International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 11 , November - 2025

Speech Enhancement Filters: A Comparative Study in Diverse Noise Environments

Vighya Duhan, Ritu Boora, Manisha Jangra
(Department of EEE, Guru Jambheshwar University of Science and Technology, Hisar, India)

Abstract - Speech enhancement is a critical process aimed at improving the quality and intelligibility of speech signals that have been corrupted by various forms of noise. This is particularly vital in real-time processing for applications like automatic speech recognition (ASR) , where background noise significantly degrades system performance. A significant challenge is effectively removing noises that change their characteristics over time, which commonly occurs in real-world environments. Many traditional techniques run the risk of introducing signal distortion, leading to a loss of clarity and naturalness. A technique that aggressively removes all background noise might make the speech sound artificial or hard to understand. Conversely, a method that perfectly preserves the natural sound of a voice might leave too much distracting noise in the background. This work evaluates the effectiveness of these denoising methods in improving speech quality, intelligibility, and noise suppression. The results indicate that there is no single technique to enhance speech in a diverse noise environment. The optimal solution is always a context-aware decision based on a trade-off between competing performance goals. Subsequently, hybrid methods that use more than one type of filter have also been evaluated in the proposed work. The results indicate that NLMS and NLM-based methods often excel in noise suppression and speech intelligibility (STOI), whereas LMS and Gaussian filters frequently achieve superior perceived speech quality (PESQ). The NLMS filter is an adaptive method that dynamically adjusts its parameters, making it highly effective against non-stationary noise that changes over time.

The core conclusion drawn from the systematic evaluation of various speech enhancement methods is that no single filter method is universally optimal across all noise scenarios and performance metrics. This finding is crucial because it confirms that different denoising approaches inherently possess distinct advantages and limitations.Consequently, the process of selecting the best enhancement technique must move beyond seeking a universal best solution. Instead, the most effective outcome requires a deliberate, context-aware decision. This decision must always be based on a clear trade-off between competing performance goals and the primary objective of the intended application (e.g., perceived speech quality, intelligibility, or aggressive noise suppression).

Keywords: **Speech Enhancement, Speech Denoising, Adaptive Filters, Non-stationary Noise, Normalized Least Mean Squares (NLMS), Non-Local Means (NLM), Least Mean Squares (LMS) filter, Normalized Least Mean Squares (NLMS) filter, Gaussian filters ,Wavelet Denoising, Spectral Subtraction, Adaptive Filters**

1. INTRODUCTION

Speech enhancement [1], also known as speech denoising, is a fundamental subfield of digital signal processing that involves the estimation and recovery of a clean speech signal from a noisy signal [2]. The primary objective is to improve perceptual and quantitative attributes of the speech signal [3]. Principally its quality i.e. the naturalness and pleasantness of the sound and its intelligibility i.e. the clarity and understanding ability of the spoken content, which are often degraded by the presence of noise. The characteristics of noise can range from stationary to highly unpredictable, making the denoising process a challenge. For example, noise from a busy restaurant, a busy street with unexpected sounds, home appliances such as a mixer, TV, air conditioner, etc, add a diverse set of noises to the speech signal. This diversity makes it difficult for a single, fixed method to consistently separate speech from the background clutter. An algorithm optimized for a consistent, stationary hum will likely fail in the chaotic, variable environment of a restaurant, and vice-versa [4]. This forces a constant struggle between creating specialized solutions for certain problems versus developing robust, general-purpose algorithms that can handle a wide variety of unforeseen noise types.

A significant and generalized difficulty is that the very process of trying to remove noise can damage the signal or introduce new, unwanted sounds. The sources identify two main forms of this issue: Signal distortion and Signal degradation. While more sophisticated algorithms are developed to overcome the limitations of simpler ones, they often introduce their own set of generalized difficulties [5]. To mitigate these issues, modifications to NLM adaptive filtering have been proposed, such as non-uniformly smoothing weight values through energy-dependent scaling, which helps address the time-varying nature of speech[6].

Hybrid techniques, like Wavelet-Nonlocal Means (W-NLM), have also been developed to combine the strengths of NLM with other methods, aiming to minimize individual limitations [7,8].

This study aims to systematically evaluate several conventional and advanced digital filters designed for speech enhancement. By testing these filters against a diverse set of real-world and synthetic noise types, this study seeks to identify their strengths and weaknesses, offering practical recommendations for their application based on specific performance objectives. The evaluation uses standard objective metrics to provide an enhanced comparison of the filter performance based on clarity, naturalness, and intelligibility of the speech. This paper is further organised as follows: Section 2 discusses the Literature, Section 3 presents the Datasets used, and denoising methods are presented in Section 4. The performance of the filters is evaluated in Section 5.

2. LITERATURE

Noise suppression remains a critical challenge in speech processing, where signals are often degraded by Gaussian and impulsive noise [9]. Traditional filters offer simple yet effective solutions for many applications. These include Spectral Subtraction, a simple and widely used method that estimates the noise power spectrum and subtracts it from the noisy speech spectrum [10]. Other traditional methods include Moving Average (MA)[11,12], Gaussian filter[14], . Savitzky-Golay[15] filter and many more. Moving Average smooths signals by averaging over a window, effectively reducing high-frequency and Gaussian noise [13]. Gaussian Filter [14] provides isotropic smoothing and is widely used in image and speech denoising, though it struggles with impulsive disturbances. Savitzky-Golay [15] filter improves upon these methods by fitting local polynomials, preserving important signal features such as peaks and formants. However, its sensitivity to outliers limits its robustness under impulsive noise conditions. Adaptive filtering methods that could dynamically adjust to changing noise, like Least Mean Squares (LMS) [16,17] algorithm adjusts filter weights iteratively to minimize error. LMS[16,17] is computationally efficient but suffers from slow convergence. Normalized Least Mean Squares (NLMS)[18] improve convergence speed and stability by normalizing or weighting step sizes, making them effective for non-stationary environments and their derivatives, have continuously evolved over time. Hybrid methods were also developed, combining the strengths of different techniques, such as Wavelet-Nonlocal Means (W-NLM)[19] or NLM combined with Spectral Subtraction[20], to minimize individual limitations. Spectral subtraction[20] remains a cornerstone for speech denoising, where the estimated noise spectrum is subtracted from the noisy signal. Though effective for Gaussian noise, it often introduces artifacts such as musical noise. Collectively, these traditional methods establish the foundation for advanced denoising research.

Modern approaches utilizing the emergence of deep neural networks have also developed recently. These deep learning-based speech enhancement models employ methods such as recurrent neural networks (RNNs)[21] and convolutional neural networks (CNNs)[21,22] to learn complex transformations aimed at converting noisy audio into clean speech. However, when considering practical implementation, factors such as the real-time applicability and computational complexity of these enhancement methods must be investigated [23,24]. Traditional and advanced digital filters remain essential, often dominating performance in specific, practical real-world scenarios due to established strengths and the inherent drawbacks of machine learning models. Conversely, conventional and advanced digital filters, which are traditional signal processing techniques including Spectral Subtraction[20], Least Mean Squares (LMS)[16,17], Normalized Least Mean Squares (NLMS)[18], and Gaussian filters, have been systematically evaluated for noise reduction in several works. These conventional methods offer a wide array of options for real-world systems. The systematic evaluation conducted on conventional and advanced digital filters provides crucial insights for actual application scenarios where practical constraints often necessitate robust, efficient methods. Unlike some complex models, traditional techniques are inherently designed to operate efficiently on the noisy audio alone[25,26].

3. DATASETS

In this work, Noizeous dataset [27] has been utilized to create a noisy dataset to systematically test the true capabilities of different speech enhancement filters. Some real -world noise corrupted signals have been taken from the Noizeous dataset[27], 30 IEEE sentences (produced by three male and three female speakers) corrupted by different real-world noises at different SNRs. All audio signals have been sampled at 16000 Hz for consistency. It also contains the clean, uncorrupted speech files, which serve as the reference signal for the experiments. By comparing the denoised audio against the original clean version, it's possible to objectively measure that how effectively a filter preserves speech quality and intelligibility while suppressing noise. The selection of distinct noise from the noise dataset enables evaluation in predictable scenarios to chaotic real-world environments and synthetic noises, where the noise level can be varied easily. From the above dataset, a diverse set of six distinct noise environments has been utilized that includes time-invariant train noise, impulse gunshot noise, real-world restaurant noise, stationary low-frequency noise, additive white Gaussian noise, and steady hum noise.

The selection criteria for the six distinct noise datasets used in the study were based on the need for diversity to systematically test the strengths and weaknesses of different digital filters across a range of acoustic environments. The *Train Noise* represents a predictable and relatively consistent acoustic environment. The input Signal-to-Noise Ratio (SNR) for this environment is approximately 10 dB. It is significant for evaluating filters that perform under stable noise conditions. This allows researchers to assess the baseline effectiveness of a filter when the noise characteristics are not changing rapidly. The *Impulse Gunshot Noise*, characterized by sudden, transient events, has been injected into random segments of clean speech. This is crucial for testing the robustness of denoising methods against abrupt, non-stationary interference. It helps determine if a filter can handle sudden loud noises without introducing significant distortion to the speech signal. *The Stationary Low-Frequency Noise* generated at 60 Hz as sine wave, mimics the constant hum found in industrial settings or vehicle interiors. Its use is significant for assessing a filter's ability to target and remove a persistent, single-frequency without affecting the broader speech spectrum. This is a common challenge in applications like in-car communication systems. The *Additive White Gaussian Noise (AWGN)* is synthetically generated background hiss with well-understood statistical properties. AWGN is a standard benchmark in signal processing. Its consistent characteristics make it an ideal baseline for evaluating and comparing the fundamental noise suppression capabilities of different algorithms. *The "Steady Hum (Air Conditioner) Noise"* is a synthetically generated, stationary signal designed to model quasi-periodic ambient noise sources, such as electronic appliances. Its significance within the study is to serve as a benchmark for evaluating the performance of speech enhancement filters against constant, yet spectrally complex, background noise. Unlike a simple, stationary low-frequency sine wave, this modulated hum allows for a more nuanced assessment of a filter's ability to suppress persistent ambient sounds that exhibit more intricate temporal characteristics. The *Real-World Restaurant Noise* contains pre-existing recordings from a restaurant, representing a complex and dynamic acoustic environment. It presents a challenging, non-stationary, and highly variable noise scenario that simulates real-world conditions where many voices and sounds overlap. This tests a filter's performance in the most difficult and unpredictable environments where speech enhancement is often needed, such as in hearing aids or voice-command systems.

4. DENOISING METHODS

Filters must be able to adapt to noise that changes its characteristics over time without causing significant distortion or loss to the speech signal itself. The fundamental goal of a denoising filter is to take a noisy speech signal and process it to remove the unwanted background sounds while preserving the original speech. The filters implemented in this study are discussed in this section.

A. Moving Average Filter

The moving average (MA) [11,12] filter, represented in equation (1), is a fundamental digital signal processing technique used for smoothing a signal to reduce noise. It operates by calculating the average of a signal's values over a defined, fixed-size window. As this window "slides" along the signal, a new average is computed for each point, resulting in a smoothed output signal. A larger window size results in more smoothing but can also cause more blurring or loss of detail in the signal. This process effectively attenuates random, short-term fluctuations (like noise) while retaining the slower, underlying trends of the signal. The slower, more persistent components of the signal.

$$y[n] = \frac{1}{W}\sum_{k=-\frac{W-1}{2}}^{\frac{W-1}{2}} x[n+k] \qquad (1)$$

In this equation, $x[n]$,$y[n]$, represents the input noisy signal and smoothed or denoised signal value of the filter at the current time index *n,* respectively. W is the window size, where k is the summation index of the window. It iterates from $-\frac{W-1}{2}$ to $\frac{W-1}{2}$, which defines the symmetric window of samples around *n*.

B. Gaussian Filter

The Gaussian filter [14], given in equation (2), is a digital signal processing technique used for smoothing signals and reducing noise. It operates through a process called convolution, where the input signal is processed using a specific weighting function known as a Gaussian kernel (ref. equation (3)). The shape of this kernel is a bell curve, which means that the filter gives more weight to the central data point (the one being calculated) and progressively less weight to points that are further away. This characteristic allows the filter to effectively average out random, high-frequency fluctuations, such as Additive White Gaussian Noise (AWGN), resulting in a smoother output signal.

$$y[n] = \sum_k x[n-k] \cdot g[k] \tag{2}$$

where the Gaussian kernel g[k] is defined in equation (3)

$$g[k] = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{k^2}{2\sigma^2}} \tag{3}$$

In these equations, σ (sigma) is the standard deviation of the Gaussian distribution. It is a crucial parameter that controls the width of the bell curve of the kernel. A larger value of σ results in a wider kernel, which leads to more smoothing (and potentially more blurring of the signal). A smaller σcreates a narrower kernel, resulting in less smoothing.

C. Savitzky-Golay Filter

The Savitzky-Golay filter[15] is a digital filtering technique used for smoothing a signal to reduce noise while preserving its key features, such as the height and width of each peak, better than a simple moving average filter. It operates by fitting a low-degree polynomial to a small, sequential window of the data using the method of least squares. This polynomial is then used to estimate the smoothed value for the central point of that window. As the window slides along the signal, this process is repeated, effectively replacing each data point with a smoothed value derived from the local polynomial fit. The polynomial coefficients of Savitzky-Golay filter are estimated from equations (4) while

$$p(i) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}} c_k i^k \tag{4}$$

$p(i)$: represents the estimated polynomial coefficients where as c_k is the individual coefficient for each term of the polynomial of order k. The polynomial order N that defines the degree of the polynomial used for the fitting process. A higher order can capture more complex signal features, but may also fit to the noise if set too high.

The least-squares solution for smoothed values is stated in equation (5)

$$x_{\text{new}} = (A^T A)^{-1} A^T b \tag{5}$$

This equation is the mathematical tool for estimating the coefficients c_k in Equation (4) by minimizing the squared error between the polynomial fit and the actual noisy data within the window. x_{new}(or sometimes represented as c) gives the best-fit polynomial coefficients c_k .Here x_{new} is the vector of new, smoothed data points. It represents the final output of the filter after the least-squares fitting is complete. This equation uses A as the polynomial basis matrix (also known as a design matrix). Each column of this matrix corresponds to a power of the independent variable (time or position), from 0 up to the polynomial order N. Its structure is determined by the window size and the polynomial order. A^T is the transpose of the matrix A and b is the data vector, which contains the original, noisy signal values within the current window. The filter attempts to find the polynomial that best fits this vector.

D. LMS Filter

The Least Mean Squares (LMS)[16,17] filter is a type of adaptive filter used in digital signal processing for tasks like noise cancellation and speech enhancement. Adaptive filters are particularly useful for handling non-stationary noise that changes its characteristics over time. These filters can dynamically adjust their internal parameters to track and minimize the noise. LMS is mathematically represented by equations (6),(7), and (8).

Equation (6) states Filter output computation. The filter output $y[n]$ is obtained as the dot product between the current filter coefficient vector h and the input vector $x[n]$.It represents how the filter processes the noisy input signal at the current time step.

$$y[n] = \sum_{k=0}^{M-1} h_k[n] x[n-k] \tag{6}$$

Equation (7) states Error signal calculation. Once $y[n]$is obtained, it is compared with the desired reference signal r[n].

$$e[n] = \mathrm{r}[n] - y[n] \tag{7}$$

The difference between $\mathrm{r}[n]$and $y[n]$ gives the error signal $e[n]$.This error directly indicates how far the current filter output is from the desired signal. Equation (8)states Adaptive weight update. Error e[n]e[n]e[n] is then used to update the filter coefficients.

$$h[n+1] = h[n] + 2\mu e[n]x[n] \tag{8}$$

$2\mu e[n]$ determines the direction and magnitude of the adjustment. If the error is large, the update is larger; if the error is small, the update is smaller. This iterative update allows the filter to learn and adapt its weights to minimize the error over time.

E. NLMS Filter

The Normalized Least Mean Squares (NLMS)[18] filter is a powerful type of adaptive filter used for speech enhancement and noise cancellation. The NLMS filter modifies the LMS filter normalizing the learning rate with the power of the input signal. The Normalized filter equation is given in (9).

$$h[i+1] = h[i] + \frac{\mu}{x[i]^T x[i] + \epsilon} e[i]x[i] \tag{9}$$

Here $[i]^T x[i] + \epsilon$, the component in denominator, distinguishes NLMS from LMS. By dividing the learning rate by the input signal's power, the filter adjusts its adaptation speed based on the signal's energy, making it more stable and robust. Futher ϵ is a small positive constant added to the denominator. Its purpose is purely practical: to avoid division by zero in cases where the input signal power $\mathrm{x[i]^T x[i]}$ is zero or extremely small.

F. Non-Local Means Filter

The Non-Local Means (NLM) filter is an advanced denoising technique, originally developed for image processing, that has been adapted for audio signals like speech. Its core principle is to leverage redundant information or similar patterns found within the signal itself to distinguish the genuine signal from unwanted noise. Unlike local filters (like a moving average) that only consider immediate neighboring points, the NLM filter takes a non-local approach by searching for similar sections or patches across the entire signal to denoise a specific point. The denoised value is then calculated as a weighted average of these similar patches, where patches that are more similar are given a higher weight. Non-Local Means filter is mathematically represented by equations (10-13).

$$V_n = \sum_{m \in \Omega_n} f_m w(n,m) \tag{10}$$

V_n: This represents the denoised value of the signal at the central point n of the target patch. This is the final, smoothed output value for that specific point. Here f_m is the value of the signal at point m within the search window.

$$w(n,m) = \frac{1}{N_n} \exp\left(-\frac{||P(v_n) - P(v_m)||^2}{h^2}\right) \tag{11}$$

Here $w_x(n,\, m)$ is the weight assigned to the patch centered at point m with respect to the target patch centered at point n. This weight determines how much influence the patch at m will have on the final denoised value at n. A higher weight means the patches are more similar. N_n is the normalization factor, which ensures that the sum of all weights equals one. It is calculated by summing all the exponential terms in the weight equation over the search window.

$$||P(v_n) - P(v_m)||^2 = \sum_{k=1}^{|v|} \left(f_n(k) - f_m(k)\right)^2 \tag{12}$$

(v_n): This represents the target patch, which is a small window of signal values centered around the point n that is currently being denoised.

(v_m): This represents a comparison patch centered at another point m within the search window.

$||(v_n) - P(v_m)||^2$: This is the squared Euclidean distance between the target patch $P(v_n)$ and the comparison patch $P(v_m)$. It is a measure of how similar the two patches are. A smaller distance means the patches are more similar.

h: This is the smoothing parameter or filter parameter. It is a crucial variable that controls the decay of the exponential function and, therefore, the degree of smoothing. A larger h results in more aggressive smoothing, while a smaller h preserves more detail but may leave more noise.

|v|: This represents the size of the patch (i.e., the number of samples in the patch vector P).

k: This is the summation index used to iterate through the individual samples within a patch to calculate the distance.

The modified weights of the filter are given by equation (13)

$$w_x(n,m) = \begin{cases} \exp\left(-\frac{r[n,m]^2}{2Pk_1\sigma}\right), r[n,m] \geq E_{th} \\ \exp\left(-\frac{r[n,m]^2}{2Pk_2\sigma}\right), otherwise \end{cases} \qquad (13)$$

G. *Median Filter*

The Median filter is a non-linear digital filtering technique used for noise reduction, particularly effective against impulse noise. Unlike linear filters such as the moving average or Gaussian filter which compute a weighted average, the median filter works by replacing each data point with the median value of its neighboring entries within a defined sliding window. The smoothed value after the median calculation is given in equation (14).

$$y[n] = \mathrm{Median}\{x[n - \lfloor K/2 \rfloor], \dots, x[n], \dots, x[n + \lfloor K/2 \rfloor]\} \qquad (14)$$

Here , K is the window size or kernel size.

H. *Spectral Subtraction Filter*

The Spectral Subtraction [20] filter is a traditional and widely used digital signal processing technique for speech enhancement. It operates in the frequency domain, meaning it analyzes and modifies the frequency components of a signal rather than its time-domain waveform.

I. *Wavelet Denoising Filter*

Wavelet Denoising[19] is a sophisticated, non-parametric method used for speech enhancement that leverages the multi-resolution properties of the Wavelet Transform (WT) to separate a speech signal from noise. The core principle behind this technique is that the energy of a speech signal, when transformed into the wavelet domain, becomes concentrated into a few large coefficients, while the energy of noise (like white noise) remains spread out across many small coefficients. Given it's equation are (15),(16),(17).

Let $x(t)$ is the noisy signal, then the initial step is to transform the noisy signal from the time domain into the wavelet domain using the Discrete Wavelet Transform (DWT) as given in equation (15). This process separates the signal into high-frequency components (known as detail coefficients) and low-frequency components (known as approximation coefficients).

$$Y = W(x) \qquad (15)$$

Here W is the Wavelet Transform operator and Y is the resulting wavelet coefficients of the noisy signal.

After decomposition, A threshold value (λ) is applied to the decomposed signal wavelet coefficient for noise removal. Coefficients with magnitudes below this threshold are considered noise and are typically modified or set to zero. The sources state that this is commonly done using soft thresholding or hard thresholding techniques. This is mathematically given in equation(16).

$$Z = D(Y, \lambda) \qquad (16)$$

Here D is the thresholding function (e.g., soft or hard thresholding) while Z denotes the new vector of thresholded wavelet coefficients after filtering.

The final step is to transform the filtered wavelet coefficients back into the time domain using the Inverse Wavelet Transform W^{-1}, which results in the denoised speech signal. This reconstructed signal,Ŝ(t) is given in equation (17).

$$\hat{S}(t) = W^{-1}(Z) \tag{17}$$

J. Combined Filters

Combined filters, also referred to as hybrid approaches, are advanced denoising techniques that integrate two or more individual filtering methods in sequence. The primary goal of this strategy is to leverage the distinct strengths of different algorithms while minimizing their individual weaknesses, leading to improved overall performance in speech enhancement. By applying filters one after another, a hybrid method can tackle different aspects of noise more effectively than a single filter could alone. It's equation is (18).

$$F_{2,1}(x) = F_2\big(F_1(x)\big) \tag{18}$$

$F_{2,1}(x)$ represents the final output of the combined filter after both filtering operations have been applied to the input signal *x*. $F_1(x)$ is the first filter in the sequence and is followed by $F_2(x)$.

5. RESULTS AND PERFORMANCE EVALUATION

The performance of speech denoising filters is evaluated using a suite of objective metrics designed to quantify different aspects of the processed audio. These parameters are used to provide a comprehensive and robust comparison of how well each filter achieves its goal of improving speech quality and intelligibility. Here are the definitions of the key performance evaluation parameters mentioned in the sources:

Perceptual Evaluation of Speech Quality (PESQ): This metric assesses the perceived quality of the speech signal. It aims to measure how natural and pleasant the processed speech sounds to a human listener. Higher PESQ scores indicate better perceived speech quality. For an application like a high-quality voice messaging app or a podcast recording, where the goal is a pleasant listening experience, a filter with a high PESQ score like LMS would be the superior choice, even if it leaves a little residual, non-intrusive noise.

Short-time Objective Intelligibility (STOI): This parameter measures the intelligibility of the speech, which is how easy it is to understand. It correlates well with how a human listener would comprehend the spoken words in the processed signal. Higher STOI scores signify greater intelligibility. For applications like voice-command systems or critical communications for first responders, a filter with the highest STOI score is the best choice because it maximizes the chances that the message is understood correctly.

Output Signal-to-Noise Ratio (Output_SNR): This metric quantifies the level of noise suppression achieved by the filter. It measures the ratio of the power of the speech signal to the power of the remaining background noise in the filter's output. A higher Output_SNR value, measured in decibels (dB), indicates more effective or aggressive noise removal. For forensic audio analysis, where the primary goal is to isolate a voice from overwhelming background noise for transcription, a filter like NLMS or NLM_SpecSub (which frequently scored highest on Output_SNR) would be chosen for its aggressive noise removal capabilities.

Log Likelihood Ratio (LLR): This parameter measures the spectral distortion of the processed signal compared to the original clean signal. It evaluates how closely the frequency spectrum of the denoised speech matches that of the ground truth. A lower LLR value indicates a stronger similarity and less distortion. In an automatic speech recognition (ASR) system, preserving the spectral shape of vowels and consonants is essential for accurate transcription. Therefore, a filter like Wavelet with a very low LLR would be the ideal choice as a preprocessing step, as it cleans the signal without changing the fundamental characteristics the ASR system relies on.

A. Performance in time-invariant Noise Environment

This environment features speech signals corrupted by a consistent, predictable noise at an input SNR of approximately 10dB. Figure 1(a),(b) shows the graph of filters when the signal is corrupted with time-invariant noise.

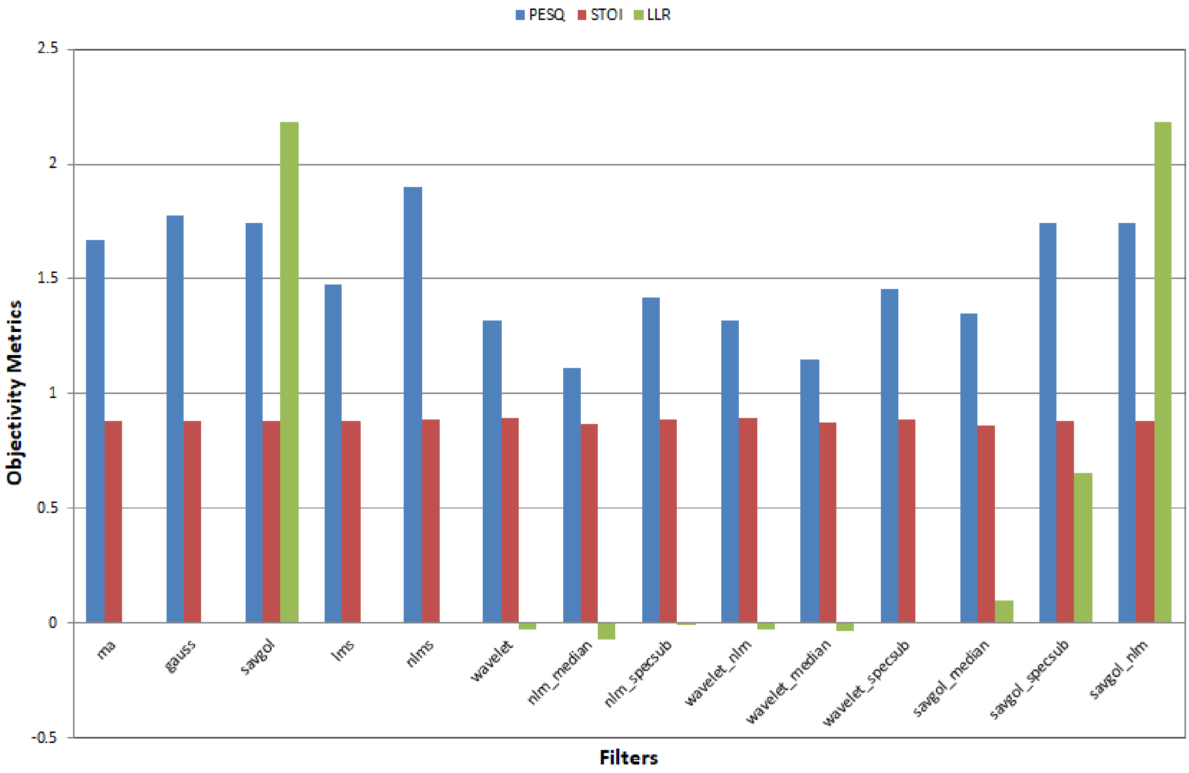

Figure 1(a): Filter response of time-invariant noise.

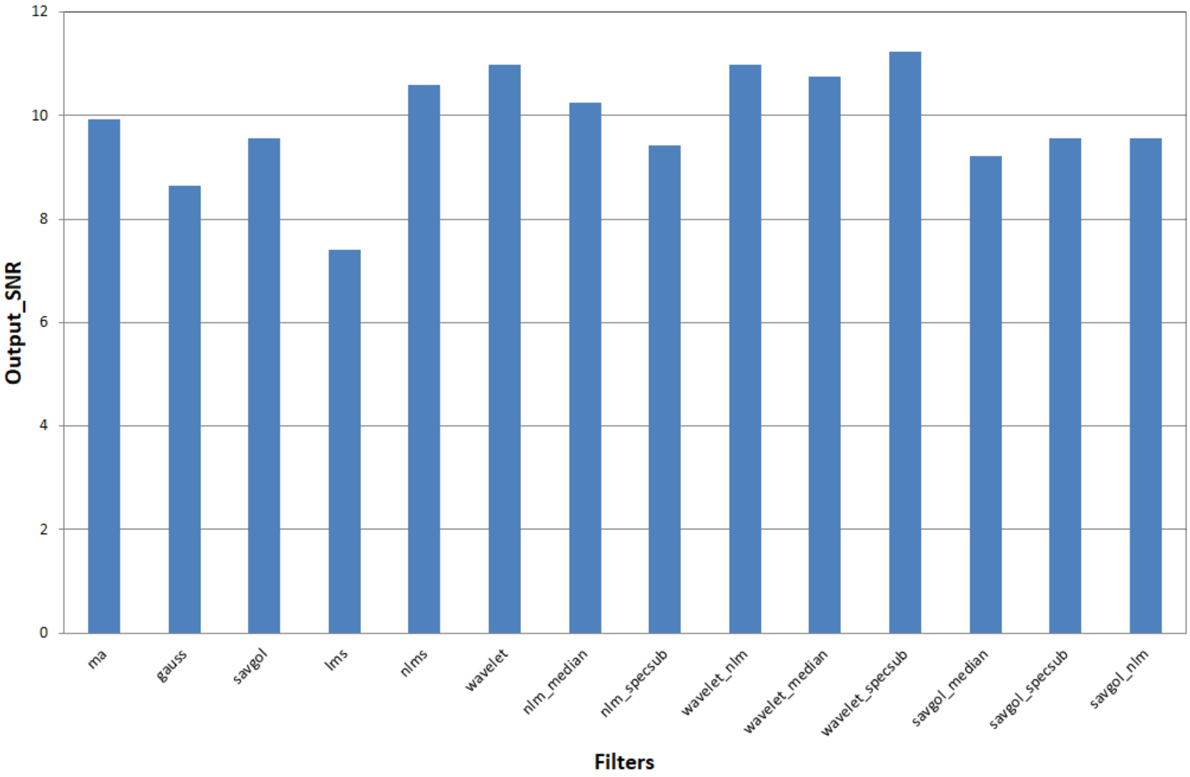

Figure 1(b): Output SNR of the filters in time-invariant noise.

From Figure 1(a), we can it is evident that the NLMS filter showed high performance with quality (PESQ 1.901) with STOI of 0.884. However, wavelet and wavelet_nlm methods achieved the best STOI of 0.891, but with PESQ 1.320, which is lower than that of NLMS. The hybrid wavelet_specsub delivered the highest Output_SNR (11.237 dB)(refer figure 1(b)) but compromised the PESQ value. The LMS filter performed the worst in noise suppression (Output_SNR 7.400 dB), while the nlm_median filter provided the worst PESQ (1.110). The cumulative results show that STOI value is approximately the same for all the filters, whereas output_SNR shows drastic changes.

B. Performance in Real-World Restaurant Noise Environment

This environment assesses denoising performance in a challenging non-stationary background noise from a "restaurant" setting, with an input SNR of approximately 10 dB. Figure 2(a),(b) shows the graph of filters when the signal is corrupted with real-world restaurant noise.

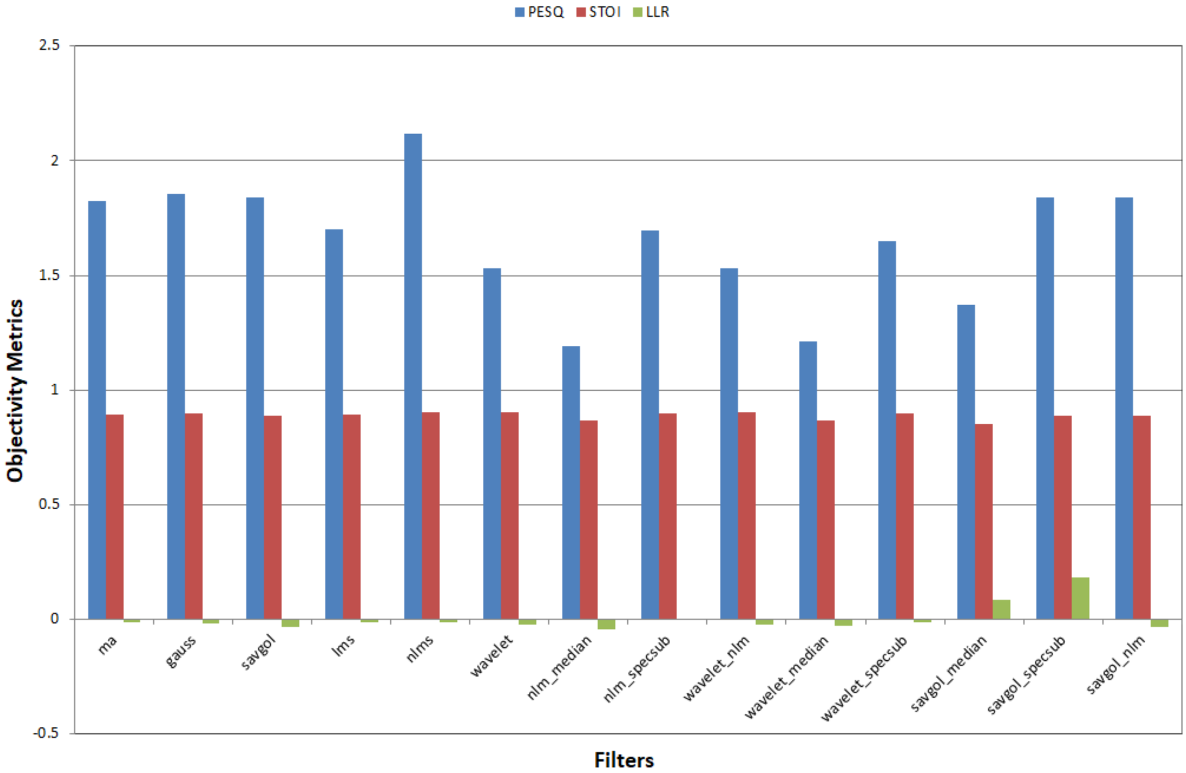

Figure 2(a): Filter response of restaurant noise.

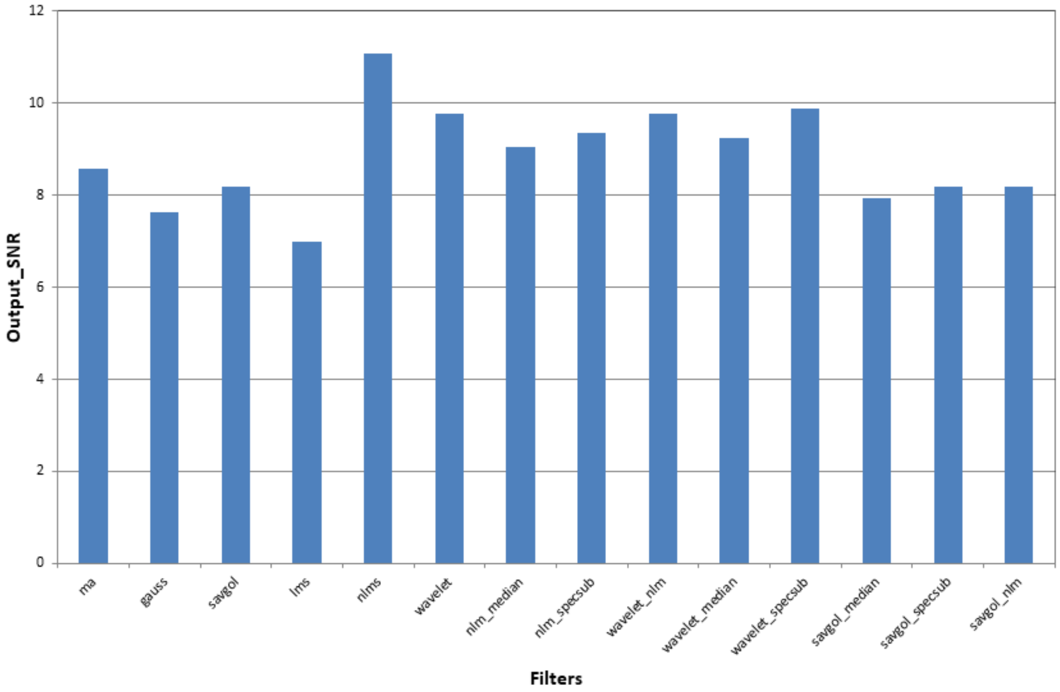

Figure 2(b): Output SNR of the filters in restaurant noise.

This environment represents a challenging, dynamic, and non-stationary scenario, typical for applications like hearing aids or voice-command systems in complex settings. The NLMS filter demonstrated superior overall performance. NLMS delivered the best result across all primary metrics: the highest Perceptual Evaluation of Speech Quality (PESQ) (2.117), the highest Short-time Objective Intelligibility (STOI) (0.905) refer figure 2(a) for both, and the strongest Noise Suppression (Output_SNR) (11.093 dB) refer figure 2(b). In contrast, the nlm_median method provided the worst result for perceived quality (PESQ 1.192), while the LMS filter gave the worst Output_SNR (6.997 dB), indicating the least effective noise removal. Compared to invariant noise in Table 1, Table 2 shows more variation in STOI and greater variation in Output_SNR. We can clearly observe that PESQ also varies from 1.2 to 2.1 for different filters.

C. Performance in Additive White Gaussian Noise Environment

This data provides average performance metrics across 10 different speech files contaminated with Additive White Gaussian Noise (AWGN), a common noise type in signal processing, at an average input SNR of approximately 10dB. Figure 3(a),(b) shows the graph of filters when the signal is corrupted with Additive White Gaussian Noise (AWGN).

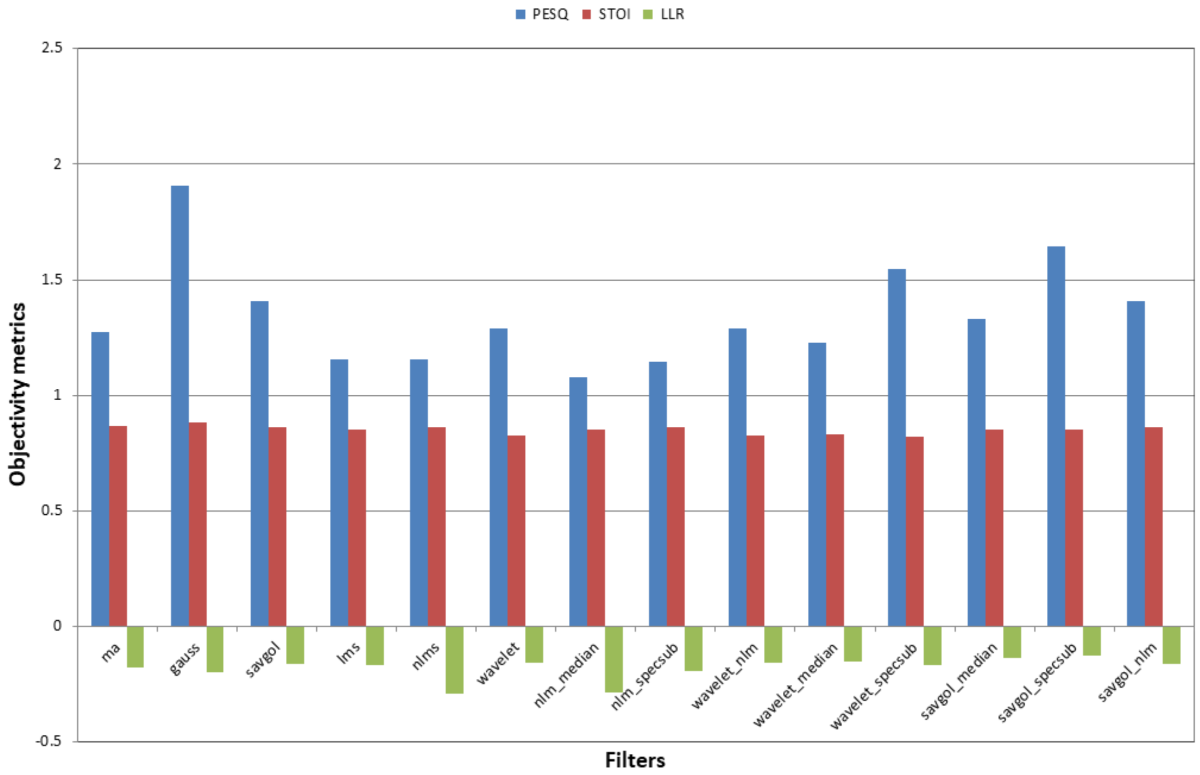

Figure 3(a): Filter response of Additive White Gaussian Noise.

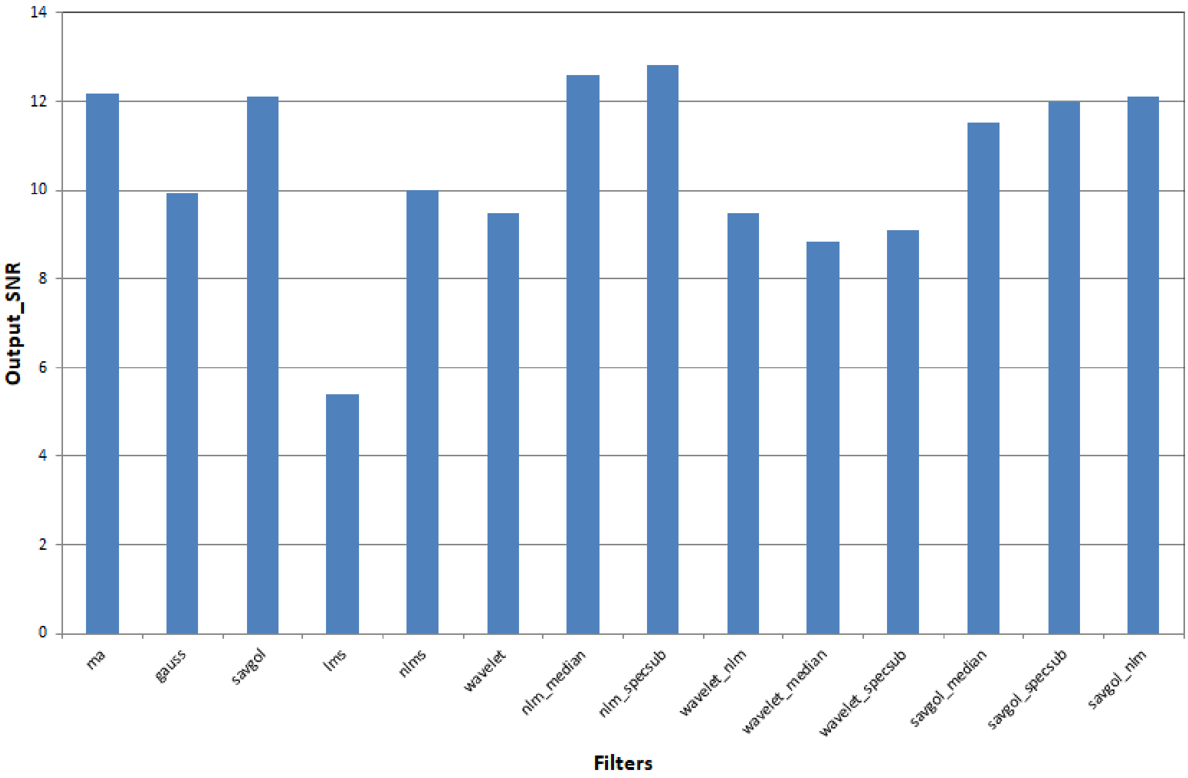

Figure 3(b): Output SNR of the filters in Additive White Gaussian Noise.

This environment represents applications like hearing aids or voice-command systems in complex settings. The NLMS filter demonstrated superior overall performance. NLMS delivered the best result across all primary metrics: the highest Perceptual Evaluation of Speech Quality (PESQ) (2.117), the highest Short-time Objective Intelligibility (STOI) (0.905) refer 3(a), and the strongest Noise Suppression (Output_SNR) (11.093 dB)refer 3(b). In contrast, the nlm_median method provided the worst result for perceived quality (PESQ 1.192), while the LMS filter gave the worst Output_SNR (6.997 dB), indicating the least effective noise removal. It indicates that the combined filters have degraded the signal quality. We can observe change from 1.0 to 1.9 in PESQ whereas STOI shows slight variation but output_SNR shows variation from 5.3 to 12.8 which is a noticeable variation.

D. Performance in Impulse Gunshot Noise Environment

This data presents results for speech signals affected by gunshot impulse noise. The Input_SNR varies significantly across the tested speech files. Figure 4(a),(b) shows the graph of filters when the signal is corrupted with impulse gunshot noise.

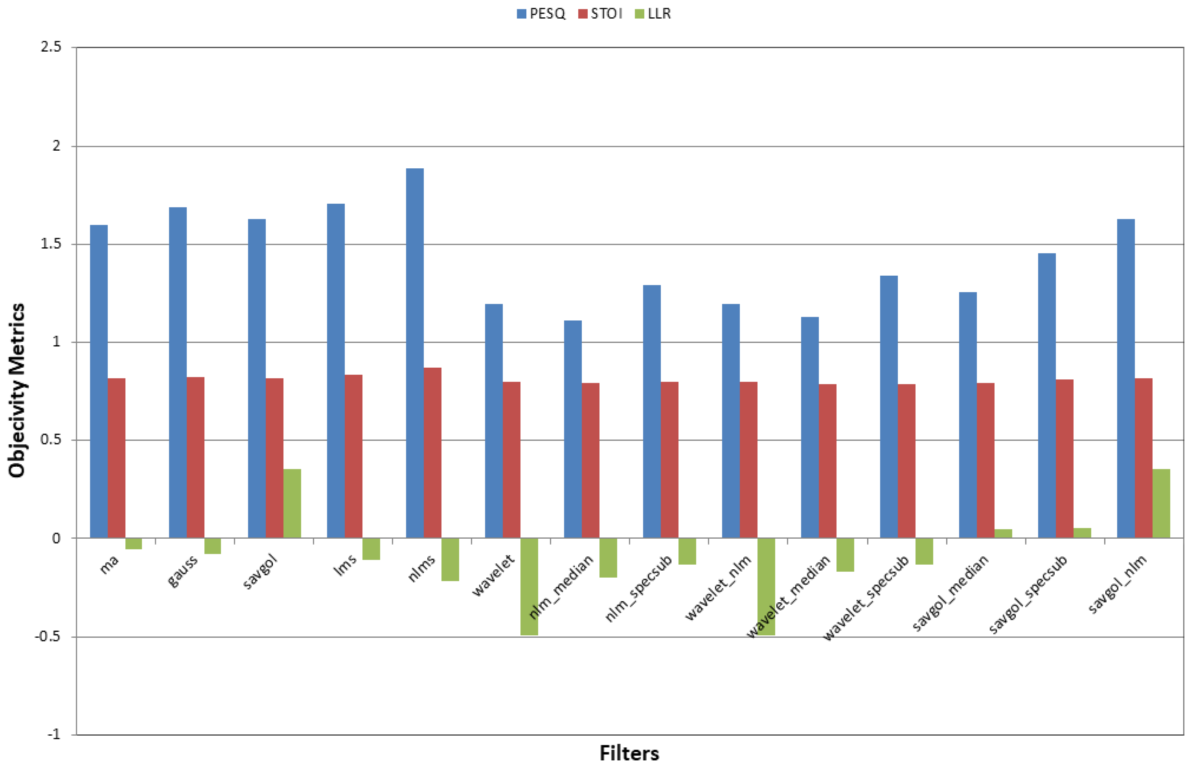

Figure 4(a): Filter response of gunshot impulse noise.

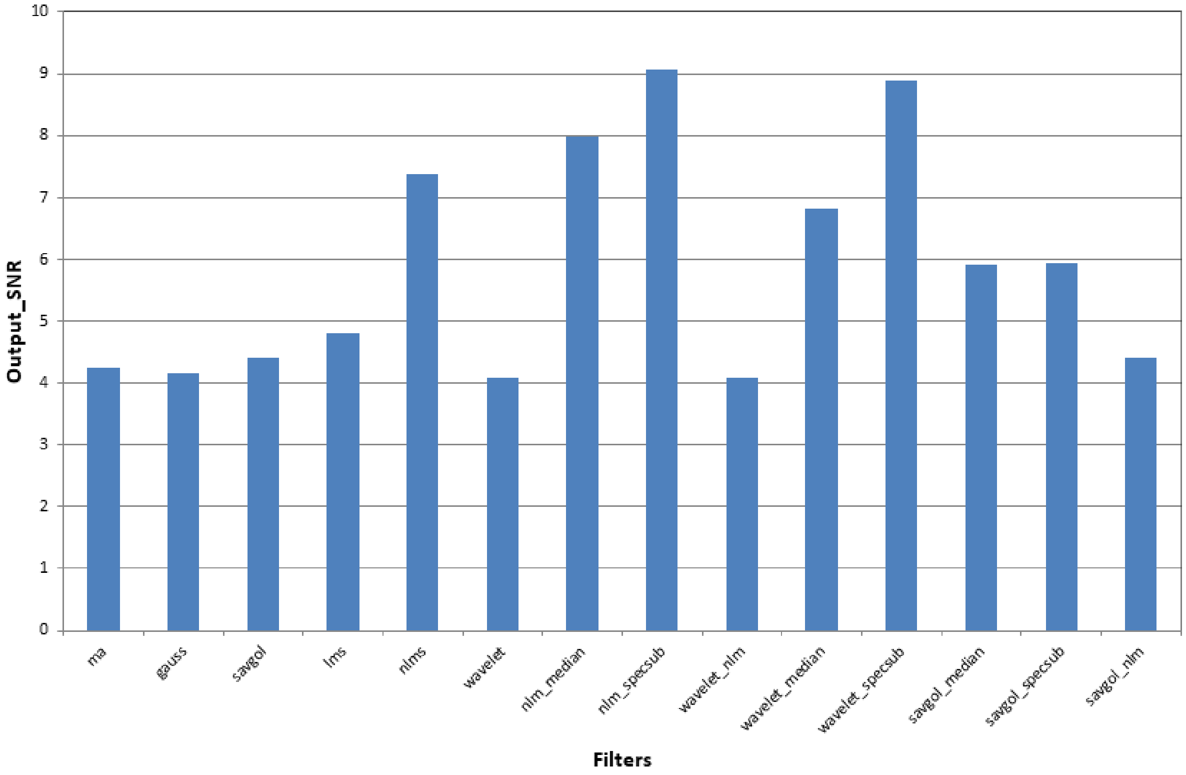

Figure 4(b): Output SNR of the filters in gunshot impulse noise.

This environment represents a challenging, dynamic, and non-stationary scenario, typical for applications like hearing aids or voice-command systems in complex settings. The NLMS filter demonstrated superior overall performance. NLMS delivered the best result across all primary metrics: the highest Perceptual Evaluation of Speech Quality (PESQ) (2.117), the highest Short-time Objective Intelligibility (STOI) (0.905) refer figure 4(a), and the strongest Noise Suppression (Output_SNR) (11.093 dB) refer figure4(b). In contrast, the nlm_median method provided the worst result for perceived quality (PESQ 1.192), while the LMS filter gave the worst Output_SNR (6.997 dB), indicating the least effective noise removal. We can observe change in STOI from 0.78 to 0.87, which is the highest as of now. The output_SNR also shows variation from 4.0 to 9.0 for different methods.

E. Performance in Stationary Low-Frequency Noise Environment

This data shows the performance on speech signals with stationary low-frequency noise at an input SNR of approximately 10 dB. Figure 5(a),(b) shows the graph of filters when the signal is corrupted with stationary low-frequency noise.

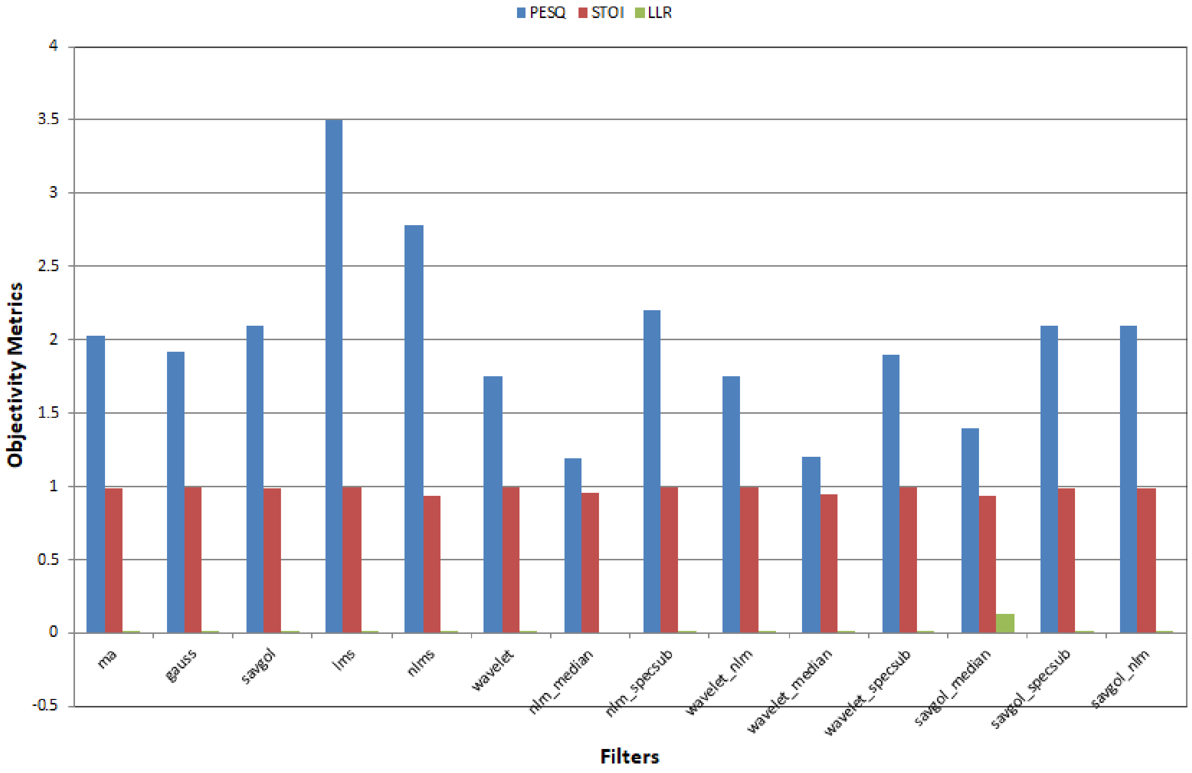

Figure 5(a): Filter response of stationary low-frequency noise.

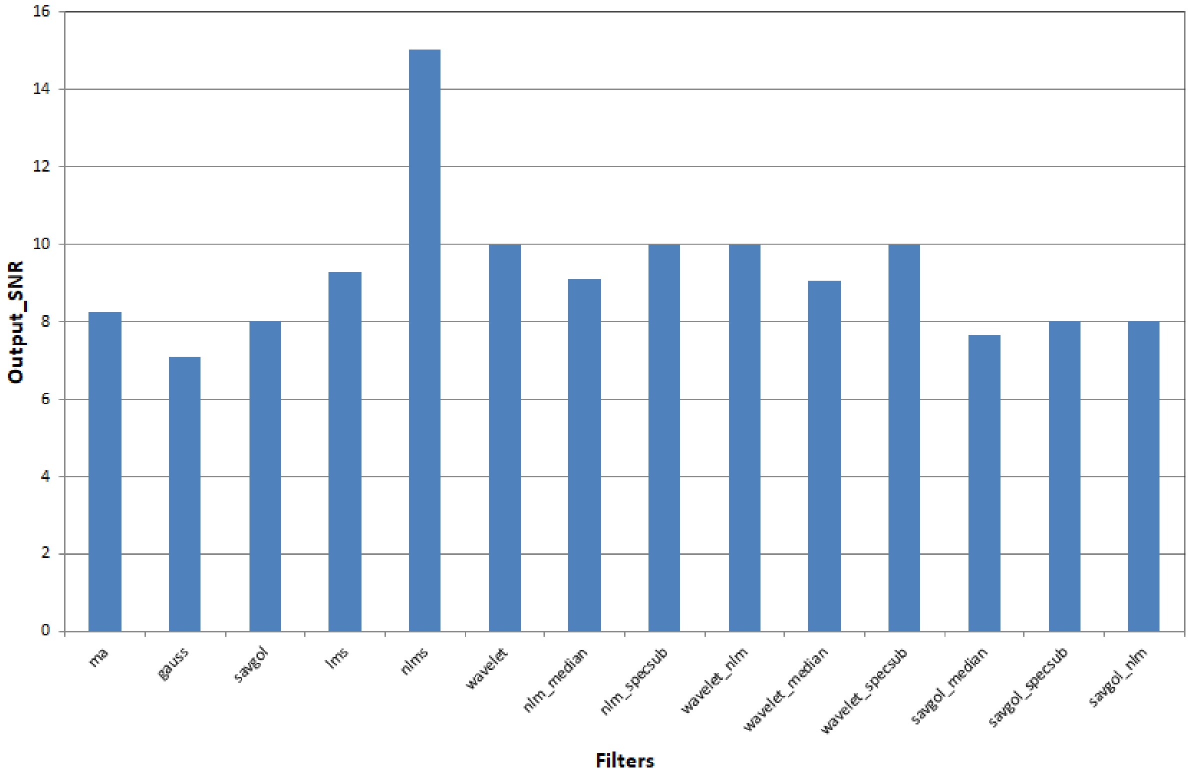

Figure 5(b): Output SNR of the filters in stationary low-frequency noise.

Simulating a persistent, single-frequency drone found in places like industrial settings or vehicle interiors, this environment showed a clear trade-off between quality and suppression. The LMS filter provided the best result for quality and fidelity, achieving an exceptional PESQ (3.497) refer figure 5(a). However, the NLMS filter delivered the best noise suppression, reaching a significantly high Output_SNR (15.027 dB) refer figure 5(b). The nlm_median filter yielded the worst PESQ (1.195), and the Gaussian filter provided the worst Output_SNR (7.088 dB). The best PESQ variation can be observed in this noise from 1.1 to 3.4, and drastic variation in output_SNR from 7.0 to 15.0.

F. Performance in Steady Hum (Air Conditioner) Noise Environment

This data evaluates denoising methods against a "steady hum" from an air conditioner, representing a stationary background noise type, with an input SNR of approximately 10 dB. Figure 6(a),(b) shows the graph of filters when the signal is corrupted with steady hum of air conditioner noise.

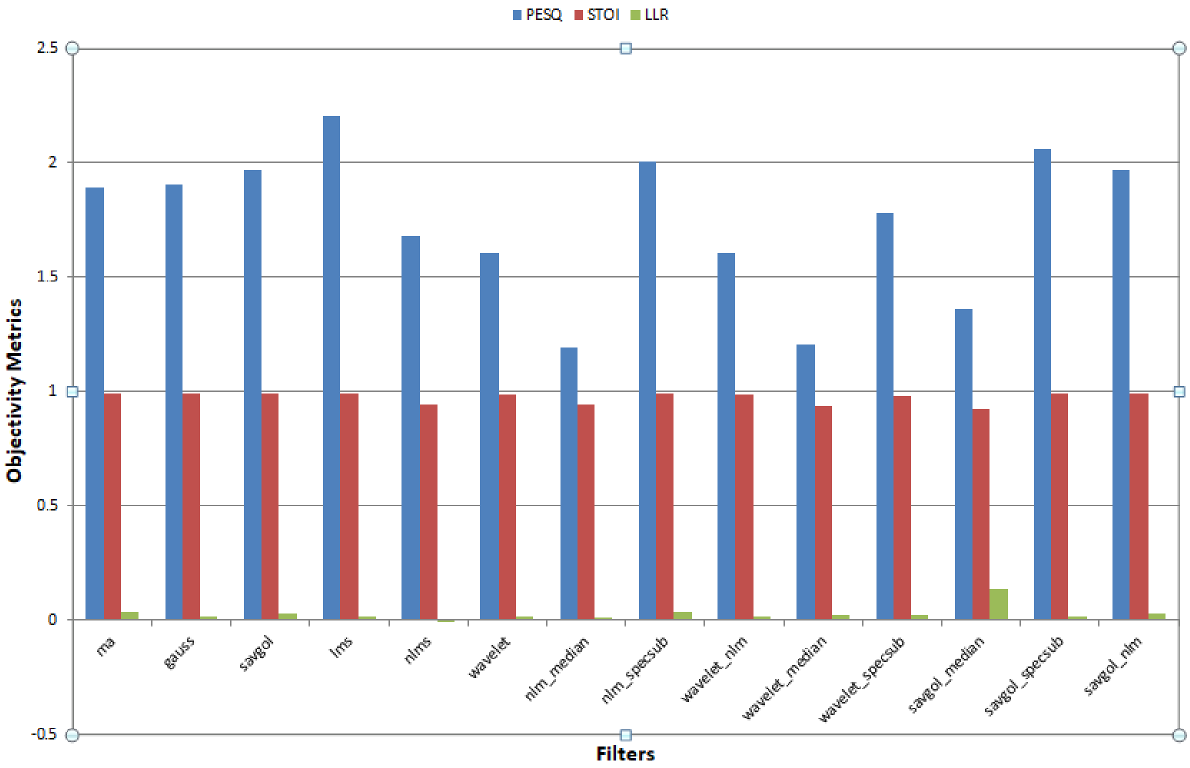

Figure 6(a): Filter response of stationary low-frequency noise.

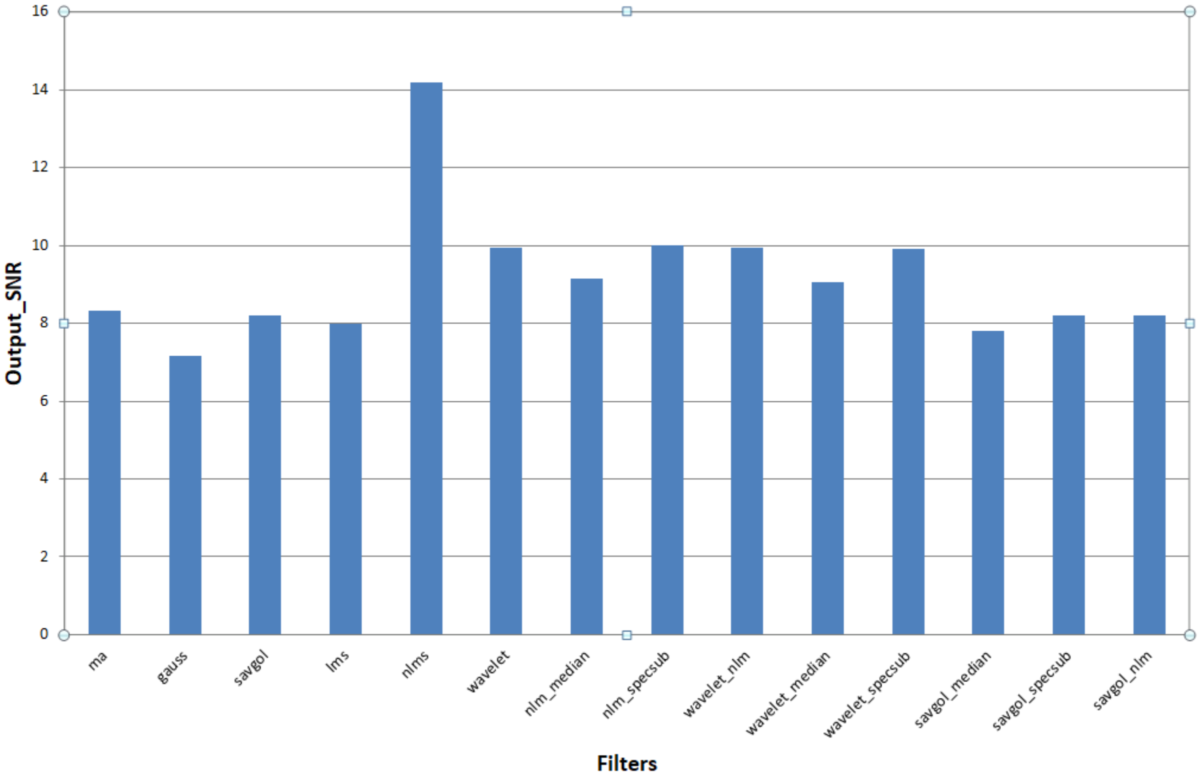

Figure 6(b): Output SNR of the filters in low-frequency noise.

This stationary background noise, simulating ambient sounds from electronic appliances, favoured methods that preserve speech naturalness. The LMS filter provided the best result for perceived quality, achieving the highest PESQ (2.206) refer figure 6(a). Conversely, the NLMS filter delivered the best noise suppression with the highest Output_SNR (14.200 dB) refer figure 6(b). The nlm_median filter recorded the worst PESQ (1.194), and the Gaussian filter yielded the worst Output_SNR (7.146 dB) refer figure 6(b). Not much variation in PESQ and STOI is observed but output_SNR varies from 7.1 to 14.2.

This comprehensive evaluation of digital filtering techniques for speech enhancement unequivocally demonstrates that **no single method achieves universal optimality**. The selection of an appropriate filter is not a one-size-fits-all decision but rather a context-aware optimization problem, guided by the statistical properties of the noise environment and the specific performance objectives of the target application.

• For **dynamic, complex, or transient noise scenarios** where maximizing speech intelligibility (STOI) and achieving aggressive noise suppression (Output_SNR) are paramount, **adaptive filters like NLMS and NLM-based algorithms are the most effective solutions**. Their ability to dynamically adjust to changing conditions and exploit signal redundancies makes them superior for real-world, unpredictable environments.

• For **stationary and predictable noise conditions** where preserving perceptual speech quality (PESQ) and overall naturalness (COVL) is the primary goal, **linear filters such as LMS and Gaussian are the preferred choice**. They provide effective smoothing and noise reduction without introducing significant signal distortion, leading to a more subjectively pleasing output.

Ultimately, this research provides a nuanced framework for the practical application of speech enhancement technologies. By mapping filter strengths to specific acoustic challenges and performance metrics, these findings can inform the design of more effective and contextually aware systems across a wide array of fields, including telecommunications, hearing aids, forensics, and automatic speech recognition.

6. CONCLUSION AND FUTURE WORK

This paper presents a systematic and comprehensive evaluation of speech enhancement techniques, investigating their performance across a diverse and challenging set of acoustic environments. For dynamic, complex, or transient noise scenarios where maximizing speech intelligibility (STOI) and achieving aggressive noise suppression (Output_SNR) are paramount, adaptive filters like NLMS and NLM-based algorithms are the most effective solutions. stationary and predictable noise conditions where preserving perceptual speech quality (PESQ) and overall naturalness (COVL) is the primary goal, linear filters such as LMS and Gaussian are the preferred choice. Wavelet-based methods demonstrated strong performance in specialized areas,it showed strong performance in speech intelligibility (STOI).

By subjecting a range of digital filters to distinct noise types this work showcases the critical principle that there is no universally optimal solution for speech denoising. The study's core contribution is its rigorous demonstration that the efficacy of any speech enhancement method is fundamentally tied to the specific characteristics of the noise and the desired application priorities. Ultimately, this paper illustrates that the challenge of speech enhancement is not about finding a single superior algorithm, but about making an informed, context-aware selection.

Future works can aim for true adaptivity. Instead of a fixed filter choice, research could focus on developing a hybrid algorithm that first classifies the acoustic environment (e.g., stationary hum, transient impulse, complex babble) and then dynamically deploys the most suitable denoising technique or a combination thereof, based on the findings of this study. In future we can involve integrating and evaluating deep learning-based speech enhancement models against the same diverse set of noise environments and objective metrics used in the current study.

REFERENCES

[1] Yuliani, A., Amri, M., Suryawati, E., Ramdan, A., & Pardede, H. (2021). Speech Enhancement Using Deep Learning Methods: A Review. *Jurnal Elektronika dan Telekomunikasi*. https://doi.org/10.14203/jet.v21.19-26.

[2] Kheddar, H., Hemis, M., & Himeur, Y. (2024). Automatic Speech Recognition using Advanced Deep Learning Approaches: A survey. *ArXiv*, abs/2403.01255. https://doi.org/10.1016/j.inffus.2024.102422.

[3] Steeneken, H., & Houtgast, T. (2024). SUBJECTIVE AND OBJECTIVE SPEECH INTELLIGIBILITY MEASURES. *Reproduced Sound 1994*. https://doi.org/10.25144/20265.

[4] Vihari, S., Murthy, A., Soni, P., & Naik, D. (2016). Comparison of Speech Enhancement Algorithms. *Procedia Computer Science*, 89, 666-676. https://doi.org/10.1016/J.PROCS.2016.06.032.

[5] Iqbal, Y., Zhang, T., Gunawan, T., Pratondo, A., Zhao, X., Geng, Y., Kartiwi, M., Saleem, N., & Bourious, S. (2025). A Hybrid Speech Enhancement Technique Based on Discrete Wavelet Transform and Spectral Subtraction. *IEEE Access*, 13, 39765-39781. https://doi.org/10.1109/ACCESS.2025.3546434.

[6] Vumanthala, S., & Kalagadda, B. (2021). Modified Nonlocal Means Filtering for Speech Enhancement and Its FPGA Prototype. *Circuits, Systems, and Signal Processing*, 40, 6035 - 6049. https://doi.org/10.1007/s00034-021-01750-5.

[7] R, L., R, S., Kumar, B., Ibrahim, M., & , S. (2023). Hybrid Threshold Speech Enhancement Scheme Using TEO And Wavelet Coefficients. *2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, 01-05. https://doi.org/10.1109/ICEEICT56924.2023.10156921.

[8] Acarbay, E., & Özkurt, N. (2023). Performance analysis of the speech enhancement application with wavelet transform domain adaptive filters. *International Journal of Speech Technology*, 26, 245-258. https://doi.org/10.1007/s10772-023-10022-3.

[9] Mafi, M., Martin, H., Cabrerizo, M., Andrian, J., Barreto, A., & Adjouadi, M. (2019). A comprehensive survey on impulse and Gaussian denoising filters for digital images. *Signal Process.*, 157, 236-260. https://doi.org/10.1016/J.SIGPRO.2018.12.006.

[10] P. Papadopoulos, C. Vaz, and S. S. Narayanan, "Noise Aware and Combined Noise Models for Speech Denoising in Unknown Noise Conditions," *Interspeech 2016*, pp. 2866–2869, 2016. doi: 10.21437/Interspeech.2016-501. ISCA Archive

[11] 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, Sep. 2003. doi: 10.1109/TSA.2003.811544

[12] Raghudathesh G. P., Chandrakala C. B., Dinesh Rao B., and Thimmaraja Yadava G., "Noise estimation based on optimal smoothing and minimum controlled through recursive averaging for speech enhancement," *Intelligent Systems with Applications*, vol. 21, Article 200310, Mar. 2024. doi: 10.1016/j.iswa.2023.200310

[13] B. G. Gowri and K. P. Soman, "Enhancement of white Gaussian noise affected speech using VMD-ℓ1 trend filter method," *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 3, pp. 1701–1711, 2018. doi: 10.3233/JIFS-169463

[14] Mafi, M., Martin, H., Cabrerizo, M., Andrian, J., Barreto, A., & Adjouadi, M. (2019). A comprehensive survey on impulse and Gaussian denoising filters for digital images. *Signal Process.*, 157, 236-260. https://doi.org/10.1016/J.SIGPRO.2018.12.006.

[15] Cao, R., Chen, Y., Shen, M., Chen, J., Zhou, J., Wang, C., & Yang, W. (2018). A simple method to improve the quality of NDVI time-series data by integrating spatiotemporal information with the Savitzky-Golay filter. *Remote Sensing of Environment*. https://doi.org/10.1016/J.RSE.2018.08.022.

[16] A. Savitzky and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964. doi: 10.1021/ac60214a047

[17] R. M. S. Pimenta, M. R. Petraglia, and D. B. Haddad, "Stability Analysis of the Bias Compensated LMS Algorithm," *Digital Signal Processing*, vol. 147, 2024, article 104395. doi: 10.1016/j.dsp.2024.104395

[18] A. Aouane and T. Laroussi, "A new peak width NLMS filtering scheme for a DVB-T based Passive Bistatic Radar," *Physical Communication*, Jul. 2025. Available: ScienceDirect.

[19] Bhobhriya, R., Boora, R., Jangra, M. *et al.* W-NLM: a proficient EMG denoising technique. *Int. j. inf. tecnol.* **15**, 2517–2527 (2023). https://doi.org/10.1007/s41870-023-01324-5

[20] Youngberg, J., Petersen, T., Boll, S., & Cohen, E. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27, 113-120. https://doi.org/10.1109/TASSP.1979.1163209.

[21] Z. Momynkulov, N. Omarov, and A. Altayeva, "CNN-RNN Hybrid Model for Dangerous Sound Detection in Urban Areas," in *Proc. 2024 IEEE 4th International Conference on Smart Information Systems and Technologies (SIST)*, 2024. doi: 10.1109/SIST61555.2024.10629358

[22] Manjunath Jogin, Divya G D, Mohana, Meghana R K, Madhulika M S, Apoorva S, "Feature Extraction using Convolution Neural Networks (CNN) and Deep Learning," in *Proceedings of the 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, Bengaluru, India, May 2018, pp. 2319-2323. [Online]. Available: https://doi.org/10.1109/RTEICT42901.2018.9012507

[23] Y. Zhang and D. Wang, "Compact Deep Neural Networks for Real-Time Speech Enhancement," *Signal Processing*, vol. 202, no. 1, pp. 107–118, 2023. doi: 10.1016/j.sigpro.2022.107118

[24] Z. A. Balogh and B. J. Kis, *"*Comparison of CT noise reduction performances with deep learning-based, conventional, and combined denoising algorithms,*" Medical Engineering & Physics*, vol. 109, Article 103897, Nov. 2022. doi: 10.1016/j.medengphy.2022.103897.

[25] Pauline, S., & Dhanalakshmi, S. (2022). A low-cost automatic switched adaptive filtering technique for denoising impaired speech signals. *Multidimensional Systems and Signal Processing*, 33, 1387 - 1408. https://doi.org/10.1007/s11045-022-00849-5.

[26] Naderahmadian, Y., Ghorshi, S., & Panahi, I. (2025). Single Microphone Speech Denoising Using Wavelet Thresholding and RLS Algorithm. *2025 International Conference on Electronics, Information, and Communication (ICEIC)*, 1-4. https://doi.org/10.1109/ICEIC64972.2025.10879774.

[27] P. C. Loizou, "NOIZEUS: A noisy speech corpus for evaluation of speech enhancement algorithms," *University of Texas at Dallas*, [Online]. Available: https://ecs.utdallas.edu/loizou/speech/noizeus/