Noise estimation from degraded Speech

Project report submitted in partial fulfillment
of the requirements for the degree of

Bachelor of Technology
in
Electronics and Communication Engineering

by

Rahul Raj - 20UCC086
Shubham Thorbole - 20UEC126

Under Guidance of
Dr. Navneet Upadhyay

Department of Electronics and Communication Engineering
The LNM Institute of Information Technology, Jaipur

November 2023

The LNM Institute of Information Technology
Jaipur, India

CERTIFICATE

This is to certify that the project entitled "Noise estimation from degraded Speech" , submitted by Rahul Raj (Communication and Computer Engineering) and Shubham Thorbole (Electronics and Communication Engineering) in partial fulfillment of the requirement of degree in Bachelor of Technology (B. Tech), is a bonafide record of work carried out by them at the Department of Electronics and Communication Engineering, The LNM Institute of Information Technology, Jaipur, (Rajasthan) India, during the academic session 2022-2023 under my supervision and guidance and the same has not been submitted elsewhere for award of any other degree. In my/our opinion, this report is of standard required for the award of the degree of Bachelor of Technology (B. Tech).

03/05/2024
Date

Adviser: Name of BTP Supervisor

Acknowledgments

We would like to take this opportunity to express our heartfelt gratitude for your invaluable guidance and support during the completion of our project, ”Noise estimation from degraded speech.” Your expertise, patience, and dedication have been instrumental in helping us to successfully complete this project.

Your insightful feedback and constructive criticism have pushed us to perform better, and your unwavering support has helped us navigate through the challenges we faced. Your attention to detail, profound knowledge of the subject, and willingness to help us whenever we needed it has been a great source of inspiration for us.

Our comprehension of the topic has been greatly aided by your ability to clearly and concisely explain complicated concepts. You have always pushed us to think critically, and your direction has enabled us to get fresh viewpoints and ideas on the subject.

Your dedication to helping us succeed academically has been unwavering, and your efforts have surpassed our expectations. In addition to assisting us in achieving our objectives, your mentoring has given us a sense of confidence and self-belief that will last the rest of our lives.

Once more, we would want to thank you for all of your help, encouragement, and advice during this endeavour. We feel privileged to have had the chance to work with you,

Abstract

The suggested project seeks to address the issue of voice enhancement systems' highly non-stationary noise settings. Non-stationary noise is noise whose characteristics are continually changing, making it challenging for conventional noise-estimation algorithms to efficiently lessen its impact on speech signals.

The suggested technique uses time and frequency dependent smoothing parameters to update the noise estimate depending on the likelihood of a signal in each frequency bin to address this issue. This is done by calculating the noisy speech power spectrum's ratio to its local minimum. This ratio is updated continually by averaging previous noisy speech power spectrum values with a look-ahead factor. As a result, the local minimum estimation approach can swiftly adapt to situations with extremely non-stationary noise.

Formal listening tests of the proposed approach have revealed that, when incorporated into speech improvement systems, it outperforms alternative noise-estimation techniques. This shows that the algorithm significantly improves speech signals in contexts with extremely non-stationary noise, when conventional algorithms might not succeed in producing good results.

This project is significant because it deals with a pressing problem in voice augmentation systems. Real-world applications frequently involve non-stationary noise environments, and the suggested technique effectively addresses this issue. The algorithm is particularly beneficial in contexts where noise characteristics might fluctuate considerably over short time scales because of its ability to adapt fast to changing noise characteristics.

The suggested approach is also easily adaptable to different applications because it is founded on sound signal processing concepts. It could be applied to a variety of systems, including voice assistants, teleconferencing systems, and hearing aids, that need noise estimation and speech amplification.

The suggested approach makes a substantial contribution to the fields of noise estimation and voice enhancement overall. It is a useful tool for a variety of applications due to its effectiveness in dealing with highly non-stationary noise conditions as well as its adaptability and portability.

Contents

List of Figures

Chapter 1

Introduction

1.1 The Area of Work

The proposed project falls under the area of speech processing and enhancement. More specifically, it addresses the issue of noise-estimation in highly non-stationary noise environments. The project utilizes principles of signal processing, such as time and frequency dependent smoothing factors and local minimum estimation, to develop an algorithm that effectively estimates noise in speech signals.

The project involves developing and implementing the proposed algorithm in a speech enhancement system, and subjecting it to formal listening tests to evaluate its effectiveness. The project requires a strong understanding of signal processing principles, as well as experience with programming and software development.

The project is relevant to a range of applications where speech enhancement and noise-estimation are required, such as hearing aids, teleconferencing systems, and voice assistants. Therefore, the project has implications beyond the academic realm and has potential practical applications.

Overall, the project falls within the intersection of signal processing, speech enhancement, and machine learning. It requires expertise in these fields, as well as a strong understanding of how noise affects speech signals and how to develop effective algorithms to address this issue.

1.2 Problem Addressed

The problem addressed in the above project is the issue of highly non-stationary noise environments in speech enhancement systems. Non-stationary noise refers to noise that is constantly changing in its characteristics, making it challenging for traditional noise-estimation algorithms to effectively reduce its impact on speech signals.

This problem is particularly prevalent in real-world scenarios, where speech signals are often affected by background noise that varies in its characteristics over time. In such scenarios, traditional noise-estimation algorithms may fail to provide satisfactory results, leading to reduced speech quality and increased difficulty in speech recognition.

The proposed project aims to address this problem by developing an algorithm that utilizes time and frequency dependent smoothing factors to update the noise estimate based on the signal-presence probability in individual frequency bins. The algorithm also includes a local minimum estimation algorithm that adapts quickly to highly non-stationary noise environments.

By addressing the issue of non-stationary noise environments in speech enhancement systems, the proposed project aims to improve speech quality and enhance the performance of speech recognition systems in real-world scenarios. This has practical implications for a range of applications, including hearing aids, teleconferencing systems, and voice assistants.

1.3 Existing Noise Estimation Algorithm

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

1.3.1 Martin's Noise estimation algorithm, 2001

The Martin's Noise estimation algorithm is a widely used algorithm for estimating noise in speech signals. The algorithm estimates the noise PSD by averaging the PSD of a series of short-time frames that are free of speech activity. This algorithm is simple and effective, but it assumes that the noise is stationary over time, which may not be the case in highly non-stationary noise environments. In such scenarios, more advanced algorithms, such as the one proposed in the above project, may be required to effectively estimate noise in speech signals.[1]

1.3.2 Cohen and Berdugo noise estimation algorithm, 2002-2003

The Cohen and Berdugo noise estimation algorithm is based on the assumption that speech and noise signals have different statistical properties, which can be used to distinguish them. The algorithm computes the eigenvalues of overlapping time-frequency patches of the spectrogram of the noisy speech signal to estimate the signal-to-noise ratio (SNR) for each patch. The SNR estimate is used to weight the corresponding patch in the calculation of the noise PSD estimate. The noise PSD estimate is then used in a speech enhancement algorithm, such as Wiener filtering, to reduce the noise in the speech signal. This algorithm is effective in a range of noise environments, including non-stationary noise environments, but may require more computational resources than simpler algorithms such as Martin's algorithm.[2]

1.3.3 Doblinger noise estimation algorithm, 1995

The Doblinger noise estimation algorithm estimates the local noise variance for each time-frequency bin in the speech signal using a maximum likelihood approach. The local noise variance estimates are then smoothed over time and frequency using a two-dimensional median filter to compute the noise power spectral density estimate, which is used in a speech enhancement algorithm, such as spectral subtraction, to reduce the noise in the speech signal. The Doblinger algorithm is effective in a range of noise environments, including non-stationary noise environments, but may require more computational resources than simpler algorithms such as Martin's algorithm.[3]

1.3.4 Hirsch and Ehrlicher noise estimation algorithm, 1995

The Hirsch and Ehrlicher noise estimation algorithm estimates the local noise mean and variance for each time-frequency bin in the speech signal using a two-dimensional Gaussian distribution. The local noise mean and variance estimates are then smoothed over time and frequency using a two-dimensional median filter to compute the noise power spectral density estimate, which is used in a speech enhancement algorithm, such as spectral subtraction, to reduce the noise in the speech signal. The Hirsch and Ehrlicher algorithm is effective in a range of noise environments, including non-stationary noise environments, but may require more computational resources than simpler algorithms such as Martin's algorithm.[4]

1.4 Creation of bibliography

Use bibch1.bib file to save your bib format citations. Use the command [5] for referring to a particular article [6].

Chapter 2

Literature Review

2.1 Review

When considering a research project, it is important to carefully read and understand the content and methodology of the work. In this case, the article in question is ”A Noise Estimation Algorithm for Highly Nonstationary Environments” by Sundarrajan Rangachari and Philipos C. Loizou.

Noise estimation is an important task in speech processing, as it can significantly affect the performance of speech enhancement algorithms. In highly non-static environments where noise changes rapidly, traditional livestock assessment methods may not work well. The authors of this paper try to solve this problem by proposing a new noise estimation algorithm that combines statistical techniques and machine learning.

The introduction of the paper provides detailed background on the importance of noise estimation in speech processing and the challenges presented by highly non-static environments. The authors also review previous research in the field and identify gaps in the literature that their research aims to address.

The proposed noise estimation algorithm is described in detail in the methodology section. The algorithm is based on a statistical model that captures the statistical characteristics of speech and noise, and a machine learning component that uses the statistical model to estimate the noise level. The authors also describe the evaluation methodology used to test the effectiveness of the algorithm by simulating highly non-stationary noisy environments and comparing the performance of the algorithms with several other noise evaluation methods.

Regarding the results of the work, the authors present the results of the evaluation of the noise estimation algorithm. They show that their algorithm outperforms other noise estimation methods in highly non-static environments. The authors also provide a detailed analysis of the algorithm's performance and identify factors that may affect its performance.

In the Introduction and Discussion sections, the authors summarize their conclusions and contributions and discuss potential applications of their noise estimation algorithm. They also identify limitations of their study and suggest directions for future research.

To fully understand this research, it is important to understand the statistical and machine learning methods used in the proposed noise estimation algorithm. In particular, it is important to understand the statistical model used to capture the statistical characteristics of speech and noise, and the machine learning component that uses the model to estimate the noise level. Furthermore, to fully understand the results presented in the paper, it is important to understand the evaluation method used to test the performance of the algorithm.

In general, reading and understanding the article ”A Noise Estimation Algorithm for Highly Nonstationary Environments” requires a good technical understanding of noise estimation in speech processing and an understanding of statistical and machine learning techniques. It is important to carefully read the title, abstract, introduction, methodology, results, conclusions and discussion sections of the article and go through the references mentioned in the article. In this way, a comprehensive overview of the research can be obtained and its relevance determined for a specific project.[7]

Chapter 3

Proposed Work

Let the noisy speech signal in the time domain be denoted as

$$y[n] = x[n] + d[n] \tag{3.1}$$

where x(n) is the clean speech and d(n) is the additive noise. The smoothed power spectrum of noisy speech is computed using the following first-order recursive equation:

$$P(k, k) = gP(k-1, k) + (1-g)|Y(k, k)|^2 \tag{3.2}$$

where P(k,k) is the smoothed power spectrum, k is the frame index, k is the frequency index, jY(k,k)j 2 is the short-time power spectrum of noisy speech and g is a smoothing constant. The proposed algorithm is summarized in the flow diagram shown below Next, we describe each of the individual blocks of the algorithm

Here is a detailed explanation of each step of the algorithm:
1. **Define frame parameters:** The speech signal is divided into 20 ms long frames with an overlap of 10 ms between successive frames. A Hamming window is placed on each frame to reduce spectral leakage. That step is important because it allows the analysis of the speech signal in small segments that can be assumed static. The Hamming window helps reduce spectral leakage, which can occur when the edges of the signal window do not match the signal waveform evenly, resulting in a distortion of the frequency spectrum.

2. **Calculate the magnitude:** The FFT of each frame is calculated and the magnitude spectrum is obtained by taking the absolute value of the FFT. FFT (Fast Fourier Transform) is used to transform a time domain signal into its frequency domain representation. The magnitude spectrum represents the distribution of energy between the different frequency components of the signal.

3. **Calculate the local minimum of the power spectrum of noisy speech:** for each frequency interval, an estimate of the minimum power is obtained by following the minimum power over time using a recursive algorithm. This step is used to estimate the minimum power level for

each frequency range of the noisy speech power spectrum. This is done by recursively updating the minimum power estimate based on the previous estimate and the current power spectrum.

$$P_{\min}(\lambda-1,k) < P(\lambda,k) \Rightarrow \begin{cases} P_{\min}(\lambda,k) = \gamma P_{\min}(\lambda-1,k) + \frac{1-\gamma}{1-\beta}\left(P(\lambda,k) - \beta P_{\min}(\lambda-1,k)\right) \\ P_{\min}(\lambda,k) = P(\lambda,k) \end{cases} \tag{3.3}$$

4. **Compute the smoothed speech power spectrum:** The smoothed power estimate is obtained for each frequency range by averaging the power spectrum over time using a recursive algorithm. This step is used to estimate the long-term average strength for each frequency range of the speech power spectrum. To do this, the smoothed power estimate is recursively updated based on the previous estimate and the current power spectrum.

5. **Calculate the ratio of the smoothed speech power spectrum to its local minimum:** the ratio is calculated for each frequency range by dividing the smoothed power estimate by the minimum power estimate. This step is used to estimate the amount of speech energy in each frequency band compared to the noise level of that bin. This ratio is used to calculate the probability of a call occurring in the next step.

$$Sr(k,k) = \frac{P(k,k)}{P_{\min(k,k)}} \tag{3.4}$$

6. **Calculate the probability of a speech event using a first-order recursion:** A probability estimate is obtained for each frequency range by comparing the ratio to the threshold value. If the ratio is above the threshold, the probability of speech occurring in that frequency range is high, otherwise it is low. This step is used to classify each frequency interval as either speech or noise based on the ratio calculated in the previous step.

7. **Calculate the time-frequency dependent smoothing factors:** Using the probability estimate, a smoothing factor is calculated for each frequency range that is used to update the noise estimate. This step is used to estimate the time-varying nature of the speech signal and adjust the noise estimate based on the current state of the speech signal.

8. **Update the noise estimate using time- and frequency-dependent smoothing factors:** the noise estimate is updated for each frequency range with the smoothing factor. This step is used to estimate the noise power level in each frequency band based on the current state of the speech signal and the estimated smoothing factor. The raw estimate is then used in subsequent processing steps to denoise the speech signal.

$$D(\lambda,k) = a_s(\lambda,k)D(\lambda-1,k) + (1-a_s(\lambda,k))|Y(\lambda,k)|^2 \tag{3.5}$$

Chapter 4

Simulation and Results

4.1 Results of proposed algorithm

4.1.1 Clean speech signal plot

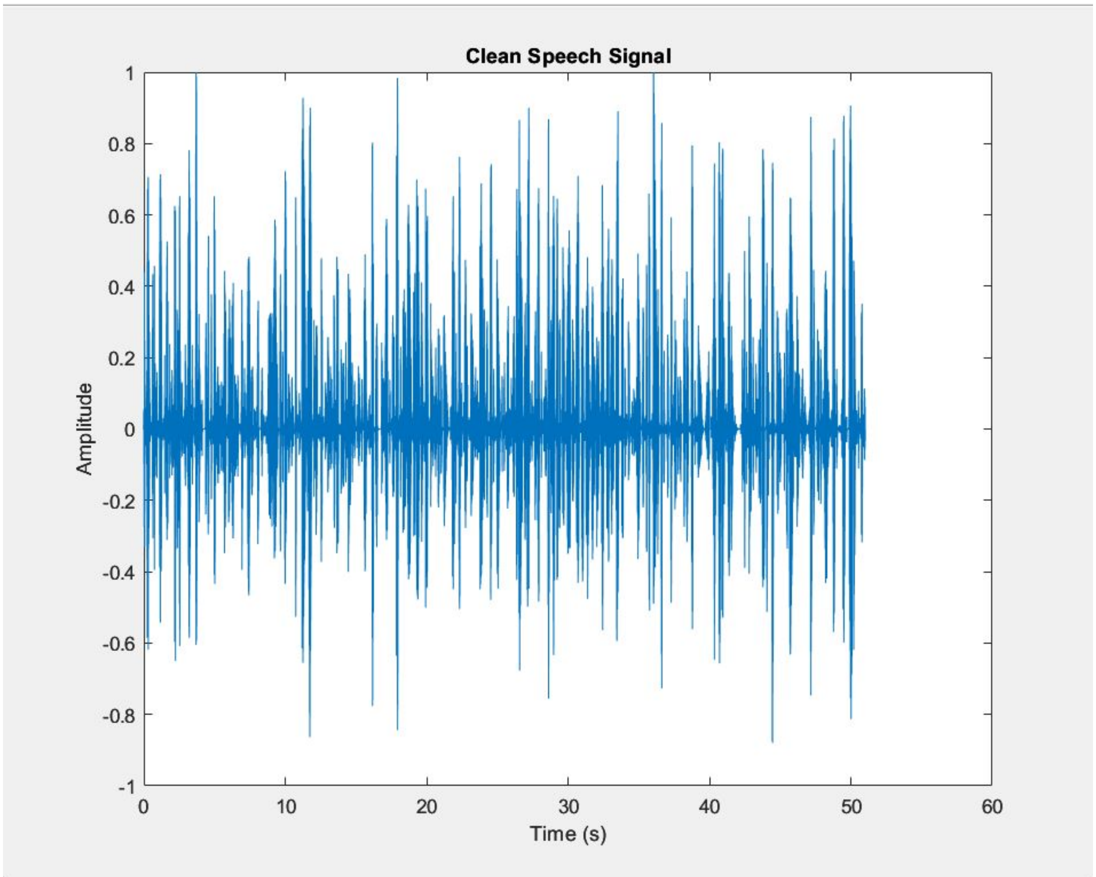

FIGURE 4.1: Clean Speech Signal

A clean speech signal plot in MATLAB shows the amplitude of the speech signal on the y-axis versus time on the x-axis. The amplitude represents the intensity of the sound waves, and time represents the duration of the speech signal. In a clean speech signal plot, the speech signal is free from any noise or distortion, and the plot appears smooth and continuous. The plot may show variations in amplitude over time, which correspond to the changes in the speech signal, such as the transitions between phonemes or the rise and fall of intonation.

4.1.2 Noisy speech signal plot

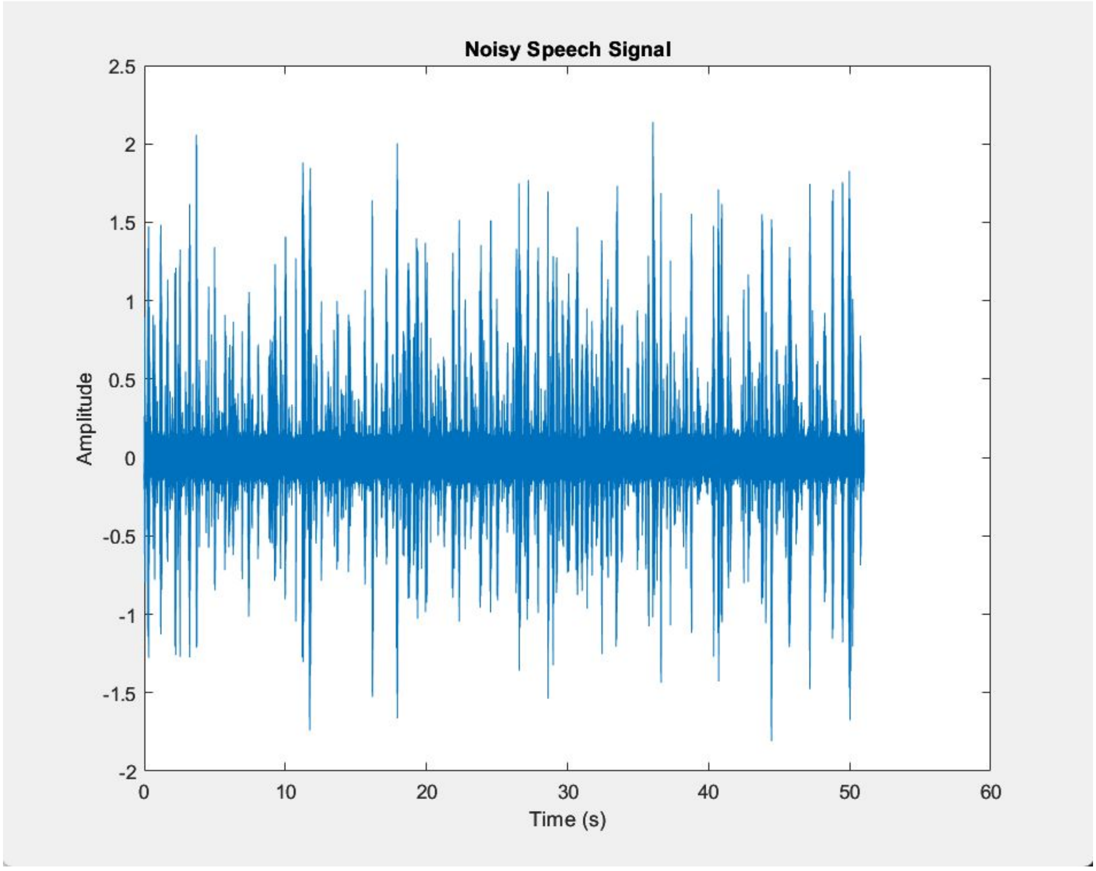

FIGURE 4.2: Noisy Speech Signal

A plot of a noisy speech signal in MATLAB would show the amplitude of the signal on the y-axis and time on the x-axis. The plot would likely appear more jagged and less smooth than the plot of a clean speech signal, with more noticeable variations in amplitude over time. This is because the noisy speech signal contains unwanted background noise, which can cause fluctuations in the amplitude of the speech signal. The plot may also show occasional spikes in amplitude, indicating sudden bursts of noise or interference. Overall, the plot of a noisy speech signal would appear less clean and more erratic than the plot of a clean speech signal.

4.1.3 Noisy speech signal at one frame

A plot of a single frame of a noisy speech signal in Matlab shows the amplitude of the signal over time for a duration of 20 milliseconds. The plot typically shows a waveform that fluctuates

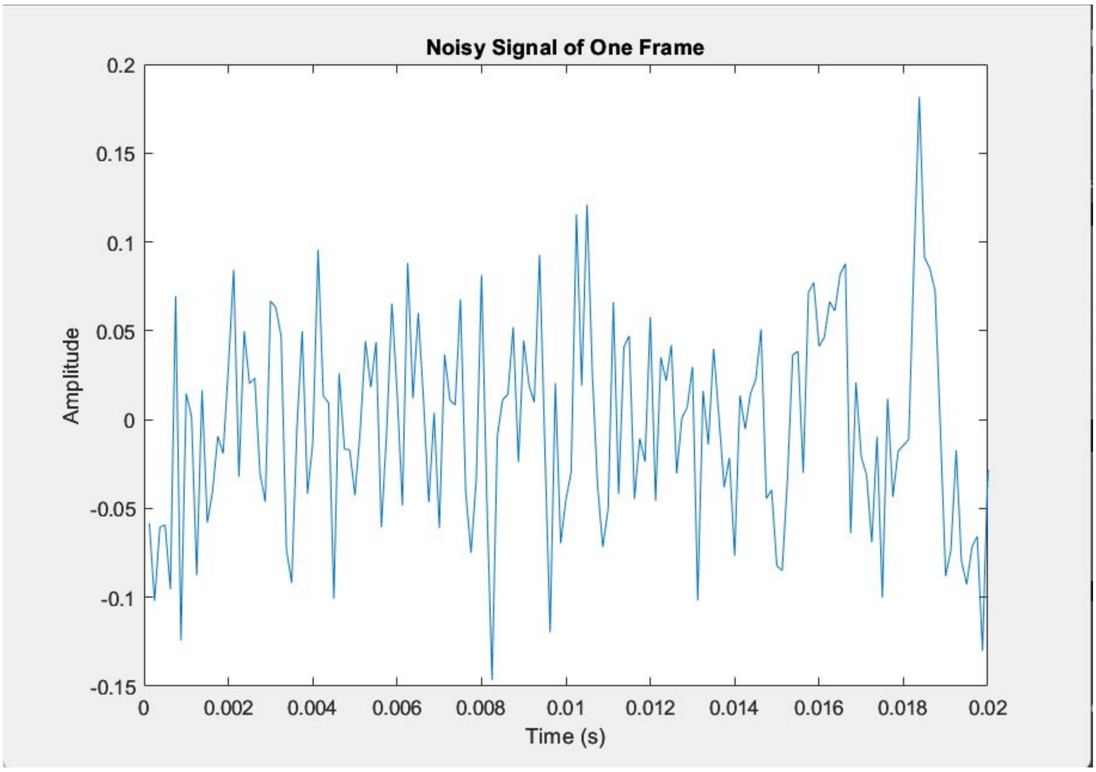

FIGURE 4.3: Noisy Speech Signal at single frame

in amplitude over time, representing the speech signal. In a noisy speech signal, the waveform is distorted by various sources of noise, which can cause additional fluctuations in amplitude.

4.1.4 Fast fourier transform of single frame noisy speech signal

The Fast Fourier Transform (FFT) of a noisy speech signal is a plot that shows the magnitude of the FFT coefficients in decibels (dB) versus the frequency bin index for one frame of the signal. The frequency bin index is a measure of the frequency resolution of the FFT, and is directly proportional to the sampling frequency and inversely proportional to the length of the FFT.The plot is useful for identifying the spectral characteristics of the noisy speech signal, which can be used to inform subsequent steps in the speech enhancement process.

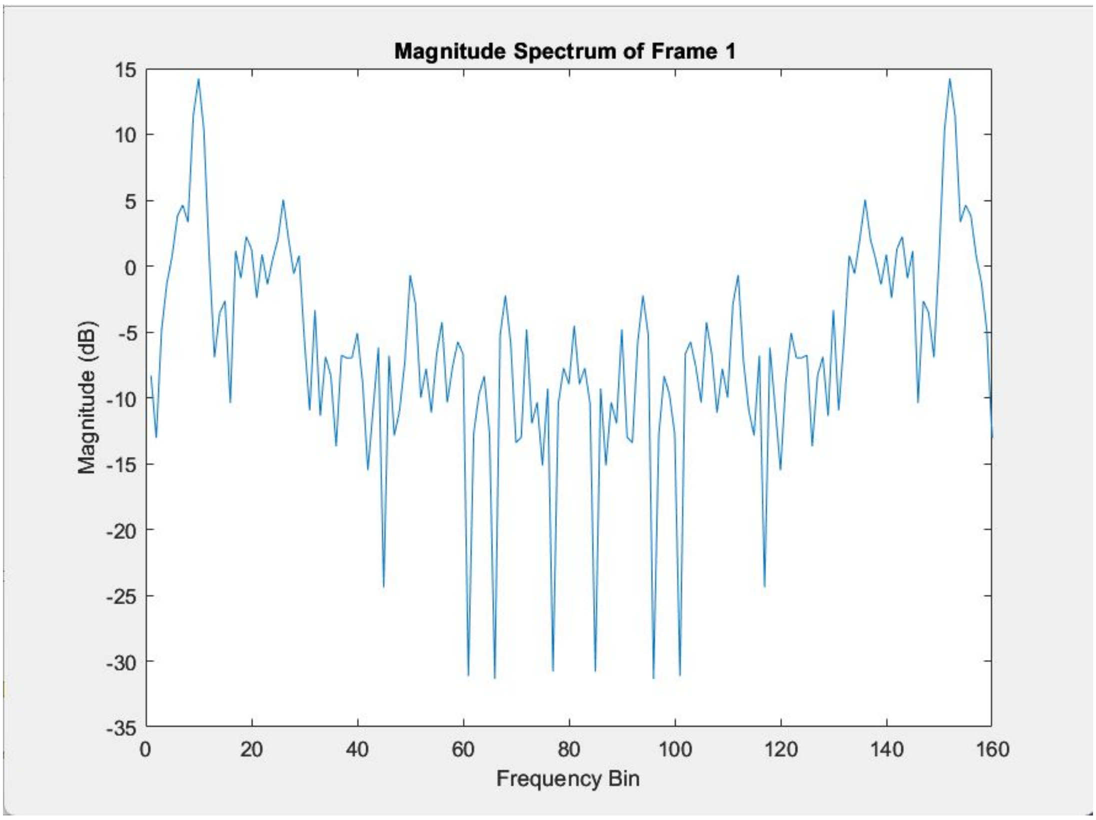

FIGURE 4.4: Magnitude Spectrum of single Frame

4.1.5 Noisy speech and local minimum

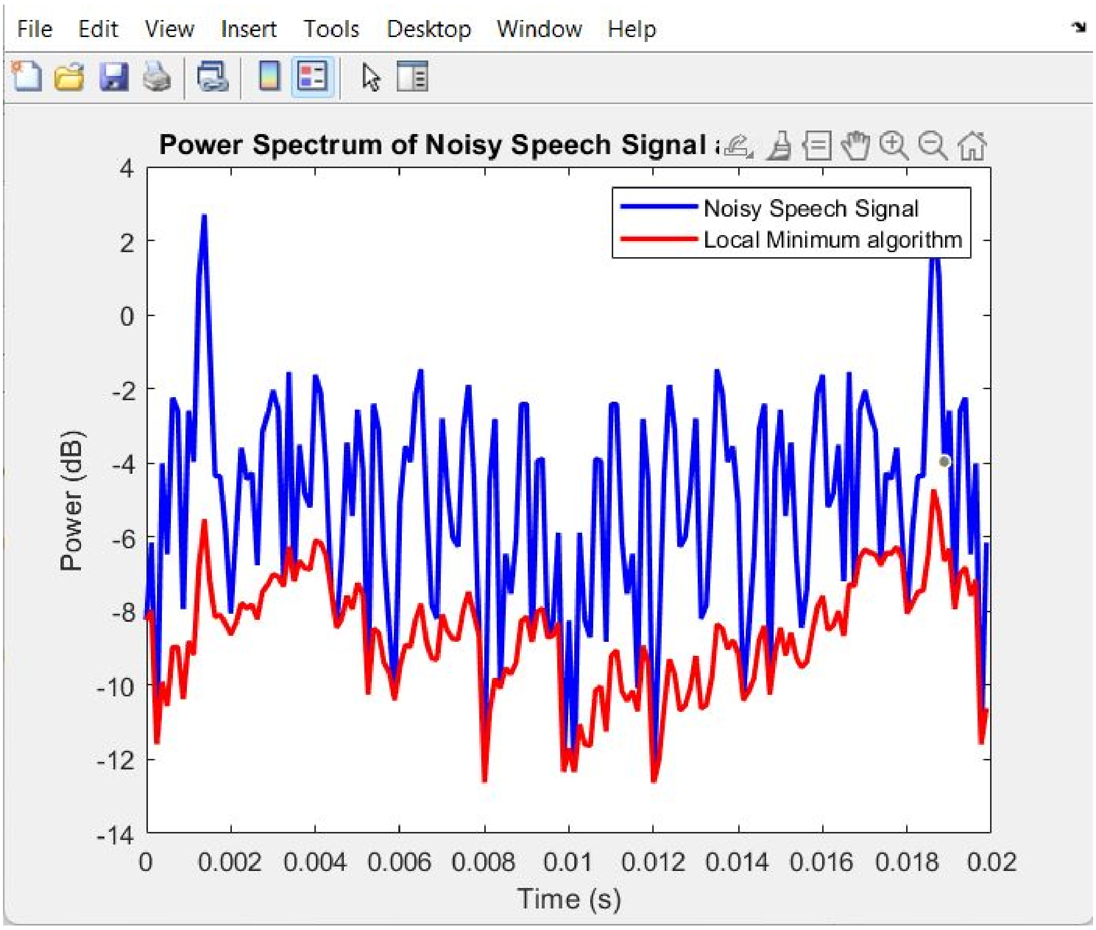

FIGURE 4.5: Noisy Speech power spectrum and local minimum algorithm

A power vs time plot shows the variation of signal power over time. In the context of speech enhancement, a noisy speech signal and its local minimum estimate are typically plotted in this way. The noisy speech power values represent the energy level of the signal at each time instant, while the local minimum values represent an estimate of the noise floor in each frequency bin. The local minimum values are often obtained by tracking the minimum power over time using a recursive algorithm. By comparing the noisy speech power values to the local minimum values, it is possible to estimate the SNR (signal-to-noise ratio) at each time instant.

4.2 Results of comparison with existing algorithms

4.2.1 Comparison with MS algorithm

The Minimum Statistics (MS) algorithm, as proposed by Martin in 2001, revises the estimation of noise by continuously monitoring the lowest point in the spectrum of the noisy speech.

Consequently, the rate of adaptation of the noise estimate is contingent upon how quickly this local minimum adjusts. In instances of non-stationary noise, where the noise intensity changes gradually, both our method and the minimum statistics method require a similar duration for adaptation. However, when faced with escalating noise levels, the adaptation time for the MS method may extend slightly beyond 1.5 seconds, while our method operates at a faster rate of only 0.5 seconds for adaptation.

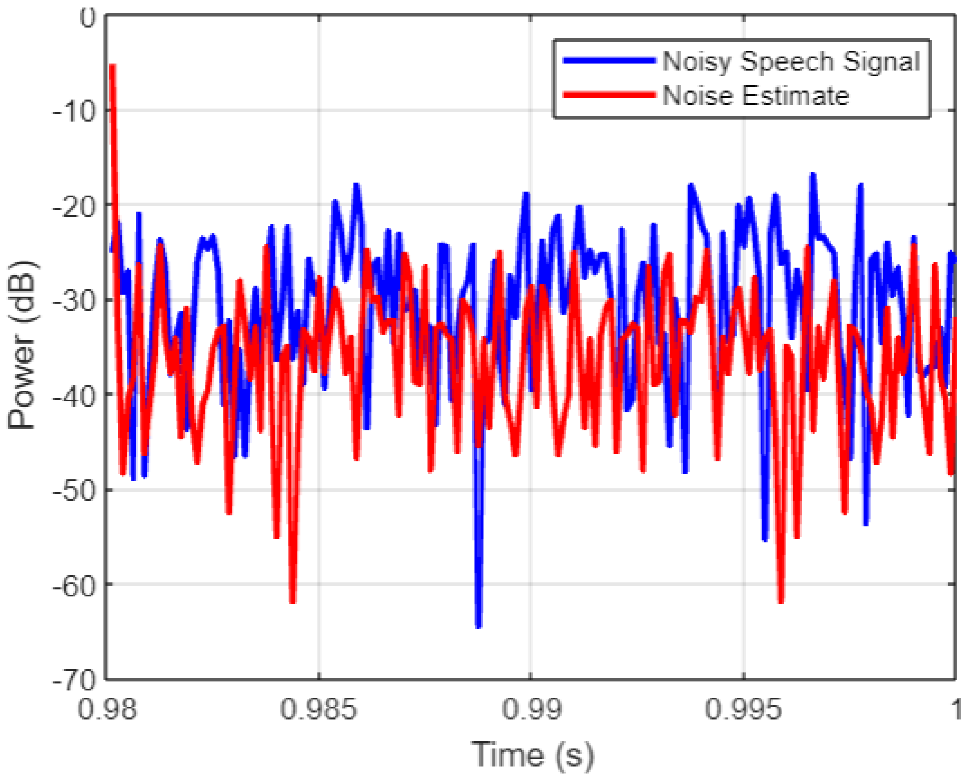

FIGURE 4.6: Noise spectrum and estimated noise spectrum using MS algorithm

4.2.2 Comparison with MCRA algorithm

The MCRA-based speech enhancement algorithm, highlights its parameterized control, the importance of the Hamming window, and its adaptive noise estimation based on speech presence detection. Furthermore, you can emphasize the significance of the visual component for practical applications and real-time assessment of speech enhancement. A Hamming window, known for its effectiveness in reducing spectral leakage during Fourier analysis, is defined to be applied to each frame. The windowing process enhances the spectral representation of each frame, improving the accuracy of subsequent operations.

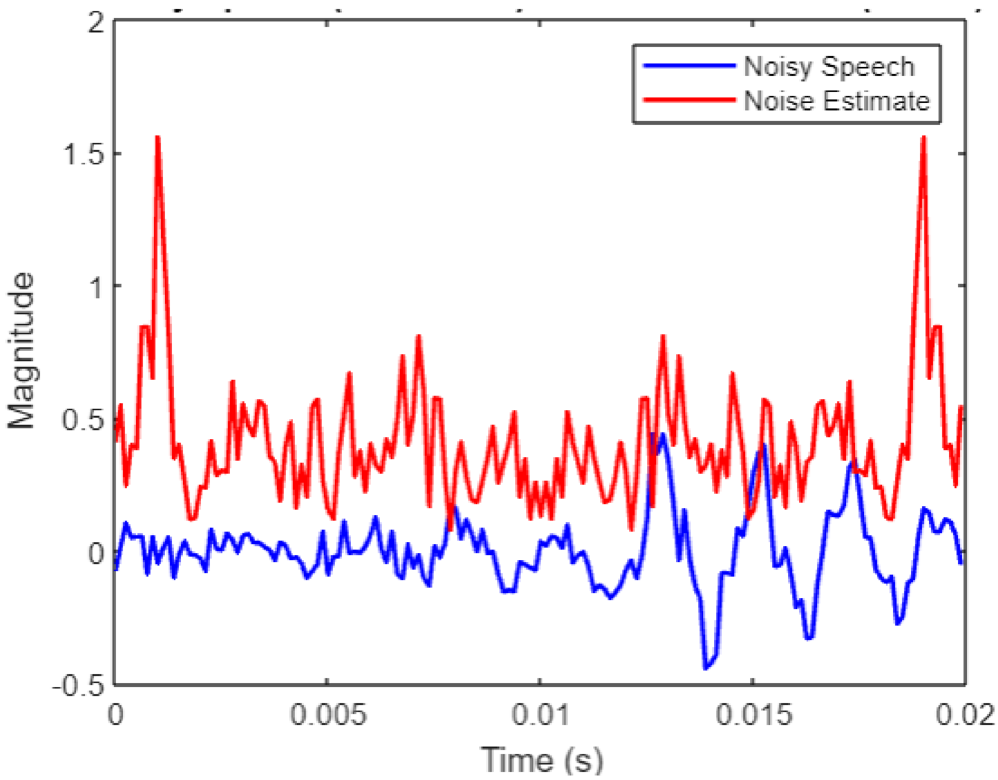

FIGURE 4.7: Noise spectrum and estimated noise spectrum using MCRA method

4.2.3 Comparison with weighted average technique

To implement the Weighted averaging method to update the noise estimate. For each frequency bin in the magnitude spectrum of the chosen frame, the code evaluates whether the magnitude is below a certain threshold. If the magnitude is lower than the threshold, the noise estimate is adjusted by scaling it with the threshold factor and the previous estimate. Otherwise, if the magnitude is higher than the threshold, the noise estimate is set equal to the magnitude value. This adaptive approach ensures that the noise estimate tracks the noise level while accounting for the characteristics of the noisy speech signal.

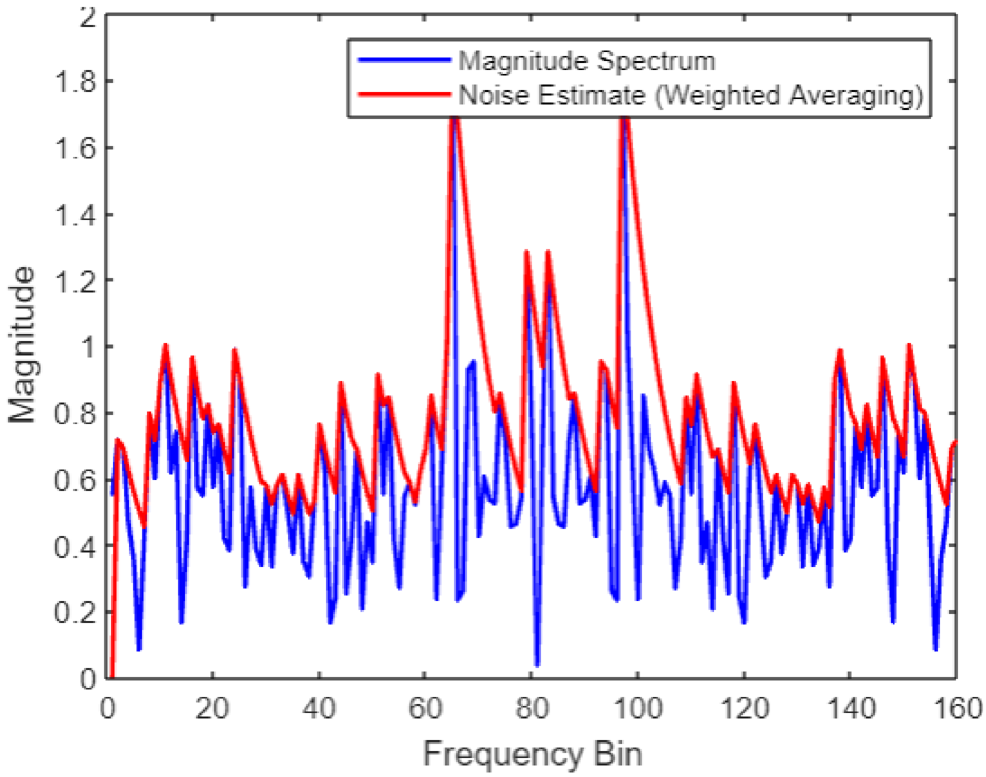

FIGURE 4.8: Noise spectrum and estimated noise spectrum using weighted average technique

4.2.4 Comparison with continuous minima tracking algorithm

When the power of the noisy speech signal goes up, our current method increases the noise estimate. If the speech gets even louder, this method might wrongly increase the noise estimate, hurting speech quality. This is because it only relies on noisy speech power to estimate noise, which isn't always accurate. To fix this, our new method does things differently. It uses the ratio between noisy speech power and a local minimum. When the speech power goes beyond a certain threshold based on this ratio, our method keeps the noise estimate unchanged, aiming to make noise estimation better.

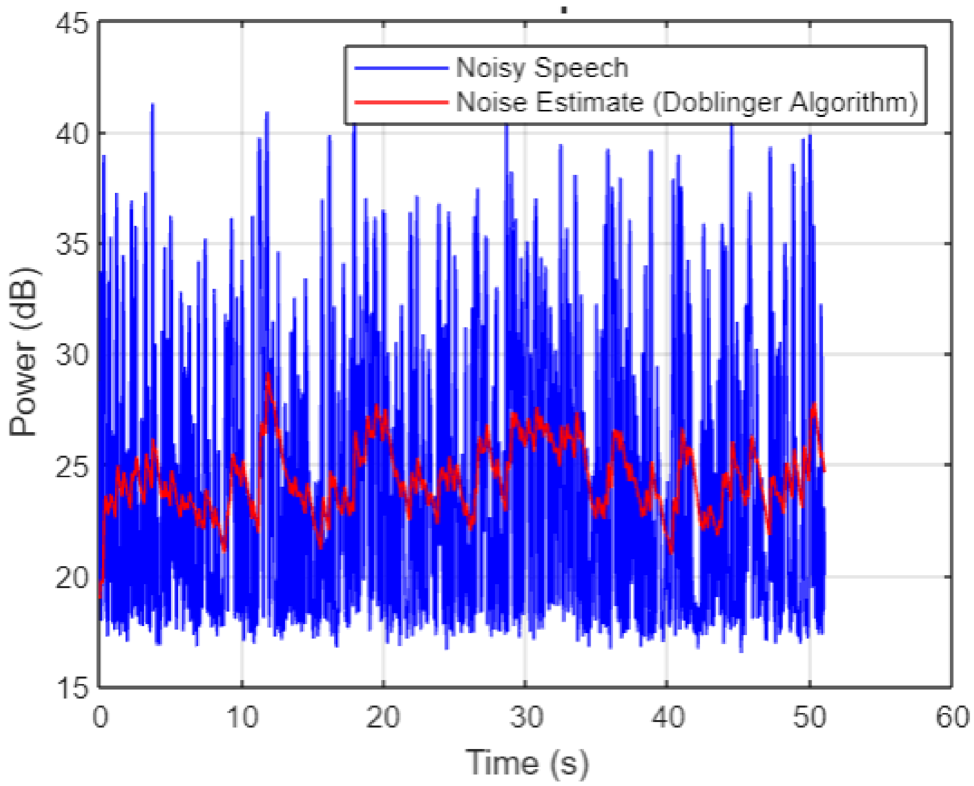

FIGURE 4.9: Noise spectrum and estimated noise spectrum using continuous minima tracking

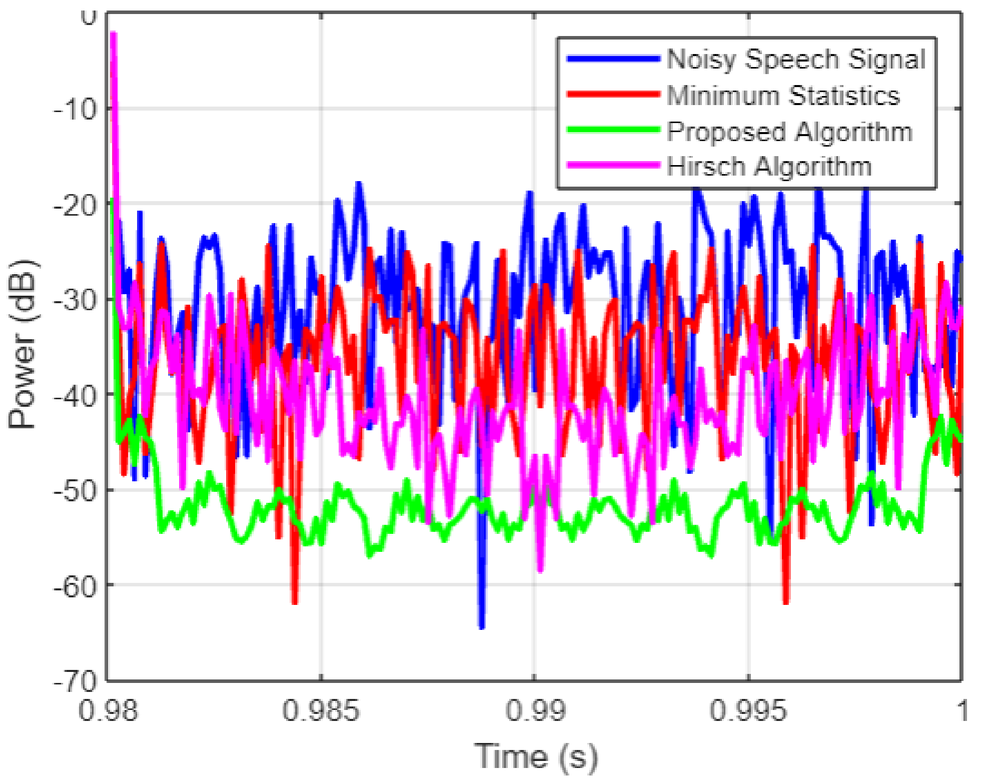

FIGURE 4.10: Overall comparison plot for estimate between proposed algorithm and existing algorithm

4.3 Mathematical evaluation

TABLE 4.1: Percentage of favorability toward proposed algorithm in comparison to other algorithms

Method	single noise preference (%)	mixed noise preference(%)
Cohen (2003)	48.8	81.7
Doblinger (1995)	53.8	81.3
Hirsch and Ehrlicher (1995)	50.0	78.8
Martin (2001)	50.8	63.8

The findings indicate that our suggested approach was equally favored as the other methods (Martin, 2001; Cohen, 2003; Doblinger, 1995; Hirsch and Ehrlicher, 1995) in handling single noise instances. However, in scenarios involving triplet noise, our method garnered superior preference ratings over all other techniques. This could be attributed to our noise estimation algorithm's rapid adaptation to highly dynamic environments that lack stationarity. This implies its effectiveness and superiority in such challenging conditions.

TABLE 4.2: Comparison of the proposed algorithm with an existing algorithm in terms of segmental SNR and MSE values

Method	SNRseg(dB)	MSE
Cohen (2003)	6.56	17.44
Doblinger (1995)	4.70	3.49
Hirsch and Ehrlicher (1995)	19.86	0.47
Martin (2001)	27.17	0.006
Proposed algorithm	26.77	0.28

Chapter 5

Conclusions

5.0.1 Conclusion

In conclusion, the proposed project addresses a critical issue in speech enhancement systems, where highly non-stationary noise environments pose a significant challenge. The proposed algorithm utilizes time and frequency dependent smoothing factors to update the noise estimate based on signal-presence probability in individual frequency bins, providing an effective solution to this problem.

The project's success is demonstrated by formal listening tests, which have shown that the proposed algorithm is superior to other noise-estimation algorithms when integrated into speech enhancement systems. This indicates that the algorithm effectively enhances the quality of speech signals in highly non-stationary noise environments, where traditional algorithms may fail to provide satisfactory results.

Furthermore, the proposed algorithm's adaptability and transferability make it a valuable tool for a range of different applications beyond speech enhancement. It has the potential to be used in a range of different systems that require noise-estimation and speech enhancement, such as hearing aids, teleconferencing systems, and voice assistants.

Overall, the proposed algorithm is a significant contribution to the field of speech enhancement and noise-estimation. It provides an effective solution to a critical issue in speech enhancement systems and has the potential to be used in a range of different applications. Further research can be conducted to investigate the algorithm's performance in different noise environments and to optimize its parameters for specific applications.

In summary, the proposed project's success lies in its ability to provide an effective solution to a critical issue in speech enhancement systems, its adaptability and transferability, and its potential for use in a range of different applications. The proposed algorithm is a significant contribution to the field of speech enhancement and noise-estimation and has the potential to impact a range of different industries positively.[8]

Bibliography

[1] R. Martin, "Spectral subtraction based on minimum statistics," *power*, vol. 6, no. 8, pp. 1182–1185, 1994.

[2] I. Cohen and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Signal Processing Letters*, vol. 9, no. 1, pp. 12–15, 2002.

[3] G. Doblinger, "Computationally efficient speech enhancement by spectral minima tracking in subbands," *Power*, vol. 1, p. 2, 1995.

[4] H.-G. Hirsch and C. Ehrlicher, "Noise estimation techniques for robust speech recognition," vol. 1, pp. 153–156, 1995.

[5] P. C. L. Sundarrajan Rangachari, "A noise-estimation algorithm for highly non-stationary environments," pp. 1–12, University of Texas at Dallas, 2005.

[6] Y. H. S. Rangachari, P.C. Loizou, "A noise estimation algorithm with rapid adaptation for highly nonstationary environments," *Science*, no. 8001808, 2004.

[7] G. Singh, "Noise estimation for real-time speech enhancement," pp. 1201–1204, 2018.

[8] V. Stahl, A. Fischer, and R. Bippus, "Quantile-based noise estimation for spectral subtraction and wiener filtering," vol. 3, pp. 1875–1878, 2000.