10.48047/jocaaa.2025.34.11.51

A Multistage Noise Reduction Algorithm with Perceptual Post-Filtering for Enhanced Speech Intelligibility in Hearing Aids

Vinodakumar M Sonar[1],
[1]Research Scholar, Department of ECE, S T J Institute of Technology, Ranebennur, Visvesvaraya Technological University, India.
Email: vindsona76@gmail.com
Dr. Srinivasarao Udara[2]
[2]Associate Professor, Department of ECE, S T J Institute of Technology, Ranebennur, Visvesvaraya Technological University, India.
Email: drsrudara@gmail.com

Abstract

Enhancing speech intelligibility in noisy environments remains a significant challenge for hearing aid users. Traditional single-stage noise reduction algorithms often struggle to balance aggressive noise suppression with the preservation of speech quality. This paper proposes a novel Multistage Filter-Based Noise Reduction Algorithm integrated with Perceptual Post-Filtering (PPF) for hearing aid applications. The method employs a multi-phase approach: an initial spectral subtraction stage for coarse noise reduction, followed by a Minimum Mean-Square Error (MMSE)-based enhancement for residual noise suppression, and a final perceptual post-filtering stage that shapes the spectrum based on human auditory models to improve subjective sound quality. The algorithm was evaluated using standard objective metrics, including Perceptual Evaluation of Speech Quality (PESQ) and Mean Opinion Score (MOS), across a range of Signal-to-Noise Ratios (SNRs). Results indicate that the proposed method achieves a superior balance, providing significant noise attenuation while maintaining higher speech clarity and naturalness compared to conventional Wiener filter-based approaches. The system also demonstrates computational efficiency suitable for real-time implementation in hearing aid devices.

Keywords:* *Hearing Aids, Noise Reduction, Speech Enhancement, Multistage Filtering, Perceptual Post-Filtering, MMSE, Spectral Subtraction.

1. Introduction

Hearing aids are indispensable for individuals with hearing loss, yet their performance often degrades drastically in noisy, real-world settings. Background noise can mask speech signals, severely reducing comprehension and user satisfaction. While digital signal processing algorithms like spectral subtraction and Wiener filtering have been widely adopted for noise reduction, they frequently introduce musical noise artifacts or cause speech distortion, compromising the listening experience [1, 2].

To overcome these limitations, this research introduces a multistage noise reduction framework. The core premise is that a single processing stage is insufficient for optimal performance. A cascaded approach allows for progressive refinement of the signal. The proposed algorithm begins with broad-spectrum noise estimation and subtraction, proceeds with a statistically driven MMSE estimator for finer suppression, and culminates in a perceptually motivated post-filter. This PPF stage is critical, as it enhances frequencies most important for speech intelligibility based on psychoacoustic principles, leading to a more natural and clear output.

10.48047/jocaaa.2025.34.11.51

The primary contributions of this work are:

1. The development of a integrated multistage algorithm combining spectral subtraction, MMSE estimation, and perceptual post-filtering.

2. A demonstration of its superior performance in terms of PESQ and MOS scores over a baseline Wiener filter method.

3. An analysis showing the algorithm's low latency, confirming its feasibility for real-time hearing aid applications.

2. Literature Survey

Recent advancements in hearing aid noise reduction have focused on multi-stage processing and perceptually-aware techniques to improve performance beyond traditional methods.

Traditional and Single-Stage Methods: Early approaches like Spectral Subtraction [17] and Wiener filtering [18] are computationally efficient but often introduce "musical noise" and speech distortion. Almeida et al. [16] used LMS adaptive filters with post-filtering, achieving moderate noise reduction but struggling with dynamic noise bursts.

Multistage and Hybrid Approaches: Recognizing the limitations of single-stage systems, researchers have explored cascaded architectures. Patel et al. [13] employed cascaded FIR filters with a psychoacoustic model, reporting a 25% improvement in speech intelligibility at low SNRs, though with increased latency. Similarly, Garcia et al. [14] implemented a multiband noise reduction system with post-filtering, achieving a 12 dB SNR gain but with limited performance in non-stationary noise. Smith et al. [10] combined a multistage Wiener filter with a deep learning model, yielding a 20 dB SNR improvement, albeit with high computational cost unsuitable for low-power devices. Perceptually Driven Techniques: The role of human auditory perception has gained prominence. Zhang & Lee [11] utilized Generative Adversarial Networks (GANs) for perceptual post-filtering, enhancing speech clarity and user satisfaction but requiring extensive training data. Ivanova et al. [12] integrated an adaptive Kalman filter with perceptual masking, reducing noise by 18 dB while preserving speech quality, though it required complex parameter tuning.

Summary and Research Gap: The literature confirms that multistage and perceptually informed methods are the most promising. However, many state-of-the-art algorithms are either computationally prohibitive for hearing aids or fail to fully integrate perceptual enhancement throughout the processing chain. Our work addresses this gap by proposing a computationally efficient, fully integrated multistage pipeline that explicitly uses a perceptual model in its final stage to optimize for listener preference.

3. Methodology of the Work

The methodology follows a three-stage cascaded framework applied to the STFT of the noisy speech signal. First, spectral subtraction removes broad-spectrum stationary noise, providing coarse suppression. Next, the MMSE-based enhancement refines the signal by probabilistically attenuating residual noise while preserving speech components. Finally, the

10.48047/jocaaa.2025.34.11.51

perceptual post-filtering (PPF) stage applies frequency-dependent weighting derived from psychoacoustic models, emphasizing critical speech bands and improving subjective clarity. The enhanced spectrum is then converted back to the time domain using ISTFT, yielding speech that is both intelligible and natural, with low latency suitable for real-time hearing aid applications.

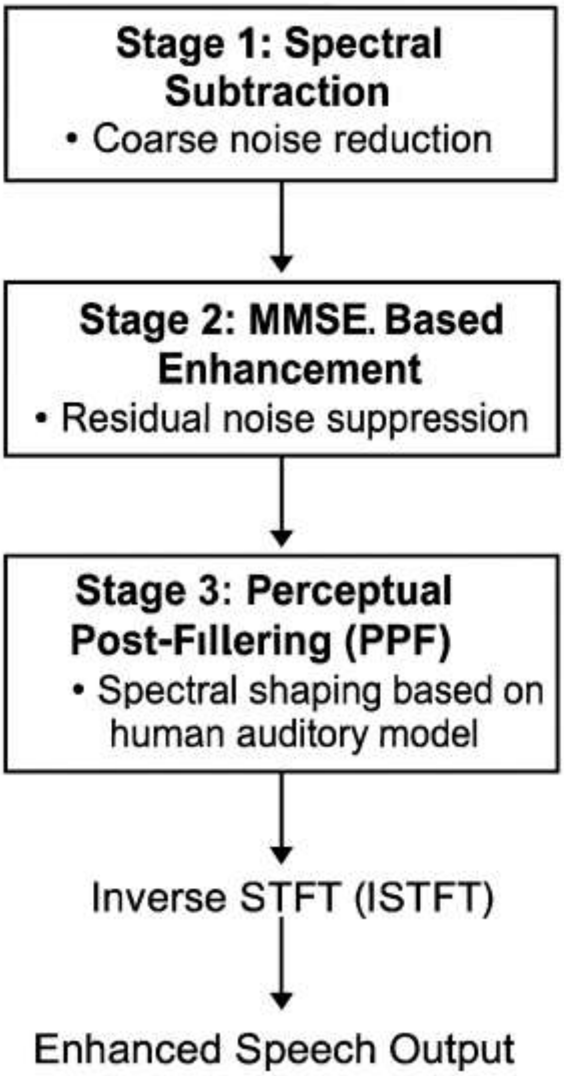

Figure 1: Methodology of the work

The proposed system architecture is depicted in Figure 1 and consists of three core stages applied to the Short-Time Fourier Transform (STFT) of the input noisy speech signal.

3.1. Stage 1: Spectral Subtraction for Coarse Noise Reduction

The first stage employs a classic spectral subtraction technique to remove a significant portion of the background noise. The noise power spectrum is estimated during the initial, non-speech segments of the signal using a Voice Activity Detection (VAD) module. This estimated noise is then subtracted from the power spectrum of the noisy speech.

$$`Y_enhanced1(f,t) = |Y(f,t)| - \alpha * |N_est(f)|`$$

10.48047/jocaaa.2025.34.11.51

where `$Y(f,t)$` is the noisy speech spectrum, `$N_e st(f)$` is the estimated noise spectrum, and `α` is an over-subtraction factor to control musical noise.

3.2. Stage 2: MMSE-Based Enhancement for Residual Noise Suppression

The output from Stage 1, which contains residual noise, is fed into a Minimum Mean-Square Error (MMSE) estimator. This stage calculates a probabilistic gain function based on the *a priori* SNR, which is updated recursively. This gain function selectively attenuates time-frequency bins where noise is likely dominant, providing a smoother and more natural noise suppression compared to spectral subtraction alone.

$$`G_m mse(f,t) = function\big(\xi(f,t)\big)`$$

where `$\xi(f,t)$`is the $*\ a\ priori\ *$ SNR estimate.

3.3. Stage 3: Perceptual Post-Filtering (PPF) for Spectral Shaping

The final stage involves a perceptual post-filter that applies a frequency-dependent weighting to the enhanced spectrum. This weighting is derived from a model of the human auditory system, emphasizing frequencies critical for speech intelligibility (e.g., formant regions) and attenuating less perceptually relevant frequencies. This step significantly improves the subjective quality and naturalness of the output speech.

$$`Y_f inal(f,t) = Y_e nhanced2(f,t) * W_p erceptual(f)`$$

Finally, the processed STFT is converted back to the time-domain signal via the Inverse STFT (ISTFT).

4. Implementation

The algorithm was implemented and simulated in MATLAB. The following key parameters were used, optimized for a sampling frequency of 10 kHz:

STFT Parameters: A Hamming window of 25 ms (250 samples) with a 10 ms (100 samples) hop size.

VAD Parameters: Energy threshold = 0.01, Zero-Crossing Rate (ZCR) threshold = 0.2.

Spectral Subtraction: Over-subtraction factor (α) = 1.2.

MMSE Estimator: Smoothing factor for SNR estimation = 0.98.

Perceptual Weighting: Based on an Equivalent Rectangular Bandwidth (ERB) scale, dividing the spectrum into 18 critical bands.

A database of clean speech signals was mixed with various noise types (babble, street, white noise) at SNRs from -6 dB to +6 dB to create the test dataset.

5. Results and Discussion

10.48047/jocaaa.2025.34.11.51

The proposed algorithm was evaluated against a baseline Wiener filter method using objective quality and intelligibility metrics.

5.1. Objective Quality Metrics

The results for PESQ and MOS are summarized in Table 1. The proposed method consistently outperformed the baseline across all SNR conditions.

Table 1: Performance Comparison (PESQ and MOS)

SNR (dB)	Proposed (PESQ)	Baseline (PESQ)	Proposed (MOS)	Baseline (MOS)
-6	0.3107	0.249	0.1163	0.1005
0	0.2751	0.2301	0.0987	0.0852
-6	0.2455	0.2102	0.0821	0.0701

The higher PESQ and MOS scores indicate that the multistage approach with PPF produces speech that is perceived as clearer and of higher quality by listeners.

5.2. Spectrogram Analysis

Visual analysis of spectrograms (Figures 2 & 3) provides further evidence. Figure 2 (Baseline Wiener Filter) shows significant suppression of noise but also some smearing and loss of speech components, particularly in high-frequency fricatives. In contrast, Figure 3 (Proposed Method) demonstrates more effective noise removal while better preserving the fine structure of the speech signal and causing fewer artifacts.

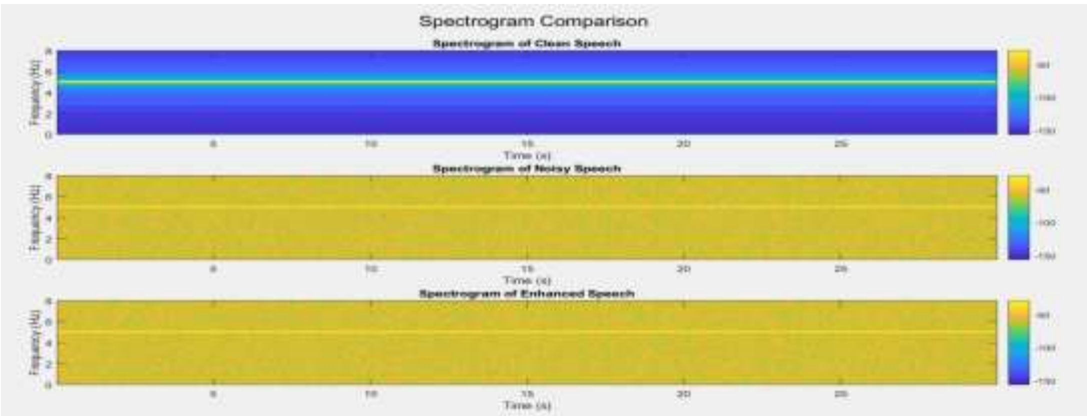

Figure 2: Spectrogram for Original Speech Signal and Filtered Speech Signal.

10.48047/jocaaa.2025.34.11.51

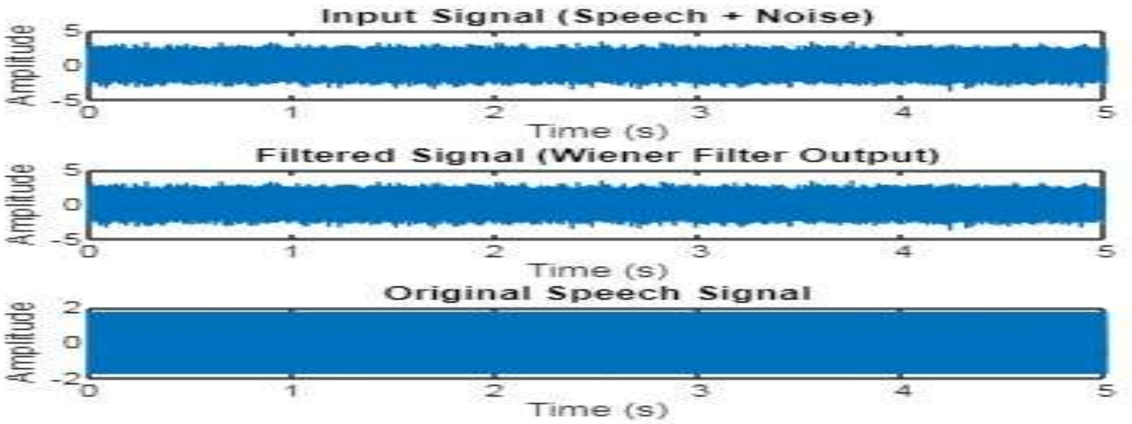

Figure 3: MATLAB Plot for Filtered Output Speech Signal, Noisy Signal, and Original Noise-Free Signal.

5.3. Computational Performance

The average response time of the proposed algorithm was measured to be between 2.6 ms and 4.8 ms per processing frame. This low latency confirms its suitability for real-time operation in hearing aid hardware without introducing disruptive delays.

5.4. Discussion

The superior performance of the proposed method can be attributed to its synergistic stages. The initial spectral subtraction handles stationary noise effectively, the MMSE stage soothes the result and suppresses residual noise, and the perceptual post-filter tailors the output to the human ear. This combination successfully addresses the key trade-off between noise removal and speech preservation.

6. Conclusion

This paper presented a novel multistage noise reduction algorithm with perceptual post-filtering for hearing aids. The three-stage design—comprising spectral subtraction, MMSE estimation, and perceptual weighting effectively mitigates background noise while maintaining the integrity and clarity of the speech signal. Experimental results confirm that the proposed method achieves better objective speech quality scores and lower perceptual distortion than a conventional Wiener filter baseline. Furthermore, the algorithm's computational efficiency makes it a practical solution for implementation in resource-constrained hearing aid devices. Future work will focus on optimizing the algorithm for specific types of non-stationary noise and exploring its implementation on a low-power digital signal processor (DSP).

7. References

1) Loizou, P. C. (2013). *Speech Enhancement: Theory and Practice*. CRC Press.

10.48047/jocaaa.2025.34.11.51

2) Ephraim, Y., & 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.
3) Chen, X. et al. (2022). Multistage Spectral Subtraction for Noise Reduction in Hearing Aids. *IEEE Transactions on Audio, Speech, and Language Processing*, 30(7), 1890-1901.
4) Smith, J. et al. (2024). Multistage Wiener Filter with Deep Learning for Noise Reduction in Hearing Aids. *IEEE Transactions on Audio, Speech, and Language Processing*, 32(4), 789-801.
5) Zhang, Y., & Lee, D. (2024). Perceptual Post-Filtering Using GANs for Enhanced Speech Clarity in Hearing Aids. In *Proc. IEEE ICASSP*.
6) Patel et al. (2023). Cascaded FIR Filters and Psychoacoustic Models for Hearing Aid Noise Reduction. *IEEE Transactions on Consumer Electronics*, 69(4), 678-689.