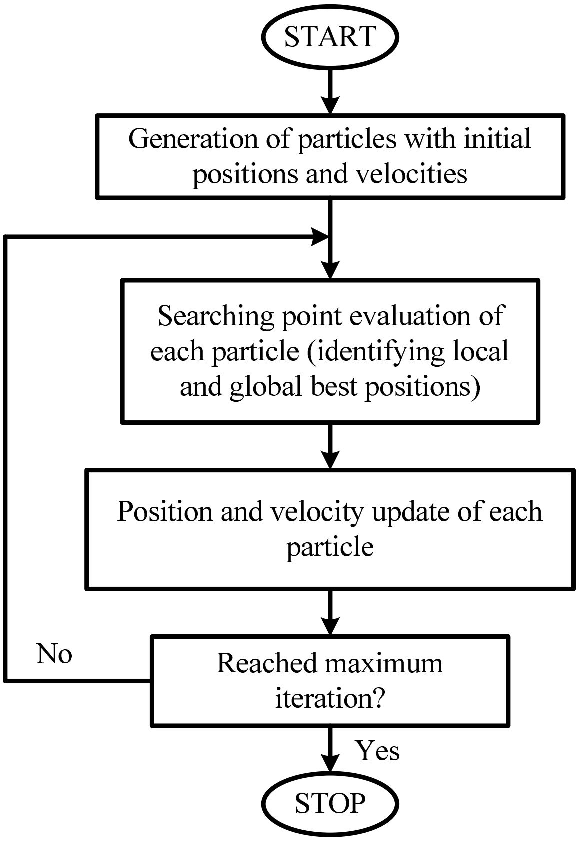
START
Generation of particles with initial positions and velocities
Searching point evaluation of each particle (identifying local and global best positions)
Position and velocity update of each particle
Reached maximum iteration?
No
Yes
STOP

Fig. 3.3: Flowchart of PSO steps

In order to obtain a set of filter coefficients from a given frame, the PSO algorithm performs a number of iterations and in each iteration two major parameters of the particles are updated namely the position vector and the velocity vector. In case of modeling the echo cancellation problem with PSO, the position vector represents the vector of unknown room response which are to be estimated adaptively by the proposed algorithms. At t-th iteration the i-th particle P_t^i has a position vector $\mathbf{w}_t^i$ and a velocity vector $\mathbf{v}_t^i$, where $\mathbf{w}_t^i = (w_t^i(1), w_t^i(2), ..., w_t^i(k), ...w_t^i(p))$ and $\mathbf{v}_t^i = (v_t^i(1), v_t^i(2), ..., v_t^i(p))$. Here, p is the number of unknowns filter coefficients. At each iteration the particle learns from its own previous best position λ_t^i and the best position of all the other particles χ_t^i in the swarm, and updates its' velocity and position. The update equation for the i-th particle in order to obtain the k-th unknown filter coefficient at the $(t+1)$-th iteration can be written as

$$\begin{aligned} v_{t+1}^i(k) &= \Delta_t v_t^i(k) + c_1 r_{1t}^i(k)(\lambda_t^i(k) - w_t^i(k)) \\ &\quad + c_2 r_{2t}^i(k)(\chi_t^i(k) - w_t^i(k)), k = 0, 1, \ldots, p \qquad (3.7) \\ w_{t+1}^i(k) &= w_t^i(k) + v_{t+1}^i(k), k = 0, 1, \ldots, p \qquad (3.8) \end{aligned}$$

where c_1 and c_2 are the cognitive and social scaling parameters, respectively, and $r^i_{1t}(k)$ and $r^i_{2t}(k)$ are random numbers in the range of [0, 1] generated at the t-th iteration. In (3.7), the inertia weight Δ_t at the t-th iteration is used to maintain the particles' momentum. The update equation for the inertia weight at the $(t+1)$-th iteration can be expressed as

$$\Delta_{t+1} = \Delta_{initial} - \frac{\Delta_{initial} - \Delta_{final}}{t_{max}}.t, \tag{3.9}$$

where t_{max} represents the maximum number of iterations. $\Delta_{initial}$ and Δ_{final} are the maximum and minimum values of the inertia weight, respectively.

The role of the inertia weight Δ_t, in (3.7), is considered critical for the PSOs convergence behavior. The inertia weight is employed to control the impact of the previous history of velocities on the current one. Accordingly, the parameter w regulates the trade-off between the global (wide-ranging) and local (nearby) exploration abilities of the swarm. A large inertia weight facilitates global exploration (searching new areas), while a small one tends to facilitate local exploration, i.e., fine-tuning the current search area [20]. A suitable value for the inertia weight Δ_t usually provides balance between global and local exploration abilities and consequently results in a reduction of the number of iterations required to locate the optimum solution. The parameters c_1 and c_2, are not critical for PSOs convergence. However, proper fine-tuning may result in faster convergence and alleviation of local minima [76]. As default values, $c_1 = c_2 = 2$ were proposed, but experimental results indicate that $c_1 = c_2 = 0.5$ might provide even better results. Recent work reports that it might be even better to choose a larger cognitive parameter, c_1, than a social parameter, c_2, but with $c_1 + c_2 \leq 4$ [86]. The parameters $r^i_{1t}(k)$ and $r^i_{2t}(k)$ are used to maintain the diversity of the population.

As there are very few parameters to be adjusted in the PSO algorithm compared to other evolutionary optimization algorithms and the updating procedure involves only simple arithmetic operations, it is a good choice for fast optimization [15] [20].

In Fig. 3.4, a detailed view of the position and operation of the PSO-TD algorithm block in the proposed scheme is shown. The PSO-TD algorithm block takes a frame from the current input signal $y(n)$ and another reference frame from the previously enhanced signal $\widehat{s}(n-k_0)$. Its position and velocity vectors are randomly initialized, which means at the beginning the algorithm considers a random estimate

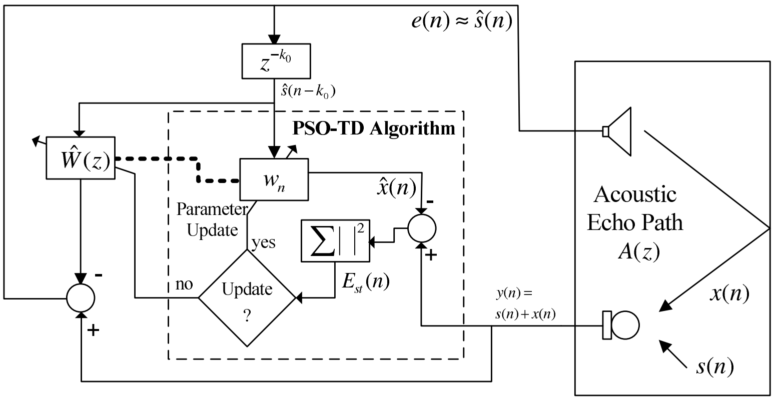

Fig. 3.4: An insight of the PSO-TD algorithm block of the proposed single channel AEC.

of the room response. From this estimate, an error $E_{st}(n)$ is calculated using equation (3.4). The PSO-TD algorithm then updates the velocity and position vectors of each particles and again calculates the sum square error $E_{st}(n)$ for all the particles. This iterative process of error calculation and parameter update continues until a maximum number of iteration is reached or the difference between two successive updates become stable for a certain number of iteration. The best positional value, i.e. the best estimate of the room response filter coefficient thus obtained, is transferred to the $\widehat{W}(z)$ block to be used for final echo suppression, as shown by the dotted lines.

It is to be mentioned that the main controlling parameters of the PSO algorithm affecting the echo cancellation performance are the number of particles, maximum particle velocity, and number of iterations. Initially a set of randomly chosen values within a certain range depending on the characteristic of the room response is assigned to position and velocity vectors of each particle. These values of position and velocity vectors are updated iteratively to find the best position among all the particles which corresponds to the optimum filter parameters. Setting a narrower search region by intelligent decision may enhance the iterative procedure. It is expected that the PSO algorithm will converge quickly to the desired values with a very low level of estimation error.

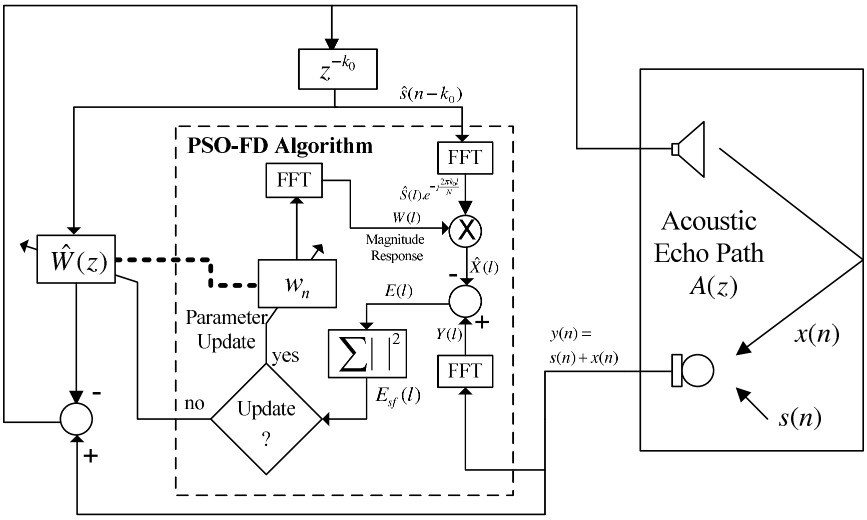

Fig. 3.5: Schematic of proposed PSO based frequency domain echo cancellation scheme

3.1.2 Proposed PSO-Based Frequency-Domain AEC

In many practical applications, it is found that the frequency domain analysis becomes more insightful and provides ease of operation. Proceeding in a similar fashion as followed in the case of time domain analysis, in what follows our objective is to develop the proposed PSO based AEC scheme (PSO-FD) frequency domain.

In the proposed PSO based frequency domain analysis the discrete fourier transform (DFT) of each frame of input data $y(n)$ is performed which is defined as

$$Y(l) = \sum_{n=-\infty}^{+\infty} y(n)e^{-j\frac{2\pi f n}{N}l} \tag{3.10}$$

In Fig. 3.5 a schematic diagram of the proposed frequency domain AEC method is shown. A frame of the input signal $y(n)$ is supplied to the PSO-FD algorithm block while another frame of echo suppressed signal $\widehat{s}(n-k_0)$ is supplied as reference. The particles are initialized with random position and velocities. The position vector, the current frame and the reference frame - all are transformed into the frequency domain by discrete fourier transform and are represented as $W(l)$, $Y(l)$ and $\widehat{S}(l).e^{-j\frac{2\pi k_0 l}{N}}$, respectively. As the time domain convolution becomes multiplication in frequency domain, the new estimate of the echo in frequency domain $\widehat{X}(l)$ can be denoted as

$$\widehat{X}(l) = \widehat{W}(l).\widehat{S}(l).e^{-j\frac{2\pi k_0 l}{M}}, \tag{3.11}$$

Thus, the error $E(l)$ is defined in the discrete frequency domain as

$$E(l) = Y(l) - \widehat{X}(l), \tag{3.12}$$

Now, the objective function for optimization can be defined as the sum of the square of the error $E_{sf}(l)$, i.e,

$$E_{sf}(l) \quad = \quad \sum_{l=0}^{N-1}(|Y(l) - \widehat{X}(l)|)^2$$

The PSO adaptive algorithm tries to minimize this error by varying the position of its particles, i.e. by varying the estimated echo path filter coefficients. Here it can be seen that, though the proposed method calculates the mean square error in the frequency domain, it updates the time domain form of the FIR filter $\widehat{w}_n$ not its frequency response. The update of the velocity and position vectors of the particles is an iterative process. The PSO-FD block takes a frame of input signals and calculates the sum square error for the present positions of all the particles. Then according to the rules of the PSO algorithm update, it updates the velocity and position of all the particles. The sum square error is calculated again for the new position vectors and the position update and error calculation is repeated unless a predefined maximum number iteration is reached or the difference between two consecutive updates are stable for a certain number of iterations. The error function is minimized when the filter coefficients of the model echo path $\widehat{w}_n$ are perfectly tuned with the room impulse response a_n. The updated position w_n which is finally obtained is then used as the estimated filter coefficients $\widehat{W}(z)$ of the room response (as shown by the dotted line between the two blocks in Fig. 3.5)to minimize the effect of echo from the current input signal frame. Since time domain convolution becomes multiplication in the frequency domain approach, it is expected that frequency domain modeling of the echo cancellation problem will involve less computation resulting in faster convergence.

3.2 Simulation Results

Extensive experimentation has been carried out in order to investigate the echo cancellation performance of the proposed PSO based time domain (PSO-TD) and

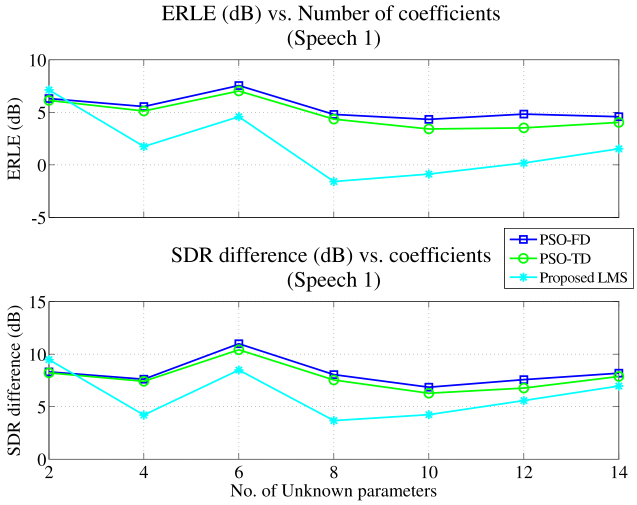

Fig. 3.6: Performance comparison of the proposed PSO-TD, PSO-FD and LMS algorithms with increasing coefficients in terms of ERLE (dB) and SDR difference (dB).

frequency domain (PSO-FD) approaches and results are compared with the state-of-the-art adaptive LMS filer algorithm based scheme proposed in chapter 2. Thus, for the purpose of simulating various acoustic environments, the room impulse response defined in the simulation section of chapter 2 is considered. The two test speech samples are also the same as in chapter 2. The improvement of signal to distortion ratio (SDR) in dB and the average echo return loss enhancement (ERLE) in dB are used as basis for performance evaluation.

3.2.1 Performance Comparison

In Fig. 3.6 and Fig. 3.7 echo cancellation performance obtained by different methods are presented considering the number of unknown coefficients incrementing from 2 to 14. As performance measurement criteria, we consider the ERLE (dB) and the SDR difference (dB). In our experimentation, different sets of PSO parameters have been used. However, the results obtained from the proposed methods in these figures utilize the following PSO parameters: number of particles is 10 for PSO-TD and 40 for PSO-FD, tolerance between two consecutive global best values $= 10^{-30}$, search range for coefficients is $[-1\ 1]$, maximum number of iterations is 40, $c_1 = 2$, $c_2 = 2$, $w_{initial} = 0.9$, $w_{final} = 0.4$, and maximum particle velocity $v_{max} = 0.2$.

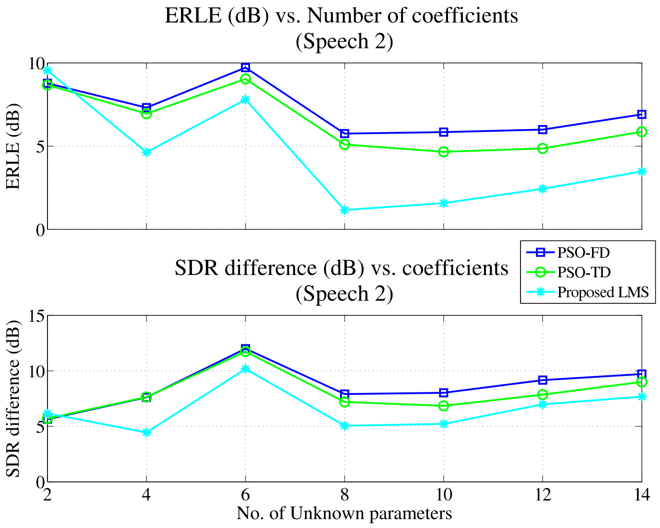

Fig. 3.7: Performance comparison of the proposed PSO-TD, PSO-FD and LMS algorithms with increasing coefficients in terms of ERLE (dB) and SDR difference (dB).

Iterations terminate if tolerance is stable for 5 iterations. From the figures, it can be observed that, the performance in terms of the ERLE (dB) is consistently good for the proposed PSO-TD method than that of the LMS algorithm. Moreover, the proposed PSO-FD approach is found to be superior to other two methods in terms of both ERLE and SDR difference.

3.2.2 Performance Analysis by Varying PSO Parameters

It is found that among all the PSO parameters, the most influential one is the number of particles. In Fig. 3.8, the effect of variation in number of particles on the ERLE obtained for both the proposed PSO-TD and PSO-FD approaches are shown. Similarly the effect of variation of number of particles on SDR improvement is shown in Fig. 3.9. In both cases, the number of filter coefficients was set to 14, keeping the other parameters same as before. The experiments show that the values of both ERLE and SDR difference remain quite stable after a certain number of particles. Also, it is obvious that with the increasing number of particles, the processing time per iteration would increases, which is illustrated in Fig. 3.10. Thus, the number of particles to be chosen to obtain a better performance is governed by the time constrain, if any. It is evident from Fig. 3.10 that PSO-FD is a faster algorithm

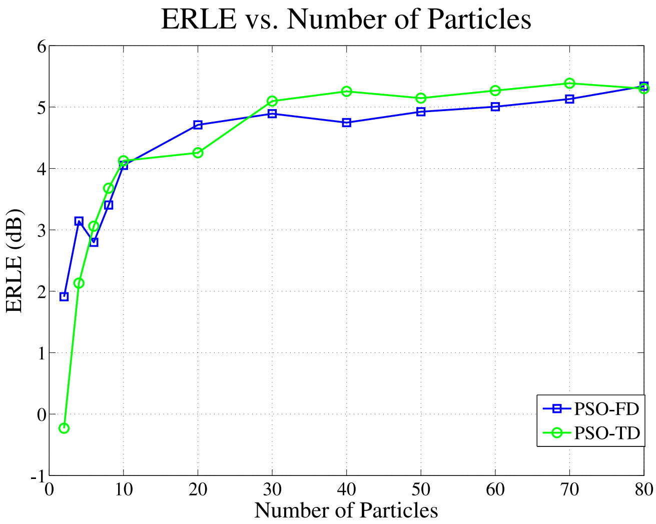

Fig. 3.8: Effect of variation of number of particles on ERLE (dB)

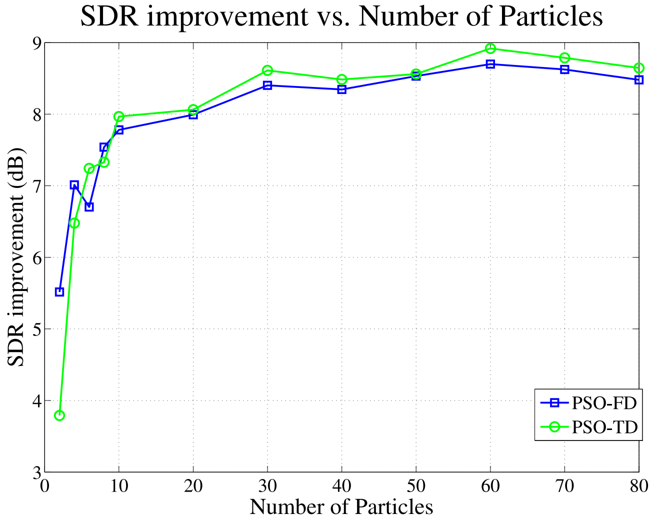

Fig. 3.9: Effect of variation of number of particles on SDR improvement (dB)

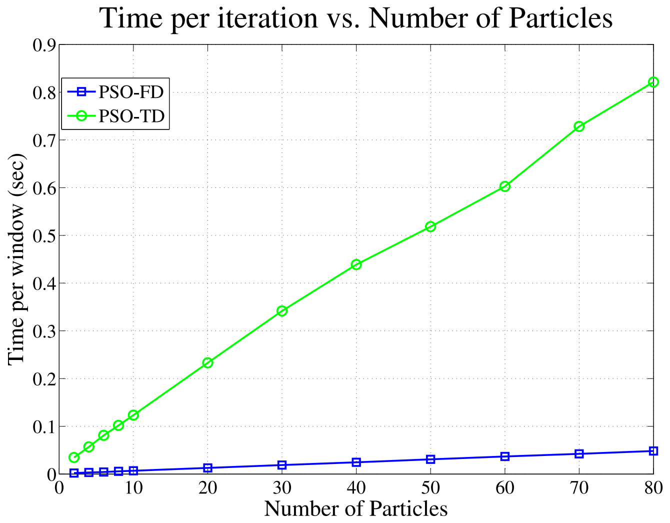

Fig. 3.10: Processing time per iteration vs. number of particles

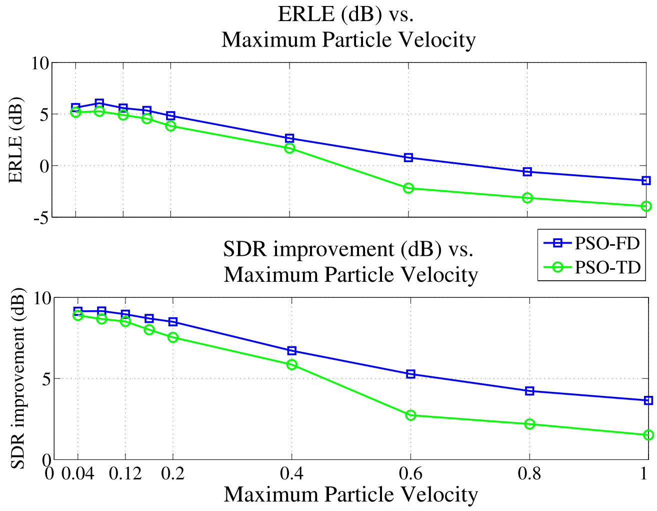

Fig. 3.11: Effect of variation of maximum particle velocity on (a) ERLE (dB) and (b) SDR difference

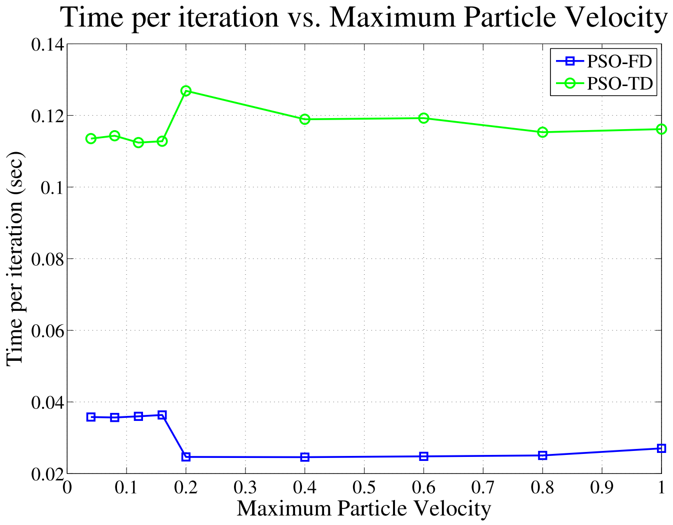

Fig. 3.12: Effect of variation of maximum particle velocity on time per iteration (sec)

than PSO-TD. Which allows us to take the liberation of using a good number of particles when using PSO-FD and still get a very good cancellation performance within a short time. In Fig. 3.11, the effect of variation in the maximum particle velocity on the ERLE (dB) and also the SDR difference (dB) obtained for both the proposed PSO-TD and PSO-FD approaches are shown. It is evident from the figure that smaller values of maximum particle velocity produces better results. However, it is understandable that the smaller the velocity the more time will the particles take to converge. So, for a limited number of iterations a decision has to be made about choosing a proper maximum particle velocity. Logically, the change in maximum particle velocity has no effect on the time per iteration as can be seen from Fig. 3.12.

3.3 Conclusion

In this chapter, a novel approach of single channel acoustic echo cancellation scheme using an optimization algorithm is proposed. Then a frame by frame overlap-add method is illustrated for window based processing. For fast and accurate optimization the renowned particle swarm optimization(PSO) algorithm is chosen for the scheme. PSO is a population based search technique which is superior to some other evolutionary algorithms, namely genetic algorithm, taboo search etc. After a brief introduction on PSO, two different PSO based schemes were proposed for single channel AEC. One is based on time domain operation of PSO while the other is based on frequency domain transformation of the signals before applying optimization algorithm. In the simulation section, the performance of the proposed schemes were evaluated in terms of ERLE and SDR difference by comparing with one another and with the previously described LMS based single channel AEC technique. It is found that the performance of the PSO based frequency domain single channel echo cancellation scheme (PSO-FD) outperforms both PSO-TD and traditional LMS and offers a great reduction of estimation error. Also, an analysis of the parameters of PSO showed that the echo cancellation performance becomes quite stable after a certain number of particles. An analysis of time per iteration revealed that PSO-FD is much faster than PSO-TD and thus it can easily fulfill many time constraints when PSO-TD may fail.

Chapter 4

Single Channel Integrated Acoustic Echo and Noise Cancellation Based on Adaptive LMS Algorithm

In this chapter we deal with a very difficult single channel scenario of acoustic echo cancellation (AEC) where apart from the echo signal environment noise is added with the input speech signal. As a result the problem is not only to remove the echo signal but also to reduce the environmental noise. Unlike conventional adaptive echo cancellation schemes where it is assumed that two channels are available, the problem at hand becomes extremely difficult as both echo and noise have to be suppressed given only a single channel input data. We propose a two stage scheme where first the echo cancellation is performed based on adaptive least mean square (LMS) algorithm and then a spectral subtraction based single channel noise suppression technique is incorporated in the design to cancel out the noisy parts of the echo suppressed signal and to produce an enhanced speech [87]. It is to be mentioned that the LMS based adaptive filter algorithm developed for adaptive echo cancellation under noise free condition in chapter 2 has been modified considering the effect of additive noise and corresponding adaptive characteristics have been further investigated. Performance of the proposed scheme has been tested for various echo corrupted speech signal under different noisy conditions.

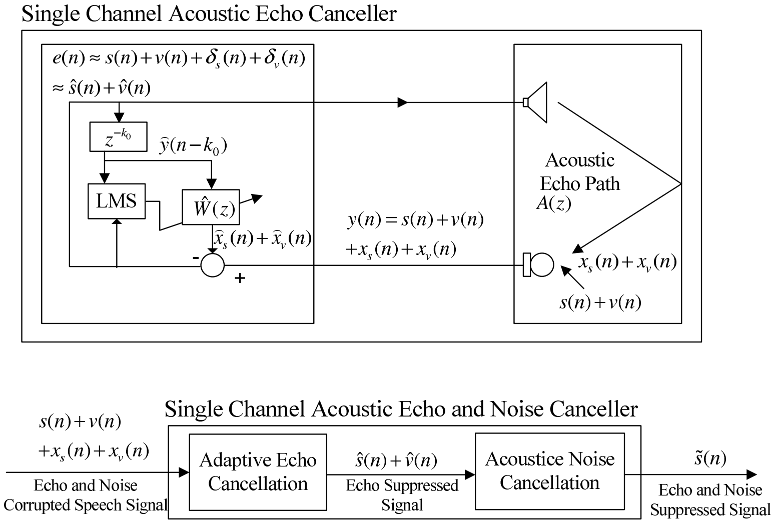

Fig. 4.1: LMS based single channel integrated acoustic echo and noise cancellation in room environment.

4.1 Single Channel Integrated Echo and Noise Canceller

4.1.1 Problem Formulation

In this section an effective solution of the rarely addressed problem of single channel acoustic echo cancellation at noisy room environment using gradient based adaptive filtering methods is developed. The scenario is extremely difficult to handle as it assumes that no reference channel is available for neither echo nor noise cancellation. Conference room environment and hearing aid systems are examples of such situations, where the speech signal itself is reflected and their attenuated version is fed back to the sole microphone as echo. In addition, environmental noise, which degrades the intelligibility of the speech, is needed to be suppressed too. In Fig. 4.1 a schematic diagram demonstrating the proposed single channel integrated echo and noise cancellation scheme (AENC) is shown. As seen from the figure, the input speech $s(n)$ is contaminated with environmental noise $v(n)$. In addition, the echo signals are fed to the input microphone. Note that, both speech and noise signals will be reflected back and produce echo signals independently considering the principle of linearity holds. The echo signals corresponding to $s(n)$ and $v(n)$, denoted as $x_s(n)$ and $x_v(n)$ in Fig. 4.1, being mixed with the noise corrupted input signal results in the combined input signal given by

$$y(n) = s(n) + v(n) + x_s(n) + x_v(n). \tag{4.1}$$

The main idea of the proposed echo cancellation algorithm under noisy condition is to adaptively obtain estimate of the the echo portions of the combined input signal $y(n)$, namely $\widehat{x}_s(n) + \widehat{x}_v(n)$. In this regard our task is to produce optimum values of unknown filter coefficients $\widehat{\mathbf{w}}_n$ from given previous echoless samples of the noisy speech, such that the resulting signal $\widehat{x}_s(n) + \widehat{x}_v(n)$ closely matches $x_s(n) + x_v(n)$. The error signal $e(n)$ thus obtained is given by

$$e(n) \;\; = \;\; y(n) - [\widehat{x}_s(n) + \widehat{x}_v(n)], \tag{4.2}$$

where the estimate of the echo signal can be expressed as

$$\begin{aligned} \widehat{x}_s(n) + \widehat{x}_v(n) &= \widehat{\mathbf{w}}_n^T[\widehat{\mathbf{s}}(n-k_0) + \widehat{\mathbf{v}}(n-k_0)] && (4.3) \\ &= \sum_{k=1}^{k=p} \widehat{w}_n(k)[\widehat{s}(n-k_0-k) + \widehat{v}(n-k_0-k)]. && (4.4) \end{aligned}$$

One possible way is to develop a gradient based adaptive algorithm to obtain an accurate estimate of the room response coefficient by minimizing the error $e(n)$. Resulting echo free signal is then fed to the noise cancellation block to reduce the effect of $v(n)$. However, the complete elimination of the echo part under noisy condition may not be possible which may leave a residual echo signal along with the noise corrupted speech. In this case, the error signal in 4.2 can be expressed as

$$\begin{aligned} e(n) &= s(n) + v(n) + x_s(n) + x_v(n) - \widehat{x}_s(n) - \widehat{x}_v(n) && (4.5) \\ &= s(n) + v(n) + \delta_s(n) + \delta_v(n) && (4.6) \\ &= (s(n) + \delta_s(n)) + (v(n) + \delta_v(n)) && (4.7) \\ &= \widehat{s}(n) + \widehat{v}(n) && (4.8) \end{aligned}$$

Here, the terms $\delta_s(n)$ and $\delta_v(n)$ are introduced to represent the residual echo of the speech and noise portions of the input signal, respectively and it is assumed that these signals exhibit the properties of white gaussian noise. Next, this error signal is fed to the noise subtraction block. In the proposed method, a spectral subtraction based noise suppression algorithm is used, where an accurate estimation of noise spectral floor is a difficult task depending on the noise characteristics. If $v(n)$ is

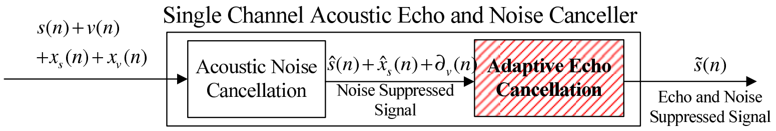
Single Channel Acoustic Echo and Noise Canceller
$s(n)+v(n)$
$+x_s(n)+x_v(n)$
Acoustic Noise Cancellation
$\hat{s}(n)+\hat{x}_s(n)+\partial_v(n)$
Noise Suppressed Signal
Adaptive Echo Cancellation
$\tilde{s}(n)$
Echo and Noise Suppressed Signal

Fig. 4.2: Single channel noise canceller followed by an echo canceller.

assumed to be a white Gaussian noise, it is expected that the outcome of the noise suppression operation would be $s(n) + \Psi(n) \approx \tilde{s}(n)$, which is close to the clean speech $s(n)$ as the other three parts of equation (4.6) would be minimized. Since the amount of residual echo and noise, denoted as $\Psi(n)$, is very small after the operation of the noise suppressor, the audience would experience echo and noise-less speech of good quality.

In case of single channel AENC schemes, an important issue is the order of performing the tasks of echo cancellation and noise reduction. In the proposed scheme, as shown in Fig. 4.1 acoustic echo cancellation operation is performed before the noise cancellation. As an alternative approach, the noise reduction operation can be performed before the acoustic echo cancellation (shown in Fig. 4.2), which is being used in some dual channel noise-sensitive adaptive echo cancellation applications [35].

However, for single channel AENC scheme, as shown in Fig. 4.2, because of possible nonlinearities introduced by the prior noise reduction block, no proper reference would be available for the AEC block to cancel echo from the outputs of the noise reduction block. On the contrary, when the acoustic echo cancellation is performed before the noise reduction operation there is no chance of introduction of nonlinearity prior to AEC operation. Moreover in this approach, the noise reduction block will serve as a post-processor for attenuating the residual echo. Therefore, in the proposed single channel AENC scheme a noise reduction block is used after the adaptive echo cancellation block.

4.1.2 Dual Channel vs. Single Channel AENC

In Fig. 4.3 a conventional dual channel integrated echo and noise cancellation block is introduced [35] to remove the effect of echo from the echo corrupted signal,

$$y_1(n) = s_1(n) + v_1(n) + x_2(n), \tag{4.9}$$

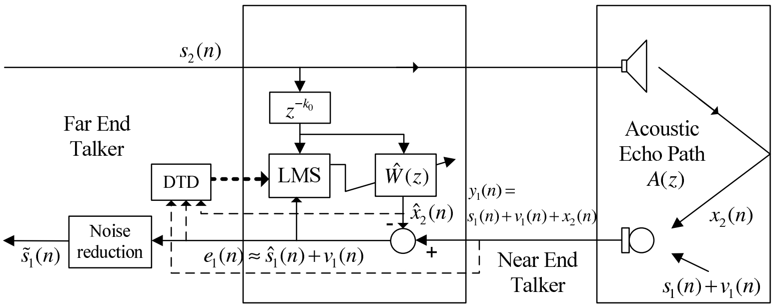

Fig. 4.3: LMS adaptive algorithm for dual channel AENC scheme in communication system.

where $s_1(n)$ is the input speech of the near-end speaker, $v_1(n)$ is the input environmental noise at the near end and $x_2(n)$ is the echo of the far end signal $s_2(n)$. For minimizing the echo signal, some adaptive filter algorithm can be used, where given a reference signal an estimate of the echo part $x_2(n)$ of $y_1(n)$ is generated based on the minimization of an error function $e_1(n)$ defined as,

$$\begin{aligned} e_1(n) &= y_1(n) - \widehat{x}_2(n) && (4.10)\\ &= s_1(n) + v_1(n) + x_2(n) - \widehat{x}_2(n), && (4.11) \end{aligned}$$

As mentioned earlier, $x_2(n)$ is an attenuated and delayed version of $s_2(n)$ which, based on linear prediction theory can be expressed as,

$$\begin{aligned} x_2(n) &= \mathbf{a}_n^T \mathbf{s}_2(n - k_0) && (4.12)\\ &= \sum_{k=1}^{p} a_n(k) s_2(n - k_0 - k), && (4.13) \end{aligned}$$

where, $\mathbf{s}_2(n-k_0) = [s_2(n-k_0-1), s_2(n-k_0-2), \ldots, s_2(n-k_0-p)]^T$ is a vector of p previous values of s_2 with predefined flat delay k_0 and $\mathbf{a}_n = [a_n(1), a_n(2), \ldots, a_n(p)]^T$ is the vector of the unknown room response coefficients. The number p of unknown attenuation coefficients $a_n(k)$ depends on the characteristics of the room.

Here, the task of an adaptive filter for echo cancellation is to produce optimum values of unknown filter coefficients $\widehat{\mathbf{w}}_n$ from given $s_2(n-k_0)$ such that the resulting signal $\widehat{x}_2(n)$ closely matches $x_2(n)$, i.e,

$$\begin{aligned} \widehat{x}_2(n) &= \widehat{\mathbf{w}}_n^T \mathbf{s}_2(n - k_0) && (4.14)\\ &= \sum_{k=1}^{p} \widehat{w}_n(k) s_2(n - k_0 - k), && (4.15) \end{aligned}$$

Here, $\widehat{\mathbf{w}}_n = [\widehat{w}_n(1), \widehat{w}_n(2) \ldots \widehat{w}_n(p)]^T$ is the estimated attenuation vector. The value of p also signifies the number of unknown parameters to be estimated from the system.

Under optimum condition, $\widehat{\mathbf{w}}_n = \mathbf{a}_n$. Next the task of the noise reduction block is to suppress the input noise $v_1(n)$ from the echo reduced signal. Finally the noise and echo suppressed signal $\widetilde{s}_1(n)$ is obtained which mostly represents the input signal $s(n)$ with some additional residual echo and noise.

An extremely important issue of designing adaptive echo cancelers for dual channel is to handle double talk, which occurs when the far-end and near-end talkers are speaking simultaneously. In this case, the far end signal consists of both echo $x_1(n)$ and far-end speech $s_2(n)$. During the double-talk periods, the error signal $e(n)$ described in Equation (4.11) contains the residual echo and the near-end speech $s_1(n)$. To correctly identify the characteristics of $A(z)$, the near-end signal must originate solely from its input signal from the far end. An effective solution, as shown in figure 2.2, is to detect the occurrence of double talk using a double talk detector (DTD) and then to disable the adaptation of $\widehat{W}(z)$ during the double-talk periods. If the echo path does not change during the double-talk periods, the echo can be canceled by the previously estimated $\widehat{W}(z)$, whose coefficients are fixed during double-talk periods.

Now, proper formulation of a single channel AENC scheme would be extremely difficult in comparison to that in dual channel case because of the following reasons,

(1) In dual channel AENC, two channels are dedicated to receive inputs from two different speakers and generally, a dual talk detector (DTD) is used. In this case, one channel carries the noisy speech signals from person 1, namely $s_1 + v_1(n)$, along with the echo signal corresponding to person 2, namely $x_2(n)$. Because of the presence of the DTD, the echo canceller can exploit the advantage of having a reference of echo-free signal from channel 2 $s_2(n)$ to cancel the echo portion of the input signal of channel 1, namely $x_s(n)$ and vice versa. On the other hand, single channel AEC deals with a one speaker and the echo itself is originated by the same speaker speaking in the microphone. Both the noisy speech and echo propagation is carried out by a single channel. The most difficulty here, unlike the dual channel case, is to obtain a separate reference signal for the AEC block to cancel out the echo

portion from the input echo-corrupted noisy signal. There is no scope of receiving reference signals for echo estimation from another channel.

(2) As a result, in the proposed single channel AENC scheme, a echo suppressed noisy speech sample is used as reference for the next samples. When suppressing echo in a certain sample, there may be some residual echo present in the cleaned speech (that is why it is denoted by $\widehat{s}(n)+\widehat{v}(n)$ rather that $s(n)+v(n)$ itself). Thus, if the echo reduced current noisy speech sample is used as reference for cancellation of echo of a future sample, it would obviously generate some error leading to suppression of speech signal along with noise in the noise reduction block. So, getting a very high degree of overall system performance using only the traditional adaptive filter algorithms may not be expected in case of single channel AENC.

(4) In case of single channel AENC, the input speech is contaminated with noise and the noise part is also reflected and fed to the microphone along with the speech. Therefore, noise reference is also needed to adaptively cancel the reflected noise $x_v(n)$ from current sample. That is why, the reference is taken before the final noise suppression. To suppress the additive noise input at the current sample, a noise reduction block is introduced which minimizes the current noise $v(n)$ as well as some parts of the residual echo $\delta_s(n)+\delta_v(n)$.

(3) In case of single channel AENC, the speech from a speaker is contaminated by attenuated previous samples of speech of the same speaker, which increases the probability of the speech and echo to be correlated to some extent. Whereas, in the case of two channel communication, since echo and noisy speech signals are coming from two different speakers, the degree of correlation would be much lower.

(7) In Fig. 2.4 the room acoustic response is denoted by $A(z)$ and its estimation for echo cancellation by the adaptive LMS filter is denoted by $\widehat{W}(z)$. The number p of unknown attenuation coefficients of the room $a_n(k)$ depends on the characteristics of the room. The duration of the room response also depends on flat delay of k_0 samples, which is the minimum time taken by a speech sample to travel from the loudspeaker to the microphone. In a real conference room environment, the value of k_0 has to be very large for human perception to distinguish an echo from the original signal because when dealing with audible frequencies, the human ear cannot distinguish an echo from the original sound if the delay is less than 100 millisecond.

Thus, the echo estimation task has to deal with a large filter on the order of thousand coefficients. However, the value of the room response $\widehat{w}_n(k)$ is generally considered to be zero for lower values of k. That is why, the variable k_0 is introduced. The value of k_0 can be thousand or more depending on the room acoustic and it symbolizes the amount of flat delay (for which the value of $\widehat{w}_n(k)$ is zero). On the other hand, in case of two channel communication, the value of k_0 may be as small as a single sample and is not significant at all. The goal of two channel echo cancellation is to cancel the echo from the other channel so that the person speaking in one channel could not hear his/her own voice through the loudspeaker while talking. It is not customary in this case for the room environment of the other end, where the signal is being fed back to the microphone from the loudspeaker, to be like a large conference hall which will produce large delay or long echo trail. Dual channel echo occurs simply when, the loudspeaker output is coupled to the microphone input in any end of the communication link in any possible way.

4.2 Analysis of the proposed single channel integrated echo and noise canceller based on LMS algorithm

The echo signals $x_s(n)$ and $x_v(n)$ corrupting the noisy input signal $s(n) + v(n)$ is generated from the delayed and attenuated version of the signals $s(n)$ and $v(n)$ respectively and can be expressed as,

$$\begin{aligned} x_s(n) &= \mathbf{a}_n^T \mathbf{s}(n-k_0) && (4.16)\\ &= \sum_{k=1}^{p} a_n(k) s(n-k_0-k) && (4.17)\\ x_v(n) &= \mathbf{a}_n^T \mathbf{v}(n-k_0) && (4.18)\\ &= \sum_{k=1}^{p} a_n(k) v(n-k_0-k), && (4.19) \end{aligned}$$

where, $\mathbf{s}(n-k_0) = [s(n-k_0-1), s(n-k_0-2), \ldots, s(n-k_0-p)]^T$ is a vector of p previous values of $s(n)$ with predefined flat delay k_0. Similarly, $\mathbf{v}(n-k_0)$ is a vector of p previous values of the input noise $v(n)$ with predefined k_0 flat delay. The number p of unknown attenuation coefficients $a_n(k)$ depends on the characteristics of the room.

The task of the adaptive filter block is to produce estimates of $x_s(n)$ and $x_v(n)$ given $y(n)$ and a reference signal. Similar to the single channel AEC problem stated in chapter 2, we intend to utilize some delayed versions of the adaptive filter output as the reference signal. The error signal which the adaptive filter tries to minimize shown in equation (4.6). Thus, in the adaptive filter algorithm, the effect of echo in $y(n)$ is iteratively minimized utilizing a certain number of previous samples of $\widehat{s}(n)$ and $\widehat{v}(n)$ as reference. The outcome of the adaptive filter block is $\widehat{s}(n) + \widehat{v}(n)$, where, $\widehat{s}(n) = s(n) + \delta_s(n)$ and $\widehat{v}(n) = v(n) + \delta_v(n)$, i.e. both the original input signal and original input noise are contaminated with some estimation error which predominantly exhibits noise-like behavior. With the increasing iterations towards an optimum solution, $\delta_s(n) + \delta_v(n)$ tends to vanish gradually resulting $\widehat{s}(n) + \widehat{v}(n) = s(n) + v(n)$.

Thus the objective function in this case can be defined as the mean square estimation of the error function, namely,

$$\begin{aligned} J_n &= E\{e^2(n)\} = E\{[y(n) - \widehat{x}_s(n) - \widehat{x}_v(n)]^2\} && (4.20)\\ &= E\{[s(n) + v(n) + x_s(n) + x_v(n) - \widehat{x}_s(n) - \widehat{x}_v(n)]^2\} && (4.21)\\ &= E\{(s(n) + v(n))^2\} + E\{[x_s(n) + x_v(n) - \widehat{x}_s(n) - \widehat{x}_v(n)]^2\} \\ &\quad +2E\{[(s(n) + v(n))(x_s(n) + x_v(n) - \widehat{x}_s(n) - \widehat{x}_v(n))]\}, && (4.22)\end{aligned}$$

where, the last term of right hand side of the objective function can be expressed as

$$\begin{aligned} & 2E\{[(s(n) + v(n))(x_s(n) + x_v(n) - \widehat{x}_s(n) - \widehat{x}_v(n))]\} \\ = \; & 2\sum_{k=1}^{k=p}\{(a_n(k) - \widehat{w}_n)(r_{ss}(k_0 + k) + r_{sv}(k_0 + k) + r_{vs}(k_0 + k) + r_{vv}(k_0 + k)) \\ & -r_{s\delta_s}(k_0 + k) - r_{s\delta_v}(k_0 + k) - r_{v\delta_s}(k_0 + k) - r_{v\delta_v}(k_0 + k)\} && (4.23)\end{aligned}$$

where, $r_{ss}(n)$ corresponds to the cross-correlation between the input signal $s(n)$ and its previous samples $s(n - k_0 - k)$. The magnitude of $r_{ss}(n)$ strongly depends on speech characteristics and the amount of flat delay k_0. Similarly, $r_{sv}(n)$ corresponds to the cross-correlation between $s(n)$ and $v(n - k_0 - k)$, $r_{vs}(n)$ corresponds to the cross-correlation between $v(n)$ and $s(n - k_0 - k)$ and $v_{sv}(n)$ corresponds to the cross-correlation between $v(n)$ and $v(n-k_0-k)$. As we have considered $v(n)$ to be a white noise, it has no correlation with itself and with $s(n)$. Thus, the terms $r_{sv}(n)$, $r_{vs}(n)$ and $r_{vv}(n)$ all tends to zero. Also, the terms $\delta_s(n)$ and $\delta_v(n)$ have noise-like characteristics, thus in equation (4.23), we can assume that $r_{s\delta_v}(n) \approx r_{v\delta_s}(n) \approx r_{v\delta_v}(n) \approx 0$. So, we can easily comprehend that optimal performance of the filter occurs when $r_{ss}(n)$ is minimum, i.e. the least possible correlation between $s(n-k_0-k)$ and $s(n)$ is desired. In that case, the correlation between reverberant and non-reverberant part of the input signal will also be minimum making the single channel echo cancellation problem easier. As a result, the objective function in equation (4.22) reduces to,

$$E\{e^2(n)\} \quad = \quad E\{(s(n)+v(n))^2\} + E\{[x_s(n)+x_v(n)-\widehat{x}_s(n)-\widehat{x}_v(n)]^2\}(4.24)$$

Minimization of the objective function (4.24) results in,

$$\frac{\delta J_n}{\delta \widehat{\mathbf{w}}_n^T} = 0 \qquad (4.25)$$

$$E\{[x_s(n)+x_v(n)-\widehat{x}_s(n)-\widehat{x}_v(n)]\sum_{k=1}^{k=p}(\widehat{s}(n-k_0-k)+$$
$$\widehat{v}(n-k_0-k))\} = 0, \qquad (4.26)$$

Now, employing the assumptions that $r_{ss}(n) = 0$ and $r_{s\delta_v}(n) \approx r_{v\delta_s}(n) \approx r_{v\delta_v}(n) \approx 0$ we obtain,

$$E\{(x_s(n)+x_v(n))(s(n-k_0-k)+v(n-k_0-k))\} =$$
$$\sum_{l=1}^{p} \widehat{w}_n(l) E\{(s(n-k_0-l)+v(n-k_0-l))(s(n-k_0-k)+$$
$$v(n-k_0-k))\}. \qquad (4.27)$$

The above equation is similar to Wiener-Hopf equation and its solution can be written as

$$\widehat{\mathbf{w}}_n = \mathbf{R}_{(s+v)(s+v)}(n-k_0)^{-1}\mathbf{r}_{(x_s+x_v)(s+v)}(n-k_0), \qquad (4.28)$$

where, $\mathbf{r}_{(x_s+x_v)(s+v)}(n-k_0)$ is the cross-correlation matrix between the echo signal $x_s(n)+x_v(n)$ and the noisy input signal $s(n)+v(n)$, while $\mathbf{R}_{(s+v)(s+v)}$ is the auto-correlation matrix of $s(n)+v(n)$. There is no doubt that $\widehat{\mathbf{w}}_n$ is the most optimum solution possible. Hence it is shown that even for a single channel noise corrupted AEC problem, the most optimum solution $\widehat{\mathbf{w}}_n$ can be achieved under the assumptions stated earlier.

4.2.1 Formulation of LMS Update Equation

From Chapter 2 we know that, adaptive filter algorithms are very popular for iterative estimation of optimal filter coefficients, which do not require any correlation measurements or matrix inversion. The update equation of the weight vector is generally expressed as

$$\widehat{\mathbf{w}}_{n+1} = \widehat{\mathbf{w}}_n - \mu \nabla \xi(n) \tag{4.29}$$

where, μ is the step factor controlling the stability and rate of convergence, $\xi(n)$ is the cost function and ∇ is the gradient operator. The LMS algorithm simply approximates the mean square error by the square of the instantaneous error, i.e. $\xi(n) = e^2(n)$. Using (4.4), (4.6) and (4.8), the gradient of $\xi(n)$ can be written as

$$\nabla \xi(n) \quad = \quad \frac{\delta \xi(n)}{\delta \widehat{\mathbf{w}}_n^T} = -2e(n)\widehat{\mathbf{y}}(n-k_0). \tag{4.30}$$

Thus, the update equation for LMS from equation (4.29) is,

$$\begin{aligned} \widehat{\mathbf{w}}_{n+1} &= \widehat{\mathbf{w}}_n + 2\mu e(n)\widehat{\mathbf{y}}(n-k_0) & (4.31)\\ &= \widehat{\mathbf{w}}_n + 2\mu e(n)(\widehat{\mathbf{s}}(n-k_0) + \widehat{\mathbf{v}}(n-k_0)) & (4.32)\end{aligned}$$

For the k-th unknown filter parameter at the n-th iteration,

$$\begin{aligned} \widehat{w}_{n+1}(k) &= \widehat{w}_n(k) + 2\mu e(n)\widehat{y}(n-k_0-k) & (4.33)\\ &= \widehat{w}_n(k) + 2\mu e(n)(\widehat{s}(n-k_0-k) + \widehat{v}(n-k_0-k)) \end{aligned}$$

where, $k = 1, 2, \ldots, p$.

4.2.2 Convergence Analysis of the LMS Update

In this section, our objective is to show that the proposed LMS update equation (4.32) for the single channel noise corrupted AEC converges to the optimum solution. In what follows, starting from the proposed update equation (4.32) we show that the average value of the weight vector $\underline{\widehat{\mathbf{w}}}_n$ converges to the Wiener-Hopf solution given by (4.28).

Considering expectation operation on both sides of equation (4.32) we obtain,

$$\underline{\widehat{\mathbf{w}}}_{n+1} = \underline{\widehat{\mathbf{w}}}_n + 2\mu E\{e(n)(\widehat{s}(n-k_0-k) + \widehat{v}(n-k_0-k))\} \tag{4.34}$$

where, $\underline{\widehat{\mathbf{w}}}_n = E\{\widehat{\mathbf{w}}_n\}$. Now, for the k-th unknown weight vector(where $k = 1, 2, \ldots, p$), using (4.2) and considering $r_{ss}(n) = 0$ the term $E\{e(n)(\widehat{s}(n-k_0-k)+\widehat{v}(n-k_0-k))\}$ of (4.34) can be written as,

$$E\{e(n)(\widehat{s}(n-k_0-k)+\widehat{v}(n-k_0-k))\} = E\{[x_s(n)+x_v(n)-\widehat{x}_s(n)-\widehat{x}_v(n)] \\ (\widehat{s}(n-k_0-k)+\widehat{v}(n-k_0-k))\} \quad (4.35)$$

Similar to the procedure followed in the previous section, employing the assumptions that $r_{ss}(n) = 0$ and $r_{ss}(n) = 0$ and $r_{s\delta_v}(n) \approx r_{v\delta_s}(n) \approx r_{v\delta_v}(n) \approx 0$ we obtain,

$$E\{e(n)(\widehat{s}(n-k_0-k)+\widehat{v}(n-k_0-k))\} = \mathbf{r}_{(x_s+x_v)(s+v)}(n-k_0) - \mathbf{R}_{(s+v)(s+v)}(n-k_0)\widehat{\mathbf{w}}_n \quad (4.36)$$

Now, using (4.36), (4.34) can be written as

$$\underline{\widehat{\mathbf{w}}}_{n+1} = \underline{\widehat{\mathbf{w}}}_n - 2\mu\mathbf{R}_{(s+v)(s+v)}(n-k_0)\underline{\widehat{\mathbf{w}}}_n + 2\mu\mathbf{r}_{(x_s+x_v)(s+v)}(n-k_0) \quad (4.37)$$

In order to obtain a homogeneous solution of equation (4.37), we consider,

$$\underline{\widehat{\mathbf{w}}}_{n+1} = \underline{\widehat{\mathbf{w}}}_n - 2\mu\mathbf{R}_{(s+v)(s+v)}(n-k_0)\underline{\widehat{\mathbf{w}}}_n \quad (4.38)$$

For correlation matrix $\mathbf{R}_{(s+v)(s+v)}$, using eigenvalue decomposition we obtain,

$$\mathbf{R}_{(s+v)(s+v)} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \quad (4.39)$$

where, each column of the matrix $\mathbf{U}$ consists of eigenvectors corresponding to eigenvalues constituting the diagonal elements of the matrix $\mathbf{\Lambda}$ and $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. Now, multiplying both sides of (4.38) by $\mathbf{U}^T$ we get,

$$\underline{\widehat{\mathbf{w}}}^U_{n+1} = \underline{\widehat{\mathbf{w}}}^U_n - 2\mu\mathbf{\Lambda}\underline{\widehat{\mathbf{w}}}^U_n \quad (4.40)$$

where, $\mathbf{U}^T\underline{\widehat{\mathbf{w}}}_n = \underline{\widehat{\mathbf{w}}}^U_n$. The k-th coefficient of the weight vector can be expressed as,

$$\underline{\widehat{w}}^U_{n+1}(k) = (1-2\mu\lambda(k))\underline{\widehat{w}}^U_n(k). \quad (4.41)$$

Hence, the homogeneous solution can be obtained as

$$\widehat{w}_{h.s} = C_k(1-2\mu\lambda(k))^n, \quad (4.42)$$

where, C_k is a constant. Next, in order to obtain the particular solution for the k-th coefficient, based on (4.37) one can get,

$$\widehat{w}_{p.s} = \widehat{w}_{p.s} - 2\mu\lambda(k)\widehat{w}_{p.s} + 2\mu r^U(k) \quad (4.43)$$

For a particular solution $\widehat{w}_{p.s} = K_p r^U(k)$ (4.43) can be written as,

$$K_p r^U(k) \quad = \quad K_p r^U(k) - 2\mu\lambda(k)K_p r^U(k) + 2\mu r^U(k) \tag{4.44}$$

which leads to $K_p = \frac{1}{\lambda(k)}$ and the particular solution,

$$\widehat{w}_{p.s} \quad = \quad \frac{1}{\lambda(k)} r^U(k) \tag{4.45}$$

Hence, the total solution of (4.40) becomes

$$\underline{\widehat{w}}^U_{n+1}(k) \quad = \quad C_k(1-2\mu\lambda(k))^n + \frac{1}{\lambda(k)} r^U(k). \tag{4.46}$$

In the iterative update procedure, obviously the homogeneous part $(1 - 2\mu\lambda(k))^n$ decays to zero with iterations. From the rest of the terms, it can be shown that,

$$\underline{\widehat{\mathbf{w}}} = \mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T\mathbf{r}_{(x_s+x_v)(s+v)} = \mathbf{R}^{-1}_{(s+v)(s+v)}\mathbf{r}_{(x_s+x_v)(s+v)}. \tag{4.47}$$

Thus, it is found that the average value of the weight vector converges to the wiener-hopf, which is the optimum solution with increasing number of iteration.

4.2.3 Noise Cancellation by Spectral Subtraction

In the previous section, after performing the proposed adaptive echo cancellation algorithm on the input signal corrupted by noise and echo, the noise part of the signal will still remain. In what follows, we develop a scheme in order to suppress that residual noise from the echo reduced signal using a spectral subtraction method, which is a popular approach for single channel noise reduction [88] - [92].

The idea of spectral subtraction is to estimate the noise spectra and then perform a frequency domain subtraction of the noise spectra from the spectra of the noisy signal. In the proposed scheme, the echo reduced noisy signal $\widehat{s}(n) + \widehat{v}(n)$ is first obtained by applying the adaptive LMS algorithm and then it is passed through a noise suppression block, as can be seen from Fig. 4.1. A frame by frame noise reduction operation is then carried out.

The outcome of the echo canceller is a noisy signal $e(n)$ which is composed of two components as described in equation (4.8). Here, the signal portion consists of the original speech signal $s(n)$ along with the residual echo of the speech signal $\delta_s(n)$

and the noise portion consists of input additive noise $v(n)$ along with the residual echo of the noise signal $\delta_v(n)$. For the i-th frame, the error signal can be written as

$$e_i(n) = \widehat{s}_i(n) + \widehat{v}_i(n). \tag{4.48}$$

In the frequency domain, the error signal can be expressed as

$$E_i(f) = \widehat{S}_i(f) + \widehat{V}_i(f). \tag{4.49}$$

Each short-time (usually overlapped) successive frames (windows of $e(n)$) needs to be processed individually, and the resulting signal will be suitably reconstructed to obtain noise reduced signal [93]. The magnitude squared spectrum of the signal $\mid \widehat{S}_i(f) \mid^2$ can be obtained as

$$\mid \widehat{S}_i(f) \mid^2 = \mid E_i(f) \mid^2 - \mid \widehat{V}_i(f) \mid^2 - \widehat{V}_i(f)\widehat{S}_i^*(f) - \widehat{S}_i(f)\widehat{V}_i^*(f), \tag{4.50}$$

where, $[\cdot]^*$ represents complex conjugation. It is desired to choose an estimate $\widetilde{S}_i(f)$ that will minimize the magnitude squared error at each frequency (f) given by

$$Err_i(f) = \mid\mid \widetilde{S}_i(f) \mid^2 - \mid \widehat{S}_i(f) \mid^2 \mid . \tag{4.51}$$

. Note that the last three terms in equation (4.50) are not known and therefore can be replaced by their expected values. Since the noise is assumed to be zero mean and uncorrelated with the signal, the expected values of the last two terms of (4.50) can be neglected. Thus, (4.51) can be expressed as

$$Err_i(f) = \mid\mid \widetilde{S}_i(f) \mid^2 - \mid \widehat{E}_i(f) \mid^2 + E\{\mid \widehat{V}_i(f) \mid^2\} \mid, \tag{4.52}$$

where $E\{\}$ represents the expectation operator. This expression of $Err_i(f)$ can be minimized by choosing

$$\mid \widetilde{S}_i(f) \mid^2 = \mid \widehat{E}_i(f) \mid^2 - E\{\mid \widehat{V}_i(f) \mid^2\}. \tag{4.53}$$

Equation (4.53) is the basic representation of spectral subtraction operation considering magnitude squared spectrum which can be generalized [92] as

$$\mid \widetilde{S}_i(f) \mid^a = \mid \widehat{E}_i(f) \mid^a - kE\{\mid \widehat{V}_i(f) \mid^a\}, \tag{4.54}$$

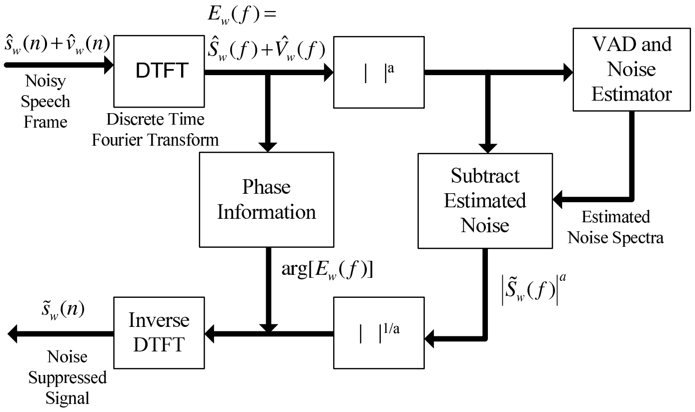

Fig. 4.4: Flow Diagram of the Spectral Subtraction Process.

where $a > 0$ and $k > 0$ are constants. Hence, an estimate of the magnitude spectrum $| \widetilde{S}_i(f) |$ of the signal can be obtained from (4.53) and (4.54) provided an estimate of noise spectrum $E\{| \widehat{V}_i(f) |^a\}$ is available. Then the signal spectrum $\widetilde{S}_i(f)$ can be computed as

$$\widetilde{S}_i(f) =| \widetilde{S}_i(f) | e^{jarg[E_i(f)]}, \tag{4.55}$$

where the phase of the signal spectrum $arg[E_i(f)]$ is assumed to be the phase of the noise corrupted signal in (4.49), which may not cause significant degradation in terms of loss of intelligibility of the speech signal [89]. Thus, our objective is now to obtain an estimate of noise spectrum $\widehat{V}_i(f)$ which is generally computed during the periods when speech is known *a priori* not to be present.

A flow diagram of the general spectral subtraction process is shown in Fig. 4.4. The output of the noise subtraction $\widetilde{s}_i(n)$, which is composed of the original input speech signal $s_i(n)$ and some residual noise-like signal $\Psi_i(n)$, is the final output of the overall AENC system. The signal $\Psi_i(n)$ is very small in magnitude, however it may still contain some signature of the input noise $v(n)$, the residual echo $\delta_s(n)$ and the residual noise $\delta_v(n)$.

A major problem of the conventional spectral subtraction method is that it may introduce a musical noise depending on the randomness of the noise [10]. It sounds metallic and distracts the attention of the listener. The fact behind this is that the noise floor which is subtracted is a smoothed estimate whereas the short time power spectrum of the actual noise may include peaks and valleys. Thus, after the noise subtraction there remains peaks in the noise spectrum. Some of those peaks, which are large spectral excursions, are perceived as time varying notes referred as musical noise. On the other hand, the wider peaks are perceived as time varying broad band noise, which is referred as residual noise. In addition to these problems, there always exists a chance of subtracting some portions of the speech signal, especially when the speech and noise characteristics are quite similar, leading to speech distortion. In order to overcome these limitations, several modified spectral subtraction algorithms are available in literature. Among them, one of the most effective methods is suggested by Berouti [10], which is capable of offering a better noise suppression than the conventional spectral subtraction technique. Moreover, it can eliminate the musical noise and can adapt to a wide range of signal to noise ratio. In this case, the main modifications made to the conventional spectral subtraction method are: (a) subtracting an overestimate of the noise power spectrum and (b) preventing the resultant spectrum from going below a preset minimum level (spectral floor). These modifications lead to minimizing the perception of the narrow spectral peaks by decreasing the spectral excursions and thus lower the musical noise perception.

The algorithm of conventional spectral subtraction from (4.53) can be written as

$$\begin{aligned} \mid \widetilde{S}_i(f) \mid^2 &= \mid \widehat{E}_i(f) \mid^2 - E\{\mid \widehat{V}_i(f) \mid^2\}, \text{ if } \mid \widetilde{S}_i(f) \mid^2 > 0 \\ &= 0, \text{ otherwise} \end{aligned} \tag{4.56}$$

Based on Berouti's modification one can rewrite the expression as

$$\begin{aligned} \mid \widetilde{S}_i(f) \mid^2 &= \mid \widehat{E}_i(f) \mid^2 - \alpha_{ss} E\{\mid \widehat{V}_i(f) \mid^2\}, \text{ if } \mid \widetilde{S}_i(f) \mid^2 > \beta_{ss}\{\mid \widehat{V}_i(f) \mid^2\} \\ &= \beta_{ss}\{\mid \widehat{V}_i(f) \mid^2\}, \text{ otherwise} \end{aligned} \tag{4.57}$$

with $\alpha_{ss} \geq 1$ and $0 \geq \beta_{ss} \leq 1$.

where α_{ss} is the subtraction factor and β_{ss} is the spectral floor parameter.

With $\alpha_{ss} > 1$ the subtraction can remove all of the broadband noise by eliminating most of the wide peaks. But deep valleys surrounding the narrow spectrum will remain in enhanced speech. This can be further reduced by filling-in the valleys. This objective is achieved by the keeping the spectral floor,

For $\beta_{ss} > 0$, the valleys between the peaks are not as deep as for the case $\beta_{ss} = 0$.

Thus the spectral excursions of noise peaks are reduced, and hence the musical noise lowered.

An important aspect is the noise power spectral density detection from speech silence. In this regard, a major task is to estimate the power spectral density of nonstationary noise when a noisy speech signal is given. In [94] a new minimum statistics noise estimator is introduced where a power spectral density smoothing algorithm is used which employs a time varying smoothing parameter. The advantage of this algorithm is that it can track the variance of the smoothed power spectral density in frequency bands, and offer a bias compensation for minimum power spectral density estimates.

4.3 Development of Adaptive Characteristics

Similar to single channel acoustic echo cancellation, the adaptive part of the integrated echo-noise canceller may suffer slow convergence and fluctuation due to the assumptions $r_{ss}(n) = 0$ and $r_{s\delta_v}(n) \approx r_{v\delta_s}(n) \approx r_{v\delta_v}(n) \approx 0$, which may not strictly hold in reality. Therefore, similar to chapter 2, we will exploit some adaptive characteristics, which along with the proposed LMS update algorithm can guarantee a better convergence performance. In view of developing such adaptive characteristics, three factors will be considered in the proposed algorithm similar to those of chapter 2: (i) the degrees of the cross-correlation term $r_{ss}(n)$ (ii) the amount of signal power for the two signals under consideration: the reference signal $s(n-k_0)$ and the current signal $s(n)$, (iii) the mean square error between consecutive estimates of the unknown filter coefficients.

In order to demonstrate the performance of the proposed LMS update algorithm, speech samples of different characteristics, such as voiced, unvoiced and pause are taken into consideration. It is found that the negligibility of the cross-correlation terms $r_{ss}(n)$, $r_{s\delta_v}(n)$, $r_{v\delta_s}(n)$ and $r_{v\delta_v}(n)$ strongly depends on the characteristics of the speech samples and the input noise. For example, because of the inherent periodicity of the voiced speech, the degree of cross-correlation between two voiced speech frames of a person becomes higher in comparison to that between two unvoiced speech frames which are random in nature. In this case, ratio of power of two

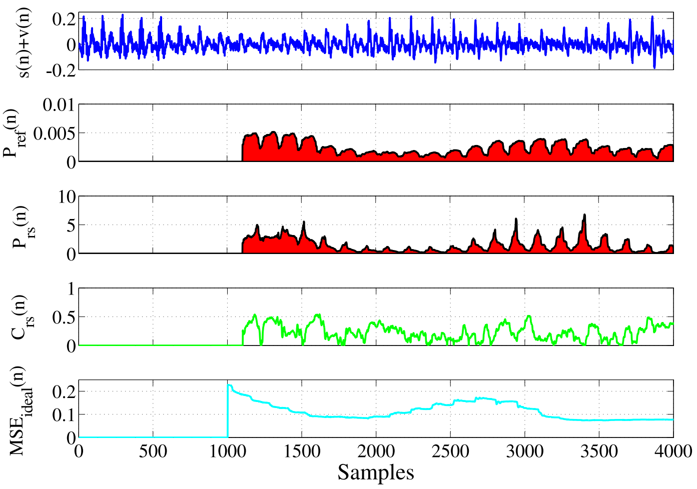

Fig. 4.5: A voiced frame followed by another voiced frame (a) Original noisy input Signal $s(n) + v(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$(e) MSE of coefficient updated from ideal values $MSE_{ideal}(n)$.

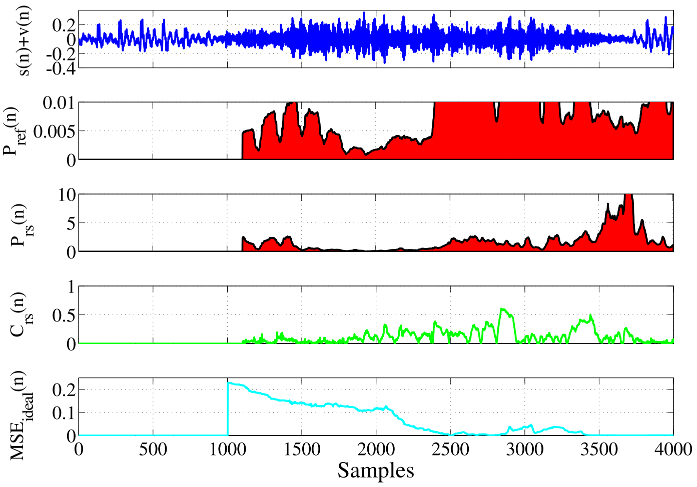

Fig. 4.6: A voiced frame followed by an unvoiced frame (a) Original noisy input Signal $s(n) + v(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$(e) MSE of coefficient updated from ideal values $MSE_{ideal}(n)$.

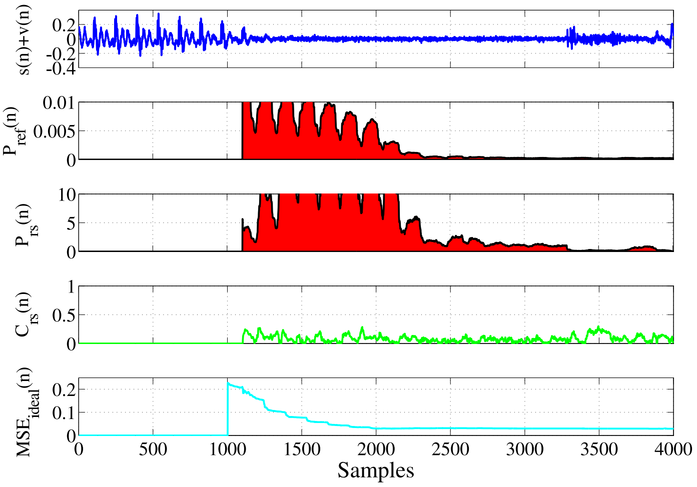

Fig. 4.7: A voiced frame followed by a pause (a) Original noisy input Signal $s(n) + v(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$(e) MSE of coefficient updated from ideal values $MSE_{ideal}(n)$.

different speech frames may also carry some significant information. For example, if we consider a voiced frame and an unvoiced frame, their power ratio is generally higher in comparison to that of two voiced speech frames. Moreover, for simplicity, let us assume that the input noise $v(n)$ is an additive white gaussian noise which emphasizes that the cross correlation of the speech signal with the noise tends to zero.

In Fig. 4.5(a), a male utterance $/iy/ - /r/$ [77] of a duration of 250 ms corrupted by 15dB white noise and with a sampling frequency of 16 kHz is shown. in this figure, a few samples of voiced phoneme are followed by another few samples of voiced phoneme. The strong periodicity of the utterance $s(n)$ clearly indicates its voiced characteristics. Considering the flat delay of $k_0 = 1000$ samples, from the starting point of $s(n)$, this utterance will act as a reference signal for the generation of echo that corrupts the current samples at or after k_0 samples. Employing the proposed LMS algorithm on the noise and echo-corrupted signal $y(n)$, an echo reduced signal $\widehat{s}(n) + \widehat{v}(n)$ is obtained. In Fig. 4.5(b), power of the reference signal $\widehat{s}(n - k_0) + \widehat{v}(n - k_0)$, namely $P_{ref}(n)$ is depicted, which is computed at every input instances considering a window of M samples and is defined as

$$P_{ref}(n) \;=\; \frac{\sum_{i=-\frac{M}{2}}^{\frac{M}{2}-1} [\widehat{s}(n-k_0+i)+\widehat{v}(n-k_0+i)]^2}{M}. \tag{4.58}$$

Here we consider $k_0 >> M$ and $M = 100$. In this connection, we also consider the average power $P_{sup}(n)$ of the last M samples of the echo suppressed speech signal $\widehat{s}(n)$, which is defined as

$$P_{sup}(n) \;=\; \frac{\sum_{j=0}^{M-1} [\widehat{s}(n-j)+\widehat{v}(n-j)]^2}{M}. \tag{4.59}$$

The ratio of $P_{ref}(n)$ and $P_{sup}(n)$ is denoted as the power ratio $P_{rs}(n)$, which is shown in Fig. 4.5(c). In Fig. 4.5(d) the cross correlation coefficient $C_{rs}(n)$ between the noisy reference signal $\widehat{s}(n-k_0)+\widehat{v}(n-k_0)$ and the current noisy signal $\widehat{s}(n)+\widehat{v}(n)$ is shown. A coefficient of correlation, $C_{rs}(n)$, is a mathematical measure of how much one number can expected to be influenced by change in another. It is defined as,

$$C_{rs}(n) = \frac{cov((\widehat{s}(n-k_0+i)+\widehat{v}(n-k_0+i))(\widehat{s}(n-j)+\widehat{v}(n-j)))}{\sigma_{\widehat{s}(n-k_0+i)+\widehat{v}(n-k_0+i)}\sigma_{\widehat{s}(n-j)+\widehat{v}(n-j)}} \tag{4.60}$$

Here, $-M/2 \leq i \leq M/2-1$ and $0 \leq j \leq (M-1)$. If $C_{rs}(n) = \pm 1$ then there is a strong positive/negative correlation between two signals. If it is zero then there is no correlation among the matrices. In order to demonstrate the performance of the proposed LMS update algorithm, in terms of convergence rate and parameter estimation accuracy, in Fig 4.5(e), the mean square error $MSE_{ideal}(n)$ between the estimated coefficients w_n and the true coefficients a_n is depicted.

In a similar fashion, in Fig. 4.6 and 4.7, first a voiced phoneme */ih/* followed by an strong unvoiced phoneme */sh/* and then a voiced phoneme */ih/* followed by pause are considered respectively for 15dB white gaussian noise at input. It is to be mentioned that in these figures Fig. 4.5-Fig. 4.7, the reference signal is always a voiced frame and the current frame is voiced, unvoiced or pause respectively. It is found that the power of the reference voiced frame is always quite high in comparison to unvoiced or pause frames. However, the power ratio not only depends on the power of the reference voiced frame but also on the power of the echo suppressed signal. If the current frame is a pause or weakly unvoiced frame then the power ratio is very high, otherwise, for voiced and strong unvoiced frames the power ratio is lower. The correlation coefficient is very small when measured between a voiced and a unvoiced frame, but is quite large for two voiced frames.

The presence of voiced frame as a reference strongly governs the rate of convergence and the estimation error of the proposed LMS algorithm. For example in Fig. 4.5, because of althrough presence of the voiced frame as reference, the convergence performance becomes very poor and even in some cases the algorithm diverges and in all cases, the estimation error was higher. On the contrary, in Fig. 4.7 it is observed that, when the current frame is pause, even in the presence of voiced reference frame a very fast convergence is obtained with a small estimation error. Moreover, in Fig. 4.7, as the current frame is unvoiced instead of pause, a slower convergence is observed with a high estimation error.

It is quite interesting that the performance characteristics of the proposed LMS update algorithm drastically changes when the reference frame is considered unvoiced, as shown in Fig. 4.8, 4.9 and 4.10. In this case a very fast convergence is obtained with a high level of estimation accuracy. The reason behind this drastic change in characteristics can be explained based on the cross-correlation that may exist between the reference frame and the current frame. In case of voiced reference frame, a strong correlation persists between each samples of the voiced frame, which makes it difficult for the LMS to estimate the room response as the assumption of the negligibility of the cross-correlation term $r_{ss}(n)$ does not hold anymore. Moreover, when the current frame has a high energy speech along with the echo, i.e. when the power ratio is lower, the convergence performance of the LMS algorithm may degrade because of the chances of suppression of the input speech. In the case when the current frame is pause, no matter whether the reference frame is voiced or unvoiced, a fast convergence with high estimation accuracy is achieved using the proposed LMS algorithm. The reasons behind are, (i) negligible cross-correlation between reference frame and current speech frame and (ii) a comparatively higher power ratio. In case of unvoiced reference frame, because of existence of a little correlation between the input and the reference frame the convergence performance of the proposed LMS algorithm is found quite satisfactory irrespective of the power of the reference signal(strong unvoiced or weakly unvoiced).

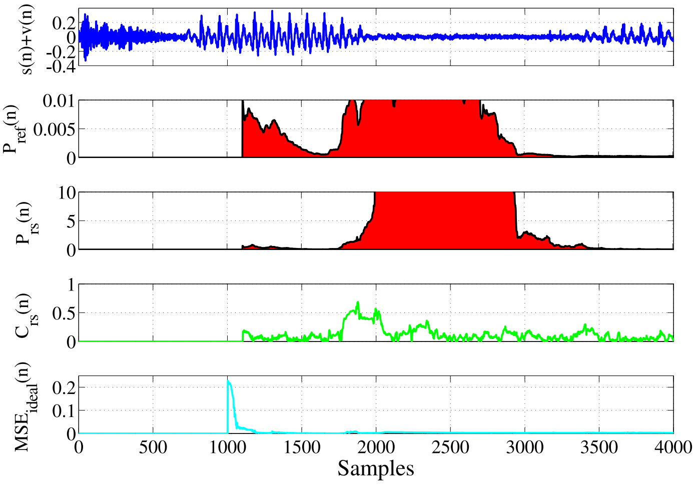

Fig. 4.8: An unvoiced frame followed by a voiced frame (a) Original noisy input Signal $s(n) + v(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$(e) MSE of coefficient updated from ideal values $MSE_{ideal}(n)$.

Finally, in Fig. 4.12, 4.11 and 4.13 the outcomes of three different cases when the references are always from a pause or stop frame are shown. Because of the additive white noise present in the input signal, the pause or stop frame may contain sufficient energy to produce a good estimation of the room response given that the power of noise is quite high. Otherwise, no significant update occurs in the proposed LMS algorithm and in some cases, as expected, the convergence performance degrades. This is because of the lack of reference data as well as signal energy, which are required for LMS updates.

4.4 Proposed Update Constraints

The insight obtained from extensive experimentation on several such case as presented in Fig. 4.5 - Fig. 4.13 are summarized in table 4.1. It is clearly observed from the table that in many cases the performance of the proposed algorithm is not satisfactory or even poor, e.g. when the reference and the input signal both are voiced frames, when the reference signal is pause etc.. In view of overcoming these

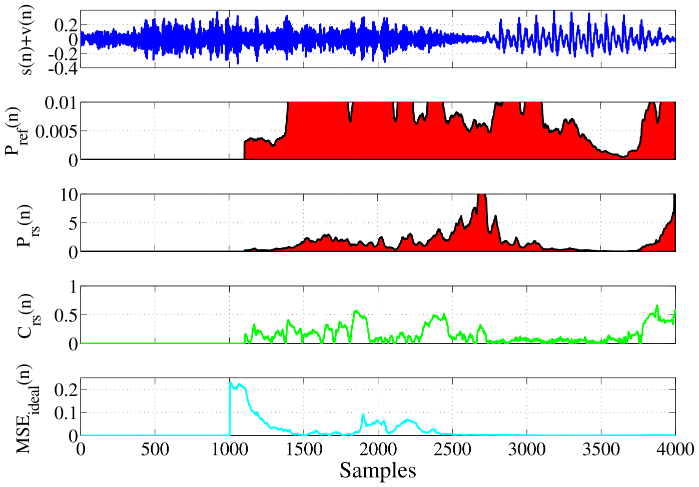

Fig. 4.9: An unvoiced frame followed by another unvoiced frame (a) Original noisy input Signal $s(n) + v(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$(e) MSE of coefficient updated from ideal values $MSE_{ideal}(n)$.

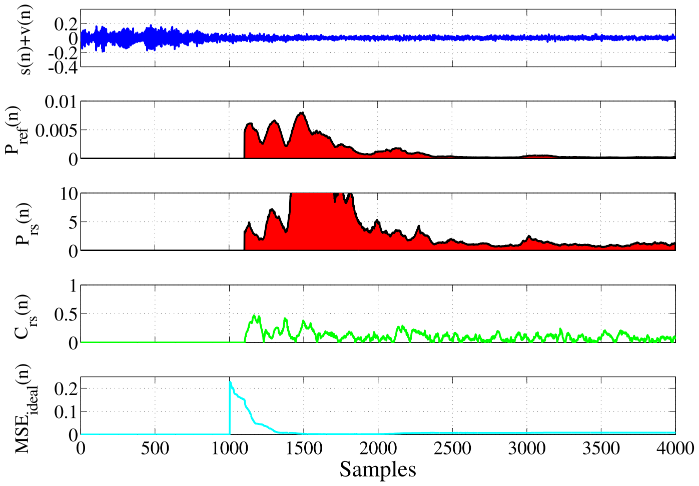

Fig. 4.10: An unvoiced frame followed by a pause (a) Original noisy input Signal $s(n) + v(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$(e) MSE of coefficient updated from ideal values $MSE_{ideal}(n)$.

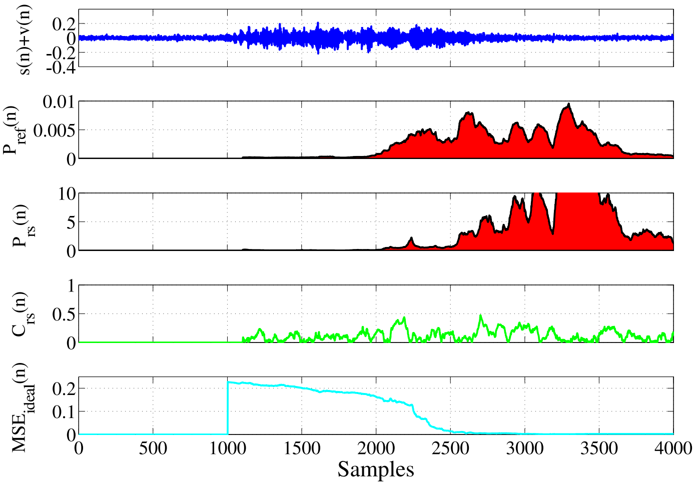

Fig. 4.11: A pause followed by an unvoiced frame (a) Original noisy input Signal $s(n)+v(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$(e) MSE of coefficient updated from ideal values $MSE_{ideal}(n)$.

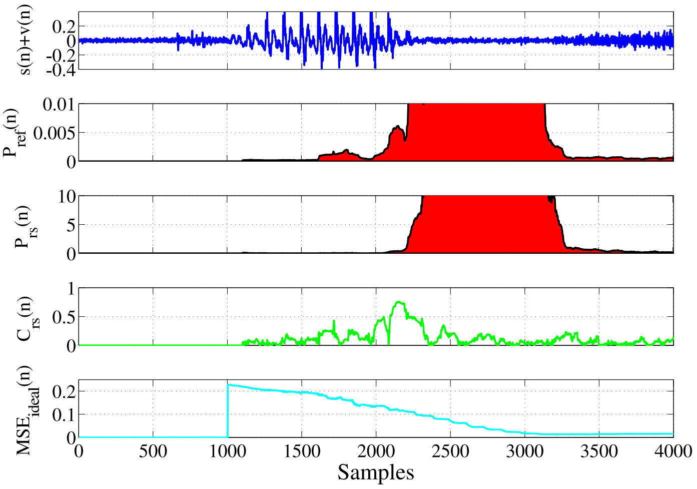

Fig. 4.12: A pause followed by a voiced frame (a) Original noisy input Signal $s(n)+v(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$(e) MSE of coefficient updated from ideal values $MSE_{ideal}(n)$.

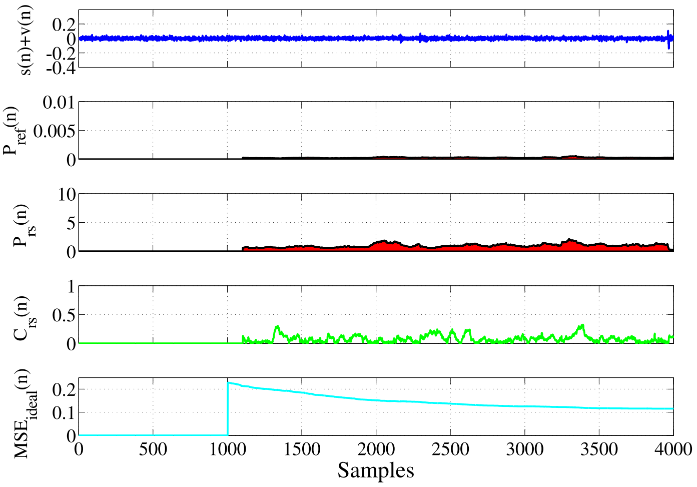

Fig. 4.13: A pause followed by another pause (a) Original noisy input Signal $s(n)+v(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$(e) MSE of coefficient updated from ideal values $MSE_{ideal}(n)$.

cases we are going to propose some conditions which will guarantee a fast convergence with a low estimation error. It is obvious that if the proposed algorithm is used, there is always a possibility to obtain poor convergence or even in some cases divergence with a high estimation error. Based on the results obtained from Table 4.1 and some more experimentation we hereby propose three conditions on LMS update, which are designed to indicate whether the updating should be carried out or halted. Implementation of these conditions in the proposed LMS update will provide assurance of fast convergence with a high estimation accuracy.

Similar to the single channel adaptive echo cancellation scheme described in chapter 2, the following conditions are proposed for constraining the LMS update,

Condition 1: $P_{rs}(n) \geq \zeta$ and $P_{ref}(n) \geq \beta$

Condition 2: $C_{xy}(n) \leq \Upsilon 1$ and $P_{ref}(n) \geq \beta$

Condition 3: $C_{xy}(n) \leq \Upsilon 2$

Condition 4: $e_{coeff}(n) \leq \aleph$

where, ζ, β, $\Upsilon 1$, $\Upsilon 2$ and $\aleph$ are threshold values and $e_{coeff}(n)$ is the mean square

Table 4.1: Dependence of LMS update on the acoustic characteristics of the frame in case of noisy input signal

Reference Speech Sample	**Speech at the noise and Echo Corrupted Sample**	**LMS Update**
Voiced	Voiced	unsatisfactory
Voiced	Unvoiced	unsatisfactory
Voiced	Pause	satisfactory/excellent
Unvoiced	Voiced	excellent
Unvoiced	Unvoiced	excellent
Unvoiced	Pause	excellent
Pause	Voiced	unsatisfactory
Pause	Unvoiced	unsatisfactory
Pause	Pause	unsatisfactory

error of the estimations of successive iterations defined as,

$$e_{coeff}(n) = \sum_{K=1}^{p} (w_n(k) - w_{n-1}(k))^2/p. \tag{4.61}$$

The only difference between these conditions and those of chapter 2 is condition 3. In case of single channel echo cancellation at noisy environment, there is always a certain level of noise present irrespective of the speech signal. The presence of noise can be used to our advantage at speech pause where in the ideal echo cancellation case we stopped the update. Now that a signal is present, the proposed algorithm can easily update its estimates. Moreover, as noise is considered uncorrelated to itself, the assumption made at frames where only noise is present would infact be quite satisfactory. As the usual performance of the algorithm is unsatisfactory when the speech pauses are taken as reference, the value of $\Upsilon 2$ should be very small so that update would take place only when the reference and the current frame are very much uncorrelated and thus it would be ensured that the estimate would not degrade. Moreover, the convergence would be much faster at noisy frames when the input noise level is high, i.e. at low SNR, then speech pauses would contain high energy noise and which would infact assist the proposed algorithm to converge to a suitable solution more quickly. The update of the proposed LMS algorithm will be carried out if any one of the first three conditions is true. The fourth condition will be checked after each estimation and if it is true the new estimation will take effect. The facts behind choosing these four conditions for our proposed method are already discussed in chapter 2.

4.5 Simulation Results and Comments

Simulations were performed on two different speech signals uttering (1) "Good service should be rewarded by big tips" by a male voice and (2) "She had your dark suit in greasy wash water all year" another male voice. Both of the speech were taken from the TIMIT database [77]. The room response illustrated in chapter 2 is adopted here for simulation. The step size for the LMS adaptive filter was varied from $1/p$ to 0.02 where p is the number of unknown coefficients of the room response.

The echo return loss enhancement (ERLE) in dB is computed as a performance measure. ERLE is a smoothed measure of the amount (in dB) that the echo has been attenuated. It is defined as the ratio of the power of the residual echo signal and the input echo signal power [2],

Another criteria for performance evaluation termed Signal to Distortion Ratio (SDR) is also computed. SDR is quite like measuring signal to noise ratio (SNR). The only difference is that in our case the output signal is contaminated not only with noise but also some residual echo, which we termed distortion in this case.

Similar to the AEC problem stated in chapter 2, in case of integrated echo-noise suppression the following values of the update constrain parameters are chosen:

- $\zeta = 2.0$
- $\beta = 0.003$
- $\Upsilon 1 = 0.25$
- $\Upsilon 2 = 0.1$
- $\aleph = 0.7e^{-4}$

Thus the update constraints for our simulation is set to,

i. $P_{rs}(n) \geq 2.0$ and $P_{ref}(n) \geq 0.003$

ii. $C_{rs}(n) \leq 0.25$ and $P_{ref}(n) \geq 0.003$

ii. $C_{rs}(n) \leq 0.1$

iii. $e_{coeff}(n) \leq 0.7e^{-4}$

Now, the first three constraints, constraints *i*, *ii* and *iii* have been simultaneously applied to the nine different cases of voiced, unvoiced and pause frames and the results are shown in Fig. 4.14 - Fig. 4.22. As can be seen from these figures, the application of condition 1 and constraint *iv* in the proposed method improved the convergence performance of the LMS algorithm to a large extent. The $MSE_{ideal}(n)$ curve is now totally nondivarging and the estimation error is minimized in every cases.

For continuous speech signals, large and abrupt fluctuation of the filter estimate increases estimation error and makes the system unstable. Thus, condition 3 is employed to suppress abrupt change in coefficient estimation of the LMS algorithm.

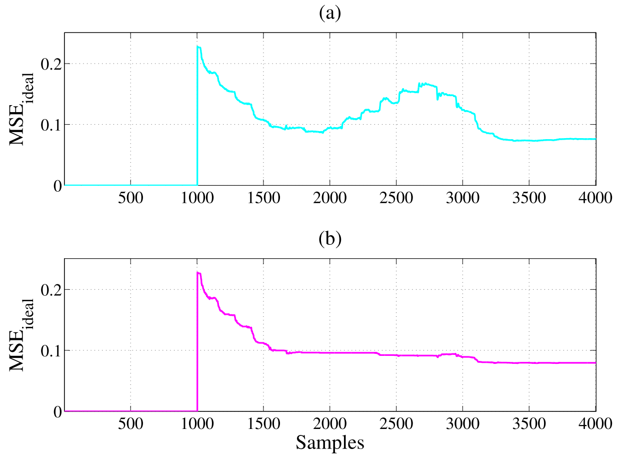

Fig. 4.14: MSE of estimation coefficients from ideal values for a voiced frame in reference and a voiced frame in current samples (a) without applying any condition (b) applying condition 1 and 2.

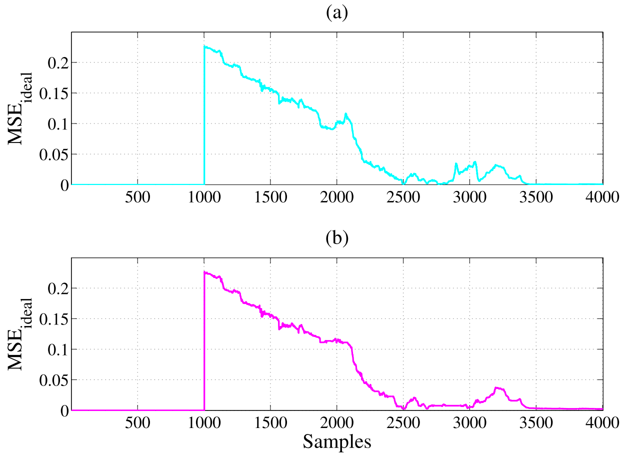

Fig. 4.15: MSE of estimation coefficients from ideal values for a voiced frame in reference and a unvoiced frame in current samples (a) without applying any condition (b) applying condition 1 and 2.

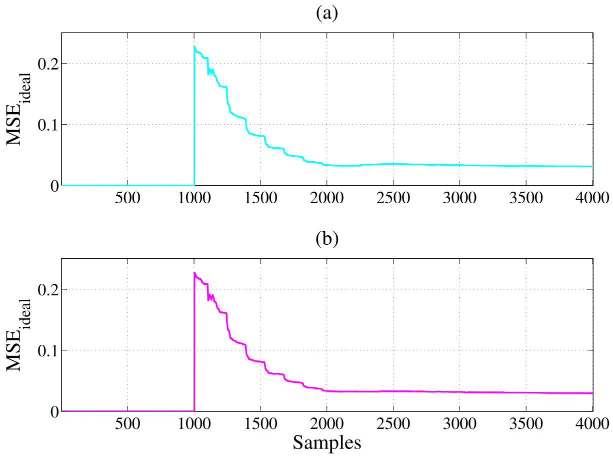

Fig. 4.16: MSE of estimation coefficients from ideal values for a voiced frame in reference and a pause frame in current samples (a) without applying any condition (b) applying condition 1 and 2.

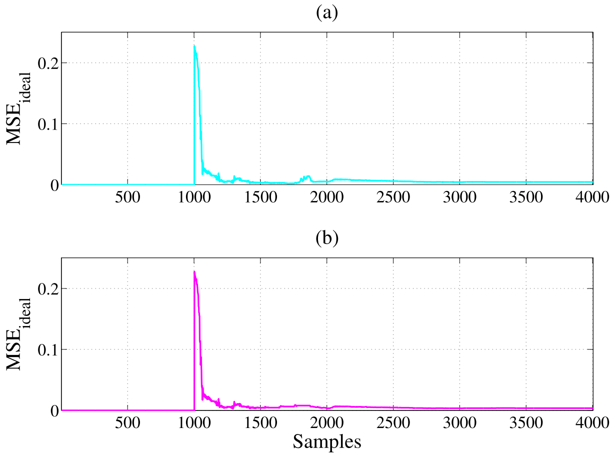

Fig. 4.17: MSE of estimation coefficients from ideal values for a unvoiced frame in reference and a voiced frame in current samples (a) without applying any condition (b) applying condition 1 and 2.

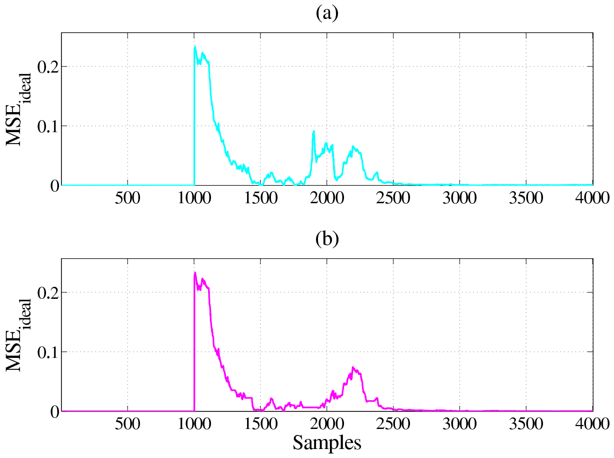

Fig. 4.18: MSE of estimation coefficients from ideal values for a unvoiced frame in reference and a unvoiced frame in current samples (a) without applying any condition (b) applying condition 1 and 2.

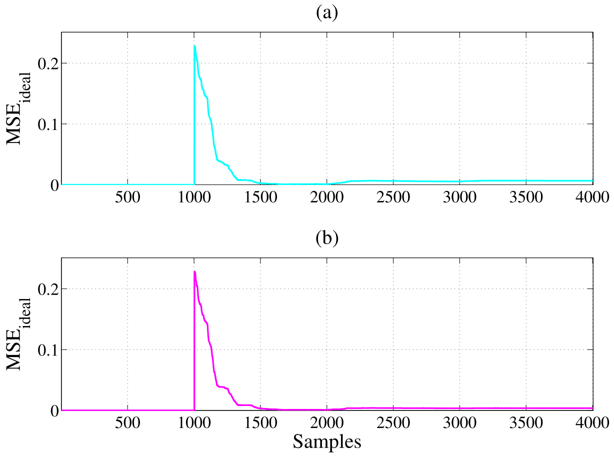

Fig. 4.19: MSE of estimation coefficients from ideal values for a unvoiced frame in reference and a pause frame in current samples (a) without applying any condition (b) applying condition 1 and 2.

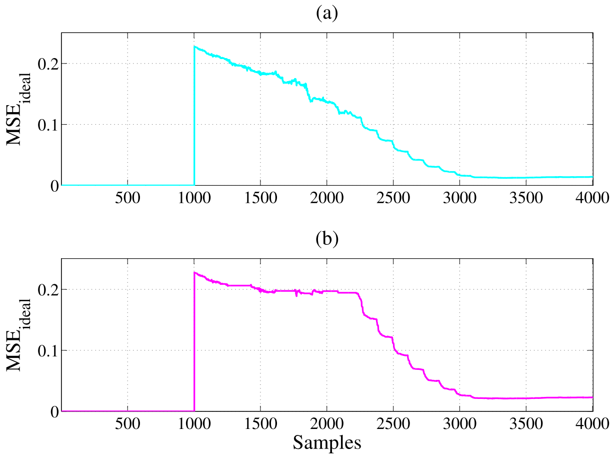

Fig. 4.20: MSE of estimation coefficients from ideal values for a pause frame in reference and a voiced frame in current samples (a) without applying any condition (b) applying condition 1 and 2.

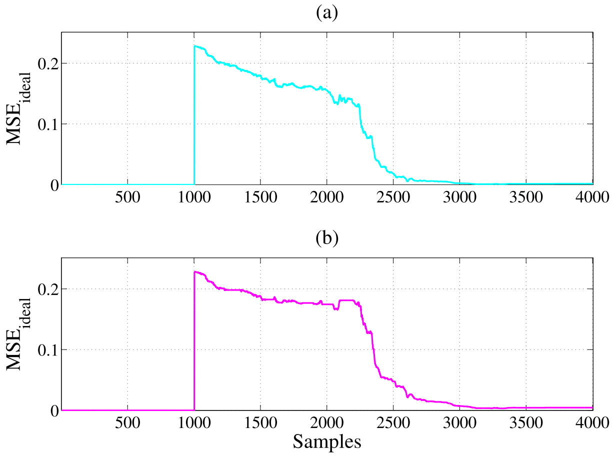

Fig. 4.21: MSE of estimation coefficients from ideal values for a pause frame in reference and a unvoiced frame in current samples (a) without applying any condition (b) applying condition 1 and 2.

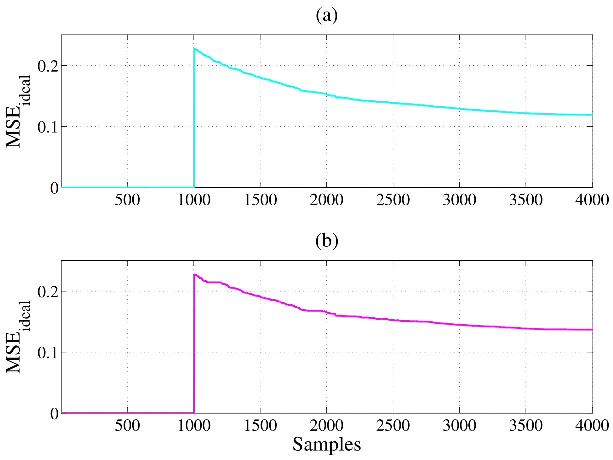

Fig. 4.22: MSE of estimation coefficients from ideal values for a pause frame in reference and a pause frame in current samples (a) without applying any condition (b) applying condition 1 and 2.

If the change in updated coefficients is smaller then a the threshold $(0.7e^{-4})$ then LMS would update, otherwise the algorithm would hold the previous coefficients.

In Fig. 4.23 the reference power, power ratio, correlation coefficients and LMS update without conditions for speech signal 1 is illustrated. In Fig. 4.24.a the MSE of coefficient estimates of LMS with respect to ideal values of unknown coefficients without applying any conditions is shown, while in Fig. 4.24.b the MSE with constraints i, ii and iii simultaneously applied is depicted. It can easily be seen that the update of LMS has been stabilized to better results robustly when the conditions are applied and degradation of estimation is prevented. Similar results are obtained for speech signal 2 as can be seen from Fig. 4.25 and 4.26. These figures also demonstrates the efficiency of the applied conditions in controlling the update of LMS algorithm for producing better results.

In Table 4.2 and Table 4.3 the performance of the proposed LMS update method with and without applying the three proposed conditions is compared in terms of SDR difference (dB) and ERLE (dB). Performance is evaluated for different number of unknown coefficients ranging from 2 to 14 for speech signal 1 and 2 respectively. The tables clearly demonstrates the superiority of the proposed LMS algorithm with update constraints over that without any constraints.

In Table 4.4 and Table 4.5 the performance of the proposed LMS update method

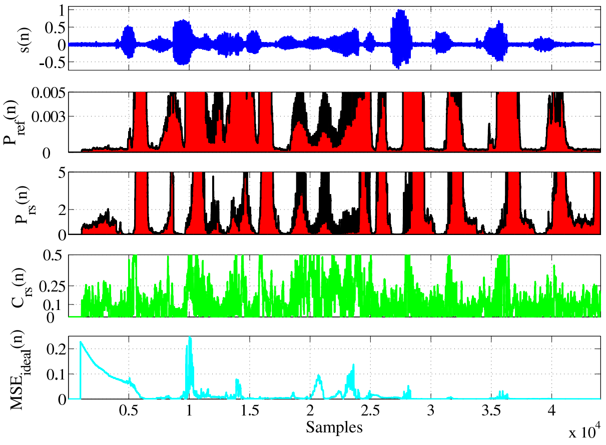

Fig. 4.23: LMS update on speech signal 1(2 unknown coefficients, no conditions applied) (a) Original Signal $s(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$(e) MSE of coefficient updated from ideal values ($MSE_{ideal}(n)$.

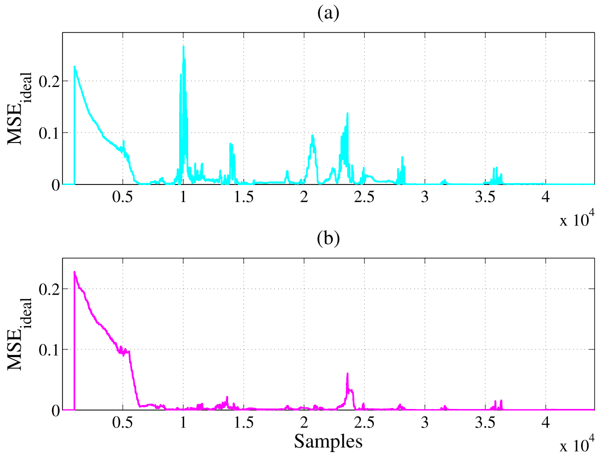

Fig. 4.24: MSE of LMS estimations from ideals on speech signal 1 (2 unknown coefficients) (a) Without any condition (b) With condition 1, 2 and 3 simultaneously applied

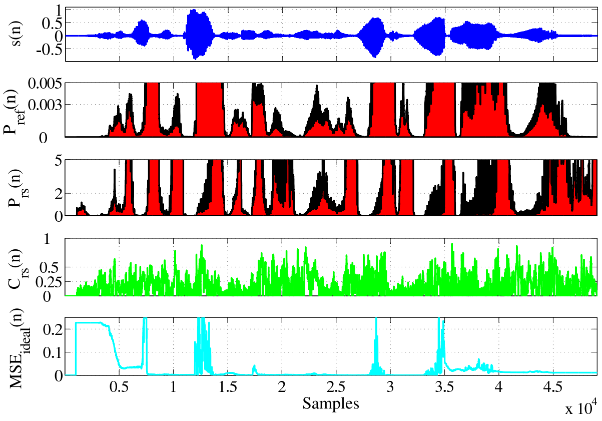

Fig. 4.25: LMS update on speech signal 2(2 unknown coefficients, no conditions applied) (a) Original Signal $s(n)$ (b) Power of the reference frame $P_{ref}(n)$ (c) Power ratio $P_{rs}(n)$(M=100) (d) Cross Correlation Coefficient between the reference frame and the current frame $C_{rs}(n)$(e) MSE of coefficient updated from ideal values $(MSE_{ideal}(n)$.

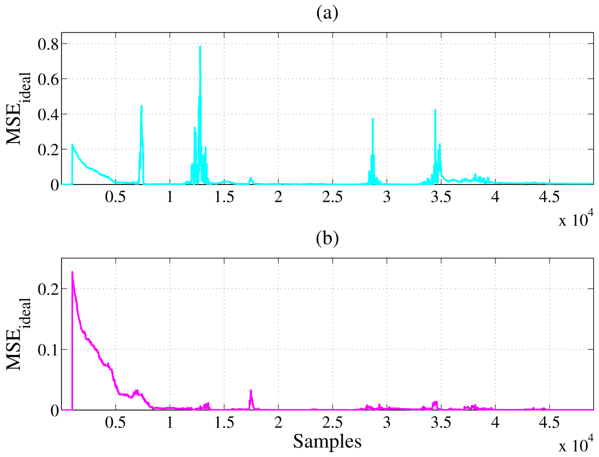

Fig. 4.26: MSE of LMS estimations from ideals on speech signal 2 (2 unknown coefficients) (a) Without any conditio (b) With condition 1, 2 and 3 simultaneously applied

Table 4.2: Performance for increasing number of unknown coefficients for the input speech signal 1 at 15dB white gaussian noise(μ is varied from $1/p$ to 0.02)

	No Conditions		With Conditions 1+2+3+4	
No. of coefficients	SDR Improvement (dB)	ERLE (dB)	SDR Improvement (dB)	ERLE (dB)
2	4.9921	8.8496	6.9848	10.6772
4	4.9027	2.0696	5.7731	2.2787
6	8.391	4.6507	9.2744	5.0313
8	6.4551	2.4214	6.3558	2.2797
10	6.3507	2.6341	6.173	2.454
12	6.7127	3.0277	7.0978	3.0048
14	7.8763	3.7481	8.2515	3.6909

Table 4.3: Performance for increasing number of unknown coefficients for the input speech signal 2 at 15dB white gaussian noise(μ is varied from $1/p$ to 0.02)

	No Conditions		With Conditions 1+2+3+4	
No. of coefficients	SDR Improvement (dB)	ERLE (dB)	SDR Improvement (dB)	ERLE (dB)
2	1.7666	7.8907	6.3291	13.3284
4	3.6895	1.715	6.6418	2.599
6	6.8115	4.1443	10.3025	5.3981
8	4.6745	2.2157	6.8105	2.7603
10	4.7391	2.4094	7.0343	2.9192
12	6.248	2.8372	8.1278	3.253
14	7.517	3.6158	8.8414	3.8667

Table 4.4: Performance for increasing input white gaussian noise level for input speech signal 1 (μ is varied from $1/p$ to 0.02)

	No Conditions		With Conditions 1+2+3+4	
Input Noise Level (dB)	SDR Improvement (dB)	ERLE (dB)	SDR Improvement (dB)	ERLE (dB)
25	7.4065	3.183	7.8189	3.0759
20	7.613	3.5382	7.9346	3.4171
15	7.8763	3.7481	8.2515	3.6909
10	8.2085	3.5999	8.386	3.5064
5	8.2434	3.0533	8.8839	3.0765
0	8.7968	2.4493	9.4557	2.542
-5	8.2259	2.0032	10.5136	1.5912

Table 4.5: Performance for increasing input white gaussian noise level for input speech signal 2 (μ is varied from $1/p$ to 0.02)

	No Conditions		With Conditions 1+2+3+4	
Input Noise Level (dB)	SDR Improvement (dB)	ERLE (dB)	SDR Improvement (dB)	ERLE (dB)
25	7.0807	3.8534	8.5601	3.7745
20	7.2175	3.8054	8.0401	3.7295
15	7.517	3.6158	8.8414	3.8667
10	7.7662	3.3768	8.8286	3.5394
5	8.584	2.8765	9.306	3.0344
0	9.1721	2.3696	9.6731	2.5163
-5	10.1367	1.8799	10.618	1.4358

with and without applying the three proposed conditions is evaluated for different level of input SNR ranging from 25dB to −5dB for speech signals 1 and 2 respectively. Stationary additive white gaussian noise is applied at the input. It can be seen that the proposed method works quite satisfactorily even for a high energy input noise level. The tables clearly demonstrates the superiority of the proposed LMS algorithm with update constraints over that without any constraints.

4.6 Conclusion

In this chapter, a novel approach of single channel acoustic echo cancellation scheme for noisy environment using gradient based adaptive LMS algorithm is proposed. The proposed scheme differs from ideal single channel AEC at noiseless environment and from dual channel adaptive algorithm in several critical issues which are highlighted in this chapter. Afterwards, the validity of the proposed scheme was proved mathematically by showing that the estimated coefficients obtained by the proposed scheme may reach wiener-hopf solution in the long run based on several critical assumptions. Later, the LMS update equation for the proposed scheme was derived and validated mathematically. The assumption on signal correlation that degrades the performance of the proposed scheme in reality is handled next by setting some constraints in the update procedure of the LMS algorithm. The constraints were obtained by analyzing the properties of speech frames and also by following the mean square change in consecutive estimations of the LMS filter. Next, a modified single channel spectral subtraction method for noise cancellation is described, which was adopted for noise cancellation for its robust performance and ability to reduce musical noise. In the simulation section, performance of the proposed scheme is evaluated based on improvement of the Signal to distortion ratio (SDR) in dB and also based on the traditional Echo Return Loss Enhancement (ERLE) parameter measured in dB. It is shown in the result section that the performance of the single channel echo cancellation scheme for increasing length of room response and for decreasing SNR is enhanced to a great extent if the proposed conditions are applied.

Chapter 5

Single Channel Integrated Acoustic Echo and Noise Cancellation Based on Particle Swarm Optimization Algorithm

Finally, the problem of single channel acoustic echo and noise cancellation (AENC) is going to be modeled by an optimization algorithm, namely the previously introduced Particle Swarm Optimization Algorithm. We already realized that there is a basic difference in the processing scheme of the gradient based approach and the optimization algorithm based approach. The gradient based approach propagates sample by sample whereas the proposed optimization based algorithm operates frame by frame. Thus, similar to chapter 3, the overlap add method would be adopted in this chapter. As before, The Particle Swarm Optimization (PSO) algorithm would be formulated for both time domain (PSO-TD) and frequency domain (PSO-FD) and the performance of these two formulations would be evaluated in the results section. In addition to the PSO based echo canceller a spectral subtraction based single channel noise suppression technique is incorporated in the design to cancel out the noisy parts of the signal and to produce an enhanced speech [95] [96]. The results would be evaluated for different input noise level and compared with those of gradient based algorithms.

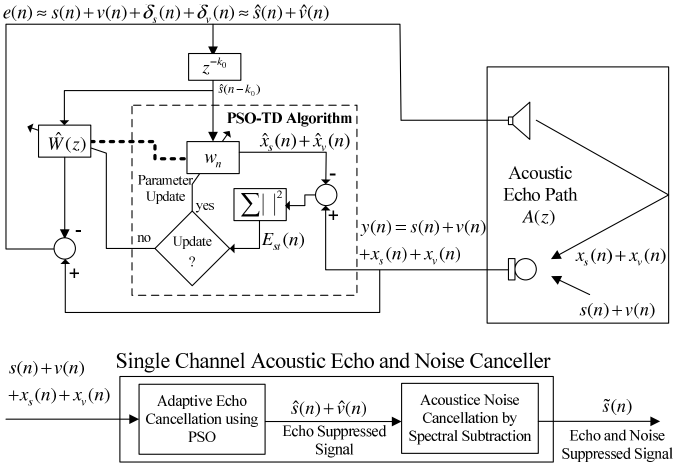
$e(n) \approx s(n)+v(n)+\delta_s(n)+\delta_v(n) \approx \hat{s}(n)+\hat{v}(n)$
z^{-k_0}
$\hat{s}(n-k_0)$
PSO-TD Algorithm
$\hat{W}(z)$
w_n
$\hat{x}_s(n)+\hat{x}_v(n)$
Parameter Update
yes
no
Update ?
$\sum|\ |^2$
$E_{st}(n)$
-
+
-
+
$y(n)=s(n)+v(n)$
$+x_s(n)+x_v(n)$
Acoustic Echo Path $A(z)$
$x_s(n)+x_v(n)$
$s(n)+v(n)$
Single Channel Acoustic Echo and Noise Canceller
$s(n)+v(n)$
$+x_s(n)+x_v(n)$
Adaptive Echo Cancellation using PSO
$\hat{s}(n)+\hat{v}(n)$
Echo Suppressed Signal
Acoustice Noise Cancellation by Spectral Subtraction
$\tilde{s}(n)$
Echo and Noise Suppressed Signal

Fig. 5.1: Basic Setup of PSO algorithm based single channel integrated echo and noise canceller for enhancing echo corrupted speech produced at noisy environment in the time domain.

5.1 Proposed Scheme of Integrated Acoustic Echo and Noise Cancellation

In Fig. 5.1, a schematic diagram of the proposed PSO algorithm based echo cancellation scheme for noisy environment (PSO-TD-AENC) is shown. Here the microphone input signal $y(n)$ consists of the input speech signal $s(n)$, the environmental noise $v(n)$ and the corresponding echo signals $x_s(n)$ and $x_v(n)$. The input environmental noise $v(n)$ is considered as white gaussian noise in this scheme. As stated in the previous section, the echo signals $x_s(n)$ and $x_v(n)$ corrupting the original input signals $s(n)$ and $v(n)$ are generated from the delayed and attenuated version of the same signals $s(n)$ and $v(n)$ and can be expressed as

$$\begin{aligned} x_s(n) &= \mathbf{a}_n^T \mathbf{s}(n-k_0) && (5.1)\\ &= \sum_{k=1}^{p} a_n(k) s(n-k_0-k), && (5.2)\\ x_v(n) &= \mathbf{a}_n^T \mathbf{v}(n-k_0) && (5.3)\\ &= \sum_{k=1}^{p} a_n(k) v(n-k_0-k), && (5.4) \end{aligned}$$

where $\mathbf{s}(n-k_0) = [s(n-k_0-1), s(n-k_0-2), \ldots, s(n-k_0-p)]^T$ and $\mathbf{v}(n-k_0) = [v(n-k_0-1), v(n-k_0-2), \ldots, v(n-k_0-p)]^T$ are vectors of p previous values of $s(n)$ and $v(n)$, respectively, with predefined flat delay k_0. The number p and the values of unknown attenuation coefficients $a_n(k)$ depend on the characteristics of the room.

The task of the adaptive filter block is to produce an estimate of $x_s(n) + x_v(n)$ given $y(n)$ and a reference signal. Since there is no scope to provide a separate reference signal in case of single channel AEC problem, we propose to utilize some delayed versions of the adaptive filter output as the reference signal. However, in case of optimization algorithm based processing, a frame by frame based operation is required. Given a flat delay of k_0 samples, the optimization process starts from k_0 samples and continues frame by frame with a certain percentage of overlap between successive frames. For a frame of N samples, the sum square error $E_{st}(n)$ between the input $(l+1)$-th frame and the corresponding reference frame that the adaptive filter tries to minimize can be defined as

$$\begin{aligned} E_{st}(n) &= \sum_{r=0}^{r=N-1} [y(n-lN-r) - \widehat{x}_s(n-lN-r) - \widehat{x}_v(n-lN-r)]^2 && (5.5) \\ &= \sum_{r=0}^{r=N-1} [s(n-lN-r) + v(n-lN-r) + x_s(n-lN-r) + x_v(n-lN-r) \\ &\quad -\widehat{x}_s(n-lN-r) - \widehat{x}_v(n-lN-r)]^2 && (5.6) \end{aligned}$$

where, $l = 0, 1, 2\ldots$ corresponds to the frame number. Here the reference signal $\widehat{x}_s(n) - \widehat{x}_v(n)$ is an estimate of the echo signal generated by the adaptive filter utilizing its estimated coefficient vector $\widehat{\mathbf{w}}_n$ and the echo suppressed input signal $\widehat{s}(n) + \widehat{v}(n)$ and can be expressed as

$$\begin{aligned} \widehat{x}_s(n) + \widehat{x}_v(n) &= \widehat{\mathbf{w}}_n^T[\widehat{\mathbf{s}}(n-k_0) + \widehat{\mathbf{v}}(n-k_0)] && (5.7) \\ &= \sum_{k=1}^{k=p} [\widehat{w}_n(k)\widehat{s}(n-k_0-k) + \widehat{w}_n(k)\widehat{v}(n-k_0-k)]. && (5.8) \end{aligned}$$

In the proposed method, unlike conventional approaches, we propose to optimize the objective function stated in equation (5.6) using the particle swarm optimization algorithm (PSO).

The basics of the PSO algorithm is already discussed in chapter 3. In Fig. 5.1, a detailed view of the position and operation of the PSO-TD algorithm block in the proposed scheme is shown. The PSO-TD algorithm block takes a frame from the current input signal $y(n)$ and another reference frame from the previously enhanced signal $\widehat{s}(n-k_0) + \widehat{v}(n-k_0)$. Its position vector $\mathbf{w}_t^i$ and velocity vector $\mathbf{v}_t^i$ are randomly initialized, which means at the beginning the algorithm considers a random estimate of the room response. From this estimate, an error $E_{st}(n)$ is

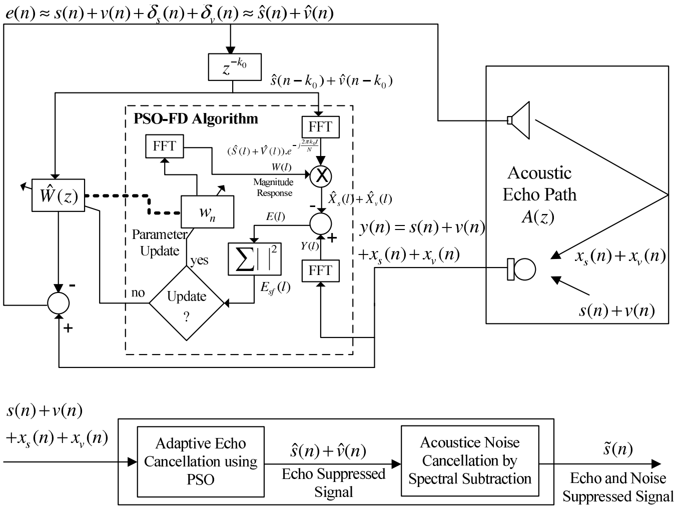
$e(n) \approx s(n)+v(n)+\delta_s(n)+\delta_v(n) \approx \hat{s}(n)+\hat{v}(n)$
z^{-k_0}
$\hat{s}(n-k_0)+\hat{v}(n-k_0)$
PSO-FD Algorithm
FFT
FFT
$(\hat{S}(l)+\hat{V}(l)).e^{-j\frac{2\pi k_0 l}{N}}$
$W(l)$
Magnitude Response
$\hat{X}_s(l)+\hat{X}_v(l)$
$\hat{W}(z)$
w_n
$E(l)$
Parameter Update
yes
$\sum|\ |^2$
$Y(l)$
FFT
no
Update ?
$E_{sf}(l)$
$y(n)=s(n)+v(n)$
$+x_s(n)+x_v(n)$
Acoustic Echo Path $A(z)$
$x_s(n)+x_v(n)$
$s(n)+v(n)$
$s(n)+v(n)$
$+x_s(n)+x_v(n)$
Adaptive Echo Cancellation using PSO
$\hat{s}(n)+\hat{v}(n)$
Echo Suppressed Signal
Acoustice Noise Cancellation by Spectral Subtraction
$\tilde{s}(n)$
Echo and Noise Suppressed Signal

Fig. 5.2: Basic Setup of PSO based single channel integrated echo and noise canceller for enhancing echo corrupted speech produced at noisy environment in the frequency domain.

calculated using equation (5.6). The PSO-TD algorithm then updates the velocity and position vectors of each particles according to the following equations,

$$\begin{aligned} v_{t+1}^i(k) &= \Delta_t v_t^i(k) + c_1 r_{1t}^i(k)(\lambda_t^i(k) - w_t^i(k)) \\ &\quad + c_2 r_{2t}^i(k)(\chi_t^i(k) - w_t^i(k)), k = 0, 1, \ldots, p \qquad (5.9) \\ w_{t+1}^i(k) &= w_t^i(k) + v_{t+1}^i(k), k = 0, 1, \ldots, p. \qquad (5.10) \end{aligned}$$

Then the sum square error $E_{st}(n)$ is again calculated for all the particles. This iterative process of error calculation and parameter update continues until a maximum number of iteration is reached or the difference between two successive updates become stable for a certain number of iteration. The best positional value, i.e. the best estimate of the room response filter coefficient thus obtained, is transferred to the $\widehat{W}(z)$ block to be used for final echo suppression, as shown by the dotted lines.

Proceeding in a similar fashion as followed in the case of time domain analysis, the proposed PSO based AENC scheme (PSO-FD-AENC) is developed in frequency domain in view of getting better performance as expected from the result section of chapter 3.

In the proposed PSO based frequency domain analysis the discrete fourier transform (DFT) of each frame of input data $y(n)$ is performed which is defined as

$$Y(l) = \sum_{n=-\infty}^{+\infty} y(n) e^{-j\frac{2\pi f n}{N} l} \qquad (5.11)$$

In Fig. 5.2 a schematic diagram of the proposed frequency domain AENC method (PSO-FD-AENC) is shown. A frame of the input signal $y(n)$ is supplied to the PSO-FD algorithm block while another frame of echo suppressed noisy signal $\widehat{s}(n-k_0)+\widehat{v}(n-k_0)$ is supplied as reference. The particles are initialized with random position and velocities. The position vector, the current frame and the reference frame - all are transformed into the frequency domain by discrete fourier transform and are represented as $W(l)$, $Y(l)$ and $(\widehat{S}(l)+\widehat{V}(l)).e^{-j\frac{2\pi k_0 l}{N}}$, respectively. As the time domain convolution becomes multiplication in frequency domain, the new estimate of the echo of the input speech and noise in frequency domain $\widehat{X}_s(l)+\widehat{X}_v(l)$ can be denoted as

$$\begin{aligned}\widehat{X}_s(l) &= \widehat{W}(l).\widehat{S}(l).e^{-j\frac{2\pi k_0 l}{M}}, &(5.12)\\ \widehat{X}_v(l) &= \widehat{W}(l).\widehat{V}(l).e^{-j\frac{2\pi k_0 l}{M}}, &(5.13)\end{aligned}$$

Thus, the error $E(l)$ is defined in the discrete frequency domain as

$$\begin{aligned}E(l) &= Y(l)-\widehat{X}_s(l)-\widehat{X}_v(l) &(5.14)\\ &= S(l)+V(l)+X_s(l)+X_v(l)-\widehat{X}_s(l)-\widehat{X}_v(l) &(5.15)\end{aligned}$$

Now, the objective function for optimization can be defined as the sum of the square of the error $E_{sf}(l)$, i.e,

$$E_{sf}(l) = \sum_{l=0}^{N-1}(|S(l)+V(l)+X_s(l)+X_v(l)-\widehat{X}_s(l)-\widehat{X}_v(l)|)^2$$

The PSO adaptive algorithm tries to minimize this error by varying the position vectors of its particles, i.e. by varying the estimated echo path filter coefficients. Here it can be seen that, though the proposed method calculates the mean square error in the frequency domain, it updates the time domain form of the FIR filter $\widehat{w}_n$ not its frequency response. The update of the velocity and position vectors of the particles is an iterative process. The PSO-FD block takes a frame of input signals and calculates the sum square error for the present positions of all the particles. Then according to the rules of the PSO algorithm update, it updates the velocity and position of all the particles. The sum square error is calculated again for the new position vectors and the position update and error calculation is repeated unless a predefined maximum number iteration is reached or the difference between two consecutive updates are stable for a certain number of iterations. The error function is minimized when the filter coefficients of the model echo path $\widehat{w}_n$ are perfectly tuned with the room impulse response a_n. The updated position w_n which is finally obtained is then used as the estimated filter coefficients $\widehat{W}(z)$ of the room response (as shown by the dotted line between the two blocks in Fig. 5.2)to minimize the effect of echo from the current input signal frame. Since time domain convolution becomes multiplication in the frequency domain approach, it is expected that frequency domain modeling of the echo cancellation problem at noisy environment will involve less computation resulting in faster convergence.

The echo suppressed noisy signal is then fed to the noise suppression block where the input noise $v(n)$ and the residual noise are suppressed as shown in Fig. 5.1 and Fig. 5.2 . The modified spectral subtraction method proposed by Berouti et. al [10], described in chapter 4, is employed as the noise cancellation method. The output of the noise subtraction $\widetilde{s}_w(n)$, which is composed of the original input speech signal $s_w(n)$ and some residual noise-like signal $\Psi_w(n)$, is the final output of the system. The signal $\Psi_w(n)$ is very small in magnitude, however it may still contain some signature of the input noise $v(n)$ and the residual echo and noise $\delta_s(n)$ and $\delta_v(n)$.

5.2 Simulation Results

Extensive experimentation has been carried out in order to investigate the echo cancellation performance of the proposed PSO algorithm based time domain (PSO-TD-AENC) and frequency domain (PSO-FD-AENC) approaches and results are compared with the state-of-the-art adaptive LMS filer algorithm based scheme proposed in chapter 4. Thus, for the purpose of simulating various acoustic environments, the room impulse response defined in the simulation section of the previous chapters is considered. The two test speech samples are also the same as in the previous chapters. The improvement of signal to distortion ratio (SDR) in dB and the average echo return loss enhancement (ERLE) in dB are used as basis for performance evaluation.

In Fig. 5.3 and Fig. 5.4 echo cancellation performance obtained by different methods are presented considering the number of unknown coefficients incrementing

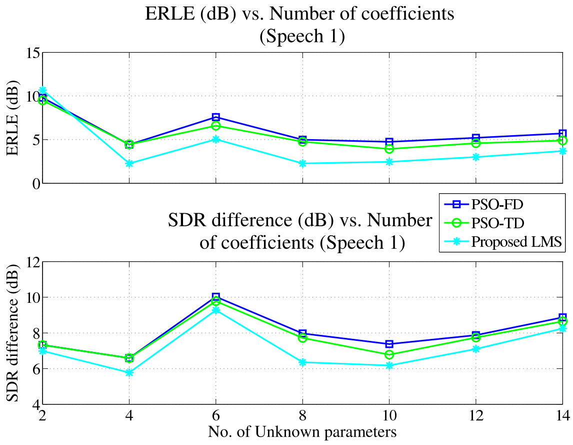

Fig. 5.3: Performance comparison of PSO-TD-AENC, PSO-FD-AENC and LMS based integrated acoustic echo and noise canceller based on ERLE (dB) and SDR difference(dB) for Speech sample 1.

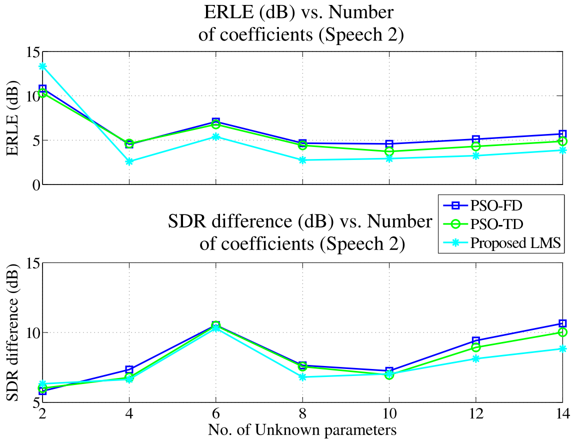

Fig. 5.4: Performance comparison of PSO-TD-AENC, PSO-FD-AENC and LMS based integrated acoustic echo and noise canceller based on ERLE (dB) and SDR difference(dB) for Speech sample 2.

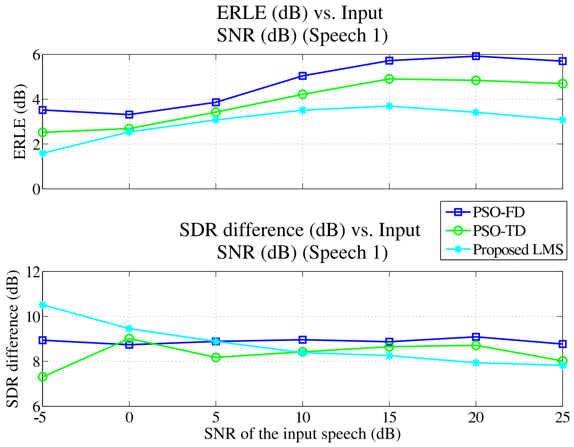

Fig. 5.5: Performance comparison of PSO-TD-AENC, PSO-FD-AENC and LMS based integrated acoustic echo and noise canceller based on ERLE (dB) and SDR difference(dB) for different input SNR (dB) for Speech sample 1.

from 2 to 14. As performance measurement criteria, we consider the ERLE (dB) and the SDR difference (dB). In our experimentation, different sets of PSO parameters have been used. However, the results obtained from the proposed methods in these figures utilize the following PSO parameters: number of particles is 10 for PSO-TD-AENC and 40 for PSO-FD-AENC, tolerance between two consecutive global best values $= 10^{-30}$, search range for coefficients is $[-1\ 1]$, maximum number of iterations is 40, $c_1 = 2$, $c_2 = 2$, $w_{initial} = 0.9$, $w_{final} = 0.4$, and maximum particle velocity $v_{max} = 0.2$. Iterations terminate if tolerance is stable for 5 iterations. The input noise was additive white gaussian and the input signal to noise ratio was 15dB. From the figures, it can be observed that, the performance in terms of the ERLE (dB) is consistently good for the proposed PSO-TD-AENC method than that of the LMS algorithm. Moreover, the proposed PSO-FD-AENC approach is found to be superior to other two methods in terms of both ERLE and SDR difference.

In Fig. 5.5 and Fig. 5.6 echo cancellation performance obtained by different methods are presented considering the input SNR being varied from 25dB to -5dB. It can be seen that the proposed PSO algorithm based methods work quite satisfactorily even for a high energy input noise level. The figures clearly demonstrate the superiority of the proposed PSO-FD-AENC algorithm over the proposed PSO-TD-AENC method and the proposed LMS method with update constraints.

In view of demonstrating the quality of the reconstructed speech obtained by using the proposed methods, in Fig. 5.7, the time waveform of the original speech

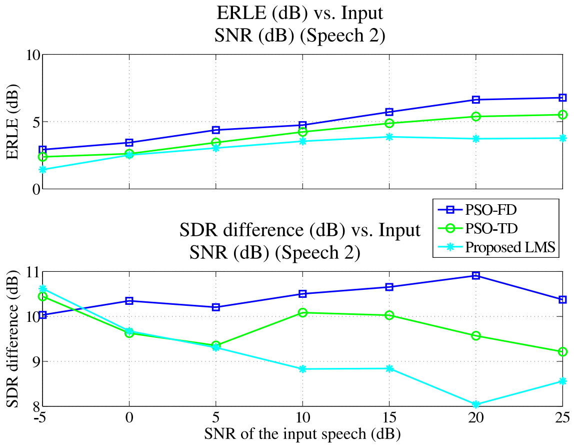

Fig. 5.6: Performance comparison of PSO-TD-AENC, PSO-FD-AENC and LMS based integrated acoustic echo and noise canceller based on ERLE (dB) and SDR difference(dB) for different input SNR (dB) for Speech sample 2.

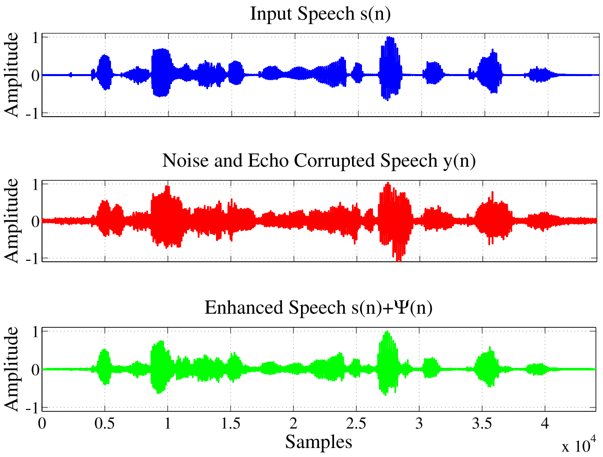

Fig. 5.7: The input speech $s(n)$, echo and noise corrupted speech $y(n)$ and outcome of the proposed PSO based single channel integrated echo-noise canceller in frequency domain $\widetilde{s}(n)$.

$s(n)$, the noise and echo corrupted speech $y(n)$ and the reconstructed speech $\widetilde{s}(n)$ obtained using the proposed PSO-FD-AENC approach are shown. The number of unknown was 14 and the input SNR was set to 15dB. The sample speech used in this experiment contains the speech sample 1. It is vividly seen from this figure that the effect of echo and noise has been completely removed in the reconstructed signal.

5.3 Conclusion

In this chapter, a novel approach of single channel acoustic echo cancellation scheme using an optimization algorithm is proposed for noisy environment. For fast and accurate optimization the renowned particle swarm optimization(PSO) algorithm is chosen for the scheme. PSO is a population based search technique which is superior to some other evolutionary algorithms, namely genetic algorithm, taboo search etc. Two different PSO based schemes were proposed for single channel AEC at noisy environment. One is based on time domain operation of PSO while the other is based on frequency domain transformation of the signals before applying optimization algorithm. The well-known spectral subtraction method is adopted for noise cancellation operation. In the simulation section, the performance of the proposed schemes were evaluated in terms of ERLE and SDR difference by comparing with one another and with the previously described LMS based single channel integrated echo and noise cancellation technique. It is found that the performance of the PSO based frequency domain single channel integrated echo and noise cancellation cancellation scheme (PSO-FD-AENC) outperforms both PSO-TD-AENC and traditional LMS and offers a great reduction of estimation error. It is shown in the result section that the performance of the single channel integrated echo and noise echo cancellation scheme for increasing length of room response and for decreasing SNR is enhanced to a great extent for the proposed PSO algorithm based approach in frequency domain.

Chapter 6

Conclusion

6.1 Summary

The main idea in this thesis work is to develop a single channel acoustic echo cancellation scheme employing the gradient based LMS algorithm and the PSO algorithm. Further improvement of the proposed scheme is suggested by incorporating frequency domain formulation of the PSO algorithm for the scenario and by extending the capability of the proposed method for noisy environment through the addition of a single channel noise suppression block. For the proposed LMS based technique, the power and correlation properties of different speech frames are taken into consideration, based on which some condition on the LMS update are proposed for further improvement in convergence performance. Next, it was shown that optimization algorithm based approach i.e. PSO based approach does not require such conditions to produce a good result, as they do not require to consider the speech properties for update. PSO only tries to match the reference speech frame with the echo corrupted speech frame by varying the position of its particles, i.e. by varying the coefficients of several estimates. In the development of the proposed scheme, effect of variation of some major PSO parameters has been also taken in consideration along with different acoustic environments. The acoustic environments were distinguished by changing the number of unknown coefficients and by employing different input SNR. In order to handle the extremely long filter response of the room, a pre-calculation of the delay has been adopted. Echo cancellation performance has been tested considering different variations in filter parameters resulting in various acoustic operating conditions and the outcomes have been compared with that of some of the standard techniques, such as the LMS and the NLMS adaptive algorithms. It has been shown that, the proposed PSO based schemes, especially the one designed in frequency domain, exhibit superior performance under various acoustic conditions in terms of time per iteration, the ERLE and the SDR difference in comparison to that obtained by other methods. Moreover, the proposed method provides a precise control on the rate of convergence, number of iterations, and quality of output, which enables the user to modify the overall system as par the requirement.

6.2 Future Design Modifications

Now a days, sub-band adaptive filtering is getting much popularity because of the huge reduction in computational burden and thus faster adaptation. The proposed problem of adaptive echo cancellation can be modeled with sub-band adaptive filters using both LMS and PSO algorithms. Performance of the proposed AEC and AENC systems will be evaluated using traditional NLMS and RLS algorithms.

On the other hand, the driving algorithm of the proposed optimization algorithm based adaptive filter in this literature is the well-known particle swarm optimization algorithm (PSO). The system will be evaluated replacing PSO by other contemporary evolutionary search algorithms such as ant colony algorithm [97] [98], bee algorithm [99] [100] etc..

The proposed AENC system is modeled to handle white noise only, which may not be the scenario in real life. Incorporation of a more efficient noise subtraction algorithm, such as MMSE estimation method [42] or an unified framework corresponding to the ML, MMSE, and MAP optimization criteria [101], for detection and suppression of non-white noise in the AENC system may enhance the performance of the proposed system.

Bibliography

[1] H. Yasukawa, I. Furukawa, and Y. Ishiyama, "Acoustic echo control for high quality audio teleconferencing," *in Proc. IEEE International Conference on Acoustic, Speech and Signal Processing*, 1989.

[2] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction.* John Wiley and Sons, 2000.

[3] "How to choose an acoustic echo canceller," *Application Note, Polycom Installed Voice Business Group*, Sep 2004.

[4] S. M. Kuo and B. H. Lee, *Real-Time Digital Signal Processing.* John Wiley and Sons, 2001.

[5] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, Inc., 1996.

[6] H. Deng and M. Dorolovacki, "Improving convergence of the pnlms algorithm for sparse impulse response identification," *IEEE Signal Processing Lett.*, vol. 12, no. 3, 2005.

[7] L. Liu, M. Fukumoto, and S. Saiki, "An improved mu-law proportionate nlms algorithm," *in Proc. IEEE International Conference on Acoustic, Speech and Signal Processing*, 2008.

[8] B. Widrow, J. R. J. Glover, J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, J. E. Dong, and R. C. Goodlin, "Adaptive noise cancelling: Principles and applications," *in Proc. IEEE*, vol. 63, no. 12, 1975.

[9] A. G. V. Turbin and P. Scalart, "Comparison of three post-filtering algorithms for residual acoustic echo reduction," 1997.

[10] M. Berouti, R. Schwartz, and J. Makhoul, “Enhancement of speech corrupted by acoustic noise,” *Proc. of the IEEE conference on Acoustics, Speech and Signal Processing*, Apr 1979.

[11] H. Yasukawa, “Acoustic echo canceller with sub-band noise cancelling,” *Electronics Lett.g*, vol. 28, no. 15, 1992.

[12] C.-T. Lin, “Single-channel speech enhancement in variable noise-level environment,” *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 33, Jan. 2003.

[13] L. B. Asl and V. M. Nezhad, “Improved particle swarm optimization for dual-channel speech enhancement,” *icsap*, 2010.

[14] J. Kennedy and R. C. Eberhart, “Particle swarm optimization,” vol. 4, IEEE Service Center, Piscataway, NJ, 1995, p. 19421948.

[15] Q. Bai, “Analysis of particle swarm optimization algorithm,” *Computer and Information Science*, vol. 3, 2010.

[16] R. Eberhart and J. Kennedy, “A new optimizer using particles swarm theory,” *Sixth Int. Symp. on Micro Machine and Human Science, Nagoya, Japan*, 1995.

[17] F. Heppner and U. Grenander, “A stochastic nonlinear model for coordinate bird flocks,” *The Ubiquity of Chaos*, 1990.

[18] C. W. Reynolds, “Flocks, herds, and schools: a distributed behavioral model,” *Computer Graphics*, vol. 4, no. 21, p. 2534, 1987.

[19] E. O. Wilson, *Sociobiology: The New Synthesis.* Cambridge, MA: Belknap Press, 1975.

[20] K. Parsopoulos and M. Vrahatis, “Recent approaches to global optimization problems through particle swarm optimization,” *Natural Computing*, vol. 1, 2002.

[21] W. T. Reeves, “Particle systems a technique for modelling a class of fuzzy objects,” *ACM Transactions on Graphics*, vol. 2, no. 2, p. 91108, 1983.

[22] H. H. Hoseinabadi, S. H. Hosseini, and M. Hajian, “Optimal power flow solution by a modified particle swarm optimization algorithm,” *43rd Int. Univ. Power Engineering Conference*, 2008.

[23] X. Hu, “Particle swarm optimization.” [Online]. Available: http://www.swarmintelligence.org/tutorials.php

[24] H. H. Balci and J. F. Valenzuela, “Scheduling electric power generators using particle swarm optimization combined with the lagrangian relaxation method,” *Int. J. Appl. Math. Comput. Sci.*, vol. 14, no. 3, 2004.

[25] T. M. Blackwell and P. Bentley, “Don’t push me! collision avoiding swarms,” *Proc. Congress Evolutionary Computation, Honolulu, USA*, 2002.

[26] A. Salman, I. Ahmad, and S. Al-Madani, “Particle swarm optimization for task assignment problem,” *Microprocess. Microsyst.*, vol. 26, no. 8, 2002.

[27] G. Schmidt, “Applications of acoustic echo control: An overview,” *in Proc. Eur. Signal Process. Conf. (EUSIPCO04), Vienna, Austria*, 2004.

[28] C. Breining, P. Dreiseitel, E. Hnsler, A.Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, “Acoustic echo controlan application of very-high-order adaptive filters,” *IEEE Signal Process. Mag.*, vol. 16, no. 4, Jul 1999.

[29] E. Hnsler, “The hands-free telephone problem: An annotated bibliography,” *Signal Process.*, vol. 27, no. 3, 1992.

[30] E. Hnsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach.* New York: Wiley, Jun 2004.

[31] V. Myllyl, “Residual echo filter for enhanced acoustic echo control,” *Signal Process.*, vol. 86, no. 6, Jun 2006.

[32] G. Enzner, “A model-based optimum filtering approach to acoustic echo control: Theory and practice,” Ph.D. dissertation, RWTH Aachen Univ., Aachen, Germany, Apr 2006.

[33] M. Ihle and K. Kroschel, “Integration of noise reduction and echo attenuation for handset-free communication,” 1997.

[34] F. Buoteille, C. Beaugeant, and P. Scalart, “Solution to acoustic echo control: pesudo APA and post-filtering,” 1999.

[35] S. J. Park, C. G. Cho, C. Lee, and D. H. Youn, “Integrated echo and noise canceller for hands-free applications,” *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 49, no. 3, Mar 2002.

[36] S. M. Kuo and J. Chen, “Analysis of finite length acoustic echo cancellation system,” *Speech Commun.*, vol. 16, no. 3, Apr 1995.

[37] Y. Grenier, M. Xu, J. Prado, and D. Liebenguth, “Real-time implementation of an acoustic antenna for audio-conference,” Sep 1989.

[38] M. Xu and Y. Grenier, “Acoustic echo cancellation by adaptive antenna,” Sep 1989.

[39] R. Le, B. Jeannes, P. Scalart, G. Faucon, and C. Beaugeant, “Combined noise and echo reduction in hands-free systems:a survey,” *IEEE Trans. Speech Audio Process.*, vol. 9, no. 9, Nov 2001.

[40] C. Beaugeant, V. Turbin, P. Scalart, and A. Gilloire, “New optimal filtering approaches for hands-free telecommunication terminals,” *Signal Process.*, vol. 64, no. 1, Jan 1998.

[41] P. J. S. Gustafsson, R. Martin and P. Vary, “A psychoacoustic approach to combined acoustic echo cancellation and noise reduction,” *IEEE Trans. Speech Process.*, vol. 10, no. 5, Jul 2002.

[42] Y. Ephraim and D. Malah, “Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator,” *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 32, no. 6, pp. 1109 – 1121, dec 1984.

[43] R. Martin and P. Vary, “Combined acoustic echo cancellation, dereverberation and noise reduction: A two microphone approach,” *Proc. Annales des Telecomm.*, vol. 49, no. 78, JulAug 1994.

[44] M. Drbecker and S. Ernst, “Combination of two-channel spectral subtraction and adaptive wiener post-filtering for noise reduction and dereverberation,” 1996.

[45] J. B. Allen, D. A. Berkley, and J. Blauert, “Multimicrophone signalprocessing technique to remove room reverberation from speech signals,” *J. Acoust. Soc. Amer.*, vol. 62, no. 4, 1977.

[46] P. Bloom and G. Cain, “Evaluation of two input speech dereverberation techniques,” vol. 1, 1982.

[47] A. Oppenheim, R. Schafer, and J. T. Stockham, “Nonlinear filtering of multiplied and convolved signals,” *Proc. IEEE*, vol. 56, no. 8, Aug 1968.

[48] D. Bees, M. Blostein, and P. Kabal, “Reverberant speech enhancement using cepstral processing,” vol. 2, 1991.

[49] B. Yegnanarayana and P. Murthy, “Enhancement of reverberant speech using lp residual signal,” *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, May 2000.

[50] K. Lebart and J. Boucher, “A new method based on spectral subtraction for speech dereverberation,” *Acta Acoustica*, vol. 87, 2001.

[51] E. Habets, “Multi-channel speech dereverberation based on a statistical model of late reverberation,” Mar 2005.

[52] J. Wen, N. Gaubitch, E. Habets, T. Myatt, and P. Naylor, “Evaluation of speech dereverberation algorithms using the mardy database,” Sep 2006.

[53] F. Forgó, “Nonconvex programming,” *Akadmiai Kiad, Budapest*, 1988.

[54] E. R. Hansen, “Global optimization using interval analysis,” *Marcel Dekker, New York*, 1992.

[55] R. Horst and P. M. Pardalos, *Handbook of Global Optimization.* London: Kluwer Academic Publishers, 1995.

[56] R. Horst and H. Tuy, *Global Optimization Deterministic Approaches.* New York: Springer, 1996.

[57] J. D. Pintér, *Global Optimization in Action.* Academic Publishers, 1996.

[58] S. S. Rao, *Engineering optimization-theory and practice.* Wiley, 1996.

[59] H. P. Schwefel, *Evolution and Optimum Seeking.* Wiley, 1995.

[60] A. Törn and A. Žilinskas, *Global Optimization.* Berlin: Springer-Verlag, 1989.

[61] T. Bäck, D. Fogel, and Z. Michalewicz, *Handbook of Evolutionary Computation.* New York: IOP Publishing and Oxford University Press, 1997.

[62] T. Bäck, *Evolutionary Algorithms in Theory and Practice.* New York: Oxford University Press, 1996.

[63] H.-G. Beyer, *The Theory of Evolution Strategies.* Berlin: Springer, 2001.

[64] H.-G. Beyer and H.-P. Schwefel, “Evolution strategies: A comprehensive introduction,” *Natural Computing*, 2002.

[65] I. Rechenberg, “Evolution strategy,” *In: Zurada JM, Marks RJ II and Robinson C (eds) Computational Intelligence: Imitating Life*, 1994.

[66] G. Rudolph, *Convergence properties of evolutionary algorithms.* Hamburg: Verlag Dr. Kovač, 1997.

[67] H.-P. Schwefel, “Evolutionsstrategie und numerische optimierung,” Ph.D. dissertation, Department of Process Engineering, Technical University of Berlin, 1975.

[68] ——, *Numerical Optimization of Computer Models.* Chichester: Wiley, 1981.

[69] ——, *Evolution and Optimum Seeking.* New York: Wiley, 1995.

[70] H.-P. Schwefel and G. Rudolph, “Contemporary evolution strategies,” 1995, p. 893907.

[71] D. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning.* MA: Addison Wesley, Reading, 1989.

[72] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs.* Berlin: Springer, 1994.

[73] W. Banzhaf, P. Nordin, R. E. Keller, and F. D. Francone, *Genetic Programming An Introduction.* San Francisco: Morgan Kaufman Publishers, 1998.

[74] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* Cambridge, MA: MIT Press, 1992.

[75] D. Fogel, *Evolutionary Computation: Towards a New Philosophy of Machine Intelligence.* Piscataway, NJ: IEEE Press, 1996.

[76] J. Kennedy and R. C. Eberhart, *Swarm Intelligence.* San Francisco: Morgan Kaufman Publishers, 2001.

[77] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue, "Timit acoustic-phonetic continuous speech corpus," *Linguistic Data Consortium, Philadelphia*, 1993.

[78] M. D. Topa, I. Muresan, B. S. Kirei, and I. Homana, "A digital adaptive echo-canceller for room acoustics improvement," *Advances in Electrical and Computer Engineering*, vol. 10, pp. 450–453, Apr. 2004.

[79] U. Mahbub and S. A. Fattah, "An acoustic echo cancellation scheme based on gradient based adaptive filtering," *Signal Processing [Submitted].*

[80] G. Sentoni and A. Altenberg, "Nonlinear acoustic echo canceller with dabnet + fir structure," *2005 IEEE Workshop on Application of Signal Processing to Audio and Acoustics*, Oct 2005.

[81] F. Guangzeng and L. Feng, "A new echo caneller with the estimation of flat delay," *IEEE Region 10 Conference. Tencon 92, Melbourne. Australia*, Nov 1992.

[82] L. B. Asl and V. M. Nezhad, "Improved particle swarm optimization for dual-channel speech enhancement," *icsap*, 2010.

[83] U. Mahbub and S. A. Fattah, “An acoustic echo cancellation scheme based on particle swarm optimization,” *EURASIP Journal on Audio, Speech, and Music Processing [Submitted].*

[84] ——, “A spectral domain acoustic echo cancellation scheme based on particle swarm optimization,” *EURASIP Journal on Advances in Signal Processing [Submitted].*

[85] Y. del Valle, G. K. Venayagamoorthy, S. Mohagheghi, J.-C. Hernandez, and R. G. Harley, “Particle swarm optimization: Basic concepts, variants and applications in power systems,” *IEEE Transactions On Evolutionary Computation*, vol. 12, no. 2, Apr 2008.

[86] A. Carlisle and G. Dozier, “An off-the-shelf pso,” 2001, p. 16.

[87] U. Mahbub and S. A. Fattah, “An integrated acoustic echo and noise cancellation scheme based on gradient based adaptive filtering,” *The IEEE Transection on Audio, Speech and Language Processing [Submitted].*

[88] M. R. Weiss, E. Aschkenasy, and T. W. Parsons, “Study and development of the intel technique for improving speech intelligibility,” *Rome Air Development Center Report*, no. RADC-TR-75-77, Mar 1975.

[89] S. Boll, “A spectral subtraction algorithm for suppression of acoustic noise in speech,” *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP ’79.*, Apr 1979.

[90] J. S. Lim, “Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 26, no. 5, Oct 1978.

[91] J. S. Lim and A. V. Oppenheim, “Enhancement and bandwidth compression of noisy speech,” *Proc. IEEE*, vol. 67, no. 12.

[92] R. J. Niederjohn, P. Lee, and F. Josse, “Factors related to spectral subtraction for speech in noise enhancement,” *Proceedings of the IEEE International Conference on Industrial Electronics, Control and Instrumentation.*

[93] S. M. McOlash, R. J. Niederjohn, and J. A. Heinen, “A spectral subtraction method for the enhancement of speech corrupted by non-white, non-stationary noise,” *IECON*, 1995.

[94] R. Martin, “Noise power spectral density estimation based on optimal smoothing and minimum statistics,” *IEEE Trans. Speech and Audio Processing*, vol. 9, no. 5.

[95] U. Mahbub and S. A. Fattah, “An integrated acoustic echo and noise cancellation scheme based on particle swarm optimization,” *Speech Communication [Submitted].*

[96] ——, “A spectral domain integrated acoustic echo and noise cancellation scheme based particle swarm optimization,” *IET Signal Processing [Submitted].*

[97] M. Dorigo, M. Birattari, and T. Stutzle, “Ant colony optimization,” *Computational Intelligence Magazine, IEEE*, vol. 1, no. 4, pp. 28 –39, Nov. 2006.

[98] G. Tambouratzis, “Using an ant colony metaheuristic to optimize automatic word segmentation for ancient greek,” *Evolutionary Computation, IEEE Transactions on*, vol. 13, no. 4, pp. 742 –753, aug. 2009.

[99] M.-M. Member and R. E. Velez-Koeppel, “Elitist artificial bee colony for constrained real-parameter optimization,” *Computational Intelligence*, pp. 18–23, 2010. [Online]. Available: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5586280

[100] M. Subotic, M. Tuba, and N. Stanarevic, “Different approaches in parallelization of the artificial bee colony algorithm,” *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 5, no. 4, pp. 191–196, 2010. [Online]. Available: http://portal.acm.org/citation.cfm?id=1863431.1863463

[101] B. J. Borgstrom and A. Alwan, “A unified framework for designing optimal STSA estimators assuming maximum likelihood phase equivalence of speech

and noise," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 8, pp. 2579 –2590, Nov. 2011.