

Noise Estimation and Speech Signal Enhancement Using Statistical Approaches

B Sai Asish

Student VI semester

Department of Information Technology
S R K R Engineering College, S R K R Marg
Bhimavaram, AP, India

D Teja Praneeth Varma

Student VI semester

Department of Information Technology
S R K R Engineering College, S R K R Marg
Bhimavaram, AP, India

Abstract— when using digital signal processing to improve the quality of voice signals that have been weakened by noise, noise estimation and speech enhancement are crucial strategies. Speech quality is measured by how much noise is present. This work helps academicians understand speech enhancement algorithms and their implementation using MATLAB. This article offers information on how to start using MATLAB to implement algorithms and carry out research for speech enhancement. Noise estimation is critical in speech enhancement. The implementation of noise estimation techniques in MATLAB is covered in this work. Moreover, statistically supported strategies for improving speech are crucial. This research examines the use of MATLAB to implement standard statistical estimating techniques. This study discusses various databases (clean, noisy voice samples) that are accessible for speech augmentation. To assess the efficiency of various speech enhancement approaches, signal-to-noise ratio (SNR), segmental SNR (Seg SNR), and perceptual evaluation of speech quality (PESQ) are utilized. Speech enhancement applications may be found in hand-free phones, hearing aids, mobile phones, personal assistants, home automation, robots, and other technology. This study also analyses the application of real-time mathematical procedures such as Windowing, averaging, minimum mean square, variance, and Fourier transformations.

Keywords—Noise Estimation, Fft, Ift, Speech Enhancement, matlab

I. INTRODUCTION

Noise estimation and speech enhancement are two important concepts in the field of digital signal processing (DSP) that are closely related to the improvement of speech quality in various communication applications. The process of evaluating the characteristics of noise present in a voice signal is referred to as noise estimation. This may be accomplished by examining the statistical data. IEEE properties of the signal, such as its power spectrum [1] or SNR. Speech enhancement is the process of increasing the quality of speech signals that have been deteriorated by various forms of noise, such as background noise, channel[2-6] noise, or electrical noise. This can be accomplished using a variety of techniques, including filtering, spectral subtraction, and Wiener filtering.

Speech enhancement approaches rely largely on good noise estimation, as their success is dependent on their ability to predict noise characteristics properly.

Speech is now mostly overshadowed by noise in our environment. The external noise we are surrounded by distorts speech, making it uncomfortable for listeners to converse. Both environmental and human-generated noise can disrupt human communication. Best noise reduction will be provided by the speech enhancement technology for effective communication. In the field of speech processing, noise estimation and speech enhancement are important strategies for improving the quality of voice signals. The purpose of noise estimation is to precisely estimate background noise characteristics so that voice enhancement algorithms can perform better. Speech enhancement tries to remove or minimize undesired noise from voice signals while keeping speech quality.

With diverse applications, including voice recognition, hearing aids, and audio communication systems, noise estimation and speech enhancement are significant subjects in audio signal processing. MATLAB is a popular tool for implementing signal processing due to its robust signal processing features and user-friendly interface, it is able to perform noise estimation and speech enhancement methods.

For speech enhancement and noise estimation, Matlab is a common tool used in speech processing. In this work, we will utilize Matlab to develop a spectral subtraction-based noise estimation technique and a Wiener filtering and Kalman filter-based speech enhancement system. Both stationary and non-stationary sounds [7] fall within this category. As non-stationary noises have a fluctuating frequency, they are harder to analyze and get rid of than stationary noises.

The frequency domain or the time domain can both be utilized to enhance speech. Signal analysis is easy, but in the time domain, signal analysis seems a bit difficult. Estimating speech noise is a very important step in the speech enhancement process. While some methods may only be appropriate for stationary work, others may provide better results for both.

In this work, we employ a few techniques for noise estimation and noise reduction. The noises that we are going to reduce may be stationary or nonstationary.

We can say that not all the techniques are suitable for reducing both kinds of noise. The noise estimation approach will use a short-time Fourier transform (STFT) to estimate the power spectral density (PSD) of the background noise before reducing the calculated PSD from the noisy speech signal to create a noise-free signal. In the speech enhancement technique, the Wiener filter is calculated using the predicted PSD and then applied to the noisy speech signal to obtain a clean speech signal. The ratio of signal to noise (SNR) and perceptual speech quality evaluation will be used to evaluate the noise estimation and speech enhancement methods (PESQ). The techniques will be evaluated as well on practical applications speech signals contaminated by many different kinds of noise, such as white noise and babble noise.

Overall, this work will clarify how noise estimation and Speech enhancement techniques may be used to increase the quality of voice signals, and they can be implemented using Matlab.

II. LITERATURE SURVEY

A. Estimating Noise Power Spectral Density Using Minimal Statistics and Optimal Smoothing:

In this work, Rainer Martin, a senior IEEE member, provides a technique for determining in loud speaking, the power spectral density of stationary noise is measured. Because nonstationary noise is not constant over time, it will vary depending on the period, and Martin's method may be used to determine the Non-stationary noise power spectral density. Due to the absence of VAD, it is more difficult than previous strategies (voice activity detector). This approach will follow the spectral minima at all frequencies without separating the spoken and silent components. The ideal smoothing component for recursive smoothing of the voice signal with noise may be identified by minimizing a conditional mean square error estimation scenario at each time step. Using the ideal smoothing power spectral density [8-10], we may create and investigate the statistics of an unbiased noise estimator's spectral minima. Moreover, real-time deployments will function flawlessly with the estimator.

B. Robust speech enhancement using minimum controlled recursive averaging:

Israel Cohen, an IEEE member, and Brauch Begrudge published this paper. To calculate the noise, he used a minimum controlled recursive averaging technique, or MCRA [11-15]. MCRA computes noise by averaging previous spectral power measurements and varying the smoothing parameter. We may alter the likelihood of a signal being present in sub bands.. We can determine if speech is present at sub bands by comparing the local energy of speech added to noise to its minimum over a specific period of time.

The most frequent method for improving speech is to calculate noise and average the noisy speech signal over non-speech data segments. Martin has already made a noise estimate method proposal, which was on minimum statistics. The variation of this noise estimator is nearly double that of a typical noise estimator and it was quite sensitive to outliers. Moreover, this technique will occasionally weaken less energetic phonemes. Cohen, on the other hand, developed the minima controlled recursive averaging (MCRA) method for evaluating noise in this study. This method computes noise by averaging previous spectral power values and a smoothing parameter that varies based on the likelihood of signal presence in the sub bands.

C. Enhanced Natural-Sounding Residual Noise Based on Linked Time-Frequency Zones of Speech Presence:

A variety of speech augmentation techniques depend on accurate noisy power spectral density (PSD) estimates for their performance. Artifacts will appear in the enhanced speech if the noise estimate differs from the actual noise. This paper's technique is based on detecting speech presence in linked regions. Using spectral and temporal masking processes in the human auditory system, we hope to reduce artefact perception in speech presence areas while eliminating artefacts in speech absence areas. In order to do this; we suppress scaled background noise that has a natural sound when speech is present in time-frequency regions. The speech estimate's artefacts will be spectrally and temporally hidden while the background noise's naturalness is preserved by the downscaled background noise. Martin has developed a bias compensation factor with theoretical validity that depends on noise PSD estimate variances [16-17], noise-smoothed speech, and the length of the smallest search window for minimal statistics noise estimation. This allows for a low-biased noise estimate that is independent of a speech existence detector. Because our proposed speech enhancement method includes a component of connected speech existence areas, we may use a novel, simple, yet efficient bias compensation method.

D. A new voice enhancement approach has been developed that is using the stationary bionic wavelet transform and the MMSE spectral amplitude estimate:

This study presents a novel technique to speech enhancement based on the Stationary Bionic Wavelet Transform (SBWT) and the Spectral Amplitude Estimate with the Minimum Mean Square Error (MMSE). The initial step in this method is to apply the SBWT to the noisy speech signal, producing eight noisy wavelet coefficients. The MMSE [18-20] Estimate of Spectral Amplitude denoising method is used to reduce noise in each of those parameters. The enhanced speech signal is then produced by applying the SBWT, SBWT inverse, to the denoised stationary wavelet coefficients. The recommended approach's performance is demonstrated by computing the Signal-to-Noise Ratio (SNR), Segmental SNR (SSNR), and Perceptual Evaluation of Speech Quality (PESQ).

An input voice signal is typically distorted by ambient noise in many speech-related applications and must be further processed using a speech enhancement approach before being used. Speech enhancement techniques are generally classified as either supervised or unsupervised. Spectral subtraction (SS), Wiener filtering, STSA estimation, and log STSA estimation are examples of unsupervised techniques.

They develop different models for noisy and clean speech signals using a series of training challenges that include codebook-based and Hidden Markov Model (HMM)-based methodologies. Well-known techniques in this direction include modulation-domain spectral subtraction, Kalman filtering, and modulation-domain Wiener filtering.

Furthermore, when compared to the Fourier transform (FT), which only considers frequency parts, the discrete wavelet transform (DWT) expression takes into account both the chronological and the frequency features of the signal to be analyzed. The DWT is a well-known method for speech analysis. Wavelet Threshold Denoising (WTD) divides a time-domain signal into sub bands using the wavelet transform. Following that, the obtained wavelet coefficients (sub-bands) are threshold. Despite its simplicity, preliminary evaluation results show that the technique proposed in allows for higher perceptual quality input signals. It was proved that this technique may be used with a variety of well-known speech augmentation methods to provide even better outcomes.

We present a unique technique for voice improvement based on the Stationary Bionic Wavelet Transform (SBWT) and Spectral Amplitude Estimation using the Minimum Mean Square Error (MMSE) in this work.

III. PROPOSED METHOD

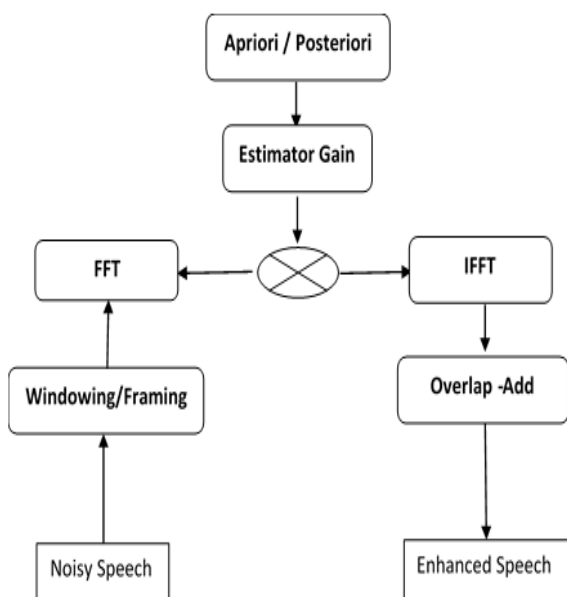


Fig. 2. Proposed method

A. Algorithm:

Step 1: Directly taking a sound recording from a database or using a microphone to convert a Sound signal to an electrical signal

Step 2: Sample the speech that is captured at a volume higher than Nyquist.

Step 3: Apply windowing and framing to the signal that was sampled.

Step 4: Apply the Fast Fourier Transform frame per frame (FFT).

Step 5: The signal is handled through frequency domain processing (SE algorithm).

Step 6: Use the IFFT, overlap, and addition approach.

Step 7. Speech quality is improved.

B. Noise estimation: Computing the SNR helps us to assess whether or not the signal contains noise (signal to noise ratio). After estimating the noise, we split the signal into frames and estimate the noise in each frame. SNR [21-22] (signal to-noise ratio): It calculates the signal- to-noise ratio's intensity. It assesses the quality of a signal in relation to the amount of noise it includes. Generally, the SNR is affected by the application as well as the signal and noise characteristics

$$SNR = 20 \log_{10}\left(\frac{A}{N}\right) \quad (1)$$

Where A denotes signal power and N denotes noise power.

This formula assumes that both A and N are expressed in the same units, such as volts or watts. The result is expressed in decibels (dB), which is a logarithmic unit that provides a convenient way to compare the relative strengths of signals.

C. *Windowing/Framing*: Windowing refers to the process of dividing an audio signal into overlapping segments, each of which is multiplied by a window function. This is done to reduce the spectral leakage caused by abrupt transitions at the edges of the frames. The window function is typically a tapered function, such as the Hamming or Hann window.

Framing, on the other hand, refers to the process of dividing an audio signal into frames of fixed duration. Each frame typically contains several 100 or 1000 samples, and 50% or more usually overlap the frames. Framing is a technique for capturing the time varying nature of speech and ensuring that any changes in speech or noise are captured within a single frame.

Windowing and framing are methods for determining the spectral properties of noise in noise estimation. The noisy voice signal may be utilized to estimate the spectral properties of the stationary noise across each frame. Overall, windowing and framing are important speech processing and enhancement approaches because they allow for exact estimate of spectral properties of speech and noise. This is required for high quality speech enhancement.

D. Noise reduction: To analyze the voice signal and determine the noise, windowing and framing are utilized. Using one of the techniques described above, the spectral characteristics of the noise are calculated from each overlapping frame of the noisy speech signal. The calculated noise is subsequently removed from the noisy voice stream, resulting in a clean speech signal. Noise is removed via spectral subtraction and filtering

E. Apriori /Posteriori SNR: Before any processing or detection, this is the signal-to-noise power ratio. It may be computed as follows

$$SNR_{apriori} = (P_s/P_n) \tag{2}$$

P_s is the signal power,
Where as P_n is the noise power.

F. Posteriori SNR: The posteriori SNR, also known as the conditional SNR, is the SNR calculated after some processing or detection has been performed. It takes into account the effect of the processing on the noise level. The formula for the posteriori SNR is:

$$SNR_{posterior} = (P_s/P_n)/(1 + N/M) \tag{3}$$

where N signifies the number of noise samples processed and M denotes the number of signal samples processed. Note that in some cases, the apriori and posteriori SNR may be the same if no processing or detection has been performed or if the processing does not affect the noise level.

G. Estimator Gain : In noise estimation, the estimator gain is used to adjust the level of the estimated noise such that it matches the actual noise level in the signal. This can be done by comparing the estimated noise with the actual noise level in a reference signal and then applying a scaling factor to the estimated noise. The estimator gain is typically a function of the signal-to-noise ratio (SNR), which is defined as the signal power to noise power ratio. The estimator gain is used in speech enhancement to increase the volume of speech in a noisy signal. This is accomplished by first calculating the signal's noise and then subtracting it from the genuine signal. Before removing the estimated noise from the original signal, the estimator gain scales it. By modifying the estimator gain, the objective is to optimum the amount of noise reduction while keeping the quality of the voice input.

A Wiener filter's gain function is defined as the product of apriori and $1 + \text{apriori SNR}$.

$$G = \text{apriori} / (1 + \text{apriori}) \tag{4}$$

G is enhanced speech

FFT: Fast Fourier Time domain to Frequency domain transformation .

The formula for computing the FFT of a windowed signal is

$$F(x[n]) = \sum_{k=0}^{N-1} (x_n) e^{-j2\pi nk/N} \tag{5}$$

Where $x[n]$ is the windowed signal

N is the length of the signal, and j is the imaginary unit.

IFFT: Inverse Fast Fourier Frequency domain to time domain transformation.

$$IFFT(X) = \frac{1}{N} \text{conj}(FFT(\text{conj}(X))) \tag{6}$$

H. Overlap-add method: Long signals can be divided into smaller parts for easier processing using the overlap-add approach.

I Enhanced speech: Enhancing the clarity and quality of voice signals using various signal processing techniques is known as enhanced speech in the field of signal processing . These techniques aim to reduce noise, reverberation, and other distortions that can affect the clarity and understand ability of speech.

IV. EXPERIMENTS & RESULTS

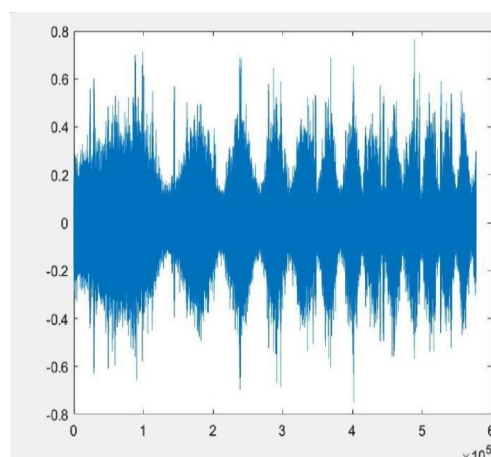


Fig 2: Input noisy speech signal

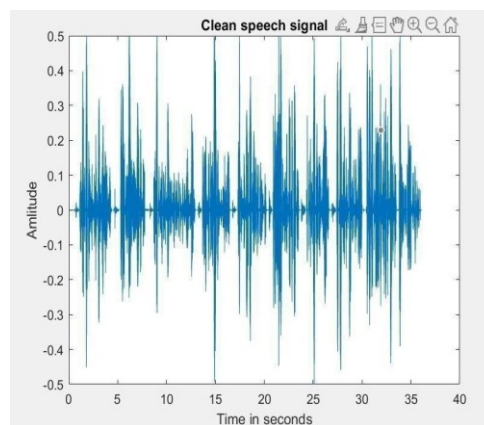


Fig 3: Output Enhanced speech signal

Noise estimation and speech enhancement are essential techniques in audio signal processing, particularly in speech communication systems. Noise estimation allows for accurate estimation of the noise present in an audio signal, which is then used in speech enhancement algorithms to remove noise from the signal and improve its quality.

Various methods have been proposed for noise estimation, including statistical methods such as spectral subtraction and Wiener filtering, as well as model-based methods such as Kalman filtering and subspace methods. Each approach has benefits and disadvantages, and the way to choose is determined by the individual application as well as the noise characteristics in the signal. Voice improvement algorithms, on the other hand, aim to improve voice quality by eliminating noise and other unwanted distortions. Among the popular methods for speech enhancement are spectral subtraction, Wiener filtering, and non-negative matrix factorization.

V. CONCLUSION

In conclusion, noise estimation and speech enhancement are critical techniques for improving the quality of audio signals, particularly in speech communication systems. These techniques are continually evolving, and new methods and algorithms are being developed to overcome the challenges posed by different types of noise and signal distortions.

A noisy speech signal is suppressed using speech enhancement. Voice recognition, hearing aids, teleconferences, and voice communication systems in noisy environments are just a few of the numerous applications for improved speech processing. The expected outcome of this research work is enhanced speech.

VI. FUTURE ENHANCEMENT

Multimodal approaches: The use of multiple modalities can benefit both noise estimation and speech enhancement. For example, one could use video information to enhance speech in noisy videos or use additional sensors to estimate noise in complex environments.

•**Robustness to different noise types:** Noise may take numerous forms, including white noise, colored noise, and impulsive noise. One could explore ways of making noise estimation and speech enhancement algorithms more robust to different types of noise.

•**Real-time implementation:** For many practical applications, real-time implementation of noise estimation and voice enhancement algorithms is crucial. One could explore ways of optimizing the algorithms for real-time implementation on embedded systems or developing hardware accelerators for faster processing.

•**Deep learning-based approaches:** Deep learning has yielded promising results in a variety of speech processing applications. Deep learning models could be used in the future to estimate noise and improve speech.

•**Dataset creation:** The availability of large and diverse datasets is crucial for the development and evaluation of noise estimation and speech enhancement techniques. Future work could focus on creating large and diverse datasets that cover different noise types, environments, and speakers.

•**Evaluation metrics:** There is a need for standardized evaluation metrics to compare different noise estimation and speech enhancement techniques. Future work could focus on developing evaluation metrics that are more representative of real-world scenarios.

•**Robustness to reverberation:** Reverberation is another challenge that affects speech enhancement. Future work could focus on developing techniques that are robust to reverberation.

REFERENCES

- [1] Martin R. "Noise Power Spectral Density Estimation based on Optimal Smoothing and Minimum Statistics", in *IEEE Transactions on Speech and Audio Processing*, pp. 50451, 2001.
- [2] Cohen L, "Noise estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement" in *IEEE Signal Processing Letters*, pp. 13-15, 2002.
- [3] Cohen I, "Noise Spectrum Estimation in Adverse Environments: Improve Minima Controlled Recursive Averaging" in *IEEE Transactions on Speech and Audio Processing*, pp. 466-475, 2003.
- [4] Philipos C. Loizou, "Speech Enhancement based on Perceptually Motivated Bayesian Estimators of the Magnitude Spectrum" in *IEEE Transactions on Speech and Audio Processing*, pp. 857-869, 2005.
- [5] Ephraim, Y. and Malah, D, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator." *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1109-1121, 1984.
- [6] Ephraim, Y. and Malah, "Speech enhancement using a minimum mean-square error log spectral amplitude estimator." *IEEE Trans. Acoust., Speech, Signal Processing*. 443-445, 1985
- [7] S. Rangachari and Loizou "An approach for estimating noise in very non-stationary situations." *Speech Communication*, 220-231, 2006
- [8] Hu, Y. and Loizou, P. (2007). "Subjective evaluation and comparison of speech enhancement algorithms," *Speech Communication*, 49, 588-601.
- [9] Ravi Kumar Kandagatla, P. V. Subbaiah "Speech Enhancement using MMSE estimation under phase uncertainty" in *International Journal of Speech Technology*, vol.20, 373-385, 2017.
- [10] Ravi Kumar Kandagatla, P. V. Subbaiah "Speech Enhancement using MMSE estimation of amplitude and complex speech spectral coefficients under phase uncertainty" in *Speech Communication*, vol.96, 10-27, 2018.
- [11] Hu, Y. and Loizou, P. (2007). "Subjective evaluation and comparison of speech enhancement algorithms," *Speech Communication*, 49, 588-601.
- [12] P. V. Subbaiah, Ravi Kumar Kandagatla "Speech Enhancement Using MMSE Amplitude Estimation and Speech Enhancement Using MMSE Amplitude Estimation and Complex Speech Spectral Coefficients Under Phase Uncertainty," *Speech Communication*, vol.96, 10-27, 2018.
- [13] Manjit Kaur. "A New Speech Enhancement Technique Based on Stationary Bionic Wavelet Transform and MMSE Estimate of Spectral Amplitude" in *HINDAWI*, 24 Dec 2021
- [14] Wei Xue's. "Neural Kalman Filtering for speech enhancement" in *ICASSP*, 13 June 2021.
- [15] Sravanthi Kantamaneni, "Speech enhancement with noise estimation and filtration using deep learning models" 12 August 2022.
- [16] Sekhar BVDS, Jagadev AK "Efficient Alzheimer's disease detection using deep learning technique", *Soft Comp*, 27, 9143-9150. 2023
- [17] Sekhar BVDS, Pamula Udayaraju, N Udaya Kumar, K Bala Sinduri, B Ramakrishna, BSSV Babu, MSSS Srinivas "Artificial neural network-based secured communication strategy for vehicular ad hoc network". *Soft Comp*, 27(1):297-309. 2022
- [18] Sekhar BVDS, Bh VS Raju, N Udaya Kumar, VVSSS Chakravarthy, "Sustainable and reliable healthcare automation and digitization using deep learning technologies". *J of Sci and Ind Res* 23:226-231. 2023
- [19] Sekhar BVDS, Reddy PP, Varma G, "Performance of secure and robust watermarking using evolutionary computing technique". *JGIM* 25(4):61-79. 2015

- [20] Sekhar BVDS, Reddy PP, Varma G, "Novel technique of image denoising using adaptive haar wavelet transformation". Irecos 10(10):1012–1017.
- [21] Udayaraju, P., Jeyanthi, P. & Sekhar, B.V.D.S, "A hybrid multi layered classification model with VGG-19 net for retinal diseases using optical coherence tomography images". Soft Comp 27, 12559–12570.
- [22] Sekhar BVDS, "Image denoising using novel social grouping optimization algorithm with transform domain technique". Int J Nat Comput Res 8(4):28–40. 2019