Sudan University of Science & Technology

College of Graduate Studies

A comparison of Improved Meta-Heuristic Optimisation Based Speech-Enhancement Systems

مقارنة نظم تحسين الكلام المبنية على طرق ما وراء الاستكشاف الأمثل

A Thesis Submitted for the Degree of Doctor of Philosophy in Computer Science

By

Taiser Mirgahni Fath Elrahman Taha

Supervisor: Prof. Amir Hussain

February 2021

ABSTRACT

Speech enhancement has become an area of interest to researchers in the field of machine learning as a result of the development of digital signal processing applications. The goal of speech enhancement is to increase the quality of these digital speech devices and update them to handle all sorts of noises. In the case of dual-channel speech enhancement systems, an Adaptive Noise Cancellation (ANC) system contains an adaptive filter and an adaptation algorithm (optimisation procedure) for adjusting their parameters. This thesis investigates

the use of meta-heuristic methods to propose a novel speech enhancement systems. This results in exploiting different meta-heuristic optimisation techniques, such as Particle Swarm Optimisation (PSO), Gaussian Par- ticle Swarm Optimisation (GPSO), Accelerated Particle Swarm Optimi- sation (APSO), Bat Optimisation (BA), Gravitational Search Algorithms (GSA). This thesis presents the formulation of an ANC system based on Butterworth, and Elliptic filters, in the form of an optimisation task. PSO, GPSO, APSO, GSA,and BA are used to find the optimal filter coefficients, that optimise the perceptual evaluation of speech quality (*PESQ*), sig- nal distortion (*C sig*), overall signal quality (*C ovrl*), and Log-Likelihood Ratio (*LLR*), for the noise-free audio

signal and the filtered signal. This made it possible to build a system capable of enhancing speech, where we employed different sentences from the NOIZEUS data-set and ARABIC speech corpus under various Signal to Noise Ratio (SNR) values. Objec-

tive and subjective evaluation tests, were conducted on the meta-heuristic speech enhancement systems and showed encouraging results. The results of the proposed speech enhancement systems revealed that PSO generally outperformed APSO and GPSO at all levels of SNR. The optimised Elliptic filter by BA showed improved scores compared to the fixed filter, and the audio-only Wiener filter at all SNR levels.

The results confirmed that meta-heuristic based acoustic noise cancellation models considered are capable of high performance.

المستخلص

أصبح تحسين الكلام مجال اهتمام الباحثين في مجال التعلم الآلي نتيجة لتطوير تطبيقات معالجة الإشارات الرقمية. الهدف من تحسين الكلام هو زيادة جودة أجهزة الكلام الرقمية هذه وتحديثها للتعامل مع جميع أنواع الضوضاء. في أنظمة تحسين الكلام ثنائية القناة، يحتوي نظام إلغاء الضوضاء التكيفي على مرشح متكيف وخوارزمية تحسين لضبط معلماتها.

تم التحقق في هذه الاطروحة من استخدام طرق ما وراء الاستكشاف لاقتراح أنظمة جديدة لتحسين الكلام. نتج عن هذا استغلال تقنيات التحسين المثلى المختلفة مثل خوارزمية تحسين سرب الجسيمات الكلاسيكي، القاوسي، والمتسارع وخوارزمية البحث عن الجاذبية وخوارزمية الخفافيش. في هذه الأطروحة تم تطوير نظم إلغاء ضوضاء تكيفية قائمة على مرشحات بتروورث والناقصية، في شكل مسائل تحسين مثلى. تقنيات التحسين كتحسين سرب الجسيمات الكلاسيكي، القاوسي، والمتسارع وخوارزمية البحث عن الجاذبية وخوارزمية الخفافيش استخدمت للعثور على معاملات المرشحات المثلى، والتي تعمل على تحسين التقييم الإدراكي لجودة الكلام وتشويه الإشارة، والجودة الإجمالية للإشارة، ونسبة احتمال تسجيل الدخول، لكلتا الإشارتين الخالية من الضوضاء والمصفاة.

أدى ذلك إلى بناء نظام قادر على تحسين الكلام حيث استخدمنا جمل كلام مختلفة من مجموعة بيانات NOIZEUS ومجموعة النصوص العربية تحت قيم مختلفة للإشارة الى نسبة الضوضاء. تم إجراء اختبارات تقييم موضوعية وباستخدام كل من مقياس التقييم الإدراكي لجودة الكلام الشائع، وقياسات موضوعية مركبة، واختبارات شخصية كمتوسط درجة الاستطلاع الرأي الشخصي. اختبارات تقييم الكلام أجريت على أنظمة طرق ما وراء الاستكشاف وأظهرت نتائج مشجعة. أظهر المرشح الإهليلجي المحسّن بواسطة خوارزمية الخفافيش أداءاً أفضل مقارنة بالمرشح الثابت، ومرشح وينر الصوتي فقط في جميع مستويات نسب الإشارة للضجيج.

أكدت النتائج أن نماذج إلغاء الضوضاء الصوتية القائمة على طرق ما وراء الاستكشاف الأمثل التي تم وضعها في الاعتبار توفر أداءاً عالياً.

Contents

List of Figures

List of Tables

Acronyms

AEC	Acoustic Echo Canceller
ANC	Adaptive Noise Canceller
APSO	Accelerated Particle Swarm Optimisation
BA	Bat Algorithm
CD	Ceptral Distance
CI	Cochlear Implant
CEP	Cepstrum Distance
CNN	Convolution Neural Network
DFBSS	Dual Forward Blind Source Separation
DFNLMS	Dual Forward Normalised Least Square
DFT	Discrete Fourier Transform
DNN	Deep Neural Network
EVD	Eigen Value Decomposition
FIR	Finite Impulse Response
GSC	Genaralised Sidelobe Cancellation
GA	Genetic Algorithm

GPSO	Gaussian Particle Swarm Optimisation
GSA	Gravitational Search Algorithm
HOT	Hirschman Optimal Transform
IIR	Infinite Impulse Response
I-MMSS	Modified Multiband Spectral Subtraction
IS	Itakura-Saito Distance
ITU-T	International Telecommunication Union
LMS	Least Mean Square
LLR	Log Likelihood Ratio
LSD	Log Spectra Distance
LPSO	Learning Based Particle Swarm Optimisation
MMSE	Minimum Mean Square Error
MOS	Mean Opinion Score
MPSO	Modified Particle Swarm Optimisation
MSA	Modern Standard Arabic
NLMS	Normalised Least Mean Square
NNSE	Neural Network Speech Enhancement
PESQ	Perceptual Evaluation of Speech Quality
PSO	Particle Swarm Optimisation
PSOGSA	Hybrid Particle Swarm Optimisation Gravitational Search Algorithm
RLS	Recursive Least Square

SegSNR	Segmental SNR
SM	System Mismatch
SNR	Signal to Noise Ratio
SPIB	Signal Processing Information Base
SPSO	Standard Particle Swarm Optimisation
SSPSO	Shuffled Sub-Swarms Particle Swarm Optimisation
STFT	Short-Time Fourier Transform
STSA	Short-Time Spectral Amplitude
SVD	Singular Value Decomposition
TFGSC	Transfer Function Generalised Sidelobe
WSS	Weighted Spectral Slope

Chapter 1

Introduction

1.1 Introduction

Enhancing speech in our daily noisy world and its modern communication systems, such as teleconferencing systems, hearing aids, mobile communications is a challenging task. Due to constant interference in the environment, the quality of communication declines and speech intelligibility is affected. Hence, speech enhancement aims to enhance the intelligibility, and quality of degraded speech signal and enable successful and effective communication. Depending on the specific application, the goal of speech enhancement is to improve the quality of the degraded speech, reduce listener fatigue, increase speech intelligibility, or improve the performance of the communication device. Over the past decades, several conventional speech-enhancement methods have been applied to enhance speech, such as spectral-subtraction(Boll, 1979), Wiener filter (Wiener, 1949),or subspace approaches (Ephraim and Van Trees, 1995), (Hu and Loizou, 2002). These qualify as one-channel speech-enhancement strategies and have the disadvantage of underestimating noise and introducing musical noise to the enhanced signal. Recently, meta-stochastic approaches have shown their utility in designing dual-channel speech enhancement systems.

1.2 Thesis motivation

The motivation for this research arose from the need to exploit meta-heuristic stochastic optimization methods for speech enhancement. Meta-heuristic optimisation methods are both popular and applicable in various fields or disciplines. According to Mirjalili et al. (2014) meta-heuristic algorithms have become prominent because they assure these properties: simplicity, flexibility, derivation-free mechanisms, and escaping local optima. First, meta-heuristic optimisation approaches are quite simple and inspired from nature (animal behaviour, physical phenomena, etc.). Second, they demonstrate flexibility, which means these algorithms can be applied to many different types of problems without much alteration. Third, most of these meta-heuristic algorithms have derivation-free processes; the algorithm starts with random solution(s), and to find the optimum, it is not required to calculate the derivative of the search spaces. Finally, these meta-heuristic methods can escape the local optima compared to conventional optimisation algorithms, because of their stochastic nature, which enables them to escape the stagnant local solutions, and exploit the entire search space.

1.3 Problem statement

There is a lack of such approaches in the literature on dual-channel speech enhancement based on meta-heuristic stochastic techniques. The commonly used methods are the gradient descent optimisation techniques, such as least mean-squares (LMS), normalised LMS (NLMS) and recursive least squares (RLS) (Widrow and Stearns, 1985), (Gorriz et al., 2009), (Guopin et al., 2013). These methods have weaknesses in the sense that they are local and unable to converge and reach the global optimal solution (get stuck at local minima). To address the problem of getting stuck at local minima and convergence, approaches based on global meta-heuristic optimisation techniques have been used in the literature [Karabu Ga and Cetinkaya (2011),Prajna et al. (2014b), Prajna et al. (2014a), (Tripathi and Ikbal, 2015), Loubna et al. (2018), Ghibeche et al. (2019) Liu et al. (2020)]. These methods seek to minimise the mean square error. To the best of our knowledge, no method in the literature used an objective function that related to signal quality measurements, such as the perceptual evaluation of speech quality(PESQ), log-likelihood ratio (LLR), signal distortion level $C\,Sig$. On the other hand, the methods using meta-heuristic algorithms have not provided clear methodology for how the parameters of their meta-heuristic methods are selected.

1.4 Research aims and objectives

The aim of this research is to develop a novel speech-enhancement system by using meta-heuristic optimisation techniques. Hence, the objectives of this thesis are as follows:

- To investigate the use of meta-heuristic optimisation methods for speech enhancement, and to propose a new speech-enhancement methods based on optimisation.
- To develop, formulate and utilise the Adaptive Noise Canceller problem as an optimization task.
- To develop and formulate a fitness function based on signal quality measurements.
- To evaluate and compare the performance of the proposed techniques with existing sate-of-art methods, by conducting both objective and subjective tests.

1.5 Methodology

We formulate a fitness (cost) function that depends on signal quality measurements including PESQ, LLR, $C\,Sig$ and $C\;ovrl$. In order to find the optimal filter for noise cancellation, five meta-heuristic approaches are developed to adjust the Butterworth and the Elliptic filter coefficients.These optimisation techniques are: the particle swarm optimization and its variants, the gravitation search algorithm, and the Bat algorithm. The grid search algorithm will be used to find the optimal set of parameters of the meta-heuristic algorithms. To evaluate the performance of the proposed systems the following are to be used:

1. Objective evaluation measurements, which include the PESQ and composite measures. To find if the mean of the optimised signal is significant to the mean of the clean signal, the T test is used at the significant level of 0.05.

2. Subjective evaluation measurements, where human volunteers are recruited to listen to the filtered sentences, through the Mean Opinion Score system.

1.6 Thesis contributions

1. A novel speech enhancement based on meta-heuristic optimization techniques was developed.

2. A fitness function based on speech quality measurements is integrated.

3. The grid search algorithm is used to select the parameters of meta-heuristic methods in order control the learning process of the proposed algorithms.

1.7 Thesis scope

- Conventional methods have many flaws; meta-heuristic expound on these to look for optimal solutions but very few are used in the literature.

- Global optimisation techniques are successful in many applications, but few were used for speech enhancement.

- The type of noise considered in this thesis is additive noise.

- The method of speech enhancement considered is by adaptive filtering.

- This research only considers the meta-heuristic optimisation algorithms.

1.8 Thesis structure

This section provides the organisation of the thesis.

Chapter 2 presents and details the literature review, and provides a summary of the existing categories of audio speech-enhancement methods based on the type of algorithm used; the input of channels involved uni-modal or multi-modal speech - enhancement is given. Then types of noises are identified. Followed by a review of a number of state-of-the-art speech-enhancement techniques, such as spectral subtraction methods, statistical model-based algorithms, and subspace speech-enhancement methods. The pros and cons of these methods are also provided. The adaptive noise cancellation concept and its conventional methods are also revised. Adaptive filtering techniques, such as the least mean square and the recursive least squares algorithms, were reassessed. The last part of this chapter focuses on some of the machine learning approaches used in speech enchantment. Furthermore, a few prominent and recent optimisation techniques with regard to speech enhancement are also addressed. Strong and weak points for every method are summarised as well.

Chapter 3 details the research methodology, along with the approaches and techniques used to build the meta-heuristic speech-enhancement systems.

Chapter 4 presents the design of the speech-enhancement system, based on PSO and its variants. A review about swarm systems is given. The proposed speech-enhancement system, employing ANC based on (PSO, APSO, and GPSO) algorithms is outlined. The stages of the proposed speech enhancement are described. Furthermore, a comparison was drawn between the proposed methods and the state-of-the-art audio-only, dual speech-enhancement algorithms to determine which approach performs better.

In chapter 5 gravitation search algorithm is utilised to introduce a speech enhancement system. This chapter describes and introduces the GSA and how it works. The structure and the components of the proposed novel GSA speech enhancement are presented, where the exploration ability of the GSA is utilised in the search space. The proposed system is tested in a noisy environment at various SNR levels, to evaluate the performance of the proposed algorithm.

In Chapter 6 a speech-enhancement system based on the Bat algorithm was developed. The background and characteristics of the Bat algorithm are presented. The design and the implementation of the proposed BA speech-enhancement system are provided. The proposed system is tested in a noisy environment at various SNR levels, to evaluate the performance of the proposed algorithm.

In Chapter 8 a comparison conducted for the previous methods that were introduced in the previous chapters.

Finally, Chapter 8, provides the concluding remarks from the work conducted in this thesis, and discusses the recommendations for future research.

1.9 Publications

Taha, T. M., Wajid, S. K., & Hussain, A. Speech enhancement based on adaptive noise cancellation and Particle Swarm Optimization. Journal of Computer Science, 15(5), 691-701,2019

Taha, T. M., & Hussain, A. A Survey on Techniques for Enhancing Speech. International Journal of Computer Applications, 179(17), 1-14, 2018

Chapter 2

Review and Background to Speech-enhancement Methods

In this chapter, we focus on the techniques featured in literature to enhance the speech signal. Various methods are used including Wiener filter, statistical methods, subspace method, and spectral subtraction. We will discuss various types of speech-enhancement methods along with their advantages and disadvantages. The discussion will also review the studies conducted by other researchers on other machine learning techniques, such as the neural network, Deep Neural Network ,Convolution Neural Networks, and optimisation techniques used for the enhancement of speech.

2.1 Introduction

Speech enhancement is a vital element to communication equipment. It refines speech and reduces noise, and it is used in a various domains for example to assist in hearing and other applications such as mobile phones, teleconferencing systems, hearing aids, and voice communication systems.

Speech enhancement is closely related to speech restoration because it reconstructs and restores the signal after degradation(Banchhor et al., 2013).

However, there is a slight difference between speech restoration and speech enhancement. Speech restoration involves converting the noisy signal back to its original form, prior to noise addition. Speech enhancement, on the other hand, helps in refining the original signal. Also, an original under-girded speech signal cannot be restored, but it can be enhanced (Ravi and Subbaiah, 2016). The aim of these speech-enhancement algorithms, to improve the perceptual aspects of the speech signal, is degraded by the additive noise, such as overall quality or intelligibility with the aim of reducing listener fatigue (Loizou, 2013b) (Panchmatia et al., 2016).

Speech enhancement can be used in different settings, such as in areas where there is an interfering background noise in a building, or on noisy streets or roads where there are motor vehicles passing. These interference noises degrade the original speech quality in such a way that it does not remain clear anymore. An important context that needs to be addressed for speech enhancement includes the compression of speech bandwidth systems(D and K, 2015). This is mostly used in the decoding of digital channels of communication. This technique is also needed for the decoding of speech, which includes integration of data and voice networks, including speech bandwidth compression systems that play an important role in speech communication systems.

Authors in (Ravi and Subbaiah, 2016) conducted a survey on single channel speech-enhancement methodologies,(Dixit and Mulge, 2014) considered single and multichannel speech enhancement in their review paper. (Chaudhari and Dhonde, 2015) presented a review on the spectral subtraction method and its modification.The authors (Kulkarni et al., 2016) addressed time and transform domain speech-enhancement methods. Statistical techniques for speech enhancement are reviewed by (Sunnydayal et al., 2014).

In this chapter, the authors classify speech-enhancement methods into four categories: conventional methods, adaptive filtering methods, machine learning methods (this includes adaptive filtering using optimisation techniques), and multi-modal methods.

Figure 2.1: Popular techniques reviewed in the current chapter

This chapter is organised as follows: Section 2.1 gives an introduction to the problem and a general overview. Section 2.2 gives the reader an understanding of the types of audio- enhancement categories. Sections 2.4 - 2.6 discuss four basic approaches of speech enhancement. A Summary is then provided in Section 2.7 .

2.2 Categories of audio-enhancement methods

According to (Lim and Oppenheim, 1979a) and (Loizou, 2013b), the classification of speech enhancement techniques based on the following:

1. Type of the algorithm used, which can either be adaptive or non-adaptive.
2. The input channels involved, which can either be single, dual, or multiple.
3. Whether the speech -enhancement technique is uni-modal or multi-modal.

The following sections, state the differences between the aforementioned techniques.

2.2.1 Comparison between adaptive and non-adaptive speech enhancement

If additive noise is present in a speech signal, then common practice is to pass it through a filter that removes the noise while minimally interfering with the original signal component. This is called direct filtering. Initial work in this domain of optimal filtering, was conducted by (Wiener, 1949) and was extended and enhanced by (Singh, 2001) and others.The filters used for direct filtering can be either fixed or adaptive.

- Fixed filters: to design these, it is important to have prior knowledge of both the signal and the noise. It passes frequencies present in the signal and discards the frequency band occupied by the noise.
- Adaptive filters: can adjust their impulse response to filter out the correlated signal component in the speech input. They require almost no prior knowledge of the characteristics of signal and noise. (in case, the signal is narrow-band and the noise is broadband, or vice versa, no prior information is needed; otherwise it is necessary to know the desired response of the signal). They can adaptively track the signal in the presence of non-stationary conditions.

2.2.2 Comparison between monaural and binaural speech enhancement

Single-channel enhancement, also known as monaural enhancement, is for situations where only one input channel is present such as mobile telephony (Ravi and Subbaiah, 2016). In multichannel speech enhancement, the noisy observations are obtained from two or more sensors. If there are only two channels in the multichannel system, then it is also called binaural enhancement.

Multichannel algorithms show better performance with respect to substantial speech-reception threshold scores when the target signal and the noise source are separated (Mauger et al., 2014). However, in practical scenarios, these requirements might not always be fulfilled, and single-channel algorithms are preferred for devices, such as hearing aids in which the number of microphones is usually limited to two and the two microphones are on the same side of the head (thus recording the same signal) (Goehring et al., 2017). Multichannel speech enchantment can be employed in autonomic speech recognition, or hands-free telephone systems in cars Meyer and Simmer (1997).

2.2.3 Comparison between uni-modal and multi-modal speech enhancement

According to (Monaci, 2007), the use of internal stimuli in senses enables individuals to identify different perceptions in the environments that they live in. Humans integrate acoustic and visual signals(Driver, 1996) (McGurk and MacDonald, 1976)(Wallace et al., 2004) (Watkins et al., 2006) or tactile and visual inputs (Violentyev et al., 2005) (Bresciani et al., 2006).

If audio perception is enhanced using just the auditory sense, then this can be referred to as uni-modal audio enhancement. On the other hand, when audio perception is enhanced by other senses, such as the auditory and the visual sense, then it is referred to as multi-modal speech enhancement.

2.3 Types of noise

When processing a speech signal, we may come across a number of noise types that it may be contaminated with common types of noise that can be added to speech signals are listed by (Lakshmikanth et al., 2014):

1. Background noise: environmental distortion or noise of cars on the road for example.
2. Echo: that occurs in closed spaces with bad acoustics.
3. Acoustic: also known as audio feedback: caused by electrical reflection on the circuits.
4. Amplifier noise: if amplifier produces even a little additional thermal noise, it becomes hugely noticeable after amplification process. Such noise is called amplifier noise.
5. Quantization noise: it is created as part of the transformation process of the signal from analogue to digital domain, interference occurs in sampling while rounding up real values of analogue signal.
6. Loss of signal quality: caused by coding and speech compression.

Because of the huge amount of works reported in this field, this thesis will only consider the case when the noise is additive and independent of the clean speech. Techniques purely for echo cancellation source separation case studies are not reviewed in this research. Various speech-enhancement techniques have been implemented for the purpose of improving the perceptual aspects of a speech signal, that has been degraded by additive noise. These techniques improve overall quality and intelligibility, and reduce listener fatigue (Loizou, 2013b; Panchmatia et al., 2016)

2.4 Conventional methods of speech enhancement

This section will discuss different single-channel speech enhancement methods.

2.4.1 Spectral subtraction method (single-channel speech enhancement)

Spectral subtraction method is one of the oldest methods of single-channel speech enhancement. It is considered to be among the first algorithms in this domain (Boll, 1979). It is simple and effective in the elimination of stationary background noise. Its limitation is that it suffers from narrow-band tonal, commonly called 'musical noise' (Vaseghi, 2008). Various modifications of spectral subtraction have been proposed to improve its results (Hu and Loizou, 2002). The signal corrupted by the noise $y(n)$, composed of the clean speech signal $x(n)$ and the additive noise signal $d(n)$, i.e:

$$y(n) = x(n) + d(n) \tag{2.4.1}$$

Since the speech signal is non-stationary, the noise component is processed frame-by-frame in the frequency domain (Hussain et al., 2007). The discrete time Fourier transform for both sides yields:

$$Y(\omega) = X(\omega) + D(\omega) \tag{2.4.2}$$

To obtain the spectrum of enhanced speech, the method by (Loizou, 2013b) is used:

$$|\hat{X}(\omega)| = |Y(\omega)| - |\hat{D}(\omega)| \tag{2.4.3}$$

Where $|\hat{X}(\omega)|$ is the estimated short time speech magnitude, $|Y(\omega)|$ is the noisy short time speech magnitude and $|\hat{D}(\omega)|$ is a noise spectral magnitude estimate computed during non-speech activity. The power spectrum subtraction is then given by:

$$|\hat{X}(\omega)|^2 = \begin{cases} |Y(\omega)|^2 - |\hat{D}(\omega)|^2 & \textit{if}\ |Y(\omega)|^2 > |\hat{D}(\omega)|^2 \\ 0 & \textit{otherwise} \end{cases} \tag{2.4.4}$$

Much work has been done to suppress noise that occurs as a side product of the spectral subtraction method using varied forms of spectral subtraction (Bahoura, 2017),(Lu and Loizou, 2008), spectral over-subtraction (Berouti et al., 1979), multi-band spectral subtraction (Kamath and Loizou, 2002) non- linear spectral subtraction (Lockwood and Boudy, 1992), iterative methods (Ogata and Shimamura, 2001) and spectral subtraction based on perceptual properties (Virag, 1999).The following points explain these variants of spectral subtraction methods.

1. **Spectral over-subtraction method**: A further modification in the basic spectral subtraction method (Boll, 1979) resulted in a variation commonly known as the spectral over-subtraction method (Berouti et al., 1979). The following parameters were introduced to reduce noise:

   (a) Over-subtraction factor: which has control over the amount of noise power spectrum subtracted from the noisy speech power spectrum.

   (b) Noise spectral floor: which restricts the resultant spectral component from increasing above a pre-set minimum spectral flow value (Loizou, 2013b).

$$|\hat{X}(\omega)|^2 = \begin{cases} |Y(\omega)|^2 - \alpha|\hat{D}(\omega)|^2 & \mathit{if} |Y(\omega)|^2 > (\alpha+\beta)|\hat{D}(\omega)|^2 \\ \beta|\hat{D}(\omega)|^2 & \mathit{otherwise} \end{cases} \tag{2.4.5}$$

   where $\alpha \geq 1$ and $0 \geq \beta \geq 1$.

2. **Multi-band spectral subtraction**: The Multi-band spectral subtraction is another variation of this type, where the speech spectrum is partitioned into several non-overlapping regions, and then spectral subtraction is applied on each band separately. The clean speech spectrum is represented by the following the mathematical model (Loizou, 2013b)(Kamath and Loizou, 2002):

$$|\hat{X}_i(k)|^2 = \begin{cases} |Y_i(k)|^2 - \alpha_i\delta_i|\hat{D}_i(k)|^2 & \mathit{if}\ |Y_i(k)|^2 > 0, \\ \quad k_i < k < k_{i+1} & \\ \beta|\hat{Y}_i(k)|^2 & \mathit{otherwise} \end{cases} \tag{2.4.6}$$

where k_{i-1}, and k_{i+1} are the beginning and the ending of the frequency bins of the i^{th} frequency band, α_i the over-subtraction factor of the i^{th} band, and δ_i is a tweaking factor in the i^{th} band.

Band-specific over-subtraction is represented as a function of segmented SNR_i of the i^{th} frequency band. Its mathematical representation is provided below (Loizou, 2013b):

$$\alpha_i = \begin{cases} 5 & if\ SNR \leq -5 \\ 4 - \frac{3}{20} SSN_i & if\ -5 \leq SNR_i \leq 20 \\ 1 & if\ SNR > 20 \end{cases} \tag{2.4.7}$$

and δ_i the control of each band is calculated as:

$$\delta_i = \begin{cases} 1 & if\ f_i \leq 1kHz \\ 2.5 & if\ 1kHz \leq f_i \leq \frac{f_s}{2} - 2kHz \\ 1.5 & if\ f_i > \frac{f_s}{2} - 2kHz \end{cases} \tag{2.4.8}$$

where f_i is the upper frequency of the i^{th} band, and f_s is the sampling frequency (Kamath and Loizou, 2002; Loizou, 2013b).

3. **Non-linear spectral subtraction**: (Lockwood and Boudy, 1992) introduced a modified version of the over-subtraction by proposing a technique where the nature of the subtraction process is nonlinear and the over-subtraction factor frequency depends upon the frame SNR (Loizou, 2013b).

Over the years, many modifications have been suggested to vary the original method of the spectral subtraction algorithm in order to reduce the musical noise that occurs. (Hu et al., 2002) proposed the combination of comb filtering and spectral smoothing along with formant intensification for the enhancement of noisy speech. This brings significant improvements over the classical method in terms of perceived sound quality. A limitation was that these researchers only performed testing on two noise types (white Gaussian and car noise). A major drawback was that the simulation analysis showed a contradiction between objective and subjective measures.

A further modified multi-band spectral subtraction (I-MBSS) algorithm was proposed in (Upadhyay and Karmakar, 2013) for the enhancement of the audio speech signal in different noise conditions. Here, the noise-speech spectrum is partitioned into K non-overlapping bands and the spectral over-subtraction method was applied independently on each band. The experimental analysis was conducted on different types of added noise. The data-set used for simulations was the NOIZEUS speech corpus. However, the simulations were not conducted on extremely low and high SNR levels.

A new algorithm comprising generalized sidelobe cancellation (GSC) combined with spectral subtraction speech enhancement was put forth by (Yu and Su, 2015). Their research showed that the output signal from the GSC module removes the remaining non-coherent noise upon filtering. They selected the additive noise from NOISEX-92, and their method showed prominent improvement in speech quality. The method was feasible enough to yield stable results.

(Cao et al., 2012) designed a modulated filter bank that was over-sampled to divide the time series into equal sub-spaced bands. The authors (Cai and Hou, 2012) used a weighted recursive averaging method to approximate the noise power spectrum, after which a multi-band subtraction was applied on noise that was added to the speech signal. An auditory masking threshold was computed with the estimated speech signal. In this way, the subsequent associated subtraction factor was adjusted. Experimentation proved their algorithm to be effective enough to enhance a signal that had been corrupted by white noise and also by musical noise. The only drawback was that they did not perform testing with multiple objective measures. They solely used the takura-saito distance (IS) objective measure to evaluate the proposed method.

Another non-linear spectral subtraction technique for speech enhancement was used by (Prabhakaran et al., 2014). Three types of noise were used to evaluate the proposed approach (pink noise, white noise, and Volvo noise). These samples were taken from the data-set of the TIMIT & NOIZEUS corpus. The authors (Islam et al., 2014) conducted a study on a speech-enhancement approach that was formulated on the modified spectral subtraction process carried out on the time magnitude spectral. Extensive testing was done on the NOIZEUS database. The simulation results showed that their proposed method is only suitable for higher segmental SNR.

In (Bharti et al., 2016) an adaptive method of noise cancellation and signal estimation that is based on short term energy is presented. In this technique, the noise spectrum is continuously updated. NOIZEUS speech corpus was used for the evaluation of the proposed approach. This method works well for both stationary and non-stationary noise. The only drawback is that system performs poorly at 0 db stationary noise. Another effort was made by authors in (Zhang and Liu, 2018) to propose a multi-band spectral subtraction algorithm using mel-scale. Their algorithm outperforms the multi-band spectral subtraction with uniform segmentation and also the conventional spectral subtraction algorithm; however, they did not conduct subjective tests to evaluate the results.

2.4.2 Statistical model-based algorithms

Statistical model-based methods are considered among the common techniques for speech denoizing. The method from this type operates in the noisy domain. In this method, noise is reduced by modifying the frequency spectrum of the noise signal(Ding et al., 2004). The two algorithms of this category are:

1. **Wiener filter speech-enhancement method**: The Wiener filter operates in the frequency domain. Its modified version, called the adaptive Wiener filter, operates in the time domain. The original Wiener filter (Wiener, 1949) was introduced in 1949. It is quite similar in nature to the spectral subtraction method. It trades the subtraction step of spectral subtraction with an approximation of the signal spectrum of clean signal with a minimum mean

square error (MMSE). It also involves the computation of short-time Fourier transform (STFT). The technique reduces the MSE between the approximated signal magnitude spectrum $\hat{D}(\omega)$ and the original signal magnitude spectrum $D(\omega)$.

The optimal Wiener filter is represented by (Lim and Oppenheim, 1979b):

$$H(\omega) = \frac{D_s(\omega)}{D_s(\omega) + D_n(\omega)} \tag{2.4.9}$$

where $D_s(\omega)$ and $D_n(\omega)$ are the estimated power spectra of the noise-free signal and the background noise(noise assumed to be uncorrelated and stationary). Finally, the speech is enhanced by:

$$\hat{D}(\omega) = X(\omega)H(\omega) \tag{2.4.10}$$

(Almajai and Milner, 2011) proposed a visually derived Wiener filter for speech enhancement, which exploits the audio-visual correlation. Wiener filters have the additional point that they can be used for both single and dual/ multiple channels.

Table 2.1: Pros and cons of Wiener filtering algorithms

Pros	Cons
• The algorithm safeguards a de-reverberation performance that does not depend on the azimuth angle of the speech source • It preserves binaural cues • This algorithm is proficient enough to significantly reduce the effects of the reverberation especially in the rooms that are highly reverberant • The algorithm is less complex in terms of computing calculations	There is a wide room for improvement in its performance in the rooms with moderate reverberation

The adaptive Wiener filter is dependent on the variation of the filter transfer function from sample to sample according to speech signal statistics (mean/variance). It was proposed by [(Abd El-Fattah et al., 2008),(El-Fattah et al., 2014)], and it works in the time domain instead of the frequency domain (the original Wiener filter works in the frequency domain). A recursive noise estimation approach is used for noise estimation. Authors in (Sulong et al., 2016) combined the compressing sensing method and Wiener filter for noise reduction.

(Coto-Jimenez et al., 2018) presented hybrid method by combining the Wiener filter with a deep neural network; the hybrid system showed good results in terms of enhancing noisy speech at different SNR with different noise types. However, the main disadvantage of their proposed method is that consumes a lot of time and entails high computational costs.

(Khaldi and Touati, 2018) initiated a new speech-enhancement method based on the Wiener filter and spectral subtraction. The method achieved higher SNR improvement than when using the Wiener filter or spectral subtraction methods. However, the authors need to do subjective listening tests and evaluate it under different experimental conditions.

2. **Maximum likelihood method**: This method was brought forth by (Ephraim and Malah, 1985), (Cappé, 1994). (Ephraim and Malah, 1984) proposed a method based on the estimation of the short-time spectral amplitude (STSA). The author derived the MMSE STSA estimator, based on modelling noise and speech spectral components as statistically independent Gaussian random variables and analysed the performance of the proposed STSA estimator and compared it with Wiener-estimator-based STSA estimator.

   The MMSE STSA estimator is used to examine signals based on the quality or strengths in the deafening conditions and in the areas where there is uncertainty in the presence of the signals. To construct the enhanced signal, an MMSE STSA estimator is used with the compound exponential of the deafening segment. A priori, the probability distribution of the speech and noise Fourier expansion coefficients should be known to derive the MMSE STSA estimator. The same authors (Ephraim and Malah, 1985) also proposed the short-time spectral amplitude (STSA) estimator for speech signals and inspect it in the context of enhancing noisy speech. The results indicated that the new estimator is good in improving noisy speech. The main Ephraim and Malah (1985) noise suppression rule is expressed in the following part. Neglecting the time and the frequency indexes(l, ω) for notation limitations, the suppression value was $G(l, \omega)$ applied to each short-time spectrum value $X(l, \omega)$ to give(Ephraim and Malah, 1984):

$$G(l,\omega) = \frac{\pi}{2}\sqrt{\frac{1}{1+R_{post}}\frac{R_{prio}}{1+R_{prio}}} \tag{2.4.11}$$

$$*M\left((1+R_{post})\frac{R_{prio}}{1+R_{prio}}\right)$$

where M is a function based on the modified Bessel functions of zero and first order.

$$M[\theta] = \exp(-\frac{\theta}{2})\left[(1+\theta)I_0\left(\frac{\theta}{2}\right) + \theta I_1\left(\frac{\theta}{2}\right)\right] \tag{2.4.12}$$

The formulations of the a-priori SNR ($R_{prio}(l, \omega)$) and a-posteriori SNR ($R_{post}(l, \omega)$) respectively (for each value of the time and frequency indexes) are given below:

$$R_{post}(l,\omega) = \frac{|X(l,\omega)|^2}{D(\omega)} - 1 \tag{2.4.13}$$

$$R_{prio}(l,\omega) = (1-\alpha)P[R_{post}(l,\omega)] + \alpha\frac{|G(l-1,\omega)X(l-1,\omega)|^2}{D(\omega)} \tag{2.4.14}$$

where $D(\omega)$ is the noise power at frequency ω, with $P[x] = x$ if $x \geq 0$ and $P[x] = 0$ otherwise. ($R_{prio}(l, \omega)$) is an estimate of the SNR that takes into account the current short-term frame with weight $(1-\alpha)$ and the noise reduced previous frame with weight α(Cappé, 1994).

To improve speech intelligibility, (Jia et al., 2019) proposed a new method based on MMSE log spectral amplitude to estimate speech phases.For the purpose of experimental evaluation authors' data was selected from NOIZEUS and the noises are taken from the Noise-92 speech library at different SNR levels of 0db, 5db, 10db, and 15db. The evaluation results of their new method indicated that it can greatly improve the quality and intelligibility of speech. However, the authors did not run subjective evaluation tests.

Another contribution was made by Soni and Vaghela (2017), where the authors proposed a hybrid speech-enhancement system based on MMSE and spectral subtraction methods. The NOIZEUS database used for evaluation and different types of noises are taken from the AURORA database at various of SNR levels. The results are very promising, however the authors mentioned it produces musical noises. (Gouhar et al., 2017) introduced a new improved method for speech enhancement based on MMSE, based on the popular searching algorithm called binary search. The proposed algorithm is tested using car, babble, street, train and white Gaussian noises, which are added to sentences taken from the NOIZEUS corpus at different SNRs levels. The simulation results revealed that their proposed algorithm outperforms other MMSE algorithms.

2.4.3 Subspace speech-enhancement methods

Another type of speech-enhancement methods is when speech estimation is considered as a constrained optimisation problem. This approach was introduced by (Ephraim and Van Trees, 1995), and by (Hu and Loizou, 2002), where the noisy speech signal vector cosmos is decayed into two subspaces i.e, a signal subspace and a noise subspace. The singular value decomposition (SVD) or the eigenvalue decomposition (EVD) is used to decompose the noisy signal into a noise signal and a speech signal.

(Surendran and Kumar, 2016) proposed a signal subspace speech-improvement algorithm using the perceptual feature by using the frequency-disguising property of the human auditory system (Jabloun and Champagne, 2003). A cue to spectral deviation ratio (SSDR) standardisation is used for the reduction of the spectral misrepresentation. Samples of speech are used from the NOIZEUS database for the assessment of the introduced algorithm. The results of their experiments showed the effectiveness of their algorithm in speech enhancement compared to some benchmark speech-enhancement methods.

An approach of the subspace method on the basis of the Karhunen-Loève transform and customs principal component analysis was proposed by (Yan et al., 2013) for the reduction of noise in different noisy environments. They used objective assessment measures (including segmental SNR (SegSNR), weighted spectral slope (WSS), the log-likelihood ratio (LLR), log spectral distance (LSD) and perceptual evaluation of speech quality (PESQ) to assess the performance of their algorithms. It was shown that their algorithm was more operative for white noise than coloured noise. The Performance was not good for SNR greater than 10dB.

An effort was made by (SUN et al., 2016), to introduce an algorithm based on joint low-rank and sparse matrix decomposition (JLSMD). It is different from the preceding subspace algorithms in its decomposition nature. The results showed that their algorithm is better for improving the overall quality of the enhanced speech, however, noise reduction still had room for improvement. Table 2.2 summarises the advantages and disadvantages of the main conventional speech-enhancement methods(Hifrin et al., 2014):

2.5 Adaptive noise cancellation(ANC)

Basically, an ANC denotes the electromechanical or electro-acoustic procedure of abandoning acoustic disruption to produce a softer environment (Lakshmikanth et al., 2014). ANCs create and use an 'anti-noise' signal with the same amplitude and opposite phase. The adaptive noise canceller has been used in a number of applications such as hearing protectors, headsets, and so on. The ANC can be globalised to a multichannel system, which can be seen as a generalised beamforming system. The

Table 2.2: Advantages and disadvantages of main conventional speech-enhancement methods

Speech enhancement	Advantages	Disadvantages
Spectral subtraction	The Spectral subtraction is effective in computational and has modest contrivance to control the trade-off between speech misrepresentation and remaining noise	The introduced musical noise is disadvantage
MSSE estimator	It has fewer computational assets and resources	Absence of the mechanism to control trade-off between speech distortion and the remaining noise
Wiener filter	Reasonable computation load	Absence of the mechanism to control trade-off between speech distortion and the remaining noise
Subspace	It delivers a mechanism to control the trade-off between speech distortion and the remaining noise	It results in heavy computational loads

adaptive noise canceller was initially introduced by (Widrow and Stearns, 1985). It requires minimum two microphones founded on the basis of the obtainability of orientation channel(s) which are features of associated samples or references of the polluted noise. An estimate of the noise is produced with the help of *adaptive filter* by utilising the reference microphone output. Its output is then deducted from the primary microphone output (signal + noise). The output of the canceller is used to regulate the tap weights in the adaptive filter. With the help of an adaptation algorithm, the ANC minimises the mean square error value of the output. It generates output, which is the best approximation of the anticipated signal in the sense of the minimum mean square error (Taha et al., 2018). The ANC removes or suppresses a noisy signal by using adaptive filters and adjusting their parameters according to an optimisation algorithm. Many works reported in the literature use adaptive filters for noise reduction and cancellation (Akhaee et al., 2005), (Kalamani et al., 2014), (Nataraj et al., 2017).

Adaptive filters fine-tune their coefficients to diminish the error signal and can be grasped as finite impulse response (FIR), infinite impulse response (IIR), lattice and transform domain filters. The Least mean square (Singh, 2001) is the most common adaptive algorithm. The advantages and disadvantages of this method are listed below (Lakshmikanth et al., 2014).

Table 2.3: Pros and cons of Adaptive noise canceller

Pros	Cons
The customary wideband algorithms of ANC produce the best results in the lower frequency bands	• As the bandwidth and the center frequency of the noise upsurges, their performance depreciates quickly. • The algorithms are not appropriate for the multimodal error surface, and they provide a single likely solution for each reiteration according to the generated error. • It is necessary to have a frequency dependent noise cancellation system to avoid adversely affecting the desired signal in order to combine the ANC system with other communication and sound systems

2.5.1 Adaptive filters

An adaptive filter is a device used for computational purposes and it endeavours to create and establish the association between two signals in real time in an iterative style. An adaptive filter is defined by the following phases (Kunche and Reddy, 2016b):

1. The signal being treated by the filter.

2. The configuration that describes how the output signal of the filter is calculated from its input signal.

3. The limitations within this structure that can be iteratively altered to change the filter's input-output association

4. The adaptive algorithm that defines how the limitations are attuned from one time prompt to the next.

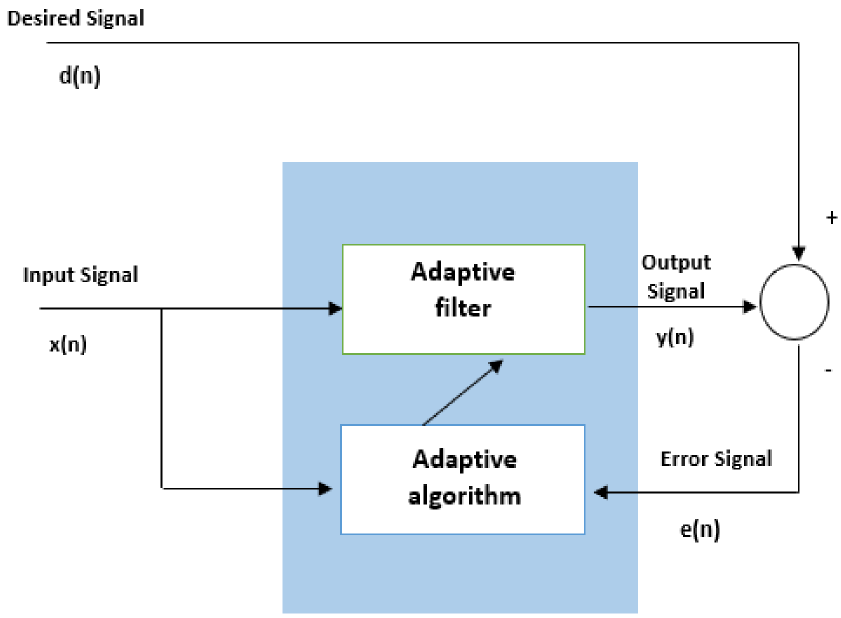


Figure 2.2: The adaptive filter (Das and Sarma, 2012)

The parameters' type and number to be adjusted are determined by selecting a certain adaptive filter form. The adaptive algorithm seeks to update the parameter values of the system: usually, it is derived as a form of optimisation operation to minimize an error. Figure 2.2 shows a block diagram of the adaptive filter, where a signal $x(n)$ is fed into a device called an Adaptive Filter, where an output $y(n)$ is computed by this device. This output signal is compared to a second signal $d(n)$, called the desired response signal, and when subtracting the two signals at time n, the error signal $e(n)$ is produced.

$$e(n) = d(n) - y(n) \tag{2.5.1}$$

Then this error signal is fed into the adaptive algorithm (optimisation procedure) to adapt or alter the parameters of the filter from time n to $(n + 1)$ in a well-defined way. This operation is referred to as "Adaptation", where by the parameters of the given system are changed from the time indexed at n to a time at $(n + 1)$.

The problem is to determine the general form of the structure of the adaptive filter, to determine the best linear relation between the input and the desired response signals. In most cases, this linear relation takes the form of FIR or IIR. An IIR is a linear digital filter that computes their output recursively and produces feedback, where the current output depends on the previous ones (Haykin, 1996).

$$y(n) = \sum_{i=0}^{L} a_i x(n-i) - \sum_{i=0}^{M} b_i y(n-i) \tag{2.5.2}$$

where a_i and b_i are the coefficients of the filter, $M \geq L$ is the filter order. The transfer function of the M^{th} order filter, is given by:

$$H(z) = \frac{A(z)}{B(z)} = \frac{\sum_{i=0}^{L} a_i z^{-i}}{1 + \sum_{i=0}^{M} b_i z^{-i}} \tag{2.5.3}$$

The characteristics of the adaptive filter have to be modified so that the output of the adaptive filter looks like the desired response as much as possible. That is why the desired response of an adaptive filter is related in some way to the input signal and is made available to the filter. as closely as possible.

The error is the difference between the adaptive and the desired and the adaptive filter response.

$$e(n) = d(n) - y(n) \tag{2.5.4}$$

An optimisation process like the mean square error or any other fitness is used, to drive the error to zero.

Conventional adaptive-filters include classic Butterworth filters, Chebyshev filters, and Elliptic filters.

A Butterworth filter provides the maximum flat response and its calculations are comparatively simpler than for other forms of filters. This factor, combined with the fact that it produces impressive performance for most applications, has made it a popular choice in the field of electronics-RF as well as with audio active filters (Adrio, 2015).

An Elliptic filter (also called a Cauer filter) has ripples in the pass-band well as in the stop-band (Adrio, 2013). Ripple levels in the pass-band and stop-band are independently adjustable during the design phase

- When a ripple in the stop band approaches zero, the filter becomes a Chebyshev type I.
- When a ripple in the pass-band approaches zero, then the filter becomes a Chebyshev type II.
- When a ripple in both, the stop and the pass-bands approaches zero, the filter becomes a Butterworth type.

2.5.2 Least mean squares (LMS) algorithm

One of the extensively used techniques for adaptive filtering is the LMS algorithm. Its foundation is credited to (Widrow and Stearns, 1985) and (Haykin, 1996). It is based on the approximation of the gradient in the direction of the optimal solution using the arithmetical properties of the input signal. A noteworthy feature of the LMS algorithm is its straightforwardness. In this algorithm filter weights are rationalised with each new sample as required to meet the anticipated output. An acoustic echo canceller (AEC) is used to remove the acoustic response from the loudspeaker to the microphone in applications, such as hands-free telephony, tele-classing, and video-conferencing.

Adaptive filters with thousands of coefficients are used for room acoustic echo cancellation. Transform domain adaptive filters result in a noteworthy decrease in computational weight. (Krishna et al., 2010) present Hirschman optimal transform (HOT) based adaptive filter for the elimination of echo from audio signals. In order to test the efficacy of the proposed method, adaptive algorithms based on LMS, Normalised least mean squares(NLMS), discrete Fourier transform(DFT)-LMS and HOT-LMS were implemented and tested in this echo cancellation application. Their experiments proved that the HOT based LMS adaptive filter is computationally effective and has fast convergence as compared to LMS, NLMS, and DFT-LMS. For the cancellation or suppression of the assorted noise, they used this spectrogram technique to sense and eradicate noise. Through the method described in this paper, 12dB or more SNR can be attained, and the noise reduction coefficient becomes greater than 0.9.(XU et al., 2017)(Sayoud et al., 2017)(Arif et al., 2017)

2.5.3 Recursive least squares (RLS) algorithm

As the adaptive filter is based on the alteration of the treated signal, it uses an adaptive algorithm for the alteration of the filter limitations and structure (Guopin et al., 2013). Normally, it is just the filter coefficients that are altered and the remainder of the filter structure is stay the same.

The RLS adaptive algorithm for noise cancellation uses the error signal to regulate the weight coefficients of the adaptive filter, and therefore attains a filter output that is an estimate of the interference signal, and then uses the mixed signal with the noise component to subtract the filter output in order to acquire the strong signal and achieve the output of eliminating the noise signal.

This method was used by (Guopin et al., 2013) to conduct research on speech enhancement using signals which had periodic noise mixed with impulse noise. A time-frequency spectrogram was used by them in order to pre-process the noisy signal, then they passed the signal through an RLS adaptive noise reduction system to terminate the noisy component. (Rakesh and Kumar, 2015) achieved the same results; this proved again that RMS is superior compared to NLMS for noise cancellation.

A new dual forward blind source separation (FBSS) algorithm was introduced by (Djendi et al., 2016) based on the use of the recursive least square algorithm to update the cross-filters of the forward structure. This algorithm combines the good features of both- the FBSS and RLS algorithms. This DFRLS algorithm was used in speech enhancement and acoustic noise reduction applications. Their method showed good results as compared to the dual forward normalised least mean square (DFNLMS) algorithm with respect to the segmental signal to noise ratio (SegSNR), the cepstral distance (CD), the system mismatch (SM), and the segmental mean square error (SegMSE). A summary is given in Table 2.4 for the advantages and dis-advantages of both the least mean square(LMS) and the recursive least square(RLS).

Table 2.4: Advantages and dis-advantages of LMS and RLS

	Pros	Cons
LMS	• The implementation of the LMS algorithm is simple (Krishna et al., 2010). • The HOT based LMS adaptive filter is computationally effective and has fast convergence as compared to LMS, NLMS and DFT-LMS (Krishna et al., 2010)	• Simple LMS has sluggish convergence and gradient noise amplification (Dewasthale et al., 2015)
RLS	• SNR can increase up to 10dB or more noise • The reduction coefficient can exceed 0.9 (Guopin et al., 2013). • Good noise reduction can be achieved.(Djendi et al., 2016) • RLS has a quicker rate of convergence as compared to LMS(Dhiman et al., 2013) • It has reduced steady-state errors(Dhiman et al., 2013) • Its spectral characteristics are better enhanced than for LMS(Rakesh and Kumar, 2015)	• Its effects are mostly restricted to periodic noise, low-frequency noise signal (Guopin et al., 2013)

2.6 Machine learning approaches to speech enhancement

2.6.1 Neural networks for speech enhancement

A speech-enhancement algorithm was evaluated by (Goehringa, et al (Goehring et al., 2017). It was based on neural networks speech enhancement (NNSE) to improve speech intelligibility in noise for cochlear implant (CI) users. The algorithm decays the noisy speech signal into time-frequency divisions, extracts a set of auditory characteristics, and inserts them into the neural network to yield an approximation of frequency channels that contain more perceptually significant statistics (higher signal-to-noise ratio). This approximation is used to reduce noise-dominated components and retain speech-dominated components for electrical stimulation. The architecture and low processing delay of the NNSE algorithm make it appropriate for application in hearing devices.

A regression-based speech-enhancement framework was presented by (Xu et al., 2014). It used deep neural networks (DNNs) with a deep architecture having multiple-layers. A restricted Boltzmann machine pre-training scheme was introduced to prepare the DNN. A huge training set is fundamental to learn the rich structure of the DNN. Using more acoustic framework statistics is shown to improve performance and make the enhanced speech less intermittent. Multi-condition training can address the speech augmentation of new speakers, hidden noise types, numerous SNR levels under different noise circumstances, and even cross-language generalisation. Compared with the SNN-based and log-MMSE methods, noteworthy enhancements were attained on the TIMIT corpus. On average, 76.35% subjective preference was attained due to the nonappearance of musical noise in improved speech.

Subsequently, the same authors introduced an altered version of this work in(Xu et al., 2015). This was an administered technique to improve speech by means of finding a mapping function between noisy and clean speech signals based on deep neural networks (DNNs). This method can well suppress extremely non-stationary noise, which is hard to handle in general. Additionally, the subsequent DNN model, trained with synthetically created data, is also effective in dealing with noisy speech data logged in real-world situations without the generation of the infuriating musical artifact usually seen in conventional enhancement methods. Multi-condition training with many kinds of noise categories can attain a good generalisation proficiency to hidden noise surroundings. By doing so, the proposed DNN framework is also influential in managing the non-stationary noises in real-world situations. Compared with the log-MMSE technique, noteworthy enhancements were attained across different hidden noise situations. The sole disadvantage was that training data was too limited to cover a wide range of various acoustic scenarios, such as speaker and language inconsistencies.

Another attempt to use DNN for speech enhancement and less aggressive Wiener filtering as an additional DNN layer was made in (Saleem et al., 2019a), (Saleem and Khattak, 2020). The same author used an ideal binary mask (IBM) as a binary classification function in DNN (Saleem et al., 2019b). Their proposed algorithm outperformed the competing methods in the literature. A model based on a signal- to-noise ratio (SNR) aware convolution neural network (CNN) was addressed for speech enhancement (SE) in (Fu et al., 2016), (Shi et al., 2018), (Pandey and Wang, 2019). This CNN model can efficiently handle the local temporal and spectral speech signals. Hence, the model can effectively separate the speech signals and noise from an input signal. Two SNR-aware algorithms were proposed using CNN with the intention of improving the generalisation capability and accuracy of these models.The first algorithm incorporates a multi-task learning (MTL) framework. The noisy speech signal is fed as input to the model. Given the input, the algorithm primarily restores noise-free speech signals. Then, the SNR level is estimated for the processed clean speech signals. The second algorithm is based on SNR adaptive denoizing. The algorithm initially computes the SNR level. Then, based on the calculated SNR level, an SNR-dependent CNN model is chosen for reducing the noise. It was found that max-pooling is not required here for speech enhancement due to its reduced capability in representing complex speech patterns. It is justified from the results that the two proposed SNR-aware CNN models outperform the deep neural networks in terms of standardised objective evaluations, provided the number of layers and nodes are defined to be the same. Additionally, the SNR-aware CNN models possess enhanced denoising potential even with unseen SNR levels. This shows promising robust potential for real-world applications.

Most recently, in 2017, another CNN model was proposed towards complex spectrogram enhancement in order to solve the difficulty in phase estimation (Fu et al., 2017). The proposed model identifies clean real and imaginary (RI) spectrograms from noisy spectrograms. These restored RI spectrograms are then utilised to generate enhanced speech waveforms. These waveforms possess phase information with high accuracy. An objective function was formulated using multi-metric learning (MML) criterion such that more than one metric is deemed. The main idea behind MML is that any signal representation can be portrayed as a function of the RI spectrograms. With optimal selection of β, MML can boost multiple objective metrics (log-spectral distortion(LSD) and segmental signal-to-noise ratio (SSNR)) concurrently. The lift in the performance can be justified by considering MML as a pseudo-layer over the original objective function. This process is believed to improve the generalisation capability of the original model. Table 2.5 presents the summary of the various kinds of neural networks in the field of speech enhancement.

Table 2.5: Pros and cons of machine learning methods

	Pros	Cons
NN	• Low computational complexity • Fewer processing delay s	• Needs improvement in accuracy • Not exceptionally good in terms of generalisation performance under unpredictable conditions
DNN	• Better performance than SNN-based and L-MMSE methods (Xu et al., 2014) • Remarkable improvements in both objective and subjective metrics when compared with the conventional MMSE-based technique • Quite effective in handling real-world distorted noisy speech in various languages and across varying recording conditions not observed during DNN training(Xu et al., 2015) • Effective suppression of highly non-stationary noise, which is usually difficult to handle.	• Improvement needed in the generalisation capability of DNN towards unseen noise (Xu et al., 2015) • Demand for large training set to provide good coverage of different acoustic environments such as speaker and language variations(Xu et al., 2015)
CNN	• Higher performance than DNN(Fu et al., 2016) • Efficient in handling local spectral and temporal structures of speech signals.(Fu et al., 2016) • Effective decomposition of the speech and noise signals from the noisy input signals.(Fu et al., 2016) • Lack of necessity for Max pooling(Fu et al., 2016). • Enhanced de-noising performance with unseen SNR levels(Fu et al., 2016) • Promising approach for real-world applications	• Computationally expensive approach(Fu et al., 2017)

2.6.2 Optimisation techniques for speech enhancement

This section reviews a few prominent and recent optimisation techniques with regard to speech enhancement. All the optimisation techniques mentioned here consider that such a dual-channel enhancement is used where one channel is for pure noise while the other is dedicated to speech distorted by noise.

1. **Particle swarm optimisation and its variants**

   (Mahbub et al., 2010) considered the variation in the total number of considered particles in different acoustic environments. They conducted research on different kinds of noise and voices, and also under varied operating conditions. They compared the results of PSO with other adaptive algorithms, namely LMS and NLMS. Their experiments showed that PSO outperforms other techniques with respect to SNR improvement, and showed a satisfactory convergence rate under different acoustic conditions. (Asl and Nezhad, 2010) proposed a modified PSO (MPSO) and compared it with PSO when used for adaptive filtering in the enhancement of speech signals. Their experimental results showed that MPSO is capable of a much faster search speed when finding an optimal solution. Moreover, MPSO improves SNR to a greater extent than the simple PSO. This improvement is more pronounced in the construction of higher order filters.

   APSO was used for speech enhancement in 2014 by (Prajna et al., 2014a). The authors conducted study on dual-channel speech enhancement and compared the results of APSO against PSO. For evaluation purposes they used objective measures of speech intelligibility (FAI), perceptual evaluation of speech quality (PESQ) and signal to noise ratio (SNR). The noise types they considered were babble and factory noise, for which APSO proved to be far superior to PSO in terms of improved speech signal quality and intelligibility. The main drawback of using standard PSO is that in some cases, its convergence speed becomes very low. Its search space is also fairly limited (Kunche and Reddy, 2016b). In (Yang, 2010) however, the authors provided a solution to these limitations by proposing another modified form of PSO, termed the accelerated PSO or APSO and. This was shown to have a comparatively simpler implementation and a much faster convergence speed. (Krohling, 2004) proposed a slightly modified MPSO technique, based on Gaussian probability distribution. It is termed Gaussian PSO or GPSO. In the standard PSO, a number of parameters, such as accelerating constants, inertia weight, maximum velocity and the number of particles, need to be initially defined, which the GPSO does not require. The sole variable that needs to be initially defined is the total number of swarm particles. Comparative simulation results showed the superiority of GPSO over the standard PSO for the data that was considered. To the best of our knowledge, GPSO has never been used before for speech-enhancement problems.

A learning-based particle swarm optimisation (LPSO), which is an improved stochastic optimisation algorithm, was introduced to devise an adaptive filter for dual-channel speech enhancement application by (Asl and Geravanchizadeh, 2010). The search of regions around the best solution is performed using a dynamic search method. The algorithm then involves adaptive local search on each particle. During the process, sub-swarms exchange the best solutions at regular intervals through a sub-population strategy. The simulation results prove that the proposed LPSO algorithm outperforms the standard particle swarm optimisation (SPSO), genetic algorithms (GA) and gradient-based NLMS algorithm with respect to SNR and stability.

During another attempt in 2010, a hybrid optimisation algorithm was suggested to boost the distorted speech signals in the framework of dual-channel speech enhancement(Osgouei and Geravanchizadeh, 2010). The proposed hybrid algorithm θ-SSPSO combines the conventional θ-PSO and the shuffled sub-swarms particle optimisation (SSPSO) technique to exploit the advantages of both algorithms. Experimental results reveal that the θ-SSPSO algorithm is highly effective in terms of global convergence for adaptive filters. Global convergence helped in achieving improved noise suppression in the candidate speech signal. The θ-PSO algorithm, however displays a better optimisation performance than the SPSO in the case of simple problems, but gets trapped in local optima when dealing with complex multi-objective functions. SSPSO overcomes this issue by increasing the diversity of particles in the search space thereby avoiding the local optima.

(Selvi and Suresh, 2016) employed a hybridization of spectral filtering and an optimisation algorithm for speech enhancement, by combining MMSE and PSO. Their proposed method yielded better evaluation results compared to the Bayesian non-negative matrix factorization (BNMF), and MMSE approaches.

A modified predator-prey particle swarm optimisation (MPPPSO) for noise cancellation has been recently proposed by (Fisli et al., 2018a), (Fisli and Djendi, 2018). The proposed algorithm showed good results compared to other methods, such as the predator-prey particle swarm optimisation (PPPSO), and the normalised least mean square (NLMS) algorithm. (Singh and Bansal, 2018) proposed a low-pass IIR filter design utilising a hybrid PSO-GSA optimisation algorithm. Their proposed algorithm shows improvement compared to traditional digital filters for the following performance evaluations: magnitude response, phase response, frequency response and group delay.

2. **BAT Algorithm** A population-based meta-heuristic approach called the Bat algorithm (BA), motivated by the hunting behaviour of bats, was devised (Yang, 2010). BA is rooted in the echolocation behaviour of micro-bats. The algorithm adopts frequency tuning to elevate the diversity of the solutions in the population. It also implements the automatic zooming characteristic of bats, such as the pulse emission rate and loudness on approaching the prey as the automatic adjustment capability in the algorithm. The capability attempts to balance exploration and exploitation during the search process by adapting from exploration to exploitation with the approaching of global optimality. This algorithm, being the first attempt to balance these important components, justifies itself

to be a very efficient optimisation technique when compared to other meta-heuristic algorithms (Prajna et al., 2014b), (Fisli et al., 2019), (Thaitangam et al., 2018a).

Yet another attempt using the Bat algorithm (BA) towards dual-channel speech enhancement systems was put forth in (Kunche and Reddy, 2016e). In this approach, the BA is utilised in determining of the weights for the adaptive filter. The methodology initially involves segmenting the input signals into frames. Then, the objective function is formulated as the mean square error between the distorted speech and the estimated noise signal in each frame. Then, the optimisation of the filter co-efficient is conducted through the BA. The results justify that the BA portrays an improved performance in terms of improved quality and intelligibility of the enhanced speech when compared to the SPSO algorithm.

The simulation results based on the BA were compared with those of the standard, accelerated PSO, gravitational search algorithm (GSA) and hybrid PSOGSA- based speech enhancement algorithms by (Kunche and Reddy, 2016b). The results evidently demonstrate the potential of the meta-heuristic BA over the other algorithms pertaining to the enhancement of speech signals.

An enhancement was formulated to the original BA in (Pérez et al., 2015). The improvement pertains to adopting a fuzzy system to dynamically adapt its parameters, such as wavelength, loudness, low frequency, and high frequency unlike the usual parameter tuning, which is performed based on trial and error. The results provide a comparison of the proposed modified algorithm with the original BA and genetic algorithms, depicting the effectiveness of the modification. Tests were also carried out with benchmark mathematical functions to demonstrate the potential of the proposed enhancement.

3. **GSA Algorithm**. An optimisation algorithm rooted in the law of gravity known as GSA was presented by (Prajna et al., 2014a). It is a population-based algorithm. Agents (individuals) are regarded as objects, and their performance

is estimated through masses. Objects attract each other due to the force of gravity. Objects with heavier mass have a higher gravitational force and tend to attract objects with lower mass. Hence, objects interact with each other by means of gravitational force. The objects with heavier mass are candidates for good solutions. These objects tend to move slower than the lighter ones, thereby improving exploitation. GSA achieves improved PESQ scores when compared to the SPSO algorithm. Although the SPSO finds good solutions, it suffers from the problem of the the local optimum. GSA yields better quality and intelligibility in the enhanced speech signals provided by SPSO algorithm.

A hybrid PSOGSA was presented to enhance the noise distorted speech signals in dual-channel systems by (Kunche et al., 2015). Each agent in the swarm, representing the filter coefficients is deemed as a candidate solution. PSOGSA is adopted to optimise these coefficients of the adaptive filter. The performance of PSOGSA excelled the performance of both the GSA and SPSO. The hybrid algorithm possesses the exploration and exploitation capabilities of the GSA and PSO, respectively. Therefore, PSOGSA suppresses the unwanted background noise signals of the noisy input speech signals more effectively. Table2.6 shows the highlights of the optimisation methods reviewed.

Having presented the machine-learning-based approaches towards speech enhancement, the following table shows some highlights of the optimisation methods for enhancing speech

Table 2.6: Highlights of optimisation methods for enhancing speech

	Highlights
LPSO(Asl and Geravanchizadeh, 2010)	• Higher performance when compared to the SPSO, GA, and gradient-based NLMS algorithm in terms of SNR improvement and stability.
θ-PSO(Osgouei and Geravanchizadeh, 2010)	• Combination of advantages from both algorithms, θ-PSO and SSPSO • Quite effective in achieving global convergence for adaptive filters • Better suppression of noise in the input speech signal • Increased diversity of particles in the search space to avoid getting caught in local optima. • Better than the standard PSO, θ-PSO, and SSPSO with respect to convergence rate and SNR improvement • Possibility of getting trapped in local minima while dealing with complex or multi-mode functions.
GSA(Prajna et al., 2014a)	• Improved PESQ scores when compared to the SPSO algorithm
PSOGSA(Kunche et al., 2015)	• Better than GSA and SPSO
BAT(Kunche and Reddy, 2016b)	• Improved quality and intelligibility of enhanced speech compared to PSO, SPSO, APSO,,GSA, PSOGSA
Modified BAT(Pérez et al., 2015)	• Better than BAT and GA

2.7 Summary

In this chapter, a survey of how researchers have tackled the issue of speech enhancement over the years has been presented. The earliest works done in this domain consist of the various kinds of spectral enhancement methods, statistical based algorithms, and subspace enhancement methods. These have performed well under test conditions, but in practical scenarios each comes with its own sets of drawbacks.

Adaptive noise cancellation is another popular domain in this regard. It continues to be a topic of interest for research by being customisable through the use of machine learning techniques of optimisation to tune its coefficients. Machine learning algorithms are quite vast in nature. It is not possible to cover them all within the scope of this chapter. We have discussed a few prominent ones and enlisted the strong points of each.

Advances in the field of artificial intelligence have yielded fruitful results in speech enhancement. Neural networks have proven to be a strong tool in this regard. After simple NN, came DNN, which had better results but showed poor real world generalisation upon encountering noise and speech signals that were unseen during the training phase. Then came the era of CNN, which has finally proven to be a reliable tool for generalisation of real world noise cancellation problems. It can effectively deal with noise signals of all kinds, whether seen or unseen during training phase.

Chapter 3

Research Methodology

3.1 Introduction

This chapter discusses the approaches and techniques that are used to establish the speech-enhancement system. These phases include the generation of noisy data applied to the audio speech corpora, filter selection, the meta-heuristic optimisation methods utilised to optimise the parameters of the filters, defining the objective function that shall be optimised. Finally, the evaluation criteria used to evaluate and assess the performance of the speech-enchantment system will be considered.

Figure 3.1 displays the schematic diagram of the proposed speech-enhancement framework, and the phases adopted to conduct this work. The phases of the overall framework will be discussed with further details in the following sections.

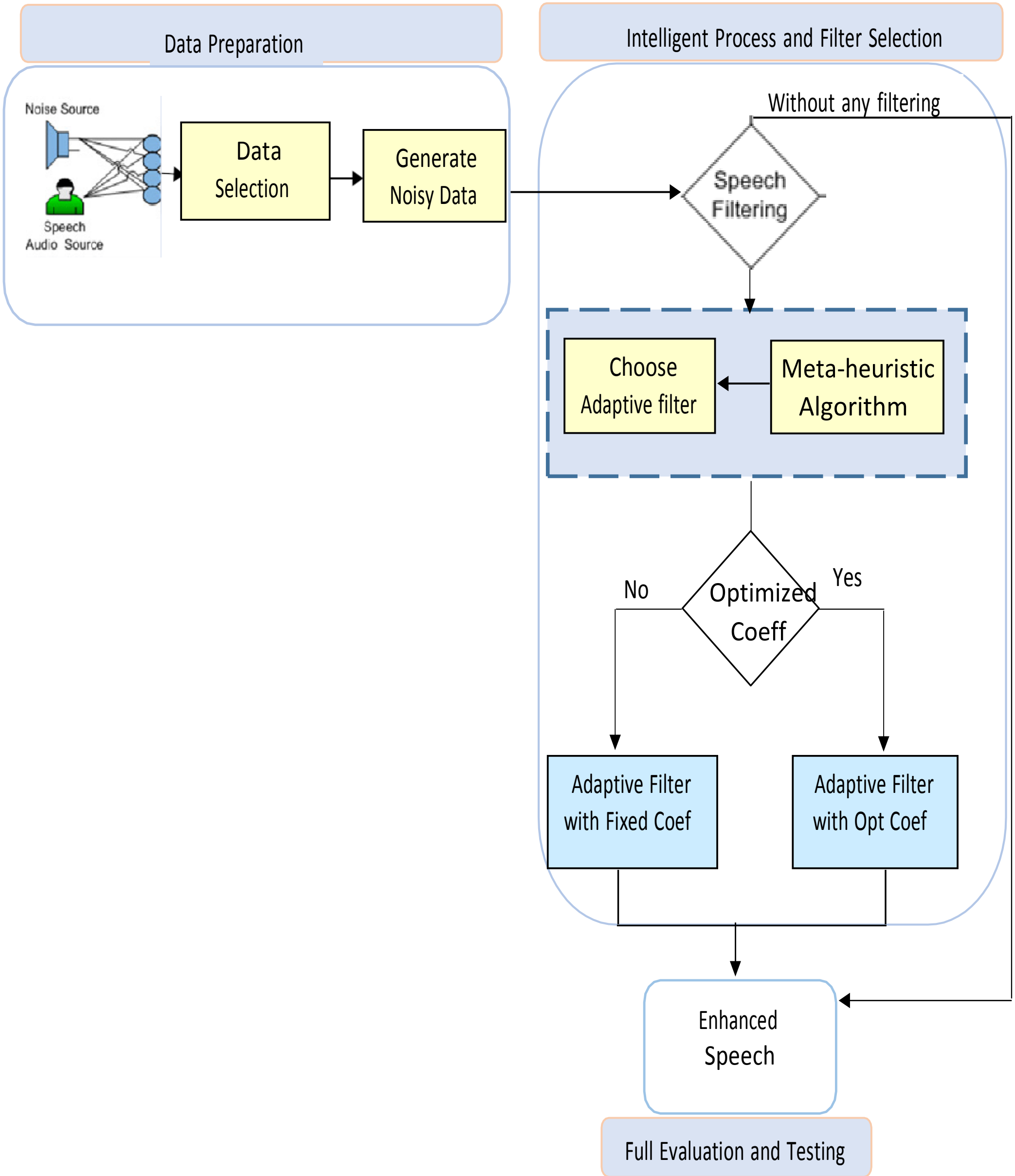


Figure 3.1: Schematic diagram of the proposed system

The rest of this chapter is structured as follows: A description of the datasets used in this research is presented in Section 3.2; this includes the NOIZEUS and the ARABIC speech corpuses. This section is then followed by how the generation of noisy data is carried out, which is discussed in Section 3.3. The phase of filters selection is summarised in Section 3.4.1. The proposed meta-heuristic optimisation speech-enhancement system and the objective function is defined in 3.4.2. An evaluation of the proposed system discussion is given in Section 3.5 followed by a summary of this chapter in Section 3.7.

3.2 Data-set description

There are several audio speech databases available to use as part of the proposed speech-enhancement system. Two different databases are selected and used for the experiments of this work; both are freely available. How these databases are recorded is provided in the next sections.

3.2.1 NOIZEUS corpus

A noisy speech corpus (NOIZEUS) was designed and developed to ease the comparison of speech-enhancement methods for researchers (Hu and Loizou, 2007). It has a total of 30 IEEE speech sentences (Rothauser, 1969), spoken by three females and three males. These sentences from the IEEE database were recorded in sound-proof booth using Tucker Davis Technologies (TDT) recording equipment. The IEEE sentences include all phonemes in American English language. The IEEE database contains phonetically balanced sentences with relatively low word-context predictability. The sentences were originally sampled at 25 kHz and down-sampled to 8kHz. Example of the sentences used are: "The birch canoe slid on the smooth planks" and " The boy was there when the sun rose ", The database is easy to download and use.

3.2.2 Arabic speech corpus

The second speech corpus used for experimenting for a proposed system was an Arabic speech corpus (Halabi, 2016): it is a modern standard Arabic (MSA) speech corpus for speech synthesis, and was recorded in South Levantine Arabic (with a Damascane accent) using a professional studio. It contains 1813 wav files containing spoken utterances.

3.3 Generation of noisy data

For this stage first we determine the type of noise to be added to the data. This type can be any one of the following: babble (crowd of people), airport, train, or factory. Then the SNR level is chosen, after which the signal is mixed with the noise at the selected SNR level after noise normalisation.

There are many types of noise as mentioned above. The babble noise is chosen from the signal processing information Base (SPIB) (Johnson and Shami, 1993; SPIB, 2013) and added to these clean signals at different SNRs for both data-sets.

Speech sentences from the NOIZEUS corpus are combined with babble noise at a variety of different SNR levels,-10dB (a loud level of noise), 0dB,+5dB, +20dB (a quiet level of noise) to construct a noisy speech. The same applies to the Arabic Speech Corpus, where the clean speech sentences are mixed with the babble noise at a variety of different SNR levels -10dB, 0dB, +5dB, +20dB to produce a noisy speech. By determining which data-set to use and the generation of noisy data, the first phase of data preparation ends. The second phase involves the intelligent process and filter selection. This phase is to be discussed in the following section.

3.4 Intelligent process and filter selection

The purpose from this phase is to produce signal with improved quality. There are many criteria for measuring the quality of a speech signal. These quality measures include: perceptual evaluation of speech quality (PESQ), log-likelihood Ratio (LLR), signal distortion level (C_sig), and scale of overall speech quality(C_ovrl). $PESQ$ returns a score value ranging from -0.5 to 4.5; the higher the value, the better the quality of the speech. The LLR is inversely related to the signal quality: the lower the value of LLR, the better signal quality. The C_ovrl is a composite measure that combines different objective measures as defined in (Hu and Loizou, 2008). There is a proportional relationship between signal quality and the C_ovrl, the higher the value of C_ovrl, the better quality of the speech signal. Finally, C_sig is also proportional to the signal quality. A higher signal quality is associated with a high value of C_{sig}.

These relationships between signal quality and the three measures of PESQ, LLR, C_sig and C_ovrl motivate us to consider a fitness function of the form:

$$C = \min \frac{1}{PESQ} + \frac{2}{C\_Ovrl} + \frac{1}{C\_Sig} + LLR \tag{3.4.1}$$

In the fitness function it is assumed that $PESQ$, C_{sig}, and LLR, have the same level of importance therefore they have the same weight in the fitness function. However, the C_ovrl is assumed to be more important, that why it has more weights in the fitness function than the measurements. This phase aims to reduce or remove the noise from the noisy signal based on the above fitness function. This is done by applying two processes (filter selection and meta-heuristic optimisation). These two processes will be detailed in the coming sections.

3.4.1 Filter selection

In this phase, the system makes choice between two filters: The Butterworth filter or the Elliptic filter. These are among the conventional ANC which remove or suppress a noisy signal, adjusting their parameters according to an optimisation algorithm. The selection of the filter parameters is a crucial process that affects the performance of the filter. Because we are not sure about the suitable choice of filter parameters, we followed the recommended values stated in the Matlab help guidelines.

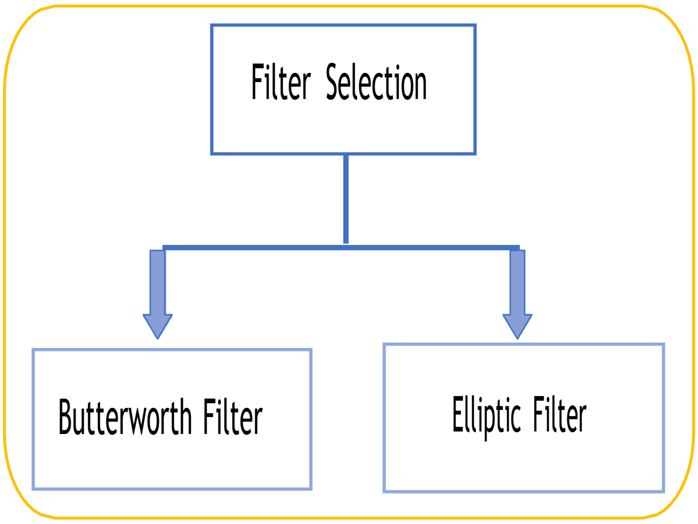


Figure 3.2: Filter selection stage of the proposed system

Hence in the case of the Butterworth filter the parameter recommended by Matlab is:

- the cut-off frequency = 0.5.

And in the case of the Elliptic filter:

- the filter order = 2.
- peak-to-peak ripple in decibels = 0.5dB.
- minimum stop band attenuation =20dB.
- passband edge frequency =0.5dB.

Hereafter, we used the term fixed filter coefficients to point to this case above. With the mentioned parameters the noisy signal is filtered. Finally, a comparison of the speech enhancement results with and without the use of optimised coefficients is carried out.

However, looking for the optimal coefficients of the Butterworth and Elliptic filters will be our main focus in the next section.

3.4.2 The meta-heuristic optimisation selection

The purpose of this process is to find the optimal filter coefficients of the adaptive filter, based on meta-heuristic optimisation algorithms.

The meta-heuristic optimization algorithms considered in this thesis include five methods. The first three methods involve the particle swarm optimisation and two of its variants. These two variants are the accelerated and Gaussian particle swarm optimisation methods. Detailed discussions are presented in Chapter 4. The fourth method is the gravitational search algorithm, which is described in more details in Chapter 5. Finally, the Bat algorithm, presented in more detail in Chapter 6.

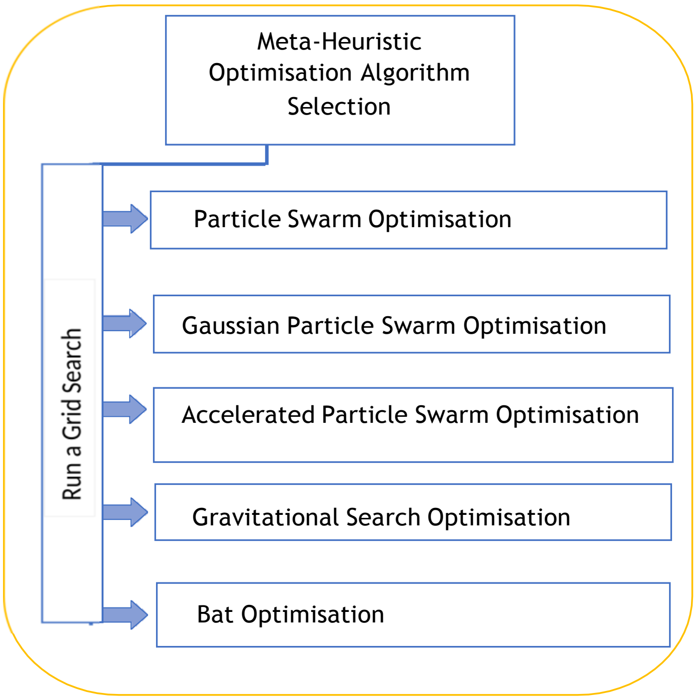


Figure 3.3: Meta-heuristic optimisation algorithms stage of the proposed system

Based on previous research in the literature, a range of values is determined for each parameter in a meta-heuristic algorithm. To select the best configuration of parameter values, a grid search method is implemented which runs through selected values for each parameter, and minimises the objective function as in Equation3.4.1.

Subsequently, the parameters of the meta heuristic algorithm are fixed, using the grid search approach. The next step is to employ it to find the best (optimised) filter coefficients. With these coefficients the noisy signal is filtered.

Finally, a comparison of the speech enhancement results with and without the use of optimised coefficients is drawn.

We have followed methods from the literature for the evaluation of our meta-heuristic system by considering both objective and subjective quality evaluation measurements.

3.5 Speech enhancement evaluation methods

In order to evaluate the performance of the speech-enhancement system proposed in this work, subjective and objective speech-quality measurements are conducted. Subjective speech quality measurements are the ideal choice, but require human volunteers to assess the quality of the speech(Hu and Loizou, 2008). Objective speech quality measurements are implemented, and they need a machine to be computed, compared to the subjective listening tests which are fast. Many different methods are available, such as PESQ (Rix et al., 2001), the log-likelihood ratio (LLR), Itakura-Saito(IS)(Hu and Loizou, 2008). Due to time limitations, for speech enhancement evaluation we only considered speech quality measures and not speech intelligibility. The coming sections describe these speech evaluation measures.

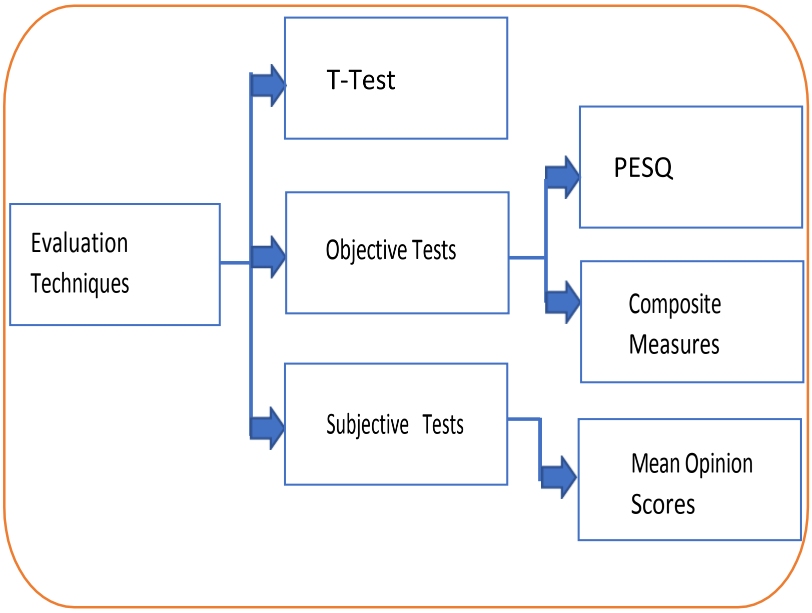


Figure 3.4: Evaluation stage of the proposed system

3.5.1 Subjective speech quality evaluation measurement

Among the natural methods for the listening tests is the subjective estimation of speech intelligibility. This subjective test is carried out by human volunteers by listening to some specific filtered speech sentences or short words and using a five-point numerical score, with 5 indicating the best and 1 indicating the worst speech quality. The listeners are requested to rate the speech utterance being tested using a five-scale category as shown in Table 3.1. The average of these scores is known as the mean opinion score (MOS). However, this MOS does not consider the distortion in the speech signal and the background noise which are introduced by speech-enhancement algorithms. In recent years, standardised approaches have been developed by the ITU-T Recommendation (Recommendation, 2001), with the aim of clearer guidance for the listeners for the evaluation of the speech materials. The ITU-T standard (P.835) addressed this problem by introducing 3 different scales to score: the speech distortion level in speech signal, the presence of back ground noise and finally overall speech quality. We will use theses scores in our evaluation.

Table 3.1: 3-Scale mean opinion scale

Rating	Speech signal quality	Background noise	Overall quality
5	Not distorted	Not noticeable	Excellent
4	Slightly distorted	Slightly noticeable	Good
3	Somewhat distorted	Noticeable but not intrusive	Fair
2	Fairly distorted	Somewhat intrusive	Poor
1	Very distorted	Very intrusive	Bad

3.5.2 Objective speech quality evaluation measurement

Although subjective listening tests are the reliable way to assess the speech quality, they are time consuming and may cause listener fatigue, another problem is finding a suitable number of listeners (Loizou, 2013a), who have no problem with hearing, and also master the language of the speech we are testing. Because of these obstacles, objective speech quality measures are useful as they only require a machine (computer) and are easy to implement compared to the subjective MOS system. The log-likelihood ratio (LLR), Itakura-Saito(IS), time-domain segmental SNR (segSNR), the Cepstrum distance (CEP), and weighted spectral slope (WSS), are listed among the objective evaluation measurements. In this work, we will be using the perceptual evaluation of speech quality (*PESQ*). PESQ is a popular and widely used objective speech measure; recommended by ITU-T recommendations P.862(Recommendation, 2001). It compares the clean signal to the degraded signal, and returns a score value ranging from -0.5 to 4.5. The higher the value, the better the quality of the speech.

Other objective quality evaluation methods are the composite measures, where different objective measures are combined. These three composite measures are defined in (Hu and Loizou, 2008). The first composite measure deals with signal distortion (C_sig), the second composite measure is noise distortion (C_bak), and the third composite measure is the overall signal quality(C_ovrl). The values of these measurements are obtained by linearly combining the existing objective measures in the following way (Hu and Loizou, 2006):

$$C_{\text{sig}} = 3.093 - 1.029 \cdot \text{LLR} + 0.603 \cdot \text{PESQ} - 0.009 \cdot \text{WSS} \quad (3.5.1)$$

$$C_{\text{bak}} = 1.634 + 0.478 \cdot \text{PESQ} - 0.007 \cdot \text{WSS} + 0.063 \cdot \text{segSNR} \quad (3.5.2)$$

$$C_{\text{ovrl}} = 1.594 + 0.805 \cdot \text{PESQ} - 0.512 \cdot \text{LLR} - 0.007 \cdot \text{WSS} \quad (3.5.3)$$

3.5.3 Statistical analysis using the t_test

To investigate whether there are any significant differences between the means of the clean speech signal, the filter with a fixed coefficient, and the filter with an optimised coefficient, the t_tests are applied to the resulting signals, at 0.05 level of significance. By letting μ_C, μ_{Fix}, μ_{Opt} be the mean of a clean signal, the application of a filter with a fixed coefficient, and a filter with an optimised coefficient respectively, the null and alternate hypothesis is tested for the case of the filter with a fixed coefficient as follows:

H_0: the means of the clean signal and the signal obtained when applying a filter to it are equal.

H_a: there is a significant difference between the means of the clean signal and a signal obtained by a filter with a fixed coefficient.

The null and the alternate hypotheses for the case of a filter with an optimised coefficient as follows:

H_0: the means of the clean signal and the signal obtained when applying a filter to it are equal.

H_a: there is a significant difference between the means of the clean signal and a signal obtained by a filter with an optimised coefficient.

3.6 Software tool

Matlab is a powerful and established high-performance language for technical computing. It involves the plotting of functions and data, graphical interactive tools, and also supplies built-in functions. The toolbox of Matlab provides specialised functionality. E.g Excel link allows data to be written in a format recognised by Excel (Houcque and Otrs, 2005)

All the experiments of this work are conducted on a computer with Intel(R) core(TM) i7-4700MQ, CPU 2.40- GHZ, RAM 8 GB, MATLAB 2017b.

3.7 Summary

This chapter discussed and presented the approaches and the methods adopted in this research. Therefore, it covered the methodology aspects that have assisted the present work. First, it gave an overview of the methodology followed in this work, followed by a description of the audio speech corpora, and generation of noisy data approach, followed by an explanation of several meta-heuristic optimisation methods, which are mainly used to tune the coefficients of the adaptive filter to reduce the noisy signal. Lastly, it is assessed how the performance and the evaluation of the research results is conducted.

Chapter 4

Speech Enhancement Based on Adaptive Noise Cancellation using Particle Swarm Optimisation

This chapter, explores the potential of different benchmark optimisation techniques, where we consider the particle swarm optimisation (PSO), and its variants in conjunction with the adaptive noise cancellation (ANC) approach, for delivering dual speech-enhancement. Hence section 4.1 introduces this chapter. Then section 4.2 presents the background and related work. section 4.3 introduces the proposed optimised speech-enhancement system. Comparative results and a discussion of the experimental set-up is presented in section 4.4. Finally, summary of this chapter is presented in Section 4.5.

4.1 Introduction

Many researchers have worked on the problem of noise cancellation over the past several decades (Mahbub et al., 2010; Aggarwal et al., 2016; Fisli et al., 2018b). Speech enhancement and noise cancellation have involved extensive applications in speech bandwidth compression, speaker verification and speech recognition (Gorriz et al., 2009),(Lin, 2003). For speech recognition and speaker identification, signal enhancement techniques improve the quality of the audio signal, which in itself is a fundamental step towards achieving correct classification.

If single channel applications are considered, spectral subtraction methods are most commonly used after noise estimation (Lin, 2003),(Lu and Loizou, 2008). In practical scenarios, however, these techniques have their own share of limitations. They can result in musical noise that might distort the signal in the process. Furthermore, such techniques are hugely dependent on properties of the noise signal, for they only work best when the additional noise is assumed to be constant or stationary. These assumptions, however, do not hold true in actual operational situations where the properties and amplitudes of additional noise signals are varying, along with external factors, such as traffic noise, factory sounds and cafeteria babble. To deal with such problems, we make use of the ANC approach. The conventional ANC comprises two channels:the first captures the reference noise signal, and the second captures the primary signal source (with noise). This enables the ANC device to sense variations in the noise amplitude quite easily. A number of different algorithms have been proposed for ANC using such a dual channel set-up, (Kunche and Reddy, 2016a).

The most commonly used methods are least mean-squares (LMS) and normalized LMS (NLMS) (Widrow and Stearns, 1985), (Gorriz et al., 2009), (Mohammed, 2007), (Bai and Yin, 2010). However, these methods are not ideal for a multi-modal error surface as they have a tendency to get stuck in local optima (Ji et al., 2008).

Stochastic optimisation algorithms have matured quite rapidly over the past few decades, and one possible application is for solving challenging noise reduction problems.

Stochastic approaches in fact, are far superior to gradient descent ones (Gentle et al., 2012). In general, there are two types of stochastic algorithms, namely, heuristics and meta-heuristics-based; heuristic means to find or to discover, whilst meta-heuristic is associated with random search algorithms (Yang, 2011a).

Popular meta-heuristic optimisation techniques include: Particle Swarm optimisation (PSO), accelerated particle swarm optimisation (APSO), and Gaussian particle swarm optimisation (GPSO). In particular, the PSO, a hugely popular optimisation technique, has been applied in a growing range of applications. The use of PSO is not restricted to a simple function optimisation, but applied in many challenging applications such as control systems and pattern classification systems (Geravanchizadeh and Asl, 2010). PSO and its variants are known for their quick convergence, robust global search and ease of implementation (Bai, 2010).

The key contribution of this chapter is to formulate an ANC system based on Butterworth, and Elliptic filters, in the form of an optimisation task.Three meta-heuristic optimisation techniques (PSO, APSO, GPSO) are used to find the optimal filters coefficients, that optimise the perceptual evaluation of speech quality (*PESQ*), signal distortion (*C_sig*), signal overall quality (*C_ovrl*), and log-likelihood ratio (*LLR*,) for the noise-free audio signal and the filtered signal. The results presented in this chapter are also published in paper by the author in (Taha et al., 2019).

4.2 Background and related Work

Swarm systems consist of nature-based computational methods (Kennedy and Eberhart, 2001) that are based on the behavior of a group of birds. Swarm systems can solve complex problems with considerable efficiency (Poli, 2008). When a group of birds solves some given problem, it is said to be due to swarm intelligence; other common examples are from colonies of social insects, such as bees, termites or ants. This section will present a review of popular meta-heuristic algorithms, namely classical PSO, and APSO, and GPSO

4.2.1 Particle swarm optimisation and its variants

PSO is an artificial intelligence technique, quite commonly used for optimisation purposes. It models the social behavior of a group of birds (a swarm) (Lee and Lee, 2013). PSO provides an appropriate and best solution for a given optimisation problem, using a population of candidate solutions (the particles are termed birds in this case). These birds then fly throughout the search space in accordance with mathematical models determining their velocity and position. One of its main advantages is that it can handle very large search spaces with little or no assumptions about the problem at hand, and does not require the problem to be differentiable. Hence it is robust enough to deal with problems that have some factors changing over time (Lee and Lee, 2013).

PSO has the ability to carry out a global search by adjusting the positions of particles (Subha and Himavathi, 2016). The position of each particle is determined by the current global best position and the personal best position.

If x_i^t and v_i^t represent the current position and velocity vector respectively for particle i, the subsequent velocity vector and the position of the particle are determined by the following equations:

$$v_i^{t+1} = wv_i^t + \alpha E_1(G_{best} - x_i^t) + \beta E_2(P_{best} - x_i^t) \tag{4.2.1}$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \tag{4.2.2}$$

where E_1 and E_2 are random numbers less than 1, α and β are the acceleration constants or the learning parameters are between $0 \leq \alpha, \beta \leq 2$, and w is the inertia weight which controls the velocity and takes a value in between [0,1].

Algorithm 4.2.1 Finding optimal solution by using PSO

```
For each particle in the population initializes positions and velocities in the search
  space
while end criteria not reached do
    for each particle i do
        Calculate velocity of the particle using Eq. 4.2.1
        Update the position of the particle using Eq. 4.2.2
        Evaluate the fitness of each particle as in Eq. 3.4.1
        if fitness is better than its pBest in the history then
            set current value as the new pBest
        end if
        if fitness is better than its gBest then
            set current value as the new gBest
        end if
    end for
end while
```

Although it has numerous advantages, PSO nevertheless has the tendency to get trapped in local minima, in some cases, converging to solutions that are far from ideal (Farooq et al., 2017).

The PSO algorithm has several parameters that are required to be appropriately set, in order to deliver a good solution. The choice of these fixed parameters is known to have a considerable effect on the quality of optimisation. Much research has been conducted to find appropriate methods which can assist in finding a suitable set of these parameters. GPSO, which is based on Gaussian distribution instead of a random distribution, enhances the convergence quality of PSO without the need for any kind of parameter adjustment, according to (Lee and Lee, 2013). Hence, the velocity equation is defined as follows (Wan et al., 2011):

$$v_i^{t+1} = v_i^t + \beta_1(G_{best} - x_i^t) + \beta_2(P_{best} - x_i^t) \tag{4.2.3}$$

where β_1 and β_2 are positive random number generated by a normal Gaussian distribution $N(0, 1)$.

The standard PSO uses both the global best and personal best position of the particles (Subha and Himavathi, 2016). The APSO algorithm is a simpler version of the PSO algorithm, which uses the global best only. Thus, in the APSO, the velocity vector is generated by the following simpler formula:

$$v_i^{t+1} = v_i^t + \alpha E + \beta(G_{best} - x_i^t) \tag{4.2.4}$$

Where the value of E is a random number between 0 and 1. α and β the learning parameters and their typical values are $0.1 \leq \alpha \leq 0.4$, $0.1 \leq \beta \leq 0.7$. The position of the particles can then be updated using equation 4.2.2. The next position of the particle is computed by combining equations 4.2.2 and 4.2.4:

$$x_i^{t+1} = (1 - \beta)x_i^t + \beta G_{best} + \alpha E \tag{4.2.5}$$

Therefore, APSO is much simpler and results in faster convergence.

In this chapter, we aim to formulate the ANC problem in the form of an optimisation task. Specifically, we optimise Butterworth and Elliptic adaptive filters for noise cancellation. Next, we outline our proposed speech enhancement system, employing ANC based on optimisation algorithms.

4.3 Proposed speech enhancement system

The purpose of this chapter is to compare the performances of PSO, APSO and GPSO for the tuning of coefficients of an adaptive filter to remove the noise from a speech signal. We are looking for the optimal set of filter parameters that optimise the perceptual evaluation of speech quality(PESQ), overall quality (C ovrl),signal distortion (C_sig) , and Log-Liklihood Ratio (LLR) for noise-free audio signal and the filtered signal.

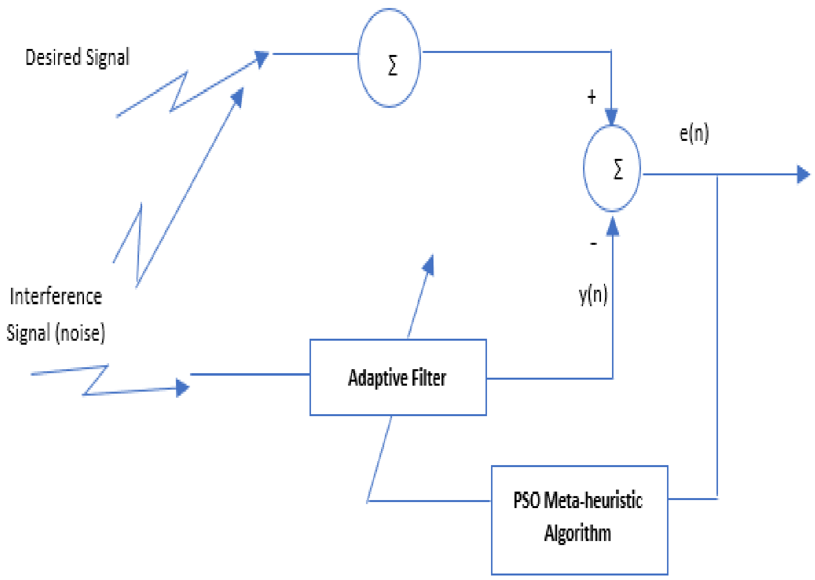


Figure 4.1: Adaptive optimised filter

Figure 4.2 explains the overall structure of the proposed speech-enhancement system. Here, the standard PSO and GPSO are utilised to obtain the optimum solution. The APSO can be obtained in the figure by ignoring the particle best (using the global best only).

4.3.1 Adaptive noise cancellation based on optimisation algorithms

Each particle in the search space is considered a possible solution representing the coefficients of the filter. The proposed optimised speech-enhancement is carried out as follows:

1. Initialize positions and velocities randomly for each particle in the search space.
2. Evaluate the fitness function for each particle using Equation 3.4.1.
3. Find the personal best, and the global best (for PSO; the global best is only for APSO).
4. Update the velocity and the position of each article using Equations 4.2.1 and 4.2.2 in the case of PSO ,4.2.4, 4.2.5 in the case of APSO, and 4.2.3 in case of GPSO.
5. Repeat steps 2-4 until the stop criteria are met (the maximum no of iteration is reached or the optimal solution is found).

4.3.2 Evaluation measurement

In order to evaluate the proposed enhancement-system, the objective PESQ measurement is used. PESQ is based on mathematical comparison of the clean and the enhanced speech signals. The composite measures described in 3.5.2 are also used along with subjective evaluation measurements where human volunteers used to listen to the filtered sentences. Previously listening tests included significantly different numbers of participants: 7 in Inai et al. (2015), 9 in Abel (2013), and 15 in Raitio et al. (2015). In this thesis, seven participants volunteer to do the test, three of them are females and four are males. A five point numerical scale is used where

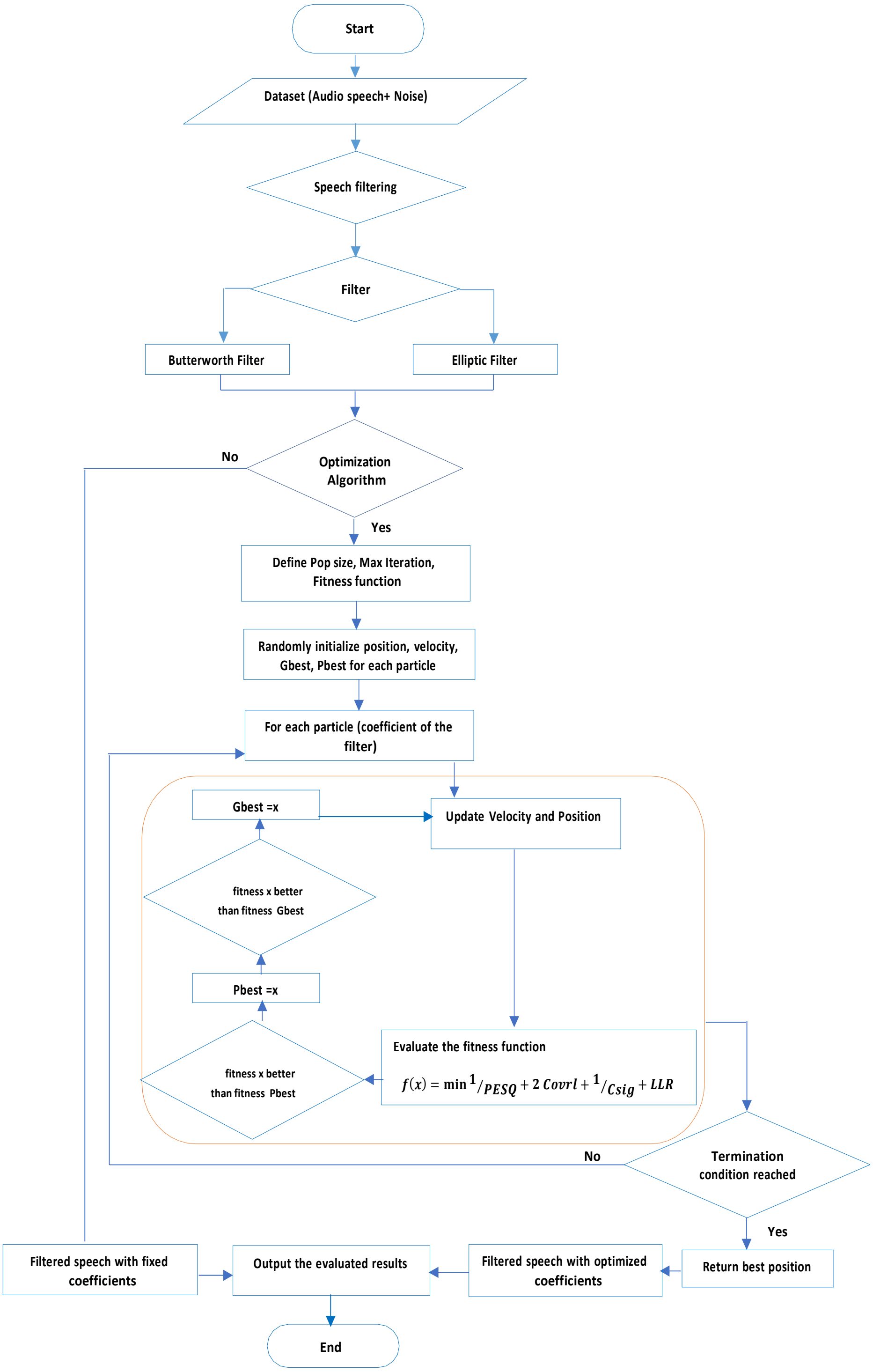


Figure 4.2: The overall structure of the proposed speech-enhancement system

five means(excellent), and one means(bad), the averaged scores are called the mean opinion score (MOS).

4.4 Results and discussion

The performance of the proposed system was examined for different SNR values (-10db, 0db, 5db, 20db), for both benchmark data-sets. Further, it was compared to that of the state-of-the-art audio only, the audio only Wiener filter (AW) (Scalart et al., 1996). Matlab implementations of the audio-only Wiener method were used from (Loizou, 2013a). To find the best values for the parameters of all of the three algorithms PSO, APSO and GPSO, we perform a grid search to configure the optimal parameters of α and β to maximize the objective function. Where the α and β were in the range $0 \leq \alpha, \beta \leq 2$ for the PSO and in the range $0.1 \leq \alpha \leq 0.4$, $0.1 \leq \beta \leq 0.7$ for APSO. The simulation conditions for all the three algorithms were as follows: the population size was set to 20, total iterations set to 50, and other parameters set as follows: $\alpha = 1.5$, $\beta = 2$ and $\alpha = 0.3$, $\beta = 0.5$ for PSO and APSO respectively. The resulting waveforms of PSO and GPSO are presented in Figure 4.3, and Figure 4.4 where an improved sound is seen to be produced when using both Butterworth and an Elliptic filters with optimised coefficients. The audio signal is corrupted by babble noise at 5 db SNR only.The spectrograms of the signals enhanced by PSO, APSO, clean speech and noise reference signals are shown Files were chosen randomly from the NOIZEUS data-set.

Table 4.1 shows the results of experiments conducted with the NOIZEUS data-set. An optimised Butterworth filter with PSO, APSO, and GPSO is applied at 20db, 5db, 0db and -10db SNRS. The averaged PESQ score were computed for all five speech-enhancement methods. The three optimised algorithms are seen to improve the PESQ score and outperform the butterworth fixed coefficient filter and the audio-only Wiener filter. Equal scores are obtained for PSO and APSO at SNRs of 20db, 5db, 0db and -10db. This trend does not remain the same for GPSO, which performs the worst at -10db among all the methods. On the other hand, the fixed coefficient filter performs better than the audio-only Wiener filter, and slightly worse than the optimised filter by PSO, APSO and GPSO.

Table 4.1: PESQ comparing filters with a fixed coefficient (Coeff), a PSO, APSO, GPSO optimised coeff, and Wiener filter (AW).For the Butterworth filter to a signal at SNRs of 20db,5db,0db and -10 db in babble noise, NOIZEUS data-set

SNR level	Fixed Coeff	PSO	APSO	GPSO	AW
20db	3.1961	3.3395	3.2789	2.9145	3.0209
5db	2.5657	2.6852	2.6852	2.7900	2.2714
0db	2.3089	2.4194	2.4194	2.1722	1.9581
-10db	1.7656	1.7890	1.7890	0.3118	1.2835

For Table 4.2 when the Elliptic filter is applied, the PSO outperforms all the other methods at all SNRs of 20db,5db, 0db, and -10db. Yet the optimised filter yields higher PESQ values compared to the audio-only Wiener filter.

We carried out experiments for the Arabic speech corpus. The results are shown in Tables 4.3 and 4.4 for different SNRs of 20db, 5db, 0db and -10db, for the case of both Butterworth and Elliptic filters. The APSO is seen to perform the best, compared to PSO and GPSO, at 0db and -10db in Table 4.3, when applying the Elliptic filter. The APSO is also seen to outperform both the PSO and APSO, at 0db and 5db.

Overall, applying optimised adaptive filter coefficients was found to enhance the results, compared to those achieved by applying a fixed adaptive coefficient filer, and state-of-the-art algorithms.

Table 4.2: PESQ comparing filters with a fixed coeff, a PSO, APSO, GPSO optimised coeff, and Wiener filter (AW). For the Elliptic filter to a signal at SNRs of 20db,5db,0db and -10 db in babble noise, NOIZEUS dataset

SNR level	Fixed Coeff	PSO	APSO	GPSO	AW
20db	3.1593	3.5096	3.5462	3.2086	3.0209
5db	2.5160	2.6015	2.5793	2.5142	2.2714
0db	2.2537	2.3144	2.2853	2.2537	1.9581
-10db	1.7018	1.8625	1.8477	1.8116	1.2835

Table 4.3: PESQ comparing filters with a fixed coeff, a PSO, APSO, GPSO optimised coeff, and Wiener filter (AW). The Butterworth filter to signal at SNRs of 20db,5db,0db and -10 db in babble noise, Arabic speech corpus.

SNR level	Fixed Coeff	PSO	APSO	GPSO	AW
20db	1.7548	1.2417	1.7554	2.1679	2.0836
5db	1.3305	1.9657	1.9697	1.2004	0.5169
0db	1.9401	2.7671	3.0092	2.0620	0.5417
-10db	1.7030	1.8513	2.0905	1.7750	0.5155

Table 4.4: PESQ comparing filters with a fixed coeff, a PSO, APSO, GPSO optimised coeff, and Wiener filter (AW). For the Elliptic filter to a signal at SNRs of 20db,5db,0db and -10 db in babble noise, Arabic speech corpus.

SNR level	Fixed Coeff	PSO	APSO	GPSO	AW
20db	1.7539	2.0834	1.7556	2.1367	2.0836
5db	1.2641	2.0975	2.3168	1.5805	0.5169
0db	1.8351	2.9864	2.9945	2.4303	0.5417
-10db	1.7030	1.8513	2.0905	1.7750	0.5155

The composite measures results which were discussed in more details in Chapter 3 utilised in this thesis as an objective measure. The results are shown in figures 4.6, 4.7 and 4.8, where C_sig is the score of the speech signal distortion, C_back is the score of the background noise intrusiveness, and C_ovrl is the score of the speech overall quality. For the speech signal distortion, it can be seen that PSO optimised by Elliptic filter scores high at all SNR values except at -10db, where the Gaussian PSO scores slightly higher than PSO, compared to the other methods and audio-only wiener filter. At 0db, the optimised filter by PSO and Ellipse shows improvement over than the other APSO and, Gaussian PSO, without filtering signal and the audio-wiener filter. However, Gaussian performs worst at both SNRs of 5db and 20 db, followed by APSO and then PSO with Butterworth filter.

The noise intrusiveness illustrated in figure 4.7 shows almost the same results.From that figure it can be seen at low SNR of -10db equal results are obtained with optimised Elliptic filter by PSO, APSO, Gaussian , without filtering signal and audio

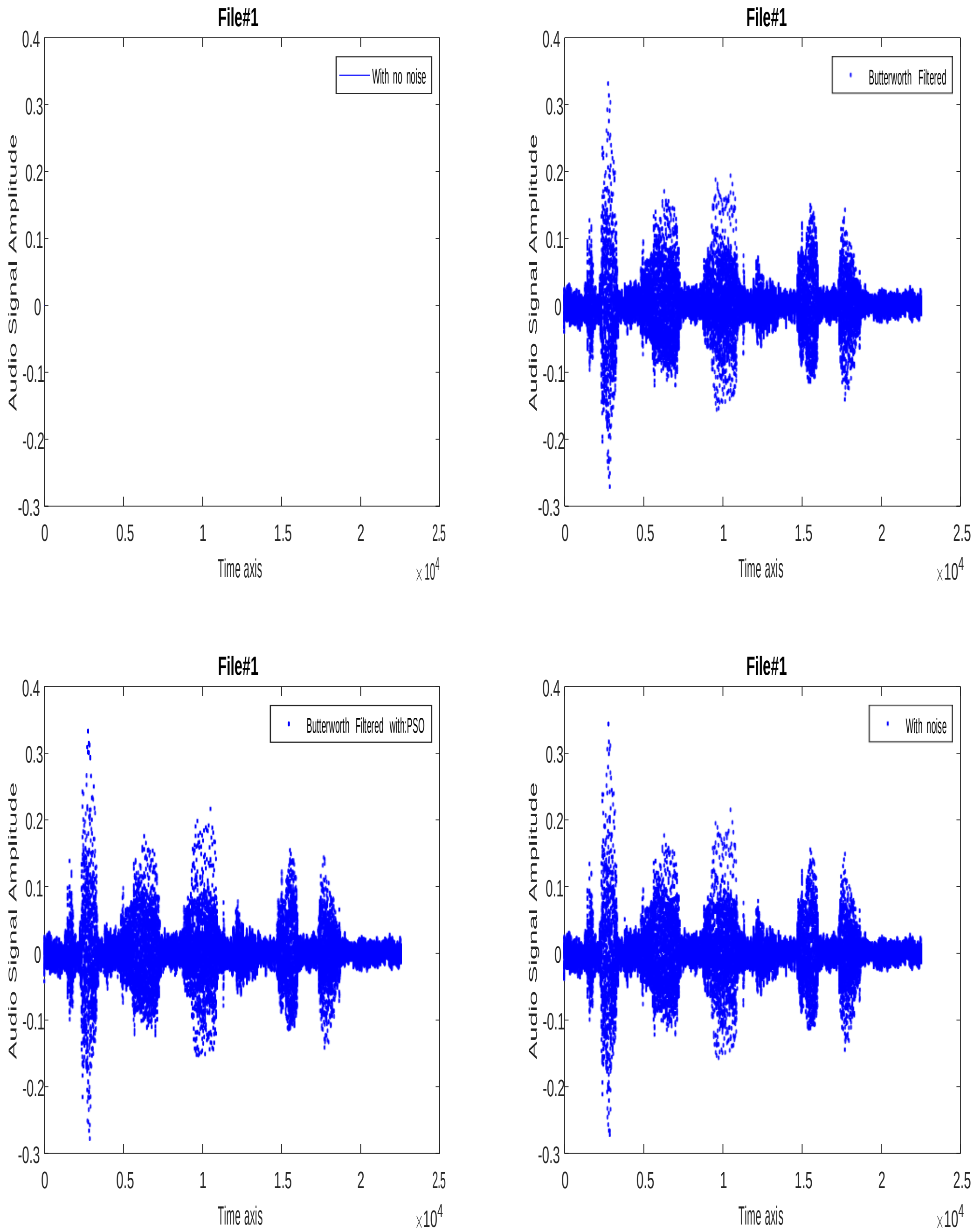


Figure 4.3: Audio signal filtered by a PSO optimised Butterworth coefficients, with babble noise at 10dB

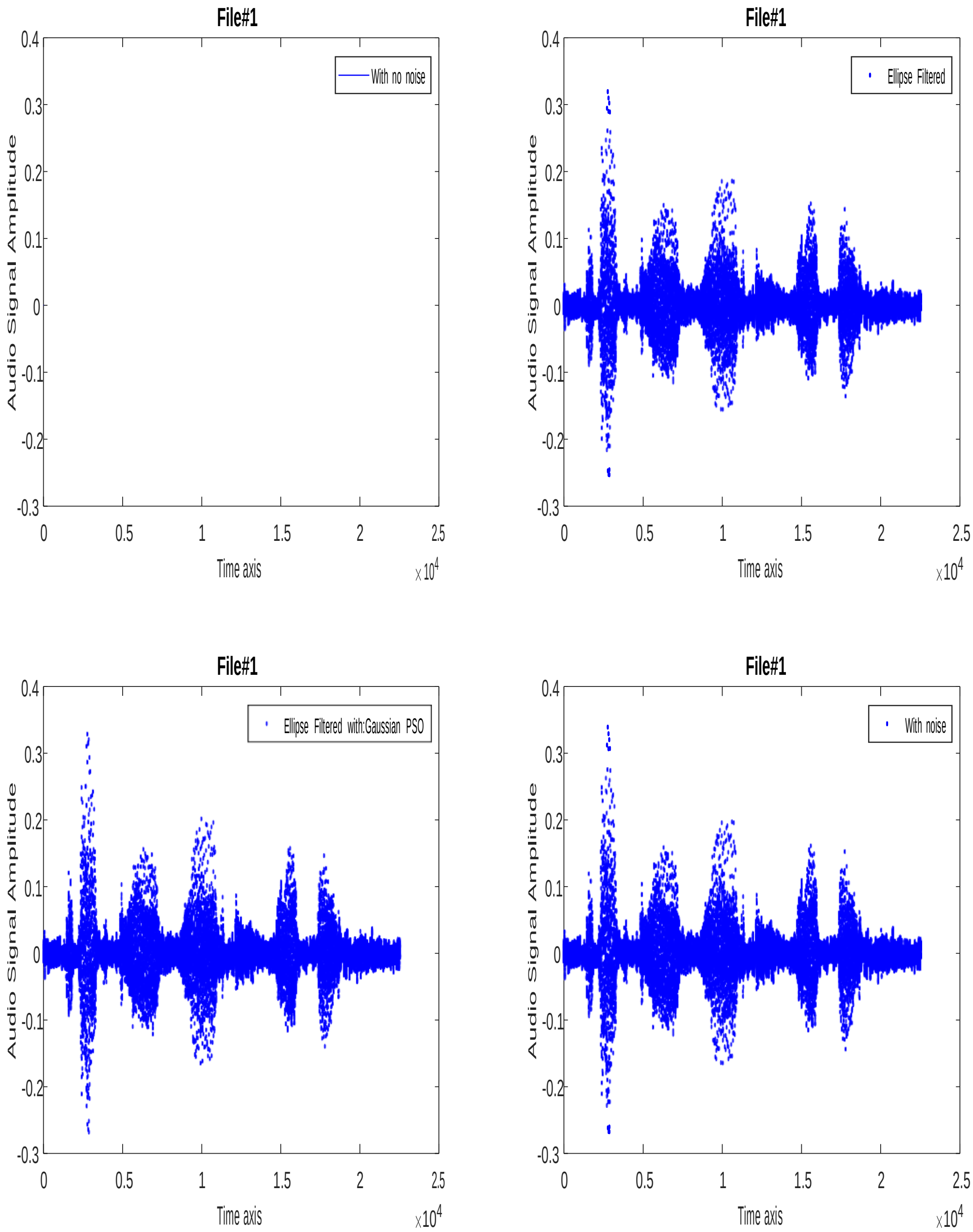
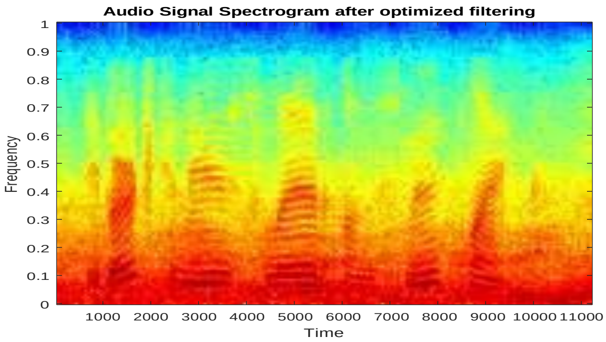
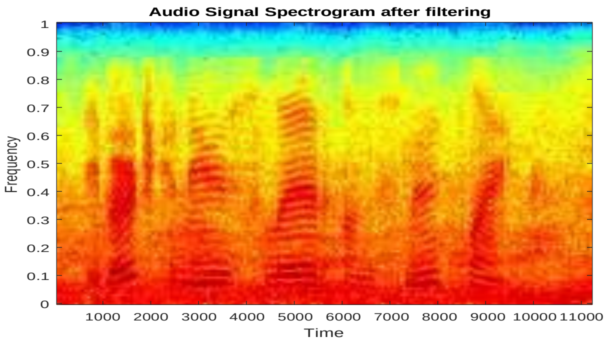


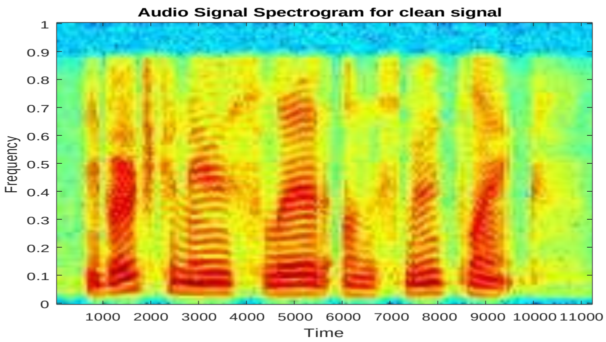
Figure 4.4: Audio signal filtered by a GPSO optimised Elliptic coefficients, with babble noise at 10dB



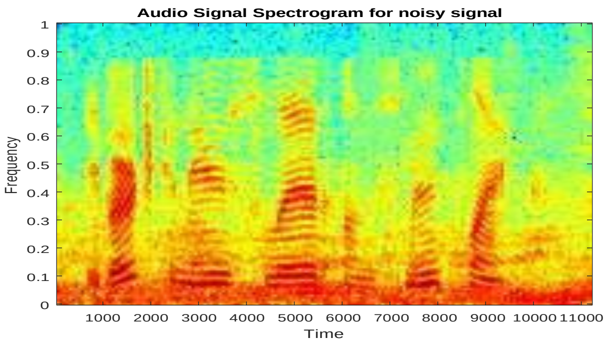
(a) Spectrogram of audio signal by an optimised PSO filter



(b) Spectrogram of audio signal by Butterworth filter



(c) Clean audio speech signal spectrogram



(d) Noise source spectrogram with babble noise at 5dB

Figure 4.5: Spectrogram of audio signal for clean, noisy and filtered signals

only Wiener algorithm, at the same SNR with optimised Butterworth filter by PSO, APSO and GPSO which performs the worst. This trend does not remain the same in the positive SNR at 0db,5db and 20db where the optimised Butterworth filter by PSO, APSO, and GPSO. The PSO and APSO outperforms the other algorithms at 20db.

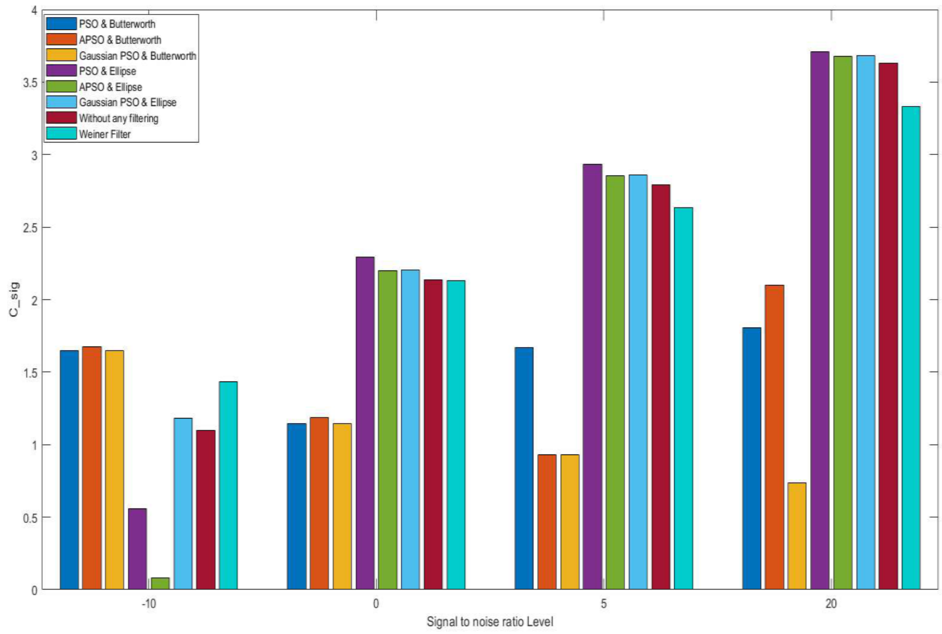


Figure 4.6: The composite objective mean score for the noise distortion for optimised filter by (PSO, APSO, GPSO, Butterworth and Elliptic), speech without filtering, Wiener filtering.

The overall scores presented in figure 4.8, are the most important scores to consider. At low SNR at -10db the optimised Elliptic filter by GPSO slightly outperforms the rest of methods (PSO, APSO, without any filtering, and audio Wiener algorithm) and outperforms the optimised filter by Butterworth. Equal scores are obtained at SNRs of 0db,5db, and 20db for the optimised Elliptic filter by PSO, APSO, GPSO and without any filtering speech. However, optimised Butterworth filter by PSO and APSO outperforms the other methods.

Subjective tests were also used in order to confirm the composite objective measures tests conducted above. Seven participants volunteered to do the listening test.

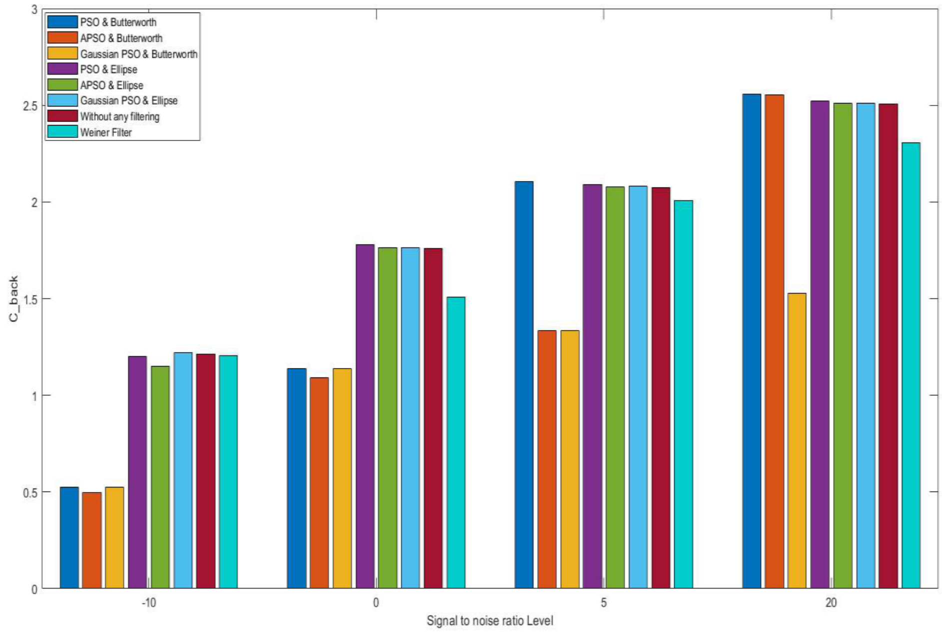


Figure 4.7: The composite objective mean score for the background noise intrusiveness for optimised filter by (PSO, APSO, GPSO, Butterworth and Elliptic) speech without filtering, Wiener filtering.

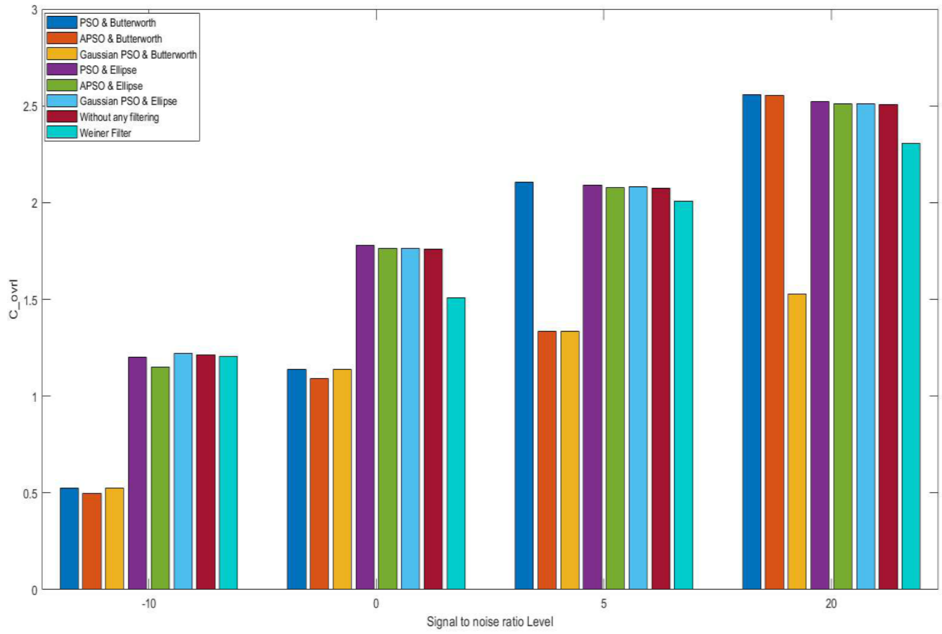


Figure 4.8: The composite objective mean score for the overall speech quality for optimised filter by (PSO, APSO, GPSO, Butterworth and Elliptic) speech without filtering, Wiener filtering.

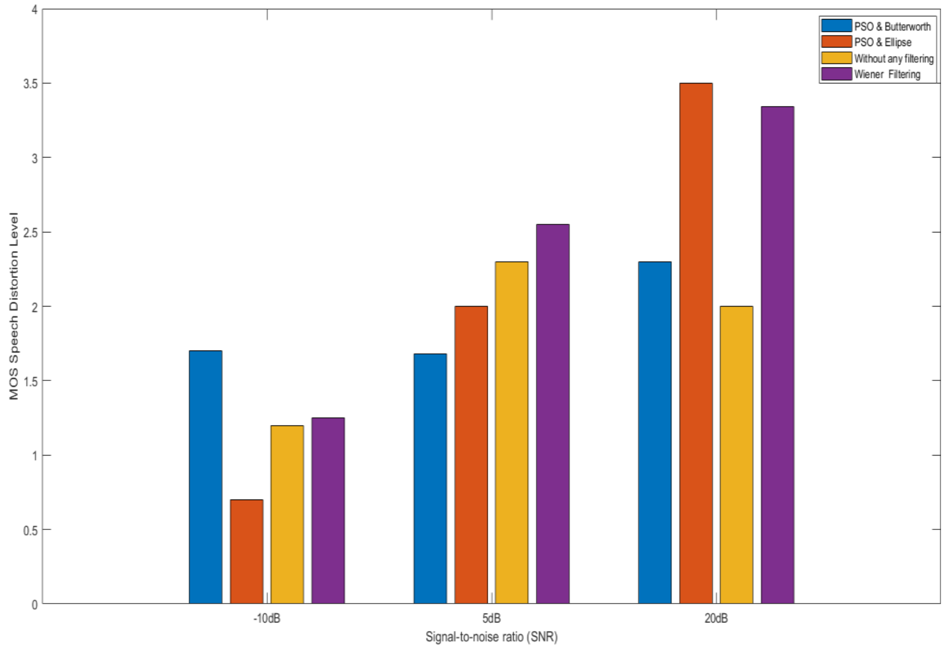


Figure 4.9: The mean opinion score for the speech signal distortion level for the optimised filter by (PSO Butterworth, and Elliptic), speech without filtering, Wiener filtering.

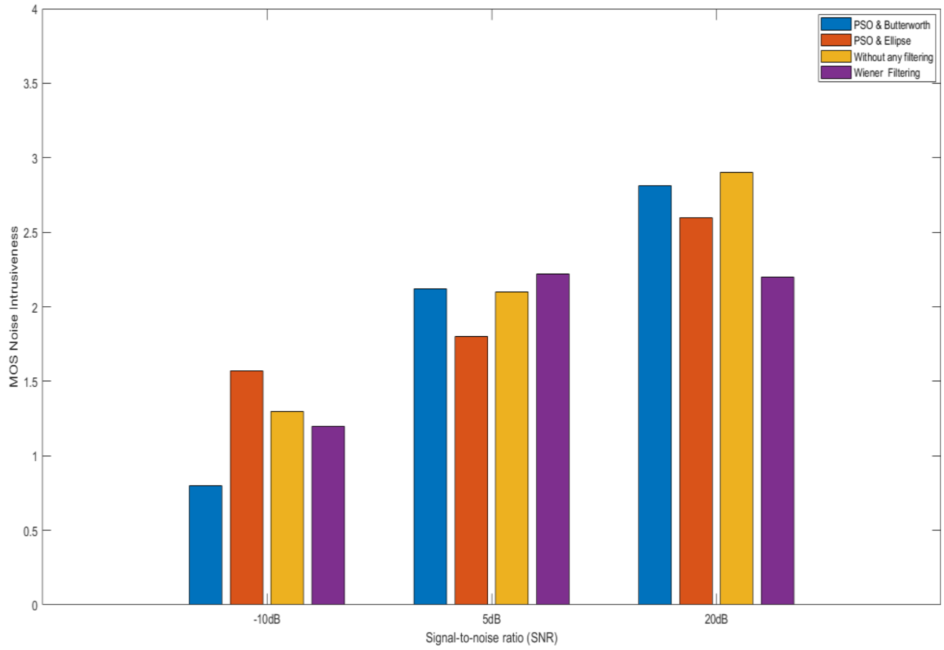


Figure 4.10: The mean opinion score for the noise intrusiveness level for optimised filter by (PSO Butterworth, and Elliptic), speech without filtering, Wiener filtering.

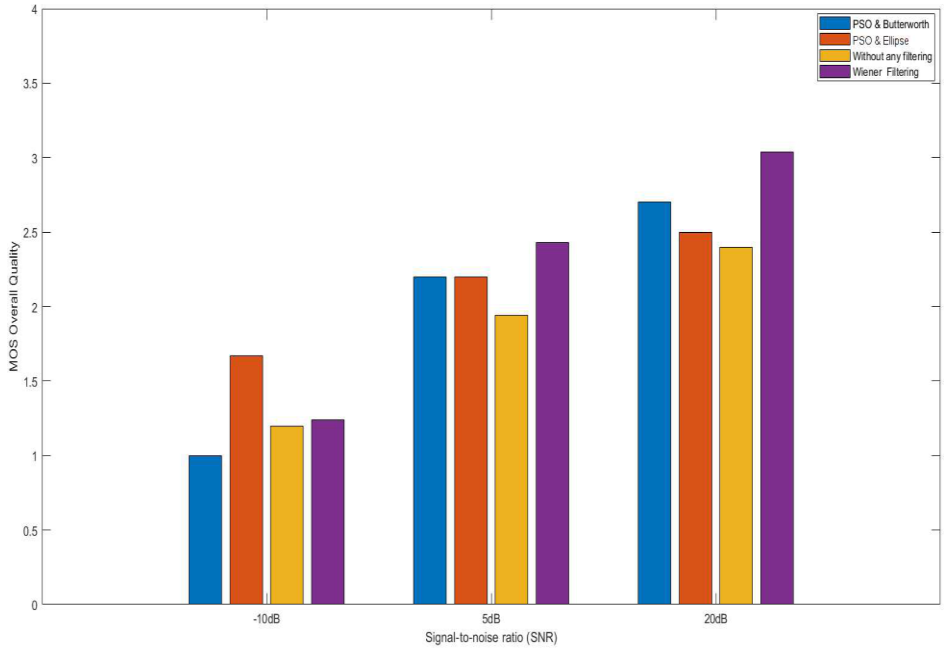


Figure 4.11: The mean opinion score for the speech overall quality for optimised filter by (PSO Butterworth, and Elliptic), Speech without filtering, Wiener filtering.

Each participant listened to the sentences at different SNRs levels(-10db, 5db, and 20db). The subjective tests took part on quit room. All participants spoke English as second language. They were three females and four males. Each participant asked to fill a form stating their age, gender, if they have abnormality in their hearing (which is a requirement). Participant were given a demo about the three sentences they will listen to .Three versions of the utterance were used ,the noisy speech sentence without any filtering, the speech sentence filtered by audio only method the Wiener Filter (Loizou, 2013a), and the speech sentence filtered by the optimised adaptive filter presented in this thesis. They asked to score between 0 and 5 using the three criteria (speech signal distortion, noise intrusiveness level, and overall speech quality). As it turned out from the composite measures scores the optimised filter by PSO outperformed the other methods, so we only apply the subjective tests on the PSO. Figures 4.9, 4.10, 4.11 illustrates the MOS results which were described in Section 3.5.1. Looking at overall speech quality first, the optimised Elliptic filter by PSO gives the best performance at -10db, and an equal performance with an optimised

Butterworth filter at 5db, while the Wiener filter outperforms the other methods at 5db and at 20db. This pattern is showed again at the speech distortion level and noise intrusiveness, where the optimised Butterworth filter and optimised Elliptic filter by PSO demonstrating large improvement at -10 db. At high SNR of 20db optimised Elliptic filter scores high, followed by the Wiener filter algorithm then the optimised Butterworth filter, this is for the speech distortion level.For noise intrusiveness levels the optimised Butterworth performs the best among the other algorithms at 20db.

4.4.1 Statistical analysis using the t_test

The performance of the proposed algorithms applied to the noisy speech signal is compared with the clean speech signal using t_test at a significant level of 0.05. That is to know, if there are any significant differences between the means of the clean speech signal, the filter with a fixed coefficient, and the filter with an optimised coefficient by PSO, APSO, and GPSO.

The null and alternate hypotheses are tested for the case of the filter with a fixed coefficient as follows:

H_0: $\mu_C = \mu_{Fix}$

H_a: μ_C != μ_{Fix}

The null and the alternate hypotheses for the case of a filter with an optimised coefficient by PSO, APSO, and GPSO are as follows:

H_0: $\mu_C = \mu_{Opt}$

H_a: μ_C != μ_{Opt}

The t_test result shown in Table 4.5 attests the significance of the optimised filters, compared to the non-optimised ones, and the noisy signal.

4.5 Summary

This chapter presented noise cancellation techniques with adaptive filter coefficients optimised using three meta-heuristic optimisation techniques, namely PSO, APSO and GPSO. The objective function is formulated such that the PESQ, Signal dis-

Table 4.5: The result of t test at the 0.05 level of significance

Dataset		Null hyp. H_0	Alternate Hyp. H_1	p_value	t_value	Decision
NOIZEUS Dataset	PSO	H_0: $\mu_C = \mu_{Opt}$	$H_a : \mu_C - \mu_{Opt} \neq 0$	0.2690	-1.1054	Accept H_0
		H_0: $\mu_C = \mu_{Fix}$	$H_a : \mu_C - \mu_{Fix} \neq 0$	0.3799	-0.8780	Accept H_0
	APSO	H_0: $\mu_C = \mu_{Opt}$	$H_a : \mu_C - \mu_{Opt} \neq 0$	0.2690	-1.1054	Accept H_0
		H_0: $\mu_C = \mu_{Fix}$	$H_a : \mu_C - \mu_{Fix} \quad 0$	0.3799	-0.8780	Accept H_0
	GPSO	H_0: $\mu_C = \mu_{Opt}$	$H_a : \mu_C - \mu_{Opt} \neq 0$	0.3242	-0.9858	Accept H_0
		H_0: $\mu_C = \mu_{Fix}$	$H_a : \mu_C - \mu_{Fix} \neq 0$	0.3799	-0.8780	Accept H_0
Arabic Corpus	PSO	H_0: $\mu_C = \mu_{Opt}$	$H_a : \mu_C - \mu_{Opt} \neq 0$	0.4982	-0.6773	Accept H_0
		H_0: $\mu_C = \mu_{Fix}$	$H_a : \mu_C - \mu_{fix} \quad 0$	0.5038	-0.6684	Accept H_0
	APSO	H_0: $\mu_C = \mu_{Opt}$	$H_a : \mu_C - \mu_{Opt} \neq 0$	0.4982	-0.6773	Accept H_0
		H_0: $\mu_C = \mu_{Fix}$	$H_a : \mu_C - \mu_{Fix} \neq 0$	0.5038	-0.6684	Accept H_0
	GPSO	H_0: $\mu_C = \mu_{Opt}$	$H_a : \mu_C - \mu_{Opt} \neq 0$	0.4955	-0.6814	Accept H_0
		H_0: $\mu_C = \mu_{Fix}$	$H_a : \mu_C - \mu_{fix} \neq 0$	0.5038	-0.6684	Accept H_0

tortion (Csig), and overall speech quality (Covrl) measures are maximized, and the log-likelihood ratio (LLR) is minimized. The algorithm searched for optimal particles over different iterations, until the optimum solution is reached or the number of iterations is exceeded.

The proposed algorithms were tested under various levels of SNR (20db,5db,0db,-10db). Benchmark NOIZEUS and Arabic data-sets were used to evaluate the proposed techniques using PESQ as a standard evaluation metric. The proposed methods were also compared with a state-of-the-art algorithm: the audio-only Wiener filter.

For the NOIZEUS data-set, and for the case of both Butterworth and Elliptic filters, the results in Tables 4.1-4.2, showed that the PSO and APSO generally performed better than GPSO at all levels of SNR. Further, the three proposed algorithms outperform the audio-only Wiener filter, except at SNR of 5db, for the case of GPSO, which performed the worst among all methods. Similarly, for the ARABIC data-set, for the case of both Butterworth and Elliptic filters, Tables 4.3 and 4.4 showed that the performance of PSO and APSO is better than the other methods in comparison with GPSO at different SNRs. However, at 5db SNR, for the case of PSO and APSO, GPSO performed the worst among all methods.

The composite measures, outlined that the optimised Elliptic filter by PSO performed better than the other methods (GPSO and APSO) and the optimised PSO Butterworth filter. For the subjective evaluation measures it seemed the optimised Butterworth by PSO at higher SNRs is generally scored the best among the other method for overall quality test, but there is no clear conclusion about which method comes second in its performance. Furthermore, a statistical analysis was carried on the means of a clean speech signal, a filter with a fixed coefficient and a filter with an optimised coefficient respectively, and on the scores collected at each SNR level. The results showed there was no statistically significant difference at ($p \leq 0.05$) amongst the enhancement methods and the clean speech.

Chapter 5

Speech Enhancement Based on Adaptive Noise Cancellation using Gravitation Search Algorithm

This chapter introduces GSA to the adaptive noise cancellation for speech-enhancement. Section 5.1 introduces this chapter and gives a motivation for using the GSA. In Section 5.2 background to the GSA and how the algorithm works are presented. Then Section 5.3 describes how the adaptive filter is modeled using the GSA optimisation method. Section5.4 then presents the results and discusses them, and finally, Section 5.5 summarises this chapter.

5.1 Introduction

Nowadays, many real-world optimisation problems are sophisticated and difficult to solve. To deal with these problems, scientists tend to use optimisation though it might not be guaranteed to provide the optimal solution. Meta-heuristic algorithms have become robust tools for optimisation problems. (Fister Jr et al., 2013) classified these meta-heuristic methods into the following two types: nature-inspired and non-nature-inspired. The nature-inspired itself is further divided into (bio-inspired algorithms and physics/chemistry based algorithms), within the known bio-inspired method are swarm-intelligence-based and evolutionary algorithms.

The Physics/chemistry algorithms, however, are inspired by the mimicking of physical/chemical laws (Rashedi et al., 2018). An example of this is simulated annealing which simulates the principle of heating and cooling metals. Another example is the gravitational search algorithm, which depends on gravity and is among the population-based meta-heuristic algorithms. The advantage of any heuristic optimisation algorithm is its ability to explore and exploit the search space efficiently. The standard PSO has the ability in exploration, but fails in the exploitation. GSA, however, is good in exploitation due to the slow movements of its heavier mass which will reach accurate solutions.

In this chapter, the authors develop an ANC system that uses GSA to optimise the filter coefficients; the ANC is based on Elliptic and Butterworth filters.

5.2 The background to GSA

GSA is in another class of optimisation techniques with a different strategy for searching (Kunche and Reddy, 2016d). The GSA is primarily rooted in the law of gravity, and the idea of mass interactions (Rashedi et al., 2009)(Jiang et al., 2014).The advantage of this technique is that it considers the distance between the neighbour agents to update the position of the currently considered agent.

In the GSA (Sabri et al., 2013), the agent constitutes four parameters, namely, position, inertial mass, active gravitational mass and passive gravitational mass. The position of the mass refers to the solution of the problem. The gravitational and inertial masses, are computed through a fitness function. The inertia mass parameter, which is used for updating the agent movement, is inversely proportional to the motion of the agent. A bigger inertia mass facilitates a slower motion of the agents in the search space. This leads to a more precise local search with increased diversity in the search space.

However, the higher the gravitational mass, the higher the attraction of the agents, leading to a faster convergence. The algorithm proceeds by adjusting these two masses, namely: the gravitational and inertia masses, for which each mass presents a solution. The masses are attracted by the heaviest mass. Hence, the heaviest mass presents an optimal solution in the search space. The masses should obey the two Newton laws, the law of gravitation, which states that each object attracts every other object in the universe with a force that is directly proportional to the product of their masses and inversely proportional to the square of the distance between their centers, and the law of motion, which states that the acceleration of any mass is equal to the force acting on the system divided by the inertia of the mass (Halliday et al., 2013). The diagram of GSA is shown in Figure 5.1.

In the Figure 5.1, *G* stands for global, *M* stands for mass, and *a* stand for acceleration. The flow chart for GSA is presented in Figure 5.1.

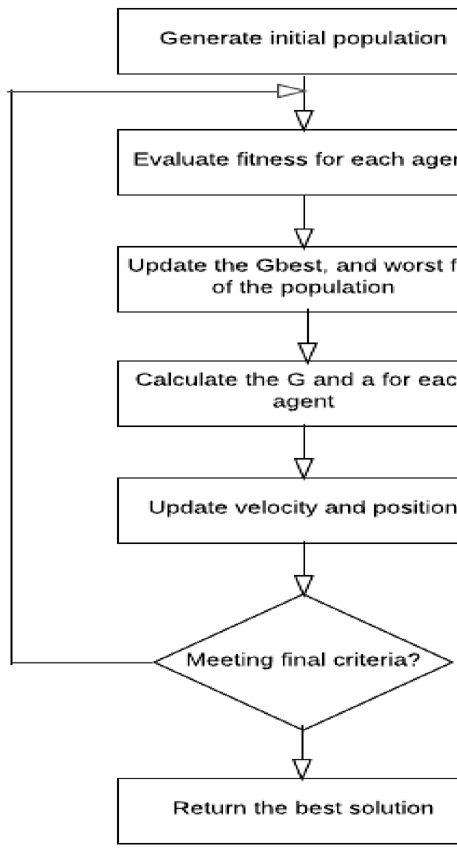


Figure 5.1: Flowchart of GSA optimisation algorithm by Sabri et al. (2013)

If GSA is a system with n agents, then GSA start randomly searching between these agents, the gravitational force is calculated as:

$$F_{ij}^{d}(t) = \frac{G(t).M_{pi}(t) \times M_{aj}(t)}{R_{ij}(t) + E}[x_j^d(t) + x_i^d(t)] \tag{5.2.1}$$

where :

- F_{ij} is the force acting on agent i from agent j at d dimension and t iteration, and is given by:

$$F_i^d = \sum_{j=1}^{N} Rand_j F_{ij}^d(t) \tag{5.2.2}$$

  where $rand_j$ is a random number in the interval $[0, 1]$.

- M_{aj} is the active gravitation mass related to agent j.
- M_{pi} is the passive gravitation mass related to agent i.
- R_{ij} is the Euclidean distance between two agents i and j at iteration t.
- $G(t)$ is the gravitational constant computed at iteration t and is calculated as:

$$G(t) = G_0 \exp \frac{\alpha t}{T} \tag{5.2.3}$$

where:

- G_0 and α are initial values decreased with time to control the search accuracy.
- t is the current iteration.
- T is the maximum no of iterations.

Then the agents acceleration at iteration t is calculated by:

$$a_i^d(t) = \frac{F_i^d(t)}{M_{ii}(t)} \tag{5.2.4}$$

Where M_{ii} is the initial mass of the agent. F_i^d is the total force acting on agent i

given in Equation 5.2.2.

After that, the next velocity and the position of the agent at the next iteration $t + 1$is computed by:

$$v_i^d(t + 1) = rand \times v_i^d(t) + a_i^d(t) \tag{5.2.5}$$

$$x_i^d(t + 1) = x_i^d(t) + v_i^d(t + 1) \tag{5.2.6}$$

The gravitational and the inertia masses, are computed for each agent at each iteration by:

$$M_{ai} = M_{pi} = M_{ii} = M_i, i = 1, 2, ..., N \tag{5.2.7}$$

$$m_i(t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)} \tag{5.2.8}$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^{N} m_j(t)} \tag{5.2.9}$$

where fit_i is the fitness of agent i at iteration t, $best(t)$ and $worst(t)$ representing the best fitness value and worst fitness value of the agent at iteration t, respectively. The values of $best(t)$ and $worst(t)$ are calculated by:

$$best(t) = \min_{j=1...N} fit_j(t) \tag{5.2.10}$$

$$worst(t) = \max_{j=1...N} fit_j(t) \tag{5.2.11}$$

this is for minimization, and given by the following for maximization:

$$best(t) = \max_{j=1...N} fit_j(t) \tag{5.2.12}$$

$$worst(t) = \min_{j=1...N} fit_j(t) \tag{5.2.13}$$

5.3 A Proposed speech enhancement system based on the GSA optimisation algorithm

As mentioned in chapter 2.5.1, it is assumed that a noisy speech signal is present in one channel, and the reference noise signal is present in the second channel. Both signals are made available to the Adaptive filter. In this chapter the Adaptive filter is modeled using the GSA optimisation method.

Each mass in the search space is considered as a possible solution representing the coefficients of the filter. The proposed GSA optimised speech enhancement system is performed by the Algorithm 5.3.1.

Algorithm 5.3.1 Finding optimal solution by using GSA algorithm

Randomly generate a group of n agents, where each agent represents coefficients of the adaptive
Evaluate the fitness function for each agent using Equation 3.4.1.
Generate randomly the initial velocity of each mass.
while the maximum no of iteration is not reached or the optimal solution is not found **do**
    Calculate the fitness of ith on time t mass by 3.4.1, then find the best mass.
    Calculate $G(t)$, $best(t)$, $worst(t)$, $M_i(t)$, $F_i(t)$, $a_i(t)$ and $v_i(t+1)$ by Equations (5.2.1) – (5.2.5)
    Update the position of each mass by Equation 5.2.6.
end while

Figure 5.2 depicts the overall structure of the proposed speech-enhancement system. Here, the GSA is used to generate the optimum solution.

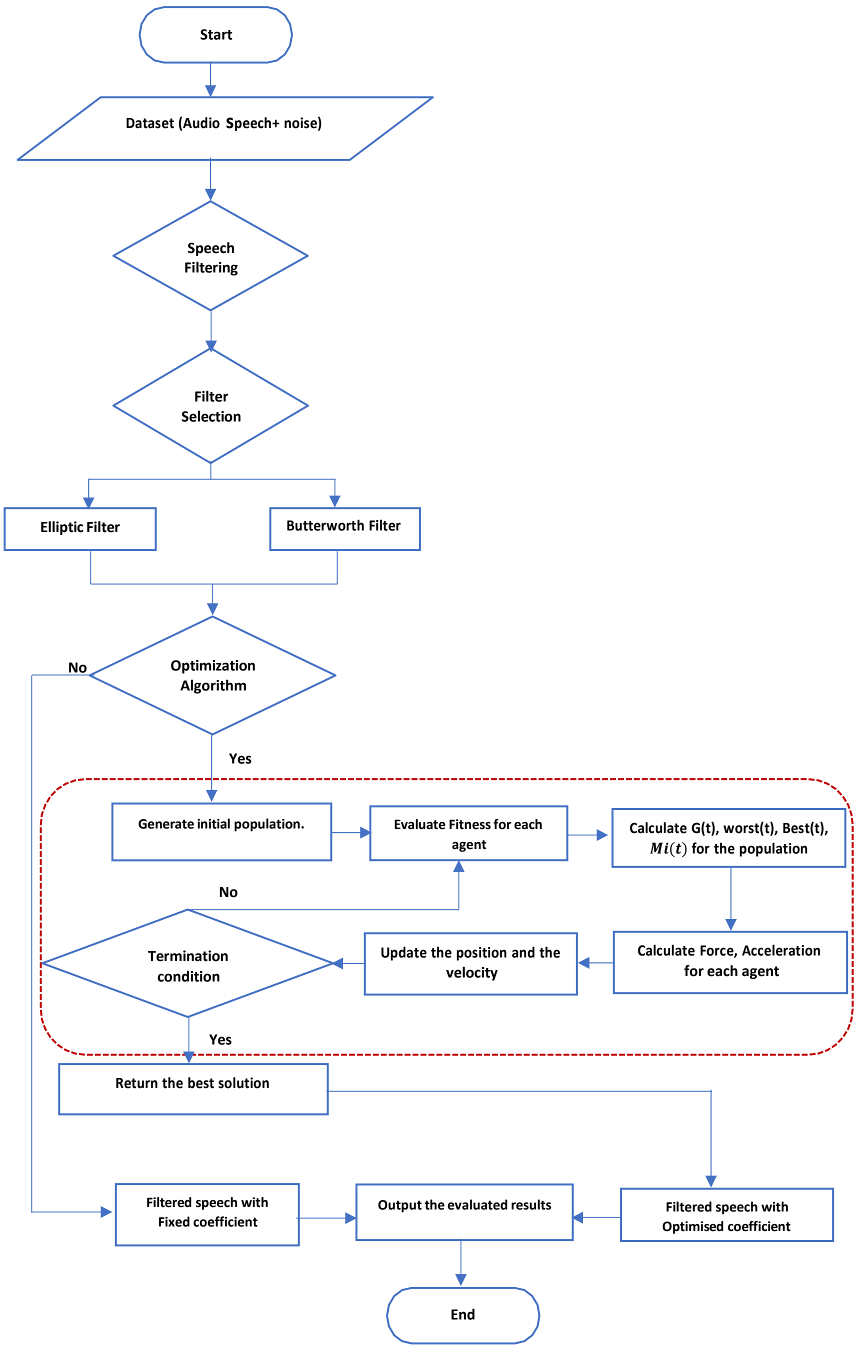


Figure 5.2: The overall structure of the proposed speech enhancement system

5.4 Results and discussion

To test the performance of the proposed system, different SNR values (-10db, 0db, 5db, and 20db) were examined for both benchmark data-sets; the two benchmarks are discussed in Chapter3. Moreover, it was compared to a state-of-the art; the audio-only Wiener filter(AW)(Scalart et al., 1996). The parameters for simulation of the algorithm tabulated in Table 5.1. The implementations of the audio-only Wiener method by Matlab from (Loizou, 2013a) are used.

Table 5.1: Simulation conditions for GSA algorithm

Algorithm	Parameter	Value
GSA	Population	30
	Iterations	50
	α	20
	G_0	100

Tables 5.2, and 5.3 shows the results of experiments for the NOIZEUS corpus, for both the Butterworth and Elliptic filters. First, the optimised Butterworth filter with GSA which was applied at 20db, 5db, 0db and -10db SNRS, The averaged PESQ score were computed for fixed filter coefficients(Fixed Coeff), an optimised filter coefficients(GSA),and AW. The GSA is seen to improve the PESQ score and outperform the audio-only Wiener speech-enhancement algorithm at SNRs of 20db,5db, 0db. Furthermore, at SNR of -10db, the fixed coefficient filter outperforms the GSA, the audio-only Wiener filter. However, the optimised filter by GSA performs better than, the audio-only Wiener filter. As for the optimised Elliptic filter in Table 5.3, the results are seen to improve the PESQ score, and outperform the Fixed Coeff, AW at SNRs of 20db,5db, 0db, and -10db.

We also carried out experiments on an Arabic speech corpus, for which the results are shown in both Table 5.4 and Table 5.5 for both the Butterworth and Elliptic filters respectively, with noises at different SNRs 20db, 5db, 0db and -10db. It is clear that at Table5.4 the optimised filter by GSA outperforms both the Fixed filter, and the audio-only Wiener Filter, at all the different SNRs 20db,5db, 0db and -10db.

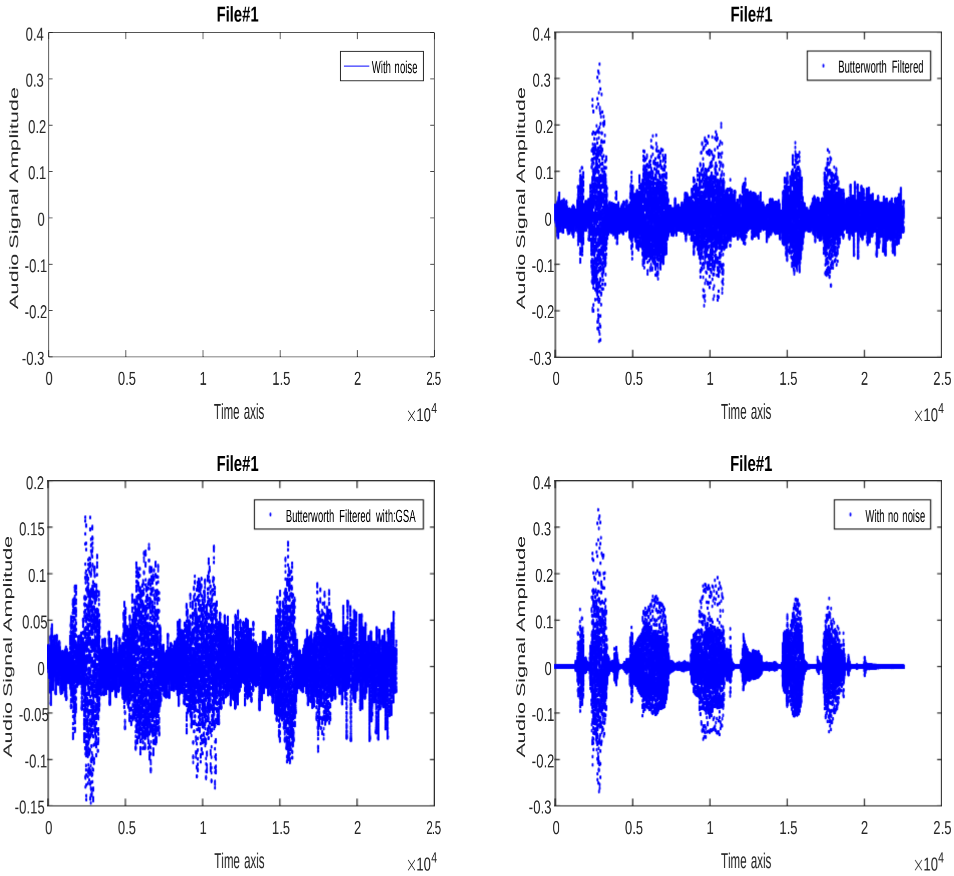


Figure 5.3: Audio signal filtered by a GSA optimised Butterworth coefficients

Table 5.2: PESQ results comparing both Butterworth filter with a fixed coefficient, optimised coeff by GSA, Wiener filter(AW), to the signal of different SNRs in babble noise, for NOIZEUS data-set

SNR level	Fixed Coeff	GSA	AW
20db	3.2084	3.3365	3.0209
5db	2.5657	2.6769	2.2714
0db	2.3098	2.4196	1.9581
-10db	1.76056	1.6380	1.2304

Table 5.3: PESQ results comparing both Elliptic filters with a fixed coefficient, optimised coeff by GSA, and Wiener filter(AW), to signal of different SNRs levels in babble noise, for NOIZEUS data-set

SNR level	Fixed Coeff	GSA	AW
20db	3.2178	3.4722	3.0209
5db	2.5160	2.5132	2.2714
0db	2.2587	2.3007	1.9581
-10db	1.7018	1.7274	1.2651

Interestingly enough at SNR of -10db the optimised filter by GSA also outperformed the fixed coeff filter and the AW. This trend remains the same for the Elliptic filter in Table 5.5, in which the optimised filter by GSA performed well compared to the Fixed coeff filter and the audio-only Wiener filter, at all SNRs: 20db,5db, 0db, and -10db. The fixed coefficient filter performs better than the audio-only Wiener filter, and slightly worse than the optimised filter.

Overall, applying optimised adaptive filter coefficients was found to enhance the results, compared to those achieved by applying a fixed adaptive coefficient filer, and state-of-the-art algorithms.

The means of the composite objective quality measures, which are the speech distortion, background intrusiveness and overall speech quality, with different SNRs levels at -10db,0db,5db and 20db, are provided in Figures 5.4, 5.5, and 5.6 respectively. As it can be seen in Figure 5.4 for the negative SNR at -10db, the optimised Butterworth filter by GSA scores higher results than the optimised Elliptic GSA , without any filtering signal and the Wiener filter. However, the optimised Elliptic filter by GSA always outperforms the optimised Butterworth filter by GSA, without any filtering and the Wiener filter algorithm at 0db, 5db and 20db. As for the background intrusiveness score it is shown in Figure 5.5, the optimised Elliptic filter by GSA and without any filtering signal scores, are very similar at all SNRs. The optimised Butterworth with GSA performed the worse with negative SNR at -10db,and at 0db. However, it performs the best compared to the other methods at high SNR value at 20db. The composite overall scores as can be seen in 5.6 are similar in that

Table 5.4: PESQ results comparing Butterworth filter with a fixed coefficient, optimised coeff by GSA, and Wiener filter(AW), to the signal of different SNRs levels in babble noise, for Arabic speech corpus

SNR level	Fixed Coeff	GSA	AW
20db	2.1396	2.1436	2.0836
5db	1.4497	1.4783	1.1630
0db	1.3515	1.5026	0.9276
-10db	1.2358	1.7747	0.1515

at 0db,5db and 20db the optimised Elliptic filter by GSA performs the best compared to the rest of the methods. The optimised Butterworth by GSA performs the worst among the other methods.

Table 5.5: PESQ results comparing both Elliptic filter with a fixed coefficient, optimised coeff by GSA, and Wiener filter(AW), to the signal of different SNRs levels in babble noise, for Arabic speech corpus

SNR level	Fixed Coeff	GSA	AW
20db	2.1396	2.1436	2.0836
5db	1.3737	1.4783	0.6647
0db	1.2715	1.4739	0.0630
-10db	1.1777	1.3410	0.1515

The composite measures are used to evaluate the noisy sentences.Three versions of sentences were compared, the sentence processed by the optimised filter by both Butterworth and Elliptic proposed in this work, the signal processed by Wiener filter (Loizou, 2013a), and the noisy signal without any filtering. Figures 5.4, 5.5, and 5.6 show the results of composite measures for the speech signal distortion (Csig), noise intrusiveness (Cback), and overall score (Covrl) respectively.

For the speech signal distortion scores which are in Figure 5.4, it can be seen at negative SNR value the optimised Butterworth filter by GSA outperforms the Wiener filter algorithm and the optimised Elliptic filter. However, at positive SNRs values 0db, 5db and 20db the optimised Butterworth filter gives the lowest scores,

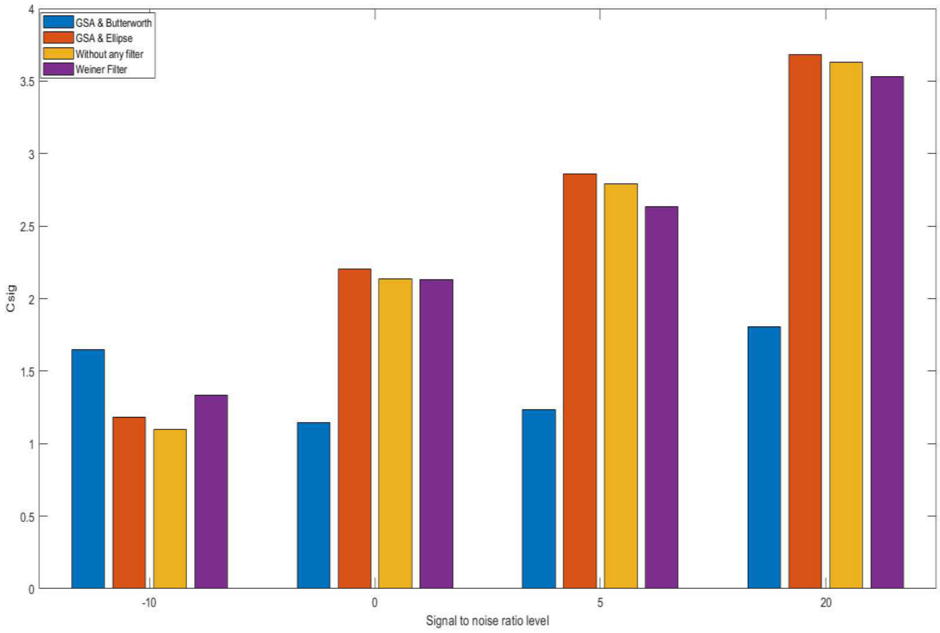


Figure 5.4: The composite objective mean score for the speech signal distortion for GSA-optimised filter (GSA ,Butterworth and Elliptic), speech without filtering, Wiener filtering.

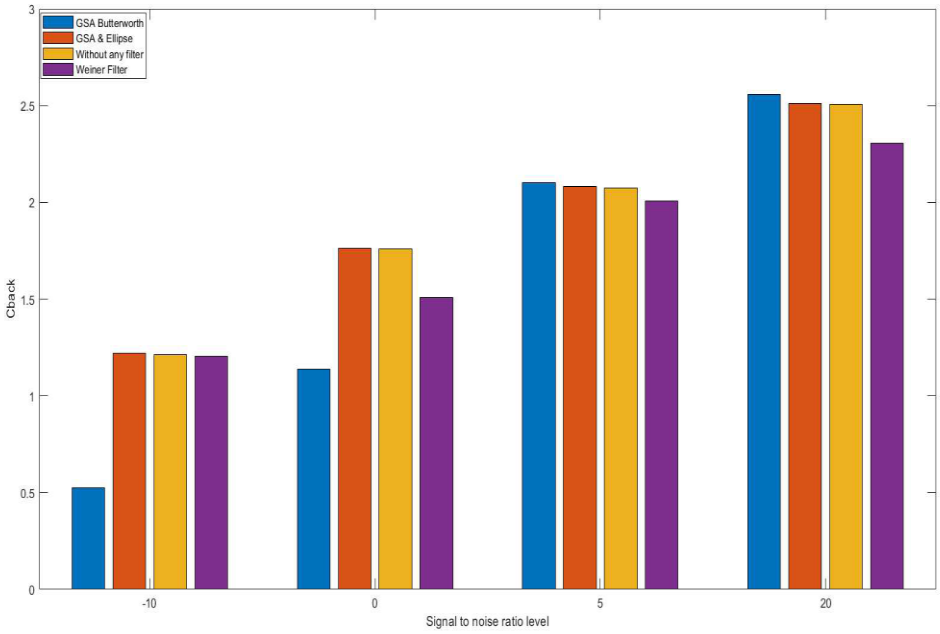


Figure 5.5: The composite objective mean score for the background noise intrusiveness for GSA-optimised filter (GSA, Butterworth, and Elliptic), speech without filtering, Wiener filtering.

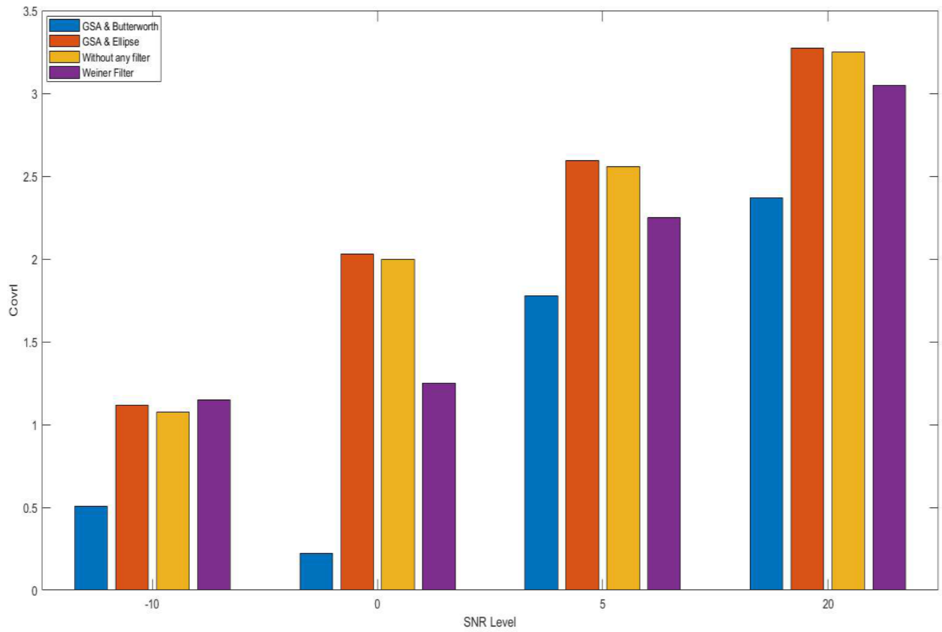


Figure 5.6: The composite objective mean score for the overall speech quality for GSA-optimised filter (GSA, Butterworth, and Elliptic), speech without filtering, Wiener filtering.

while the optimised Elliptic filter by GSA performs higher than the other methods. For the Background noise intrusiveness scores showed in Figure 5.5, the optimised Butterworth filter outperforms the other methods at high SNR value of 20db. An equal score can be seen for the optimised Elliptic filter and without filtering signal at all SNR values. For the overall scores in Figure 5.6, the optimised Elliptic filter scores the best at all SNR levels, except at negative value of -10db, where the Wiener filter algorithm slightly outperforms the optimised Elliptic filter.

Although objective tests were provided, we will nevertheless supplement it with subjective tests for further justification. By using subjective listening tests, volunteer participants were employed. Seven participants were used. Four were male and three were female. Each participant listened to the sentences at different SNRs values(-10db,5db, 20db). For a comparison purpose, three versions of the utterance were used: firstly, the noisy speech sentence without any filtering, secondly, the speech sentence filtered by the audio only method, the Wiener Filter (Loizou, 2013a), and

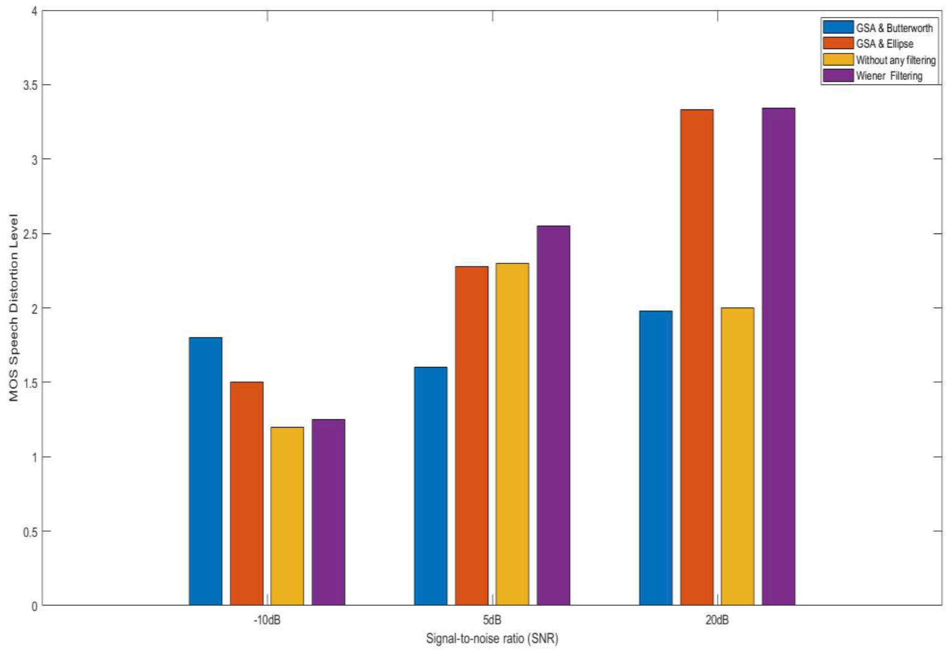


Figure 5.7: The mean opinion score for the speech distortion level for GSA-optimised filter (GSA, Butterworth, and Elliptic), Speech without filtering, and Wiener filtering.

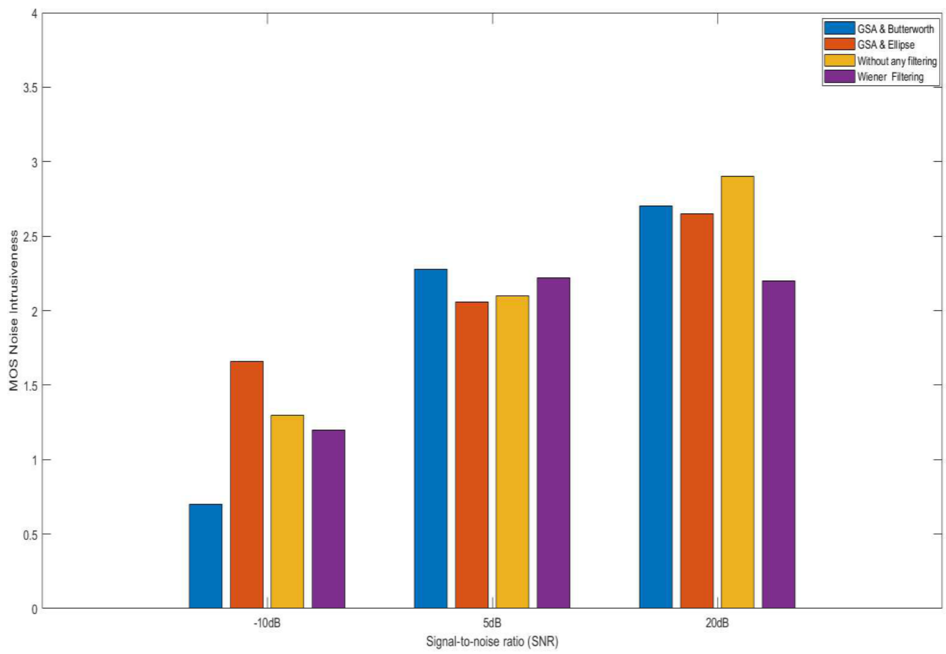


Figure 5.8: The mean opinion score for the noise intrusiveness level for GSA-optimised filter (GSA, Butterworth, and Elliptic), Speech without filtering, and Wiener filtering.

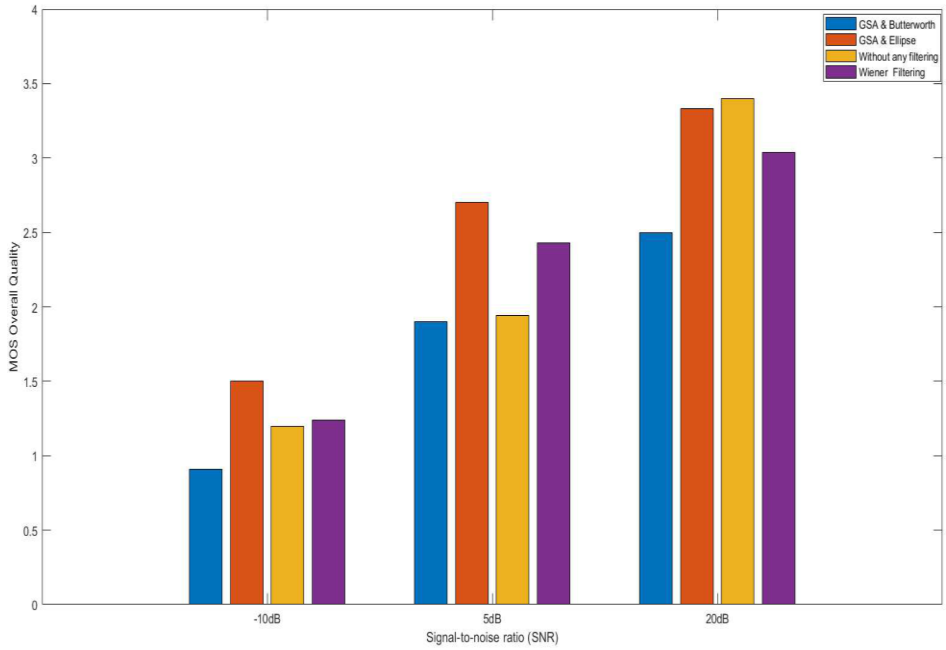


Figure 5.9: The Mean Opinion Score for overall speech quality for GSA-optimised filter (GSA, Butterworth, and Elliptic), Speech without filtering,and Wiener filtering.

thirdly, the speech sentence filtered by the optimised adaptive filter introduced in this chapter. Figures 5.7, 5.8, and 5.9 show the MOS results for the three different techniques mentioned in 3.5.1 ,the speech distortion level, the noise intrusiveness level, and the overall speech quality. As it can be seen in the speech distortion level, the optimised Butterworth by GSA outperforms the other methods at negative SNR of -10db, followed by the optimised Elliptic filter. At both SNRs of 5db and 20db the Wiener filter scores the highest, and also has an equal score with the optimised Elliptic at 20db. As for the background intrusiveness level it could be seen that the listeners were not able to listen to the sentences at low SNR of -10db, as they gave low scores for the optimised Butterworth filter by GSA. However, at low SNR of -10 the optimised Elliptic filter produces high scores as the listeners were able to identify the sentences. At the high SNR values of 5db, and 20db all of the method scores were close and comparable. When it comes to overall scores the optimised Elliptic filter showed improved results at all the SNRs levels. The results obtained by the MOS, were closed to the results found by the objective composite measures.

5.4.1 Statistical analysis using the t_test

We applied the ttest to the computed results, for both the NOISUS and ARABIC data-sets. The t test was performed using a 0.5 level of significance. That was to know, if there were any significant differences between the means of the clean speech signal, the filter with a fixed coefficient, and the filter with an optimised coefficient GSA. The null and alternate hypotheses were tested for the case of the filter with a fixed coefficient as follows:

H_0: $\mu_C = \mu_{Fix}$

H_a: $\mu_C \; != \mu_{Fix}$

The null and the alternate hypotheses for the case of a filter with an optimised coefficient by GSA were as follows:

H_0: $\mu_C = \mu_{Opt}$

H_a: $\mu_C \; != \mu_{Opt}$

Table 5.6: The results from the t_test at the 0.05 level of significance

Dataset		Null hyp. H_0	Alternate Hyp. H_1	p_value	t_value	Decision
NOIZEUS Dataset	GSA	H_0: $\mu_C = \mu_{Opt}$	$H_a : \mu_C - \mu_{Opt} \; != 0$	0.4546	-0.7467	Accept H_0
		H_0: $\mu_C = \mu_{Fix}$	$H_a : \mu_C - \mu_{Fix} \; != 0$	0.3033	-1.0239	Accept H_0
Arabic Corpus	GSA	H_0: $\mu_C = \mu_{Opt}$	$H_a : \mu_C - \mu_{Opt} \; != 0$	0.9986	0.0016	Accept H_0
		H_0: $\mu_C = \mu_{Fix}$	$H_a : \mu_C - \mu_{fix} \; != 0$	0.9999	-9.7121	Accept H_0

The t_test result shown above in Table 5.6 attests to the significance of the optimised filters by GSA. As it can be seen in this table the optimised filter by GSA generated better results compared to those produced by the non-optimised ones, and the noisy signal.

5.5 Summary

This chapter began by giving the background to the GSA and explaining where and how it had been inspired and how it currently operates. Then the meta-heuristic GSA algorithm was proposed as a means of solving the problem of adaptive noise cancellation in dual channel speech enhancement and in Section5.3, the overall structure of the proposed GSA speech enhancement system was described. The results of this comprehensive evaluation of the performance of this speech enhancement system were then presented in Section 5.4; where the objective and subjective tests used for a comprehensive evaluation of the performance of the proposed speech enhancement system in noisy environments at different SNRs values(-10db,0db,5db,and 20db), using babble noise. The results were then compared to ”without filtering” speech sentences and with a state-of-the-art algorithm: the audio-only Wiener Filter. The PESQ objective evaluation score results showed that, the optimised filter by GSA performs better than the Fixed filter at all SNRs levels, and that the Fixed filter performs better than the audio only Wiener filter. Furthermore, the composite objective measures revealed that the optimised Elliptic filter performed better than the Butterworth filter, and slightly better than the ”without filtering”. However, the audio-only Wiener filter scores higher than the optimised Butterworth filter in general, but giving lesser performance than the other methods. As for the subjective evaluation measures it seems that the optimised Elliptic generally scores the best among other methods, but there is still no clear conclusion about which method comes second in its performance.

Chapter 6

Speech Enhancement Based on Adaptive Noise Cancellation using Bat Algorithm

The Bat algorithm (BA) is a population-based algorithm that was recently developed(Yang, 2011b). BA is inspired by the hunting behaviour of bats. This chapter proposed the use of the BA in the adaptive noise cancellation for speech enhancement. Hence Section 6.1 introduces this chapter. The following section gives a brief background to the Bat algorithm and the way it works. Then, Section 6.3 presents how the BA method used to model the proposed speech-enhancement system. The results and discussion of the experiments are given in Section 6.4 and, finally Section6.5 sets out the summary of this chapter.

6.1 Introduction

The Bat algorithm was recently introduced by Yang, and it is based on the ability of bats to use echolocation to distinguish between prey, and background barriers, furthermore recognizing distance(Yang, 2011b).

The idea behind the Bat algorithm is to control the movement direction and speed by adjusting the frequency of each Bat by changing its location. The way that the

BA controls the local search, is by controlling the loudness and the emission rate. BA expounds on other meta-heuristics by having a way to automatically balance between the exploration and the exploitation of the search space.

In this chapter, we will formulate an ANC system based on Butterworth, and Elliptic filters, in the form of an optimisation task. The Bat algorithm is used to find the optimal filter coefficients. This chapter presents a novel dual speech-enhancement system based on the Bat algorithm. The results of the experiments are discussed with speech sentences from the NOIZUS database mixed with babble noise at a variety of different SNRs levels. Also, in this chapter, the performance of the proposed system is evaluated by using an objective and subjective listening test, and the results are discussed. Also it was compared to that of the state-of-the-art audio only.

The rest of this chapter is divided into the following sections: A review and a background of the Bat algorithm and flow chart of the Bat algorithm is discussed in Section 6.2. The proposed optimised speech-enhancement system and the overall structure of this system is presented in Section 6.3. The results and a discussion of the experimental set-up along with an evaluation of the performance is summarised in Section 6.4, followed by summary of this chapter in Section 6.5.

6.2 The background to Bat

The Bat algorithm is a population-based meta-heuristic approach, put forth by (Yang, 2010). The algorithm is inspired by the hunting behaviour of bats. It is rooted in the concept of the echolocation behaviour of micro bats. During the search for prey, the technique of pulse emission rate and loudness revealed by bats is mimicked in the Bat algorithm (Kunche and Reddy, 2016c). BA incorporates frequency tuning to elevate the diversity of the solution in the population, but at the same time, it adopts the automatic zooming concept and attempts to maintain a balance between exploration and exploitation during the search process. The auto zooming ability in micro bats is manifested as the automatic adjustment from exploration to exploitation upon approaching of the global optimality. The Bat algorithms is considered as one of the first type of algorithms that balance these two key components in the search process. As a result, it proves to be a very efficient optimisation technique compared to all other meta-heuristic algorithms. This algorithm demonstrates effective solutions for a variety of problems. The binary version of the algorithm has been very successful in image processing and classification. The flow diagram of the Bat algorithm is given in Figure 6.1.

The use of the Bat algorithm for enhancing speech was proposed in (Kunche and Reddy, 2016e) , who conducted a study on dual-channel speech enhancement and compared the results to APSO, GSA, and PSOGSA a hybrid algorithm. Their proposed algorithm outperforms the other algorithms in terms of improved speech signal quality and intelligibility.

(Thaitangam et al., 2018b) utilised an adaptive filter optimised by the Bat algorithm. The proposed algorithm is compared to LMS and RLS, and to an adaptive filter optimised by PSO. Their results showed that their algorithm had less time complexity, stability, and better SNR than these algorithms.

Figure 6.1 shows the flow chart of the Bat algorithm. At the beginning, the position x_i and the velocity v_i have to be initialised in the d-dimensional search space. The initial pulse frequency is updated by the following equation:

$$f_i = f_{min} + (F_{min} - f_{max})\gamma \tag{6.2.1}$$

Where $\gamma \in [0, 1]$ is a random vector following the uniform distribution; here for our problem we put $f_{min} = 0, f_{max} = 2$.

At each step, we update the velocity by:

$$v_i^t = v_i^{t-1} + (x_i^t - x_*)f_i \tag{6.2.2}$$

x_* denotes the current best global solution(location).

At each step, the new solution is x_i^t

$$x_i^t = x_i^{t-1} + v_i^t \tag{6.2.3}$$

The loudness A_i and the rate r_i of pulse emission are updated accordingly as the iterations proceed. For the local search part, using random walk, a new solution is generated randomly:

$$x_{new} = x_{old} + EA^t \tag{6.2.4}$$

$E \in [-1, 1]$ is a random number, A_i^t is the loudness of bats at time step t. According to the natural behaviour of bats, A_i decreases if the bat has found its prey, while the pulse rate r_i increases.

$$A_i^{t+1} = \alpha A_i \tag{6.2.5}$$

$$r_i^{t+1} = r_i^0[1 - \exp(-\gamma t)] \tag{6.2.6}$$

where $\alpha \in (0, 1)$ and $\gamma > 1$, are defined according to the problem, if $t \to \infty$ then

$$A_i^t \to 0, \ and \ r_i^t \to r_i^0 \tag{6.2.7}$$

$r_i^0 \in [0, 1]$ can take any value according to Equation 6.2.6. Having presented the characteristics of the BA, the steps involved in this work are presented subsequently.

6.3 Proposed speech enhancement system based on Bat optimisation algorithm

In this section, a novel optimised speech-enhancement system that employs ANC is proposed. The aim of this research is to utilise Butterworth and Elliptic adaptive filters, and use of Bat for the tuning of coefficients of an adaptive filter to remove the noise from a speech signal.

Figure 6.2 shows the overall structure of the proposed speech enhancement system. Here, the BA is utilized to obtain the optimum solution.

The position of each bat in the search space is considered as a possible solution, which represents the coefficients of the filter. The proposed optimised speech enhancement is represented in Algorithm 6.3.1.

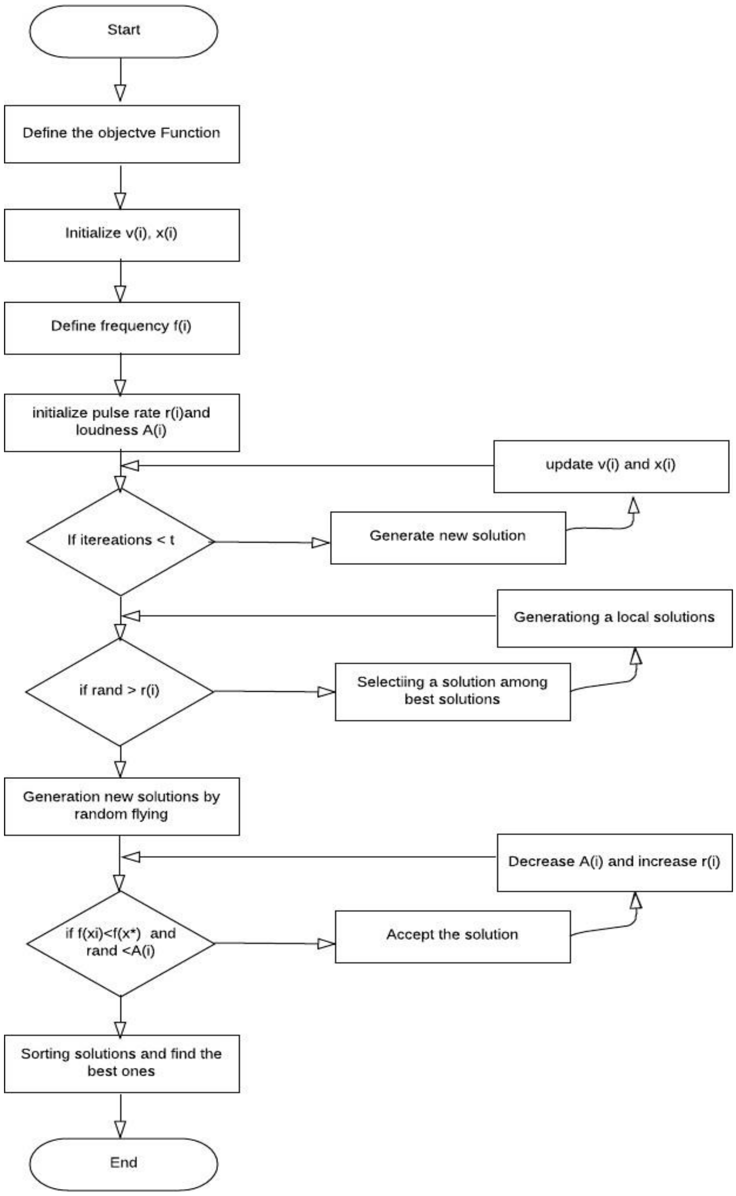


Figure 6.1: Flowchart of Bat optimisation algorithm(Hafezi et al., 2015)

6.4 Results and discussion

To see how the proposed speech-enhancement system performs, the system was examined for different SNR values at (-10db, 0db, 5db, and 20db), for both two data-sets. Further, a state-of-the art audio only, which is the Winer filter (AW) (Scalart et al., 1996), was compared with the BA. The experimental conditions for simulation are tabulated in Table6.1. The implementations of the audio only Wiener method by

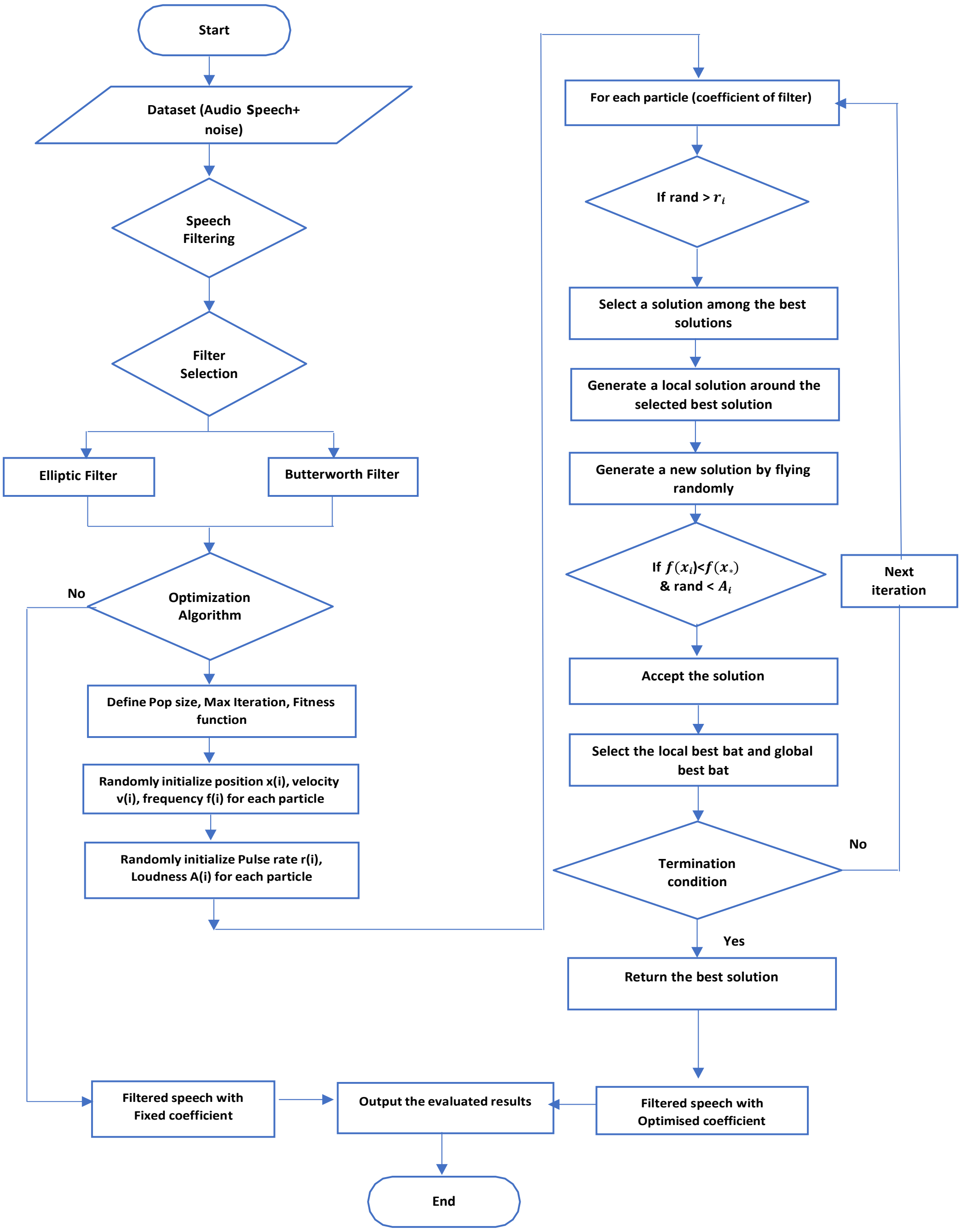


Figure 6.2: The overall structure of the proposed speech-enhancement system

---

Algorithm 6.3.1 Finding an optimal solution by using the Bat algorithm

Initialize positions x_i and velocities v_i randomly for each particle in the search space.
Initialize Frequency f_i, Pulse rate r_i and Loudness A_i.
while t < Max number of iterations **do**
  Generate new solutions by adjusting frequency,
  Update velocities and locations/solutions using equations 6.2.1 to 6.2.3
  if rand > r_i **then**
    Select a solution among the best solutions
    Generate a local solution around the selected best solution
  end if
  Generate a new solution by flying randomly
  if if rand < A_i & $f(x_i)<f(x_*)$ **then**
    Accept the new solutions
  end if
  Rank the bats and find the current best x_*
end while

Matlab from (Loizou, 2013a) are used in this chapter.

Table 6.1: Simulation conditions for the Bat algorithm

Algorithm	Parameter	Value
Bat	Population	30
	Iterations	50
	r_0	0.5
	A_0	0.5

Table 6.2 displays the results of the experiments for the NOIZEUS corpus. An optimised Butterworth filter with BA is applied at different SNRs 20db, 5db, 0db and -10db. The averaged PESQ scores were computed for the Fixed Coeff, an optimised filter coefficient (BA), and AW. The BA is seen to improve the PESQ score and outperform the audio-only Wiener filter. On the other hand, the fixed coefficient filter performs better than the Audio-only Wiener filter, and slightly worse than the optimised filter by the BA.

For Table 6.3, when the Elliptic filter is applied, the optimised filter by the BA outperforms all the other methods at all SNRs of 20db,5db, 0db, and -10db. Yet the optimised filter yields higher PESQ values compared to the audio-only Wiener filter.

Table 6.2: PESQ results comparing the filter with a fixed coefficient, a BA optimised coeff, and Wiener filter(AW), for the Butterworth filter to the signal of SNRs at 20db, 5db, 0db and -10 db in babble noise, NOIZEUS data-set

SNR level	Fixed Coeff	BA	AW
20db	3.5434	3.5631	3.0209
5db	2.4521	2.6631	2.2714
0db	2.2325	2.3347	1.9581
-10db	1.6701	1.7276	1.2304

Table 6.3: PESQ results comparing the filter with a fixed coefficient, BA optimised coeff, and Wiener filter(AW), for the Elliptic filter to the signal of SNRs at 20db, 5db, 0db and -10 db in Babble noise, NOIZEUS data-set

SNR level	Fixed Coeff	BA	AW
20db	3.2163	3.3905	3.0209
5db	2.5160	2.6454	2.2714
0db	2.2439	2.5308	1.9581
-10db	1.7332	1.8826	1.2304

The experiments are also conducted for the Arabic speech corpus, for different SNRs of 20db, 5db, 0db and -10db, and both the Butterworth and Elliptic filters. The results are shown in Tables 6.4 and 6.5. The optimised BA filter is seen to perform best, compared to the fixed coefficient filter, and audio-only Wiener filter algorithm, at 20db,0db, 5db and -10db. The optimised algorithm and the fixed filter coeff performs equally at 20db and almost equally at 5db with the Elliptic filter.

Composite measures results that are based on Equations 3.5.13.5.23.5.3, with babble noise at SNRs of -10db,0db,5db and 20db are shown in Figures 6.3, 6.4, and 6.5. Where *C sig* is the score of speech signal distortion, *C back* is the score of background noise intrusiveness, and *C_ovrl* is the score of overall speech quality. As it can be seen, the optimised Elliptic filter scores high compared to the 'without filtering' and Wiener filter in positive SNR values. For negative SNR values at -10db the optimized Butterworth filter outperforms the other methods for speech signal distortion. The optimised Butterworth filter is slightly better than the other

Table 6.4: PESQ results comparing the filter with a fixed coefficient, BA optimised coeff, and Wiener filter(AW), for the Butterworth filter to the signal of SNRs at 20db, 5db, 0db and -10 db in babble noise, for Arabic speech corpus

SNR level	Fixed Coeff	BA	AW
20db	2.1396	2.1452	2.0836
5db	1.6389	1.7297	0.5170
0db	3.1802	2.9521	0.4568
-10db	2.4537	2.7926	0.5155

Table 6.5: PESQ results comparing the filter with a fixed coefficient, BA optimised coeff, and Wiener filter(AW), for the Elliptic filter to the signal of SNRs at 20db, 5db, 0db and -10 db in babble noise, for Arabic speech corpus

SNR level	Fixed Coeff	BA	AW
20db	2.1358	2.1358	2.0836
5db	2.8507	2.8549	0.5170
0db	2.7068	3.4529	0.4568
-10db	2.6339	3.2869	0.5155

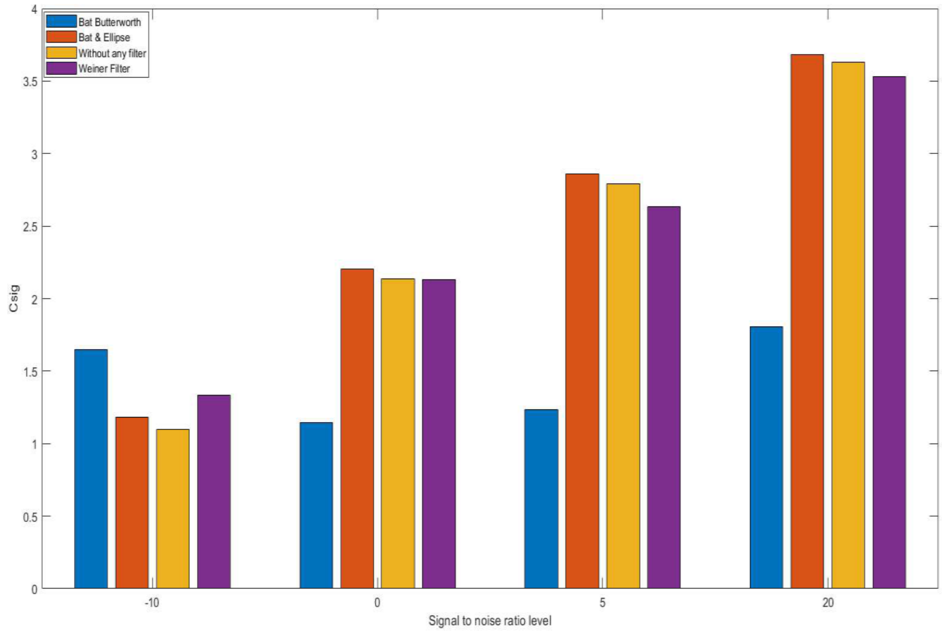


Figure 6.3: The composite objective mean score for speech signal distortion for the BA-optimised filter (BA ,Butterworth and Elliptic), speech without filtering, Wiener filtering.

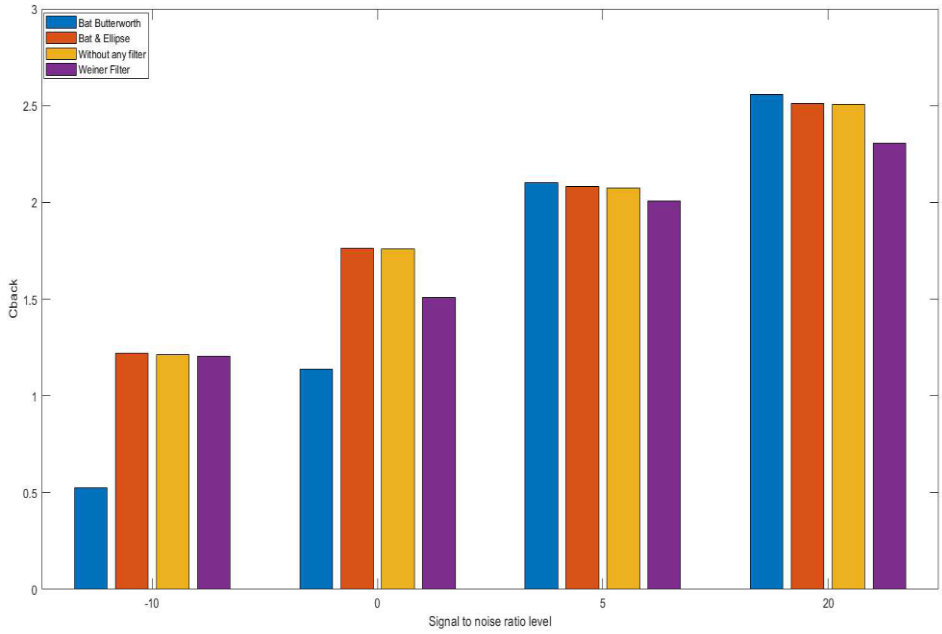


Figure 6.4: The composite objective mean score for background noise intrusiveness for the BA-optimised filter (BA, Butterworth and Elliptic), Speech without filtering, Wiener filtering.

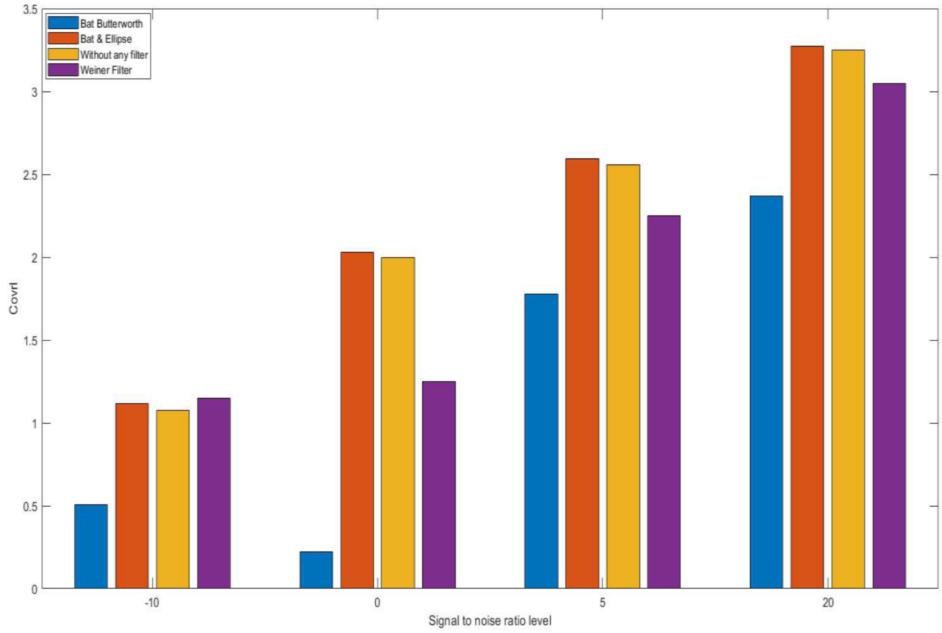


Figure 6.5: The composite objective mean score for overall speech quality for the BA-optimised filter (BA, Butterworth and Elliptic), Speech without filtering, Wiener filtering.

methods for positive SNRs at 5db and 20db, when it comes to noise intrusiveness, but it is not the case at low SNRs of 0db and -10db, where it performs the worst. The overall scores, which are presented in Figure 4.8, indicate that, the optimised Elliptic filter shows improvement compared to without filtering, and the Wiener filter in the positive SNRs. However the optimised Butterworth shows the worst scores at all SNRs levels.

To support the objective speech evaluation approaches conducted above, we also carried out subjective speech evaluation measures, where seven participants volunteered to take the test and listen to the sentences.Babble noise was added at different SNRs levels (-10db, 5db and 20db). For comparison purposes, three versions of sentences were used: the noisy speech sentence without any filtering, the speech sentence filtered by audio-only method, the Wiener filter (Loizou, 2013a), and, the speech sentence filtered by the optimised adaptive filter introduced in this chapter. Participants were asked to score between 0 and 5, using the three criteria (speech signal distortion,

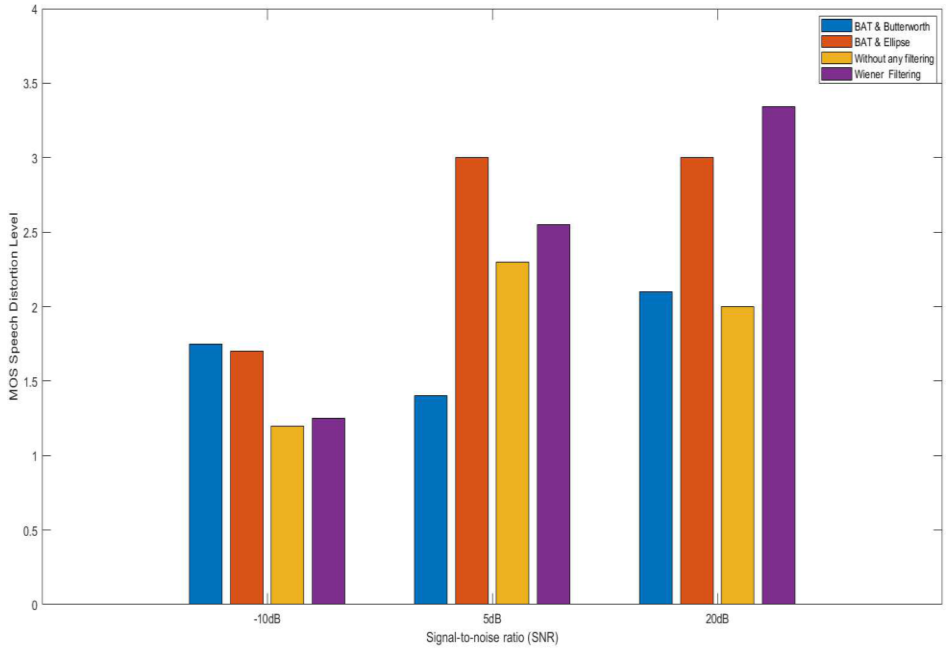


Figure 6.6: The mean opinion score for speech distortion level for the noise distortion for BA-optimised filter (Bat, Butterworth and Elliptic), speech without filtering, Wiener filtering.

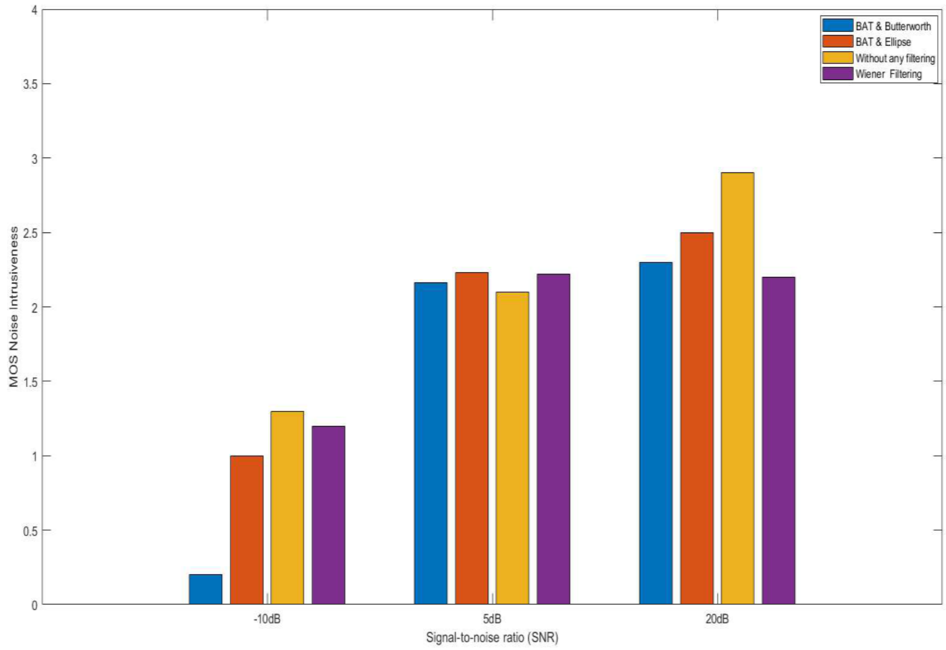


Figure 6.7: The mean opinion score for the noise intrusiveness level for BA-optimised filter (Bat, Butterworth and Elliptic), speech without filtering, Wiener filtering.

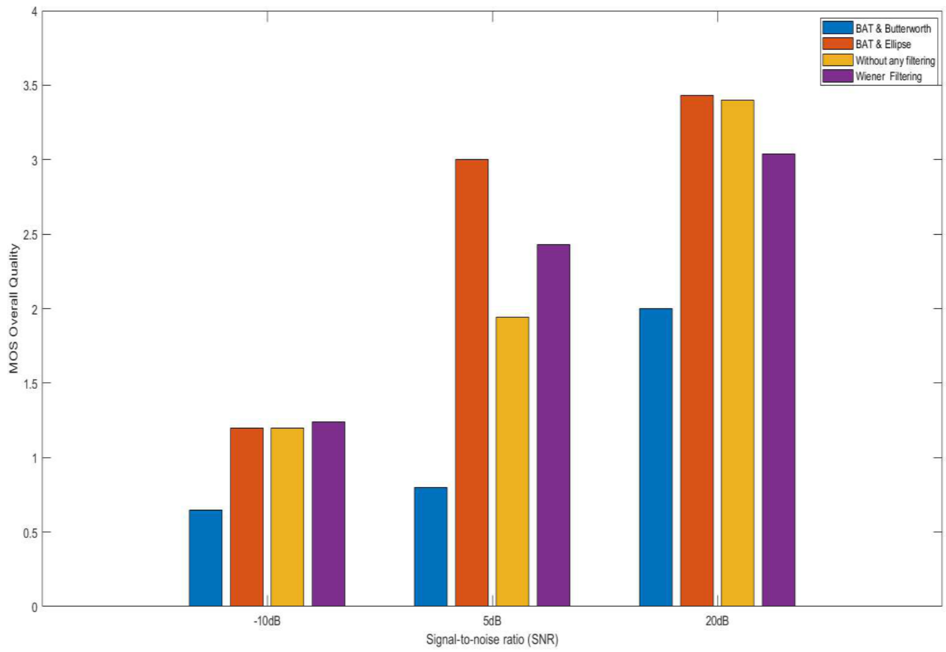


Figure 6.8: The mean opinion score for speech overall quality for noise distortion for BA-optimised filter (Bat, Butterworth and Elliptic), speech without filtering, Wiener filtering.

noise intrusiveness level, and overall speech quality).

The speech distortion, noise intrusiveness, and overall quality of the MOS results are presented in Figures 6.6, 6.7, and 6.8 respectively. As it can be seen, for speech distortion (Csig), the optimised Butterworth filter by Bat scored the highest at a low SNR of -10db, followed by the optimised Elliptic filter. For the positive SNR values the Butterworth scores poorly compared to the other methods, while the optimised Elliptic filter outperforms the other methods at 5db, and the audio only Wiener filter outperforms the methods at 20db. This is almost the same when compared to the objective composite evaluation previously done. Regarding the noise intrusiveness shown in Figure 6.7, the optimised Butterworth filter performs very poorly at negative SNR, but comparably at positive SNRs. While the optimised Elliptic filter by Bat performs equally with the audio-only Wiener at 5db. For the overall scores, the optimised Butterworth filter once again performs the worst among the methods at all SNRs values, and an equal performance is seen at -10 SNR for the optimised Elliptic filter, without any filtering speech and the audio-only Wiener filter. The results agree with the objective scores previously calculated.

6.4.1 Statistical analysis using the t_test

To investigate whether there are any significant differences between the means of the clean speech signal, the filter with a fixed coefficient, and the filter with an optimised coefficient by the BA, the authors applied t_tests to the results, at the 0.05 level of significance. The null and alternate hypotheses were tested for the case of the filter with a fixed coefficient as follows:

H_0: $\mu_C = \mu_{Fix}$

H_a: μ_C != μ_{Fix}

The null and the alternate hypotheses for the case of a filter with an optimised coefficient by Bat are as follows:

H_0: $\mu_C = \mu_{Opt}$

H_a: μ_C != μ_{Opt}

The t test result shown above in Table 6.6 attests to the significance of the opti-

Table 6.6: The results from the t_test at the 0.05 level of significance

Dataset		Null hyp. H_0	Alternate Hyp. H_1	p_value	t_value	Decision
NOIZEUS Dataset	BA	H_0: $\mu_C = \mu_{Opt}$	$H_a : \mu_C - \mu_{Opt} \mid= 0$	0.2082	-1.2583	Accept H_0
		H_0: $\mu_C = \mu_{Fix}$	$H_a : \mu_C - \mu_{Fix} \mid= 0$	0.2056	-1.2659	Accept H_0
Arabic Corpus	BA	H_0: $\mu_C = \mu_{Opt}$	$H_a : \mu_C - \mu_{Opt} \mid= 0$	0.9998	-0.0001	Accept H_0
		H_0: $\mu_C = \mu_{Fix}$	$H_a : \mu_C - \mu_{fix} \mid= 0$	0.9999	-9.7121	Accept H_0

mised filters, compared to those produced by the non-optimised ones, and the noisy signal.

6.5 Summary

In this chapter, an optimised speech-enhancement system was presented, which enhances noisy speech in different SNRs values. Being able to balance between the exploration and the exploitation during the search process makes the BA has a quick convergence rate compared to the previous algorithms. Hence, Section6.2 outlines the Bat algorithm and explains where it is inspired from. The overall structure of the proposed Bat speech-enhancement system was addressed in Section 6.3. A comprehensive evaluation of the performance of the proposed Bat speech enhancement system was presented in Section 6.4, where objective and subjective tests are both utilised, in a noisy environment, using babble noise at different SNR values of (-10db,0db,5db, and 20db). To compare the results, without any filtering speech sentences and a state-of-the-art algorithm: the audio-only Wiener Filter are used. The results outlined that objective and subjective evaluation measurements agreed that, the optimized Elliptic filter by the BA performs better than all the other filters, followed by the Wiener filter, and finally the optimised Butterworth filter by the Bat algorithm.

Chapter 7

Discussions and Analyses

7.1 Introduction

The work in this thesis utilised meta-heuristic optimisation approaches (PSO, APSO, GPSO, GSA and BA), to design an improved speech-enahancement filtering system, by making use of both Butterworth and Elliptic filters. The detailed system design was presented and discussed in chapter 4, chapter 5, chapter 6. This chapter will discuss the results obtained by the five methods used in the previous chapters.

7.2 Discussions

Tables 4.1, 5.2, and 6.2 the results of experiments conducted with the NOIZEUS data-set. An optimised Butterworth filter with PSO, APSO, GPSO, GSA and BA is applied at 20db, 5db, 0db and -10db SNRS. The averaged PESQ score were computed. As it can be seen the optimized filter by BA scored the highest value among all the methods at high SNR value of 20db, followed by the GSA optimized filter and PSO optimized filter. At SNR values of 5db and 0db, GSA outperforms the other methods followed by PSO, and at negative SNR value of -10db it seems PSO outperforms the other methods, followed by the BA algorithm. The PESQ scores in all methods, outperformed the fixed filter coefficient.

As for the optimised Elliptic filter, with the NOIZEUS data-set, the results are shown in Tables 4.2, 5.3, and 6.3 where it can be seen that, the optimized filter by PSO scores higher value of PESQ at high SNR of 20db. The optimized filter by BA seems to outperform all the other methods at the rest of SNR values(5db, 0,db, and -10db).

When it comes to the ARABIC corpus, for which the results are shown in both Tables 4.3, 5.4, and 6.4 for different SNRs of 20db, 5db, 0db and -10db, for the case of Butterworth. Again, it can be seen the filter optimsed by BA outperforms the other methods in all SNR values except at 5db, where PSO scores the best among the methods. The BA is seen to perform the best, compared to PSO and GSA in Tables 4.4, 4.4 and 6.5 for the Elliptic filter.

For C_{sig} the score of speech signal distortion the first composite measure,results are shown in Figures 4.6, 5.4, and 6.3. The PSO ,GSA and BA seems to perform equally at high level of SNR 20db, for both Butterworth and Elliptic filters, as well as without any filtering signal, and this could be filtering introduced distortion makes the signal less clear. Both GSA and BA score the same at SNR of 5bd,0db and -10db. However, it can be seen that, the PSO outperform both BA and GSA for the butterworth filter, and have the worst performance at negative SNR value at -10db.

When it comes to the background noise intrusiveness scores showed in Figures 4.7, 5.5, and 6.2. An equal scores are seen for both GSA and BA for both filters the Butterworth and the Elliptic, at all SNR values. However, the PSO scores are higher than those scores obtained by both GSA and BA, for the Butterworth filter at 20db, 5db. At negative value of -10db, the optimized filter by PSO scored the worst.

The overall speech qualityscores, which are presented in Figure 4.8, 5.6, and 6.5 indicated that both GSA and BA outperformed the PSO in all categories of SNR. But the PSO, seemed to to score high at negative SNR of -10db for the Ellipse filter. It is noticed scores were not high as expected at very low SNR levels of -10, for all the composite measures this could be due it was often impossible to recognise speech, and also because these algorithms were unable to identify sufficient level of speech in these levels to assign a quality score.

The mean opinion scores are shown at Figures 4.11, 5.9, and 6.8 , if we start with overall results, listeners were unable to hear speech filtered by Butterworth for all the three algorithms PSO, GSA, and BA at very low SNR levels. However, BA seemed to outperform PSO aand GSA and scores the best.

For speech distortion levels, Figures 4.10, 5.7, and6.6, volunteers listeners assigned a higher MOS to signal filtered by the Wiener filter and a very similar score to noisy and by the PSO, GSA and BA methods .

As for the noise intrusiveness scores, Figures 4.10, 5.8, and 6.7 listeners gave similar scores for all the three methods at high SNR value, where without any filtering speech scores the best. At SNR level of 5db optimized filter by Butterworth, scores higher than the Elliptic for PSO, GSA and BA algorithms. However at negative SNR value of -10db,the Elliptic filter outperformed the Butterworth filter. This could be because of the level of speech distortion introduced, gives a low overall score.

7.3 Summary

This chapter aimed to compare the results obtained by the different meta-heuristic speech enhancement systems The PESQ objective evaluation score results are presented as well as the composite objective measures, which are the speech signal distortion, background noise intrusiveness , and the overall speech quality. Furthermore, results obtained by MOS were presented.

Chapter 8

Concluding Remarks and Future Directions

This chapter draws the conclusions and summarises the findings of this thesis, by first giving a review of what was covered in previous chapters,. An overview of the research contributions is also given and potential future work directions are identified at the end of this chapter.

8.1 Summary

In this thesis, novel algorithms for speech enhancement have been developed by exploiting meta-heuristic approaches. We considered different techniques of meta-heuristic optimisation based on adaptive noise cancellation. Six novel methods are developed for speech-enhancement purposes. This thesis presented a review and a background to the research domain in Chapter 2. The authors categorised the speech-filtering techniques into four groups: conventional methods, adaptive filtering methods, machine learning methods (this includes adaptive filtering using optimisation techniques), and multi-modal methods. In this chapter, we also reviewed types of noises and also provided the difference between adaptive and non-adaptive, monaural and binaural, uni-modal speech-enhancement systems. The state-of-the-art speech-enhancement methods in the literature are provided. Statistical model-based methods are also examined, and adaptive noise cancellation concepts were explored. Their advantages and disadvantages are also outlined. Another category that was reviewed consisted of machine- learning approaches to speech enhancement, which include optimisation techniques for speech enhancement. Their pros and cons are also addressed.

After discussing the review and the background of this thesis, Chapter 3 presented the research methodology, the schematic overall speech enhancement system is provided, and the phases of the framework are identified. NOIZEUS and Arabic speech corpuses are chosen to develop and test the system. The aim of this thesis is to formulate the ANC problem in the form of an optimization task, hence Butterworth-filters, and Elliptic-filters are selected for this purpose. In order to perform the optimisation meta-heuristic optimisation algorithms are exploited in this research, particle swarm optimisation and its variants such as the accelerated particle swarm optimisation and the Gaussian particle swarm optimization methods are utilised. Also this research considered the gravitational search algorithm, and lastly the Bat algorithm. Furthermore, subjective and objective speech quality evaluation measurements, and statistical analyses using the t_test are discussed to assess the performance of the proposed speech-enhancement system.

The original key contributions of this thesis were given in Chapters 4, 5, and Chapter 6. In Chapter 4, a description of novel PSO and its variant speech enhancement system was detailed. The aim was to formulate an ANC system based on Butterworth, and Elliptic filters, in the form of an optimisation task and tuning of coefficients of an adaptive filter to remove the noise from a speech signal. The overall framework of the proposed speech-enhancement system is discussed at every stage on it and the algorithm is summarised. Then the results of the system were evaluated using subjective and objective tests along with the statistical analysis of the T_test. The results obtained by the filter optimised by PSO and APSO outperforms the GPSO, for both filters and both data-sets as well as for the results obtained by the fixed filter. For the composite measures, results outlined that the optimised Elliptic filter by PSO performs better than the other methods (GPSO and APSO) and the optimised PSO Butterworth filter.

In Chapter 5, a developed ANC system that utilised GSA to optimise the filter coefficients was introduced; the ANC is based on Elliptic and Butterworth filters. This GSA is based on the gravity concept, and here each mass in the search space is considered as a possible solution representing the coefficients of the filter. The proposed GSA optimised speech-enhancement system is outlined in this chapter with the overall structure of the system. Finally to test the performance of the proposed system, different SNR values (-10db, 0db, 5db, and 20db) are examined for both benchmark data-sets; also subjective and objective speech evaluation tests are conducted. The results revealed that, the optimised filter by GSA performs better than the Fixed filter at all SNRs levels, and that the fixed filter performs better than the audio only Wiener filter. Furthermore, the composite objective measures revealed that the optimised Elliptic filter performed better than the Butterworth filter, and slightly better than the ”without filtering”

Another contribution of this research is presented in Chapter 6. The application of the Bat algorithm for speech enhancement was the subject of Chapter 6. The Bat algorithm utilised the concept of the echolocation behaviour of micro bats. With the ability of auto-zooming, micro bats demonstrate automatic adjustment from exploration to exploitation when reaching global optimal solutions. Hence, in this chapter, a novel optimised speech-enhancement system based on the Bat algorithm is deployed. The overall structure of the proposed speech-enhancement system is demonstrated. The evaluation of the results of the proposed system is presented, then subjective and objective tests were discussed. The results indicated that objective and subjective evaluation measurements agreed that, the optimized Elliptic filter by the BA performs better than all the other filters, followed by the Wiener filter, and finally the optimised Butterworth filter by the Bat algorithm.

8.2 Conclusions

Based on the investigated meta-heuristic optimisation approaches , the experiments carried out with the meta-heuristic speech enhancement systems, and the evaluation of this novel speech enhancement systems, the following conclusions can be drawn:

- This thesis investigated the use of meta-heuristic optimisation methods for speech enhancement, and to propose a new speech-enhancement methods based on optimisation. A fitness function based on signal quality measurements was formulated and utilised. The performance of the proposed techniques with existing sate-of-art methods, is conducted , by using both objective and subjective tests.

- It is worthy to say that most of the SNR levels considered in this thesis were high, with which, limited improvements to the speech signal were obtained. Compared to some state-of-art the results obtained in this thesis are slightly better.

- Among the proposed filters, the Elliptic filter found to give better results than the Butterworth filter, that could be because it can provide degrees of freedom for controlling its response.

- BA algorithm also showed the best performance amongst the other approaches, and that because BA is being able to balance between the exploration and the exploitation during the search process, which makes it has a quick convergence rate compared to the previous algorithms.

- It can be seen that the meta-heuristic speech enhancement system presented in this thesis is capable of successfully enhancing noisy speech, as proven by PESQ, composite objective scores and subjective listening tests

8.3 Future directions

There are other meta-heuristic algorithms to be explored such as galactic swarm optimisation, genetic algorithms, ant colony optimisation ,equilibrium optimiser, seagull optimization algorithm, sooty tern optimisation algorithm .

Further new meta-heuristic-based speech enhancement approaches by hybridising the existing methods to enhance the search performance require further research.

In this research, the grid search is used for hyperparameter optimisation, another technique is to use automated hyper-parameter tuning (Bayesian optimization, or genetic algorithms).

Also, there is the adoption of different cost functions like the mean square error to minimise cost based on other objective methods.

Filters considered in this thesis are out band filters, where noise that does not overlap, in the frequencies of the desired signal. To overcome this, in the future a band pass or stop band filters are to be explored and utilized.

It is also possible to extend the proposed framework to explore other machine learning techniques, like neural networks or convolutional recurrent networks.

Bibliography

Abd El-Fattah, M., Dessouky, M. I., Diab, S. M., and Abd El-Samie, F. E.-S. (2008). Speech enhancement using an adaptive wiener filtering approach. *Progress In Electromagnetics Research M*, 4:167–184.

Abel, A. (2013). Towards an intelligent fuzzy based multimodal two stage speech enhancement system.

Adrio, C. L. (2013). http://www.radio-electronics.com/info/rf-technology-design/rf-filters/elliptic-cauer-rf-filter-basics.php. Accessed: 2018-11-04.

Adrio, C. L. (2015). http://www.radio-electronics.com/info/rf-technology-design/rf-filters/butterworth-rf-filter-calculations-formulae-php. Accessed: 2018-11-04.

Aggarwal, A., Rawat, T. K., and Upadhyay, D. K. (2016). Design of optimal digital fir filters using evolutionary and swarm optimization techniques. *AEU - International Journal of Electronics and Communications*, 70(4):373 – 385.

Akhaee, M. A., Ameri, A., and Marvasti, F. A. (2005). Speech enhancement by adaptive noise cancellation in the wavelet domain. In *Information, Communications and Signal Processing, 2005 Fifth International Conference on*, pages 719–723. IEEE.

Almajai, I. and Milner, B. (2011). Visually derived wiener filters for speech enhancement. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(6):1642–1651.

Arif, M., Naseem, I., Moinuddin, M., Khan, S. S., and Ammar, M. M. (2017). Adaptive noise cancellation using q-lms. In *2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT)*, pages 1–4. IEEE.

Asl, L. B. and Geravanchizadeh, M. (2010). Dual-channel speech enhancement based on stochastic optimization strategies. In *Information Sciences Signal Processing and their Applications (ISSPA), 2010 10th International Conference on*, pages 229–232. IEEE.

Asl, L. B. and Nezhad, V. M. (2010). Speech enhancement using particle swarm optimization techniques. In *Measuring Technology and Mechatronics Automation (ICMTMA), 2010 International Conference on*, volume 3, pages 441–444. IEEE.

Bahoura, M. (2017). Pipelined architecture of multi-band spectral subtraction algorithm for speech enhancement. *Electronics*, 6(4):73.

Bai, L. and Yin, Q. (2010). A modified nlms algorithm for adaptive noise cancellation. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 3726–3729. IEEE.

Bai, Q. (2010). Analysis of particle swarm optimization algorithm. *Computer and information science*, 3(1):180.

Banchhor, S., Dodia, J., and Gowda, D. (2013). Gui based performance analysis of speech enhancement techniques. *International Journal of Scientific and Research Publications*, 3(9):1.

Berouti, M., Schwartz, R., and Makhoul, J. (1979). Enhancement of speech corrupted by acoustic noise. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '79.*, volume 4, pages 208–211.

Bharti, S. S., Gupta, M., and Agarwal, S. (2016). A new spectral subtraction method for speech enhancement using adaptive noise estimation. In *Recent Advances in Information Technology (RAIT), 2016 3rd International Conference on*, pages 128–132. IEEE.

Boll, S. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27(2).

Bresciani, J.-P., Dammeier, F., and Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of vision*, 6(5):2–2.

Cai, Y. and Hou, C. (2012). Subband spectral-subtraction speech enhancement based on the dft modulated filter banks. In *Signal Processing (ICSP), 2012 IEEE 11th International Conference on*, volume 1, pages 571–574. IEEE.

Cao, L., q. Zhang, T., x. Gao, H., and Yi, C. (2012). Multi-band spectral subtraction method combined with auditory masking properties for speech enhancement. In *2012 5th International Congress on Image and Signal Processing*, pages 72–76.

Cappé, O. (1994). Elimination of the musical noise phenomenon with the ephraim and malah noise suppressor. *IEEE transactions on Speech and Audio Processing*, 2(2):345–349.

Chaudhari, A. and Dhonde, S. B. (2015). A review on speech enhancement techniques. In *2015 International Conference on Pervasive Computing (ICPC)*, pages 1–3.

Coto-Jimenez, M., Goddard-Close, J., Di Persia, L., and Rufiner, H. L. (2018). Hybrid speech enhancement with wiener filters and deep lstm denoising autoencoders. In *2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*, pages 1–8. IEEE.

D, S. and K, P. (2015). A survey on various multichannel speech enhancement algorithms. *IJARCCE*, pages 254–255.

Das, S. and Sarma, K. K. (2012). Noise cancellation in stochastic wireless channels using coding and adaptive filtering. *International Journal of Computer Applications*, 975:8887.

Dewasthale, M. M., Kharadkar, R. D., and Bari, M. (2015). Comparative performance analysis and hardware implementation of adaptive filter algorithms for acoustic noise cancellation. In *2015 International Conference on Information Processing (ICIP)*, pages 124–129.

Dhiman, J., Ahmad, S., and Gulia, K. (2013). Comparison between adaptive filter algorithms (lms, nlms and rls). *International Journal of Science, Engineering and Technology Research (IJSETR)*, 2(5):1100–1103.

Ding, G.-H., Huang, T., and Xu, B. (2004). Suppression of additive noise using a power spectral density mmse estimator. *IEEE Signal Processing Letters*, 11(6):585–588.

Dixit, S. and Mulge, D. M. Y. (2014). Review on speech enhancement techniques. *International Journal of Computer Science and Mobile Computing, IJCSMC*, 3(8):285–290.

Djendi, M., Henni, R., and Sayoud, A. (2016). A new dual forward bss based rls (dfrls) algorithm for speech enhancement. In *Engineering & MIS (ICEMIS), International Conference on*, pages 1–4. IEEE.

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature*, 381(6577):66.

El-Fattah, M. A. A., Dessouky, M. I., Abbas, A. M., Diab, S. M., El-Rabaie, E.-S. M., Al-Nuaimy, W., Alshebeili, S. A., and El-Samie, F. E. A. (2014). Speech enhancement with an adaptive wiener filter. *International Journal of Speech Technology*, 17(1):53–64.

Ephraim, Y. and Malah, D. (1984). Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(6):1109–1121.

Ephraim, Y. and Malah, D. (1985). Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(2):443–445.

Ephraim, Y. and Van Trees, H. L. (1995). A signal subspace approach for speech enhancement. *IEEE Transactions on speech and audio processing*, 3(4):251–266.

Farooq, M. U., Ahmad, A., and Hameed, A. (2017). Opposition-based initialization and a modified pattern for inertia weight (iw) in pso. In *Innovations in Intelligent Systems and Applications (INISTA), 2017 IEEE International Conference on*, pages 96–101. IEEE.

Fisli, S. and Djendi, M. (2018). Blind speech intelligiblility enhancement by a new dual modified predator-prey particle swarm optimization algorithm. *Applied Acoustics*, 141:125–135.

Fisli, S., Djendi, M., and Guessoum, A. (2018a). Modified predator-prey particle swarm optimization based two-channel speech quality enhancement by forward blind source separation. In *Natural Language and Speech Processing (ICNLSP), 2018 2nd International Conference on*, pages 1–6. IEEE.

Fisli, S., Djendi, M., and Guessoum, A. (2018b). Modified predator-prey particle swarm optimization based two-channel speech quality enhancement by forward blind source separation. In *2018 2nd International Conference on Natural Language and Speech Processing (ICNLSP)*, pages 1–6.

Fisli, S., Djendi, M., and Guessoum, A. (2019). A new dual modifted bat algorithm for design of adaptive noise canceller. In *2019 International Conference on Advanced Electrical Engineering (ICAEE)*, pages 1–6. IEEE.

Fister Jr, I., Yang, X.-S., Fister, I., Brest, J., and Fister, D. (2013). A brief review of nature-inspired algorithms for optimization. *arXiv preprint arXiv:1307.4186*.

Fu, S.-W., Hu, T.-y., Tsao, Y., and Lu, X. (2017). Complex spectrogram enhancement by convolutional neural network with multi-metrics learning.

Fu, S.-W., Tsao, Y., and Lu, X. (2016). Snr-aware convolutional neural network modeling for speech enhancement. In *INTERSPEECH*, pages 3768–3772.

Gentle, J. E., Härdle, W. K., and Mori, Y. (2012). *Handbook of computational statistics: concepts and methods*. Springer Science & Business Media.

Geravanchizadeh, M. and Asl, L. B. (2010). Asexual reproduction-based adaptive quantum particle swarm optimization algorithm for dual-channel speech enhancement. In *Communications, Control and Signal Processing (ISCCSP), 2010 4th International Symposium on*, pages 1–4. IEEE.

Ghibeche, Y., Saadi, S., and Hafaifa, A. (2019). Optimal design of iir filters based on least lpnorm using a novel meta-heuristic algorithm. *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, 32(1):e2480.

Goehring, T., Bolner, F., Monaghan, J. J., van Dijk, B., Zarowski, A., and Bleeck, S. (2017). Speech enhancement based on neural networks improves speech intelligibility in noise for cochlear implant users. *Hearing research*, 344:183–194.

Gorriz, J. M., Ram´ırez, J., Cruces-Alvarez, S., Puntonet, C. G., Lang, E. W., and Erdogmus, D. (2009). A novel lms algorithm applied to adaptive noise cancellation. *IEEE Signal Processing Letters*, 16(1):34–37.

Gouhar, T., Jaber, N., and Kuntumalla, P. (2017). Speech enhancement using new iterative minimum statistics approach. In *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 1–4.

Guopin, H., Wei, Z., and Qin, Z. (2013). Improvement of audio noise reduction system based on rls algorithm. In *Computer Science and Network Technology (ICCSNT), 2013 3rd International Conference on*, pages 964–968. IEEE.

Hafezi, R., Shahrabi, J., and Hadavandi, E. (2015). A bat-neural network multi-agent system (bnnmas) for stock price prediction: Case study of dax stock price. *Applied Soft Computing*, 29:196–210.

Halabi, N. (2016). *Modern standard Arabic phonetics for speech synthesis*. PhD thesis, University of Southampton, UK. url:https://eprints.soton.ac.uk/409695/.

Halliday, D., Resnick, R., and Walker, J. (2013). *Fundamentals of physics*. John Wiley & Sons.

Haykin, S. (1996). *Adaptive Filter Theory (3rd Ed.)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Hifrin, S., Palash, N. B., and Uzzal, S. (2014). General report on speech recognition using pattern classification methods. *Communication, Cloud and Big Data: Proceedings of CCB 2014*.

Houcque, D. and Otrs (2005). Introduction to matlab for engineering students. *Northwestern University*, pages 1–64.

Hu, H.-T., Kuo, F.-J., and Wang, H.-J. (2002). Supplementary schemes to spectral subtraction for speech enhancement. *Speech Communication*, 36(3–4):205 – 218.

Hu, Y. and Loizou, P. C. (2002). A subspace approach for enhancing speech corrupted by colored noise. In *ICASSP*, pages 573–576. IEEE.

Hu, Y. and Loizou, P. C. (2006). Evaluation of objective measures for speech enhancement. In *Ninth International Conference on Spoken Language Processing*.

Hu, Y. and Loizou, P. C. (2007). Subjective evaluation and comparison of speech enhancement algorithms. *Speech Commun*, 49:588–601.

Hu, Y. and Loizou, P. C. (2008). Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on audio, speech, and language processing*, 16(1):229–238.

Hussain, A., Chetouani, M., Squartini, S., Bastari, A., and Piazza, F. (2007). *Nonlinear Speech Enhancement: An Overview*, pages 217–248. Springer Berlin Heidelberg.

Inai, T., Hara, S., Abe, M., Ijima, Y., Miyazaki, N., and Mizuno, H. (2015). Sub-band text-to-speech combining sample-based spectrum with statistically generated spectrum. In *Sixteenth Annual Conference of the International Speech Communication Association*.

Islam, M. T., Shahnaz, C., and Fattah, S. (2014). Speech enhancement based on a modified spectral subtraction method. In *2014 IEEE 57th International Midwest Symposium on Circuits and Systems (MWSCAS)*, pages 1085–1088. IEEE.

Jabloun, F. and Champagne, B. (2003). Incorporating the human hearing properties in the signal subspace approach for speech enhancement. *IEEE Transactions on Speech and Audio Processing*, 11(6):700–708.

Ji, C., Zhang, Y., Tong, M., and Yang, S. (2008). Particle filter with swarm move for optimization. In *International Conference on Parallel Problem Solving from Nature*, pages 909–918. Springer.

Jia, H., Wang, W., Wang, D., and Zhang, X. (2019). Speech enhancement using modified mmse-lsa and phase reconstruction in voiced and unvoiced speech. *International Journal of Pattern Recognition and Artificial Intelligence*, 33(02):1958002.

Jiang, S., Ji, Z., and Shen, Y. (2014). A novel hybrid particle swarm optimization and gravitational search algorithm for solving economic emission load dispatch problems with various practical constraints. *International Journal of Electrical Power & Energy Systems*, 55:628–644.

Johnson, D. H. and Shami, P. N. (1993). The signal processing information base. *IEEE Signal Processing Magazine*, 10(4):36–42.

Kalamani, M., Valarmathy, S., and Krishnamoorthi, M. (2014). Modified noise reduction algorithm for speech enhancement. *Applied Mathematical Sciences*, 8(89):4447–4452.

Kamath, S. and Loizou, P. C. (2002). A multi-band spectral subtraction method for enhancing speech corrupted by colored noise. In *ICASSP*, page 4164. IEEE.

Karabu Ga, N. and Cetinkaya, M. B. (2011). A novel and efficient algorithm for adaptive filtering: artificial bee colony algorithm. *Turkish Journal of Electrical Engineering & Computer Sciences*, 19(1):175–190.

Kennedy, J. and Eberhart, R. C. (2001). *Swarm Intelligence*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. ISBN:1-55860-595-9.

Khaldi, K. and Touati, H. (2018). Speech enhancement in emd domain using spectral subtraction and wiener filter.

Krishna, E. H., Raghuram, M., Madhav, K. V., and Reddy, K. A. (2010). Acoustic echo cancellation using a computationally efficient transform domain lms adaptive filter. In *Information Sciences Signal Processing and their Applications (ISSPA), 2010 10th International Conference on*, pages 409–412. IEEE.

Krohling, R. A. (2004). Gaussian swarm: a novel particle swarm optimization algorithm. In *Cybernetics and Intelligent Systems, 2004 IEEE Conference on*, volume 1, pages 372–376. IEEE.

Kulkarni, D. S., Deshmukh, R. R., and Shrishrimal, P. P. (2016). A review of speech signal enhancement techniques. *International Journal of Computer Applications*, 139(14).

Kunche, P., Rao, G. S. B., Reddy, K., and Maheswari, R. U. (2015). A new approach to dual channel speech enhancement based on hybrid psogsa. *International Journal of Speech Technology*, 18(1):45–56.

Kunche, P. and Reddy, K. (2016a). Heuristic and meta-heuristic optimization. In *Metaheuristic Applications to Speech Enhancement*, pages 17–24. Springer.

Kunche, P. and Reddy, K. (2016b). *Metaheuristic Applications to Speech Enhancement*. Springer.

Kunche, P. and Reddy, K. (2016c). Speech enhancement approach based on accelerated particle swarm optimization (apso). In *Metaheuristic Applications to Speech Enhancement*, pages 39–60. Springer.

Kunche, P. and Reddy, K. (2016d). Speech enhancement approach based on gravitational search algorithm (gsa). In *Metaheuristic Applications to Speech Enhancement*, pages 61–75. Springer.

Kunche, P. and Reddy, K. (2016e). Speech enhancement based on bat algorithm (ba). In *Metaheuristic Applications to Speech Enhancement*, pages 91–110. Springer.

Lakshmikanth, S., Nataraj, K., and Rekha, K. (2014). Noise cancellation in speech signal processing: A review. *International Journal of Advanced Research in Computer and Communication Engineering*, 3(1).

Lee, J.-W. and Lee, J.-J. (2013). Gaussian-distributed particle swarm optimization: A novel gaussian particle swarm optimization. In *Industrial Technology (ICIT), 2013 IEEE International Conference on*, pages 1122–1127. IEEE.

Lim, J. S. and Oppenheim, A. V. (1979a). Enhancement and bandwidth compression of noisy speech. *Proceedings of the IEEE*, 67(12):1586–1604.

Lim, J. S. and Oppenheim, A. V. (1979b). Enhancement and bandwidth compression of noisy speech. *Proceedings of the IEEE*, 67(12):1586–1604.

Lin, C.-T. (2003). Single-channel speech enhancement in variable noise-level environment. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 33(1):137–143.

Liu, A., Li, P., Deng, X., and Ren, L. (2020). A sigmoid attractiveness based improved firefly algorithm and its applications in iir filter design. *Connection Science*, pages 1–25.

Lockwood, P. and Boudy, J. (1992). Experiments with a nonlinear spectral subtractor (nss), hidden markov models and the projection, for robust speech recognition in cars. *Speech Communication*, 11(2):215 – 228.

Loizou, P. (2013a). *Speech Enhancement: Theory and Practice, Second Edition*. Taylor & Francis. ISBN:9781466504219.

Loizou, P. C. (2013b). *Speech Enhancement: Theory and Practice*. CRC Press, Inc., Boca Raton, FL, USA, 2nd edition.

Loubna, K., Bachir, B., and Izeddine, Z. (2018). Optimal digital iir filter design using ant colony optimization. In *2018 4th International Conference on Optimization and Applications (ICOA)*, pages 1–5. IEEE.

Lu, Y. and Loizou, P. C. (2008). A geometric approach to spectral subtraction. *Speech communication*, 50(6):453–466.

Mahbub, U., Shahnaz, C., and Fattah, S. A. (2010). An adaptive noise cancellation scheme using particle swarm optimization algorithm. In *Communication Control and Computing Technologies (ICCCCT), 2010 IEEE International Conference on*, pages 683–686.

Mauger, S. J., Warren, C. D., Knight, M. R., Goorevich, M., and Nel, E. (2014). Clinical evaluation of the nucleus® 6 cochlear implant system: Performance improvements with smartsound iq. *International journal of audiology*, 53(8):564–576.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588):746–748.

Meyer, J. and Simmer, K. U. (1997). Multi-channel speech enhancement in a car environment using wiener filtering and spectral subtraction. In *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 1167–1170 vol.2.

Mirjalili, S., Mirjalili, S. M., and Lewis, A. (2014). Grey wolf optimizer. *Advances in engineering software*, 69:46–61.

Mohammed, J. R. (2007). A new simple adaptive noise cancellation scheme based on ale and nlms filter. In *Communication Networks and Services Research, 2007. CNSR'07. Fifth Annual Conference on*, pages 245–254. IEEE.

Monaci, G. (2007). On the modelling of multi-modal data using redundant dictionaries. *EPFL*.

Nataraj, V. S., Athulya, M., and Savithri, S. P. (2017). Single channel speech enhancement using adaptive filtering and best correlating noise identification. In *2017 IEEE 30th Canadian conference on electrical and computer engineering (CCECE)*, pages 1–4. IEEE.

Ogata, S. and Shimamura, T. (2001). Reinforced spectral subtraction method to enhance speech signal. In *TENCON 2001. Proceedings of IEEE Region 10 International Conference on Electrical and Electronic Technology*, volume 1, pages 242–245. IEEE.

Osgouei, S. G. and Geravanchizadeh, M. (2010). Dual-channel speech enhancement based on a hybrid particle swarm optimization algorithm. In *Telecommunications (IST), 2010 5th International Symposium on*, pages 873–877. IEEE.

Panchmatia, H., Gaikar, K., and Patel, D. (2016). Comparison of different speech enhancement techniques. *Imperial Journal of Interdisciplinary Research*, 2(5).

Pandey, A. and Wang, D. (2019). A new framework for cnn-based speech enhancement in the time domain. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(7):1179–1188.

Pérez, J., Valdez, F., and Castillo, O. (2015). Modification of the bat algorithm using fuzzy logic for dynamical parameter adaptation. In *Evolutionary Computation (CEC), 2015 IEEE Congress on*, pages 464–471. IEEE.

Poli, R. (2008). Analysis of the publications on the applications of particle swarm optimisation. *Journal of Artificial Evolution and Applications*, 2008.

Prabhakaran, Indra, and Kasthuri (2014). Tamil speech enhancement using nonlinear spectral subtraction. In *2014 International Conference on Communication and Signal Processing*, pages 1482–1485.

Prajna, K., Rao, G., Reddy, K., and Maheswari, R. U. (2014a). A new approach to dual channel speech enhancement based on gravitational search algorithm (gsa). *International Journal of Speech Technology*, 17(4):341–351.

Prajna, K., Rao, G. S., Reddy, K., and Maheswari, R. U. (2014b). Application of bat algorithm in dual channel speech enhancement. In *Communications and Signal Processing (ICCSP), 2014 International Conference on*, pages 1457–1461. IEEE.

Raitio, T., Juvela, L., Suni, A., Vainio, M., and Alku, P. (2015). Phase perception of the glottal excitation of vocoded speech. In *Sixteenth Annual Conference of the International Speech Communication Association*.

Rakesh, P. and Kumar, T. K. (2015). A novel rls based adaptive filtering method for speech enhancement. *World Academy of Science, Engineering and Technology, International Journal of Electrical, Computer, Electronics and Communication Engineering*, 9(2):624–628.

Rashedi, E., Nezamabadi-Pour, H., and Saryazdi, S. (2009). Gsa: a gravitational search algorithm. *Information sciences*, 179(13):2232–2248.

Rashedi, E., Rashedi, E., and Nezamabadi-pour, H. (2018). A comprehensive survey on gravitational search algorithm. *Swarm and evolutionary computation*, 41:141–158.

Ravi, K. K. and Subbaiah, P. (2016). A survey on speech enhancement methodologies. *International Journal of Intelligent Systems and Applications*, 8(12):37.

Recommendation, I.-T. (2001). Perceptual evaluation of speech quality (pesq): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. *Rec. ITU-T P. 862*. https://www.itu.int/rec/T-REC-P.862.

Rix, A. W., Beerends, J. G., Hollier, M. P., and Hekstra, A. P. (2001). Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs. In *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)*, volume 2, pages 749–752. IEEE.

Rothauser, E. (1969). Ieee recommended practice for speech quality measurements. *IEEE Trans. on Audio and Electroacoustics*, 17:225–246.

Sabri, N. M., Puteh, M., and Mahmood, M. R. (2013). A review of gravitational search algorithm. *Int. J. Advance. Soft Comput. Appl*, 5(3):1–39.

Saleem, N., Irfan Khattak, M., Ali, M. Y., and Shafi, M. (2019a). Deep neural network for supervised single-channel speech enhancement. *Archives of Acoustics*, 44.

Saleem, N. and Khattak, M. I. (2020). Deep neural networks for speech enhancement in complex-noisy environments. *IJIMAI*, 6(1):84–90.

Saleem, N., Khattak, M. I., and Qazi, A. B. (2019b). Supervised speech enhancement based on deep neural network. *Journal of Intelligent & Fuzzy Systems*, 37(4):5187–5201.

Sayoud, A., Djendi, M., and Guessoum, A. (2017). A two-channel fast nlms algorithm for speech enhancement and acoustic noise reduction. In *2017 5th International Conference on Electrical Engineering-Boumerdes (ICEE-B)*, pages 1–4. IEEE.

Scalart, P. et al. (1996). Speech enhancement based on a priori signal to noise estimation. In *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, volume 2, pages 629–632. IEEE.

Selvi, R. S. and Suresh, G. (2016). Hybridization of spectral filtering with particle swarm optimization for speech signal enhancement. *International Journal of Speech Technology*, 19(1):19–31.

Shi, Y., Rong, W., and Zheng, N. (2018). Speech enhancement using convolutional neural network with skip connections. In *2018 11th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, pages 6–10. IEEE.

Singh, A. (2001). Adaptive noise cancellation. *Central Elektronica Engineering Research Institute, University of Dehli*.

Singh, S. and Bansal, E. M. K. E. D. (2018). Low pass iir filter design using hybrid pso-gsa optimization algorithm. *International Journal of Research and Analytical Reviews*, 5.

Soni, B. A. and Vaghela, K. (2017). Spectral subtraction and mmse: A hybrid approach for speech enhancement. *International Reaserch Journal of Engineering and Technology*, 4(4):2340–2343.

SPIB (2013). Signal processing information base (spib). http://spib.linse.ufsc.br/noise.html. Accessed: 2018-11-04.

Subha, R. and Himavathi, S. (2016). Accelerated particle swarm optimization algorithm for maximum power point tracking in partially shaded pv systems. In *Electrical Energy Systems (ICEES), 2016 3rd International Conference on*, pages 232–236. IEEE.

Sulong, A., Gunawan, T. S., Khalifa, O. O., Kartiwi, M., and Ambikairajah, E. (2016). Speech enhancement based on wiener filter and compressive sensing. *Indonesian Journal of Electrical Engineering and Computer Science*, 2(2):367–379.

SUN, C., XIE, J., and LENG, Y. (2016). A signal subspace speech enhancement approach based on joint low-rank and sparse matrix decomposition. *Archives of Acoustics*, 41(2).

Sunnydayal, V., Sivaprasad, N., and Kumar, T. K. (2014). A survey on statistical based single channel speech enhancement techniques. *International Journal of Intelligent Systems and Applications*, 6(12):69.

Surendran, S. and Kumar, T. K. (2016). Perceptual subspace speech enhancement with ssdr normalization. In *2016 International Conference on Microelectronics, Computing and Communications (MicroCom)*, pages 1–6.

Taha, T. M., Wajid, S. K., and Hussain, A. (2019). Speech enhancement based on adaptive noise cancellation and particle swarm optimization. volume 15, pages 691–701. Science Publications.

Taha, T. M. F., Adeel, A., and Hussain, A. (2018). A survey on techniques for enhancing speech. *International Journal of Computer Applications*, 179(17):1–14.

Thaitangam, P. G., Laishram, R., Devi, K. U., Khwairakpam, R., Singh, M. R., and Oinam, C. (2018a). Speech enhancement using adaptive filter with bat algorithm. In *2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pages 1–5. IEEE.

Thaitangam, P. G., Laishram, R., Devi, K. U., Khwairakpam, R., Singh, M. R., and Oinam, C. (2018b). Speech enhancement using adaptive filter with bat algorithm. In *2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pages 1–5. IEEE.

Tripathi, S. and Ikbal, M. A. (2015). Optimization of lms algorithm for adaptive filtering using global optimization techniques. *International Journal of Computer Applications*, 975:8887.

Upadhyay, N. and Karmakar, A. (2013). An improved multi-band spectral subtraction algorithm for enhancing speech in various noise environments. *Procedia Engineering*, 64:312–321.

Vaseghi, S. V. (2008). *Advanced digital signal processing and noise reduction*. John Wiley & Sons.

Violentyev, A., Shimojo, S., and Shams, L. (2005). Touch-induced visual illusion. *Neuroreport*, 16(10):1107–1110.

Virag, N. (1999). Single channel speech enhancement based on masking properties of the human auditory system. *IEEE Transactions on speech and audio processing*, 7(2):126–137.

Wallace, M. T., Roberson, G., Hairston, W. D., Stein, B. E., Vaughan, J. W., and Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, 158(2):252–258.

Wan, C., Wang, J., Yang, G., and Zhang, X. (2011). Gaussian particle swarm optimization with differential evolution mutation. In *International Conference in Swarm Intelligence*, pages 439–446. Springer.

Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D., and Rees, G. (2006). Sound alters activity in human v1 in association with illusory visual perception. *Neuroimage*, 31(3):1247–1256.

Widrow, B. and Stearns, S. D. (1985). *Adaptive signal processing*, volume 15. Prentice-hall Englewood Cliffs, NJ.

Wiener, N. (1949). *Extrapolation, interpolation, and smoothing of stationary time series : with engineering applications*. M.I. T. paperback series. Cambridge, Mass. Technology Press of the Massachusetts Institute of Technology.

XU, W., WANG, G., and CHEN, L. (2017). Speech enhancement algorithm based on improved variable-step lms algorithm in cochlear implant. *Journal of Computer Applications.*, 37:1212–1216.

Xu, Y., Du, J., Dai, L.-R., and Lee, C.-H. (2014). An experimental study on speech enhancement based on deep neural networks. *IEEE Signal processing letters*, 21(1):65–68.

Xu, Y., Du, J., Dai, L.-R., and Lee, C.-H. (2015). A regression approach to speech enhancement based on deep neural networks. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 23(1):7–19.

Yan, W. G., Xiang, G. Y., and Qun, Z. X. (2013). A signal subspace speech enhancement method for various noises. *Indonesian Journal of Electrical Engineering and Computer Science*, 11(2):726–735.

Yang, X.-S. (2010). *Nature-inspired metaheuristic algorithms*. Luniver press.

Yang, X.-S. (2011a). Review of meta-heuristics and generalised evolutionary walk algorithm. *International Journal of Bio-Inspired Computation*, 3(2):77–84.

Yang, X.-S. (2011b). Review of meta-heuristics and generalised evolutionary walk algorithm. *Int. J. Bio-Inspired Comput.*, 3(2):77–84.

Yu, C. and Su, L. (2015). Speech enhancement based on the generalized sidelobe cancellation and spectral subtraction for a microphone array. In *2015 8th International Congress on Image and Signal Processing (CISP)*, pages 1318–1322.

Zhang, R. and Liu, J. (2018). An improved multi-band spectral subtraction using mel-scale. *Procedia computer science*, 131:779–785.