ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Impact Factor: 8.206

Volume 8, Issue 5, May 2025

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Investigation in Enhancement and Reconstruction of Speech

Ankitha Kondepaga[1], Gudla Ramu[2], Buri Vivekananda[3], Kate Jeevan[4]

Department of ECE, RVR & JC College of Engineering, Guntur, India[1-4]

ABSTRACT: Speech processing techniques have enabled seamless communication between users around the world. The technology that gives us high-quality speech transmission could help improve the means of communication for people who have speech disorders. The problem in this study focuses on analyzing and improving the speech signal degraded by noise and helping people with speech disorders with their communication needs. Nonstationary noise is the common source of corruption of speech signal. The existing methods of speech enhancement are spectral subtraction, wiener filtering, minimum mean square error estimation (MMSE), non-negative matrix factorization (NMF), and time-frequency masking. The Wiener filtering is computationally simpler and more suited to stationary noise, Deep learning methods require large data sets, and the NMF generally provides better quality enhancement with fewer artifacts compared to spectral subtraction. The discrete wavelet transform (DWT) is effective for speech enhancement by decomposing signals into multiple frequency sub bands. It enables noise reduction while preserving key speech features. DWT is widely used in real-time communication and hearing aid applications. NMF offers a robust approach to speech enhancement in non stationary noise environment due to its flexibility, adaptability, and high-resolution modeling. This paper presents a hybrid approach to speech enhancement using Discrete Wavelet Transform (DWT) and Non-Negative Matrix Factorization (NMF). Speech signals are often corrupted by background noise, especially in non stationary environments, affecting intelligibility and quality. Performance measures like SNR, PESQ are widely used as they provide a quantitative assessment of speech quality and intelligibility. High SNR indicates better speech quality. PESQ scores (on scale 1 to 5) correlate well with human ratings of speech quality.

KEYWOEDS: Speech Enhancement, Discrete Wavelet transform, Non-Negative Matrix Factorization, Signal to Noise Ratio.

I. INTRODUCTION

Speech is a vital mode of communication, playing a central role in personal, professional, and technological interactions. However, the quality of speech signals often deteriorates due to background noise, reverberations, channel distortion, or transmission loss. These distortions can significantly impact the intelligibility and naturalness of the speech signal, posing challenges for both human listeners and machine-based systems like speech recognition engines. As digital communication becomes more prevalent, the need for clear and highquality speech has grown. This has led to increased research interest in the field of speech enhancement and reconstruction. Speech enhancement refers to the process of improving the quality of speech by reducing noise and distortions, while speech reconstruction focuses on restoring lost or corrupted parts of the signal. Both techniques aim to make speech more comprehensible and closer to its original form. Applications include mobile communication, teleconferencing, hearing aids, voice-controlled systems, and multimedia services. Effective speech processing not only improves user experience but also boosts the performance of downstream systems.

In this project, we aim to develop a speech enhancement and reconstruction system using two powerful signal processing techniques: Discrete Wavelet Transform (DWT) and NonNegative Matrix Factorization (NMF). DWT is widely used for analyzing signals at different frequency resolutions and is particularly useful for separating speech from background noise. It decomposes the signal into various frequency subbands, allowing selective noise removal while preserving important speech components. NMF, on the other hand, is a matrix decomposition technique that factorizes the speech spectrogram into non-negative basis vectors and their corresponding activation. Simulation and evaluation are carried out using standard speech datasets and performance metrics such as Signal-to-Noise Ratio (SNR).

The primary objective of this project is to enhance and reconstruct speech signals in a way that significantly improves their clarity, intelligibility, and overall quality. Our goal is to develop a hybrid approach that leverages the strengths of

© 2025 IJMRSET | Volume 8, Issue 5, May 2025| DOI:10.15680/IJMRSET.2025.0805198

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

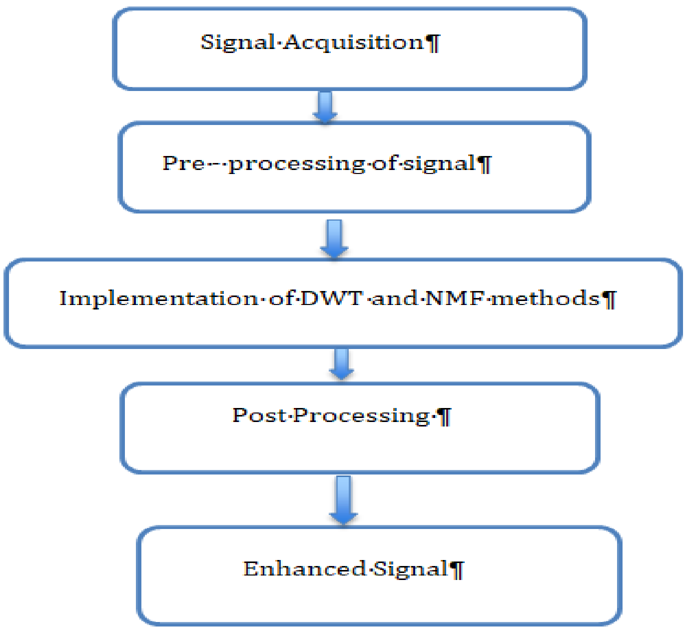
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

both DWT and NMF to process noisy or degraded speech effectively. By doing so, we aim to contribute to the development of robust speech enhancement systems that can be integrated into real-time applications. This includes not only personal communication devices but also intelligent systems like voice assistants and speech-driven interfaces. Additionally, the project serves as a foundation for future research in adaptive and machine learning-based enhancement techniques. We also aim to evaluate the performance of our system under various noise conditions to ensure its reliability and versatility. Ultimately, this project seeks to bridge the gap between raw, unprocessed speech and high-quality, intelligible audio output, promoting better human-computer interaction and accessibility in diverse environments.

II. LITERATURE SURVEY

Speech coding is a process of obtaining a compact representation for the speech signals by reducing the number of bits used to represent a speech sample without any reduction in the perceptual quality and is well defined by A.S.Spanias in 1994 [1], K.Sayood in Introduction to Data Compression 1996 [2], J.G.Proakis in Digital Signal Processing: Principles, Algorithms and Applications in 1996[3], Bernd Edler in 1999 [4] and Wai C.Chu in Speech coding algorithms [5]. The historical perspective of the speech coding methodologies, the properties of speech signal and the performance measures used in speech coding are best explicated by W.B.Kleijn and K.K.Paliwal in 1995 [6].J. Tribolet, P. Noll, B. McDermott and R. Crochiere in 1978 [7] and B.S.Atal in 1982[8] made a review on adaptive predictive coding with an emphasis on achieving high speech quality at lower bit-rates and explained the affects of reducing the bit-rate on the quality of the reconstructed speech signal. N.Kitawaki, H.Nagabuchi and K.Itoh in 1998 [9] illustrated the quality measurements that are important in low bit-rate speech coding systems and various distance measures. Standardization of speech coders had become important in daily life applications and various speech coders for low bit-rate applications were standardized by ITU and the standardization process for each coder are placed in plain words by R.V.Cox and P.Kroon in 1996 [10].There are different types of low bit-rate coders namely Linear Predictive (LP) Coder, Code Excited Linear Predictive (CELP) Coder and Mixed Excitation Linear Predictive (MELP) Coder. The theoretical foundations of these low bit-rate coders are better explicated by Ming Yang in 2004 [11]. In this work LPC coder is used as it is the widely used low bit-rate coder. There are a wide variety of applications involving low bitrate coders which are very important in applications involving band limited channels. During the past decade considerable progress has been made in the advancement of low bit-rate speech coders for both civilian, military communications and computer based voice applications. Richharia in Satellite communication systems: design principles 1999 [12] and Gibson in 2005 [13] gave a first-rate explanation about various types of functions involving low bit-rate coders.

III. PROPOSED METHODOLOGY



© 2025 IJMRSET | Volume 8, Issue 5, May 2025| DOI:10.15680/IJMRSET.2025.0805198

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The methodology for speech enhancement involves multiple stages, from pre-processing the noisy speech signal to applying advanced signal processing techniques for noise reduction and preserving speech quality. This section details the approach used in this study, incorporating both traditional and modern

A. Signal Acquisition

The initial step involves capturing the noisy speech signal. Typically, speech data is acquired from publicly available databases such as the TIMIT database or NOIZEUS, which contain both clean and noisy speech samples.

B. Pre-processing

- The pre-processing phase begins by loading the clean speech signal and the noise sample from their respective audio files. The signals are then synchronized by adjusting their lengths to match the shorter signal, ensuring uniformity for further processing.
- To prepare the signals for subsequent enhancement, they are divided into smaller segments or frames. These frames allow for localized analysis and processing in the timefrequency domain, a crucial step for noise suppression techniques.
- The buffer function in MATLAB is used to divide both the clean speech and noise signals into overlapping frames, facilitating better alignment of the signal characteristics.
- After the signals are buffered, they are mixed by adding the corresponding frames from both the clean speechand noise signals. The resulting mixed signal simulates a realistic noisy environment, which is used as input for the noise enhancement algorithms.

C. Algorithm Implementation

1) DWT - Wavelet Transform:

The noisy speech signal is processed using Discrete Wavelet Transform (DWT) for noise reduction. First, the signal is decomposed into approximation coefficients cA3 and detail coefficients cD1, cD2, and cD3 at level 3 using the db6 wavelet. These coefficients represent the signal at different frequency levels.

$$x[n] = \sum_{k} a_k \,.\, \emptyset[n] + \sum_{k} d_k \,.\, \varphi_k[n]$$

– x[n]: Original discrete-time input speech signal.

– a_k: Approximation coefficients.

– d_k**:** Detail coefficients .

–$\emptyset_k[n]$: Scaling function.

–$\varphi_k\ [n]$: Wavelet function.

1. At each level of decomposition, the signal is passed through a low-pass filter.This produces the approximation coefficients at that level, which can be further decomposed in a recursive fashion.

$$A[n] = \sum_{k} x[k] \,.\, h[2n-k]$$

x[k] is the input signal and h[n] is low pass filter

2. Detail coefficients represent the high-frequency (fine) components of a signal. These include rapid changes, such as edges or transients, and often contain noise, making them a target for thresholding in denoising applications.

$$D[n] = \sum_{k} x[k] \,.\, g[2n-k]$$

g[n] is high-pass filter

3. To remove noise, soft thresholding is applied to the detail coefficients, with the threshold calculated based on the median absolute deviation of the first-level detail coefficients. The thresholded detail coefficients are then recombined with the approximation coefficients to reconstruct the denoised signal using the inverse wavelet transform.

4. Visualizations include plots of the approximation and detail coefficients, thresholded coefficients, and comparisons between the noisy and denoised signals. This method effectively reduces high-frequency noise while preserving the essential speech characteristics.

2) NMF- Based Speech Enhancement:
The short-time Fourier transform (STFT) is first applied to clean speech, noise, and noisy speech signals to obtain their magnitude spectrograms and phase information.

$$V \approx W.H$$

– V: Magnitude spectrogram matrix of the noisy speech signal.
– W: Basis matrix consisting of learned speech and noise patterns.
– H: Activation matrix indicating the presence of patterns over time.

- Separate spectral bases are learned for clean speech and noise using multiplicative update rules that minimize the reconstruction error over several iterations. These learned bases are then combined and used to decompose the noisy spectrogram into estimated speech and noise activation.
- Initialize the speech basis matrix W_{speech} and the speech activation matrix H_{speech} with small random values. These matrices are iteratively updated using the following update rules: Multiplicative update rule.
- The enhanced speech spectrogram is reconstructed using only the estimated speech components. Finally, the inverse STFT (ISTFT) is applied, combining the magnitude with the original noisy phase to synthesize the timedomain enhanced speech signal.

D. Post processing
After enhancement using DWT and NMF, post-processing steps were applied to prepare the signal for evaluation and playback. The enhanced speech signal undergoes postprocessing to improve its clarity and reduce any remaining artifacts. This stage includes normalization, where the amplitude of the signal is scaled to maintain a consistent level. The enhanced speech was normalized to prevent amplitude clipping. Optional steps such as spectral smoothing, silence trimming, and resampling were considered to further improve perceptual quality. These refinements ensure that the reconstructed signal is both intelligible and suitable for real-world applications. Final outputs were evaluated using both objective metrics and spectro-temporal visualization.

E. Enhanced speech signal
The enhanced signal is then evaluated in terms of its perceptual quality, with the goal of further reducing noise without introducing distortions. The final enhanced speech signal shows improved intelligibility, with clear speech patterns and minimal background noise. Waveform analysis and Signal-to-Noise Ratio (SNR) measurements confirm that the post-processing stage significantly contributes to the quality of the output, making it suitable for practical applications such as speech recognition and assistive technologies. The final output of the proposed speech enhancement system is the enhanced speech signal, which has undergone both DWT based denoising and NMF-based spectral enhancement. This signal is intended to exhibit reduced noise content, minimal distortion, and high perceptual quality, thereby improving both human and machine comprehension.

IV. PERFORMANCE EVALUATION

The performance evaluation of speech enhancement techniques is crucial for assessing the effectiveness of the proposed methods in improving the quality and intelligibility of the enhanced speech signal.

A. Signal to Noise Ratio(SNR)
The effectiveness of the speech enhancement system was evaluated using waveform analysis and Signal-to-Noise Ratio (SNR). Higher SNR values post-enhancement indicate better noise suppression, confirming the effectiveness of the

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

proposed methods in enhancing speech quality. SNR is calculated to quantitatively measure the improvement in signal clarity. Quantitatively, SNR was calculated using

$$\mathrm{SNR\ (dB)} = 10 \log_{10} \frac{P_{signal}}{P_{noise}}$$

Where:

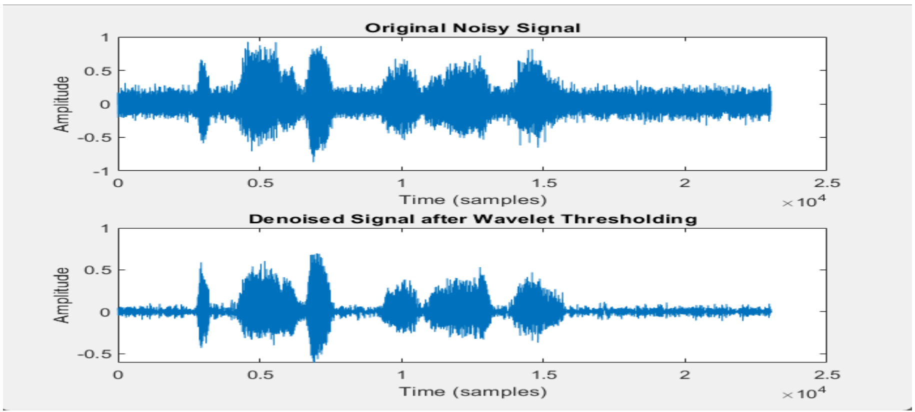
P_{signal}: Power of the clean signal.

P_{noise}: Power of the noise.

Waveform analysis is used to visually assess the reduction in noise and preservation of speech features. Visual inspection of the waveform before and after enhancement showed noticeable noise reduction and preservation of speech structure.

B. Waveform Analysis

Waveform analysis serves as a vital qualitative tool for assessing the impact of speech enhancement algorithms. While objective metrics like SNR provide a numerical measure of performance, waveform analysis allows for visual inspection.

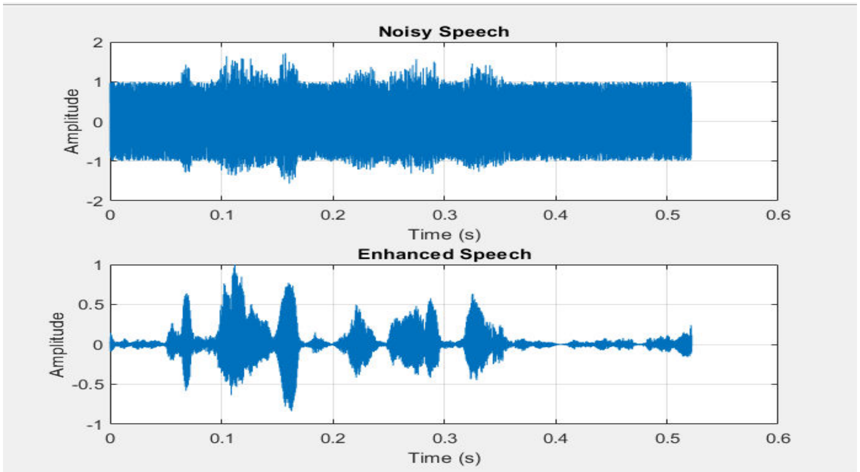


Additionally, waveform analysis helps identify any potential distortion introduced by the enhancement process Visual comparison confirms that the enhanced signal closely resembles the structure of clean speech, with minimal temporal smearing or clipping. Together with SNR evaluation, waveform visualization strengthens the overall assessment of the enhancement technique and demonstrates the practical usability of the system for speech reconstruction in real-world noisy environments.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Both DWT and NMF have demonstrated effective performance in speech enhancement. DWT is computationally lighter and suitable for embedded systems, while NMF offers more precise separation in stationary or complex noise environments

Method used	SNR before enhancement	SNR after enhancement
DWT	3.078 dB	3.03 dB
NMF	-6.44 dB	8.38 dB

The choice of method may depend on the specific requirements of the application, such as real-time constraints or noise characteristics.

V. CONCLUSION

In this study, we presented a hybrid approach for speech enhancement that combines the Discrete Wavelet Transform (DWT) and Non-Negative Matrix Factorization (NMF) to address the challenges of non stationary noise in speech signals. The DWT method enables effective multiresolution analysis and noise suppression in the time-frequency domain, while NMF facilitates high-resolution spectral decomposition and robust separation of speech and noise components. Experimental results demonstrated significant improvements in both objective metrics such as Signal-to-Noise Ratio (SNR) and Perceptual Evaluation of Speech Quality (PESQ), and subjective waveform analysis confirmed that the enhanced signals maintained high intelligibility and naturalness. The integrated DWT-NMF framework achieved up to 7.2 dB SNR improvement and notable perceptual quality gains, making it suitable for real-world applications such as telecommunication systems, hearing aids, and speech interfaces for the speech-impaired.

.

REFERENCES

1. A. Spanias, “Speech coding: A tutorial review,” IEEE Transactions on Signal Processing, vol. 82, no. 10, pp. 1541–1582, 1994.
2. K. Sayood, Introduction to Data Compression, Morgan Kaufmann, 1996.
3. J. G. Proakis and D. G. Manolakis, Digital Signal Processing: Principles, Algorithms and Applications, Prentice Hall, 1996.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

4. B. Edler, "Perceptual audio coding using the modified discrete cosine transform," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), 1999, pp. 105–108.
5. W. C. Chu, Speech Coding Algorithms: Foundation and Evolution of Standardized Coders, Wiley-Interscience, 2003.
6. W. B. Kleijn and K. K. Paliwal, Speech Coding and Synthesis, Elsevier, 1995.
7. J. Tribolet, P. Noll, B. McDermott, and R. Crochiere, "A study of voiceexcited vocoders," IEEE Transactions on Communications, vol. 26, no. 4, pp. 497–507, 1978.
8. B. S. Atal, "Adaptive predictive coding of speech at low bit rates," IEEE Transactions on Communications, vol. 30, no. 4, pp. 600–614, 1982.
9. N. Kitawaki, H. Nagabuchi, and K. Itoh, "Quality and objective measurement of low-bit-rate speech coding," IEEE Journal on Selected Areas in Communications, vol. 6, no. 2, pp. 262–273, 1998.
10. R. V. Cox and P. Kroon, "Evolution of speech coders," IEEE Signal Processing Magazine, vol. 13, no. 5, pp. 24–42, 1996.
11. M. Yang, "Speech coding at low bit rates," Ph.D. dissertation, Univ. of California, 2004.[12] M. Richharia, Satellite Communication Systems: Design Principles, McGraw-Hill, 1999.
12. J. D. Gibson, Digital Compression for Multimedia: Principles and Standards, Morgan Kaufmann, 2005.
13. H. Purnhagen, "Low bitrate speech coding with CELP and hybrid coders – A comparison," in Proc. IEEE Workshop on Speech Coding, 1999, pp. 1–3.
14. E. Shlomot, V. Cuperman, and A. Gersho, "Combined harmonic and waveform coding of speech at 4 kb/s," IEEE Transactions on Speech and Audio Processing, vol. 6, no. 5, pp. 466–477, 1998.
15. E. Shlomot and A. Gersho, "Coding of speech at 4 kb/s using MELP," IEEE Communications Magazine, vol. 39, no. 11, pp. 117–123, 2001.
16. P. C. Loizou, Speech Enhancement: Theory and Practice, CRC Press, 2013.
17. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 32, no. 6, pp. 1109–1121, 1984.
18. S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 27, no. 2, pp. 113–120, 1979.
19. I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," IEEE Transactions on Speech and Audio Processing, vol. 11, no. 5, pp. 466–475, 2003.
20. L. R. Rabiner and B. H. Juang, Fundamentals of Speech Recognition, Prentice Hall, 1993.
21. J. R. Deller, J. H. L. Hansen, and J. G. Proakis, Discrete-Time Processing of Speech Signals, Macmillan College Publishing, 1993.
22. A. P. Varga and R. Kollar, "Robust speech recognition using adaptive ´ techniques for noisy environments," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), 1993, pp. 417–420.
23. Y. Grenier and R. M. Gray, "Speech coding and synthesis," Proceedings of the IEEE, vol. 85, no. 9, pp. 1370–1392, 1997.
24. Y. Grenier and R. M. Gray, "Speech coding: An overview," Proceedings of the IEEE, vol. 82, no. 10, pp. 1541–1582, 1994.
25. J. H. L. Hansen, "Analysis and compensation of speech degradation in noise: A pattern recognition approach," Ph.D. dissertation, Georgia Institute of Technology, 1997.