Proceedings of CONF-CDS 2026 Symposium: Computer Vision-Based Multimodal Cognitive Load Estimation for Adaptive Media Communication
DOI: 10.54254/2755-2721/2026.GL35875

A Comparative Study of Audio Noise Reduction Performance Based on FIR, IIR, and Adaptive Filters

Jing Ma

Macao University of Science and Technology, Macao, China
18148255835@163.com

Abstract. To address background noise that reduces speech intelligibility in voice communications, this paper compares the noise-reduction performance of FIR (window function), Butterworth IIR, and adaptive LMS filters. Experiments using speech signals with white noise, pink noise, and real-world environmental noise under input SNRs of 0–10 dB show that at 0 dB input SNR, the LMS filter achieves an output SNR of 8 dB and MSE of 0.01, significantly outperforming FIR (6 dB, 0.07) and IIR (5 dB, 0.08). At an input SNR of 10 dB, the performance gap narrows to 1 dB. The IIR filter has the lowest order (5th) and minimal computational cost; the FIR filter (65th) offers the best linear phase; the LMS filter (32nd) exhibits the strongest adaptability at low SNRs. Overall, LMS prioritizes adaptability, IIR prioritizes efficiency, and FIR suits phase-sensitive applications. This study provides quantitative references for noise suppression in voice communication systems.

Keywords: speech noise reduction, FIR filter, IIR filter, adaptive filtering, LMS algorithm, performance comparison

1. Introduction

Audio noise reduction is a critical issue in signal processing and is widely applied in scenarios such as mobile phone calls, speech recognition front-ends, music recording, and online education. In real-world environments, background noise (such as street noise, group conversations, and wind noise) can significantly reduce speech intelligibility and the audio quality of music signals, and negatively impact the accuracy of subsequent recognition systems. To address this, researchers have proposed various noise-reduction methods, including spectral subtraction, Wiener filtering, statistical modeling, and deep neural networks. Among these, digital filters remain widely used in resource-constrained or latency-sensitive systems due to their simple structure, real-time performance, and ease of hardware implementation [1-3].

Common digital filters include FIR, IIR, and adaptive filters. This paper uses noise in everyday piano sounds as a case study to design and compare the noise-reduction performance of a window-function-based FIR filter, a Butterworth IIR filter, and an adaptive LMS filter across different noise types. Through simulation experiments under various signal-to-noise ratio conditions, we analyze the applicable scenarios, advantages, and disadvantages of each method, providing a theoretical basis and engineering reference for selecting appropriate filters in practical systems [4, 5].

DOI: 10.54254/2755-2721/2026.GL35875

2. System model

This paper considers the scenario of single-microphone speech denoising, where a noisy speech signal can be represented as the sum of clean speech and additive noise:

$$y(n) = s(n) + v(n)$$

where y(n) is the noisy speech signal captured by the microphone, s(n) is the clean speech signal, v(n) is the additive background noise, and n is a discrete-time sample point [4, 6].

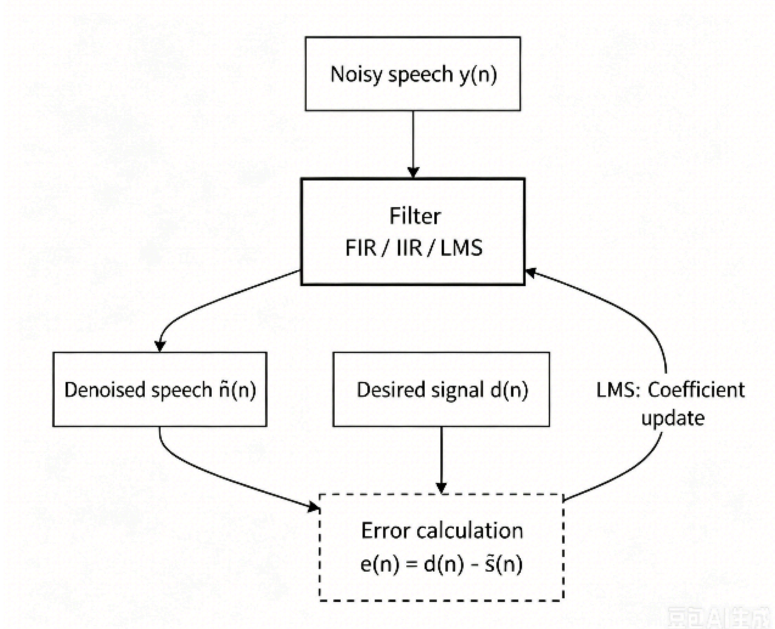

Figure 1. Block diagram of the speech noise reduction system

Figure 1 shows the block diagram of the speech denoising system. The input noisy speech signal y(n) undergoes preprocessing, including framing and windowing, before being fed into the filter module (FIR/IIR/LMS). The filter outputs the denoised speech signal s(n) based on optimization criteria or noise estimation. In the adaptive LMS filter, the error signal e(n) is used to dynamically update the filter coefficients.

3. Filter design and principles

This chapter introduces the design principles and methods for determining key parameters of FIR window-based filters, IIR Butterworth filters, and adaptive LMS filters. All filters are designed for speech noise reduction tasks, with a uniform sampling frequency of f_s = 8000 Hz. The noise is additive, and the filter's objective is to suppress out-of-band noise while preserving the main speech frequency band.

3.1. FIR filter

The unit impulse response h(n) of a FIR filter is of finite length and can achieve strict linear phase. This paper employs the windowing method to design a low-pass filter, with the target frequency response being an ideal low-pass filter:

DOI: 10.54254/2755-2721/2026.GL35875

$$H_d\left(e^{j\omega}\right) = \begin{cases} 1, & \omega_{c1} < |\omega| < \omega_{c2} \\ 0, & otherwise \end{cases}$$

where $\omega_{c1} = 2\pi \times 300/8000 = 0.075\pi$ and $\omega_{c2} = 2\pi \times 3400/8000 = 0.85\pi$. The ideal unit impulse response $h_d(n)$ is obtained through inverse Fourier transform, and the actual response is obtained by windowing [5, 6]:

$$h(n) = h_d(n) \cdot w(n)$$

The Hamming window is selected, with its time-domain expression:

$$w\left(n\right) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), 0 \le n \le N-1$$

The Hamming window has a side-lobe attenuation of approximately -43 dB, effectively suppressing spectral leakage. The transition bandwidth requirement determines the filter order: set the transition bandwidth $\Delta\omega = 0.1\pi$. The approximate transition bandwidth of the Hamming window is $\Delta\omega \approx 6.6\pi/N$, yielding $N \approx 66$, i.e., an order of 65. Fine-tuning through simulation ensures a stopband attenuation of at least 40 dB.

3.2. IIR filter

IIR filters feature a feedback structure and can achieve steep transition bands with low orders, making them suitable for resource-constrained real-time speech systems [5]. This paper uses the Butterworth bandpass prototype and employs the bilinear transformation [7]. The specifications are as follows: Digital domain specifications: Passband range 300–3400 Hz, stopband edges at 200 Hz and 3600 Hz, passband maximum attenuation $A_p = 1\,\text{dB}$, stopband minimum attenuation $A_s = 30\,\text{dB}$, sampling frequency $f_s = 8000\,\text{Hz}$. Pre-warping: For the digital angular frequency $\omega=2\pi f/f_s$, pre-warping yields the analog frequency $\Omega = \tan(\omega/2)$. Order calculation: Based on the bandpass filter order formula, substituting the values gives $n \approx 4.2$, so $n = 5$ is selected. Transfer function: The normalized analog lowpass transfer function is obtained from tables, then converted to the digital filter $H(z)$ through lowpass-to-bandpass transformation and bilinear transformation.

The final IIR filter order is 5, much lower than the FIR's 65th order, resulting in low computational cost but nonlinear phase characteristics that may introduce speech distortion.

3.3. Adaptive LMS filter

The Least Mean Square (LMS) filter can dynamically track non-stationary noise environments and is suitable for speech noise reduction [8, 9]. Its block diagram is shown in Figure 2.

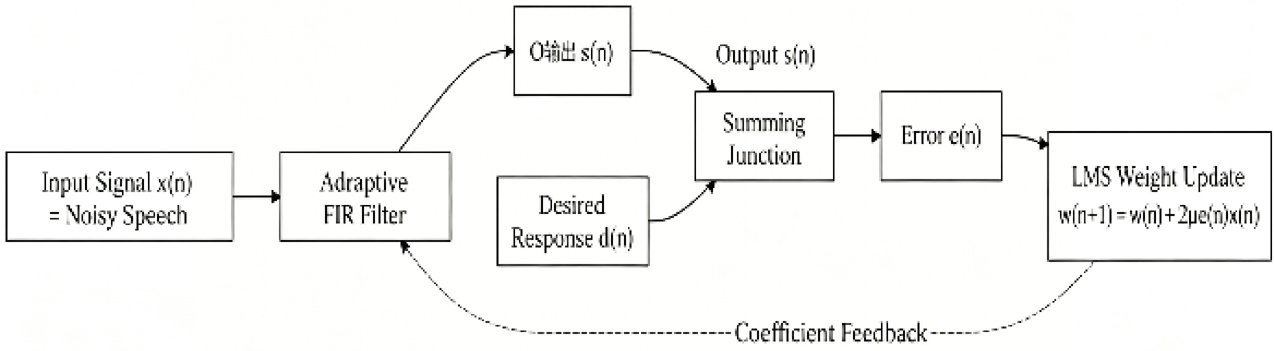

Figure 2. Schematic diagram of an adaptive LMS filter

In speech noise reduction applications, the input is $x(n) = y(n) = s(n) + v(n)$ (noisy speech), and the desired response $d(n)$ can be taken as the noisy speech itself. The filter output $\hat{s}(n)$ is the estimate of the clean speech, and the error signal is defined as:

$$e(n) = d(n) - \hat{s}(n)$$

The LMS algorithm adopts a transversal FIR structure of length L, with the weight update formula:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + 2\mu e(n)\mathbf{x}(n)$$

where $\mathbf{w}(n) = [w_0(n), w_1(n), \ldots, w_{L-1}(n)]^T$, $\mathbf{x}(n) = [x(n), x(n-1), \ldots, x(n-L+1)]^T$, and μ is the step size factor.

The step size factor μ must satisfy the convergence condition: $0 < \mu < \frac{1}{L \cdot P_x}$, where $P_x = E[x^2(n)]$ is the input signal power. This paper selects $\mu = 0.008$ and a filter order of $L = 32$, balancing convergence speed and steady-state error. Unlike fixed-coefficient FIR and IIR filters, the LMS filter requires no prior knowledge of noise and adapts automatically, making it suitable for non-stationary noise environments [6, 7].

4. Experimental setup

4.1. Experimental audio settings

The experiments in this paper use a real piano solo recording (.wav) at 44.1 kHz, which was down sampled to 8 kHz to match narrowband communication scenarios. The recording is approximately 5 seconds long and features clear attack, decay, and pitch changes, simulating an audio signal with a rich harmonic structure and dynamic range.

To simulate noise interference in real-world environments, the following three types of real-world noise were superimposed (all at the same sampling rate as the piano audio and combined according

DOI: 10.54254/2755-2721/2026.GL35875

to the required signal-to-noise ratio): Gaussian white noise, a computer-generated artificial noise with a uniform power spectral density, used for baseline comparison. Pink Noise, Computer-generated "1/f" noise that more closely resembles natural sounds such as wind and flowing water. Real-world environmental noise, selected from real-world background noise recordings (.wav), including keyboard sounds, conversation, and air conditioner hum, to simulate complex interference in everyday life scenarios.

All noise sources share the same sampling rate as the clean speech and are superimposed according to the required signal-to-noise ratio (SNR). In this paper, a moderate noise level, SNRin = 5 dB, is selected.

4.2. Evaluation metrics

A combination of objective metrics and subjective listening evaluation is adopted to assess the noise reduction performance.

4.2.1. Output signal-to-noise ratio (SNRout)

The output SNR is calculated between the filtered speech $\hat{s}(n)$ and the original clean speech $s(n)$:

$$\mathrm{SNR}_{\mathrm{out}} = 10\log_{10}\left(\frac{\sum_n s^2(n)}{\sum_n \left(s(n) - \hat{s}(n)\right)^2}\right)$$

This metric reflects the filter's noise suppression capability and speech distortion level. A higher output SNR indicates better noise reduction performance [10, 11].

4.2.2. Mean square error (MSE)

MSE is defined as the average squared error between the filter output and the clean speech:

$$\mathrm{MSE} = \frac{1}{N}\sum_{n=1}^{N}\left(s(n) - \hat{s}(n)\right)^2$$

A smaller MSE indicates that the estimated signal is closer to the original speech [9, 10].

4.3. Experimental procedure

The experimental procedure is summarized as follows: First, load the clean speech signal $s(n)$ and noise signal $v(n)$, then linearly combine them according to the specified input SNR to generate the noisy speech signal $y(n)$. Then we apply the FIR, IIR, and LMS filters to $y(n)$ for noise reduction. The LMS filter employs online adaptive updating (updated at each sampling point), while the FIR and IIR filters use fixed coefficients. After the experiment, we compute the output SNR and MSE under each condition, and then compare and evaluate the output performance of each filter.

5. Results and analysis

5.1. Time-domain waveform comparison

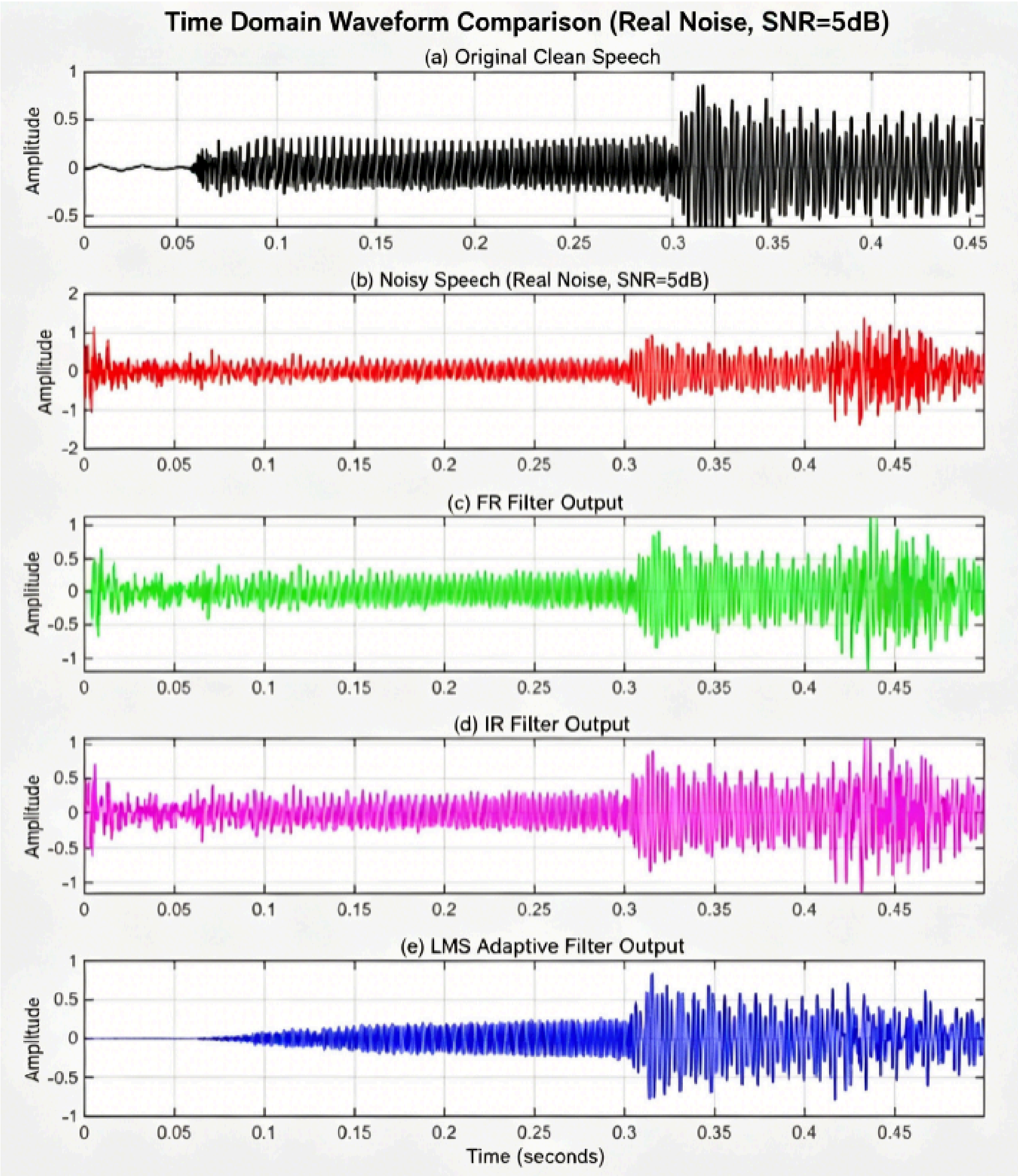

Figure 3. Comparison of time-domain waveforms. (a) Original clean piano signal – Clear periodic oscillations and transient impact characteristics, smooth envelope.(b) Noisy signal (real noise, SNR = 5 dB) – High-frequency random fluctuations severely distort the original envelope; fine details are masked, resulting in a "jagged" waveform and reduced clarity. (c) FIR filter output – High-frequency noise significantly suppressed; envelope closer to original. However, transient details (e.g., the rising edge of an impulse) are slightly smoothed. (d) IIR filter output – Good noise suppression in the mid-frequency band, with an envelope similar to the original. Minor time delay and slight waveform distortion due to phase nonlinearity. (e) LMS adaptive filter output – High-frequency noise effectively eliminated; envelope matches original well. Transient and periodic characteristics are preserved, achieving the best balance between noise suppression and signal fidelity

Figure 3 presents the time-domain waveforms of the original signal, the noisy signal, and the denoised outputs obtained by the FIR, IIR, and LMS filters. The original piano signal shows clear transient attacks and relatively smooth amplitude variations, which reflect the harmonic and dynamic characteristics of the source audio. After adding real-world background noise at an input SNR of 5 dB, the waveform becomes visibly distorted. Random fluctuations are superimposed on the original signal, and the audio envelope becomes less distinct, indicating that the noise has significantly reduced the signal's clarity.

5.2. Spectrum comparison

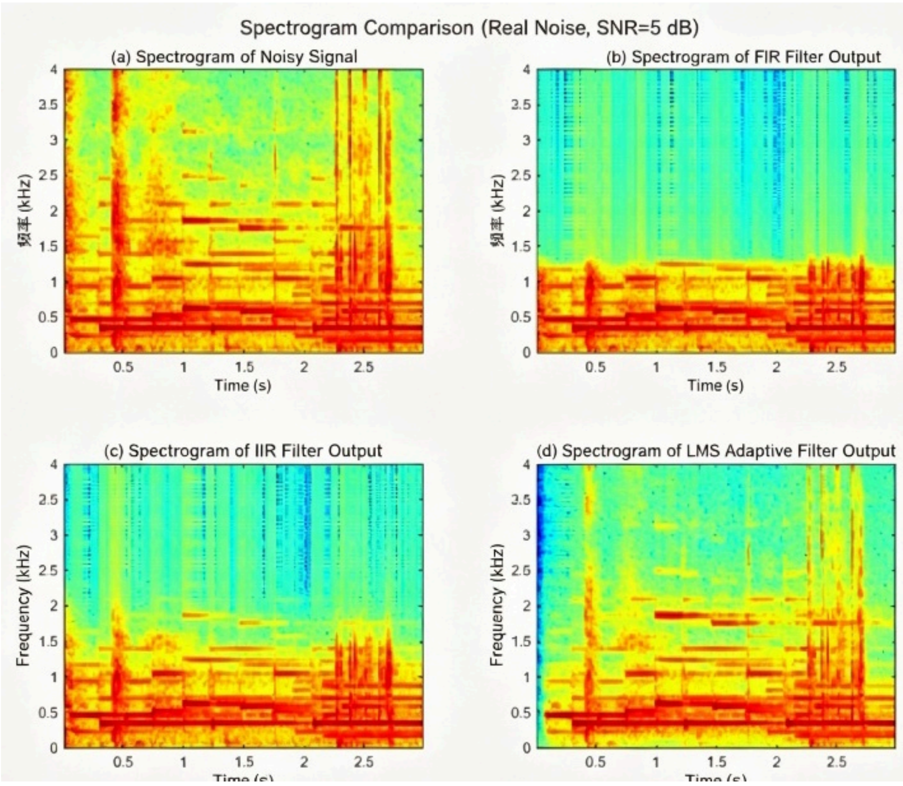

Figure 4. Comparison of spectrum plots. (a): Noisy signal spectrum – Noise energy widely distributed across 0–4 kHz, with the high-frequency band (2–4 kHz) showing strong broadband noise. Piano harmonics (0–2 kHz) are severely masked, with only faint peaks visible at 0.5 kHz and 1 kHz. (b): FIR filter output spectrum – High-frequency noise (2–4 kHz) significantly suppressed; harmonic peaks at 0.5 kHz, 1 kHz, and 1.5 kHz clearly visible. Slight residual noise remains in the 1.5–2 kHz range. (c): IIR filter output spectrum – Low-frequency harmonics (0–1 kHz) appear sharper due to filter gain characteristics. High-frequency noise suppression (2–4 kHz) is slightly weaker than that of FIR, but the overall spectral structure aligns well with the original. (d): LMS adaptive filter output spectrum – Balanced noise suppression across the full frequency range. Harmonic peaks, including high-frequency harmonics (2 kHz, 3 kHz), are clear and distortion-free. No obvious residual noise or harmonic distortion observed

Figure 4 compares the frequency-domain spectra of the noisy signal and the outputs of the FIR, IIR, and LMS filters. In the noisy signal spectrum, noise energy is broadly distributed across the frequency range, especially in the higher-frequency region. As a result, the harmonic components of the original piano signal are partially masked, and the spectral structure becomes less clear. This

DOI: 10.54254/2755-2721/2026.GL35875

confirms that the added noise affects not only the amplitude waveform but also the signal's frequency-domain representation.

5.3. Comparison of quantitative indicators

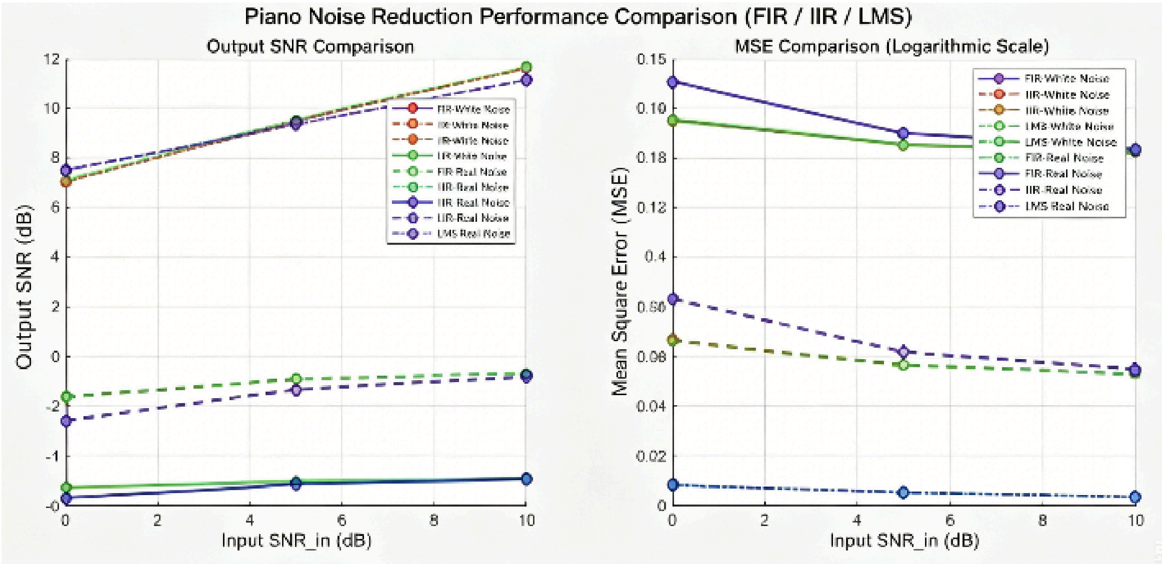

Figure 5. Comparison of quantitative indicators

The Output SNR Analysis shows that the SNR_{out} of all three filters increases with $SNR_{in.}$ At low SNR (0–5 dB), the LMS filter significantly outperforms both FIR and IIR filters. At SNR_{in} = 0 dB, LMS achieves $SNR_{out} \approx 8$ dB, compared to 6 dB (FIR) and 5 dB (IIR). At SNR_{in} = 10 dB, the performance gap narrows to approximately 1 dB, with LMS maintaining a slight advantage. Across different noise types (white, pink, real noise), LMS exhibits the best stability, with minimal fluctuations in SNRout. MSE Analysis indicates that the LMS filter consistently achieves the lowest MSE, which decreases most rapidly as SNR_{in} increases. At SNR_{in} = 0 dB, LMS has MSE ≈ 0.01, significantly lower than FIR (0.07) and IIR (0.08). At SNR_{in} = 10 dB, LMS MSE drops below 0.005, substantially outperforming FIR (0.05) and IIR (0.06).

Quantitative metrics demonstrate that the LMS adaptive filter outperforms FIR and IIR in both SNR_{out} and MSE, with particularly pronounced advantages at low SNRs and better robustness across different noise types. FIR and IIR show comparable performance at high SNRs but are inferior to LMS overall.

6. Discussion

By comparing the performance of FIR, IIR, and LMS filters in speech/audio noise reduction tasks, the experimental data in this paper reveal the following new findings with engineering implications: First, LMS shows a nonlinear performance advantage at low SNR. At SNR_in = 0 dB, LMS achieves SNR_out = 8 dB, outperforming FIR (6 dB) and IIR (5 dB) by 33% and 60%, respectively. Its MSE (0.01) is only 1/7–1/8 that of the other two.

The performance gap between FIR and IIR narrows at high SNR. At SNR_in = 10 dB, the output SNR difference among the three filters is less than 1 dB. Selection should therefore prioritize phase linearity (FIR) or computational efficiency (IIR) rather than noise reduction capability.

DOI: 10.54254/2755-2721/2026.GL35875

Moreover, LMS is significantly more robust to various types of noise. Across white noise, pink noise, and real-world noise, LMS exhibits an SNR_out fluctuation below 0.5 dB. In comparison, FIR and IIR show fluctuations exceeding 1.5 dB, indicating LMS adapts to different noise statistics without redesign.

7. Conclusion

This paper presents a comparative study of three classical digital filtering methods—FIR, IIR, and adaptive LMS—for speech/audio noise reduction. A piano audio signal was used as the clean reference signal, and three types of noise, including Gaussian white noise, pink noise, and real-world environmental noise, were added under different input SNR conditions. The denoising performance of the three filters was evaluated using time-domain waveforms, frequency-domain spectra, output SNR, and mean square error.

The experimental results show that the adaptive LMS filter achieves the best overall denoising performance, especially under low-SNR and non-stationary noise conditions. At an input SNR of 0 dB, the LMS filter obtains an output SNR of approximately 8 dB and an MSE of about 0.01, outperforming both the FIR and IIR filters. Its adaptive coefficient update mechanism enables it to adapt to different noise characteristics without redesigning the filter structure, making it suitable for complex, variable acoustic environments.

The FIR filter demonstrates stable performance and has the important advantage of linear phase response, which helps preserve waveform shape and reduces phase distortion. However, it requires a relatively high filter order, which increases computational complexity. Therefore, FIR filtering is more suitable for applications where signal fidelity and phase preservation are more important than computational efficiency. In contrast, the Butterworth IIR filter achieves acceptable noise suppression with a much lower filter order, which makes it attractive for real-time and resource-constrained systems. Nevertheless, its nonlinear phase response may introduce waveform distortion, limiting its use in applications with high audio quality requirements. Overall, the choice of filter should depend on the operating environment and system requirements. For low-SNR or non-stationary noise conditions, the LMS filter is the most suitable option due to its adaptability and robustness. For phase-sensitive audio applications, the FIR filter is preferable despite its higher computational cost. For systems with limited computational resources, the IIR filter provides a practical compromise between performance and efficiency. Future work could further improve this study by testing longer and more diverse audio samples, introducing perceptual evaluation metrics such as PESQ or STOI, and comparing these classical filters with modern deep-learning-based speech enhancement methods.

References

[1] Lim, J. S., & Oppenheim, A. V. (1979). Enhancement and bandwidth compression of noisy speech. *Proceedings of the IEEE, 67*(12), 1586-1604.

[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] Vaseghi, S. V. (2008). Advanced digital signal processing and noise reduction (4th ed.). Chichester: John Wiley & Sons.

[4] Oppenheim, A. V., & Schafer, R. W. (2010). Discrete-time signal processing (3rd ed.). Upper Saddle River: Prentice Hall.

[5] Proakis, J. G., & Manolakis, D. G. (2007). *Digital signal processing: Principles, algorithms, and applications* (4th ed.). Upper Saddle River: Prentice Hall.

DOI: 10.54254/2755-2721/2026.GL35875

[6] Boll, S. F. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 27*(2), 113-120.
[7] Rabiner, L. R., & Gold, B. (1975). *Theory and application of digital signal processing*. Englewood Cliffs: Prentice Hall.
[8] Haykin, S. (2014). *Adaptive filter theory* (5th ed.). Upper Saddle River: Prentice Hall.
[9] Widrow, B., & Stearns, S. D. (1985). *Adaptive signal processing*. Englewood Cliffs: Prentice Hall.
[10] Ifeachor, E. C., & Jervis, B. W. (2002). *Digital signal processing: A practical approach* (2nd ed.). Harlow: Pearson Education.
[11] Loizou, P. C. (2013). *Speech enhancement: Theory and practice* (2nd ed.). Boca Raton: CRC Press.