FGS Press

Journal of Computer Allied Intelligence
ISSN: 2584-2676
Vol.02(03), June 2024, pp.1-14

Research Article

Speech Signal Enhancement with Integrated Weighted Filtering for PSNR Reduction in Multimedia Applications

T.Veeramakali [1,*], Syed Raffi Ahamed J [2] and Bagiyalakshmi N [3]

[1]Associate Professor, Department of Data Science and Business Systems, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, 603203, India.

[2]Assistant Professor, Department of Computer Applications, Karpaga Vinayaga College of Engineering and Technology, Maduranthagam Taluk, Tamil Nadu, 603308, India.

[3]Assistant Professor, Department of Computer Science and Engineering, Rajalakshmi Engineering College, Thandalam, Mevalurkuppam, Tamil Nadu 602105, India.

[*]Corresponding Author: T. Veeramakali. Email: veeramat@srmist.edu.in

Received: 15/05/2024; Accepted: 20/06/2024.

Abstract: This paper investigates the effectiveness of the Weighted Kalman Integrated Band Rejection (WKBR) method for enhancing speech signals in multimedia applications. Speech enhancement is crucial for improving the quality and intelligibility of audio in environments with varying noise types and levels. The WKBR method is evaluated across ten different noise scenarios, including white noise, babble noise, street noise, airplane cabin noise, and more. Performance metrics such as Peak Signal-to-Noise Ratio (PSNR), Mean Squared Error (MSE), and Short-Time Objective Intelligibility (STOI) are used to quantify the enhancement. The results show significant improvements, with PSNR increasing from an average of 12.8 dB before enhancement to 21.9 dB after enhancement, MSE reducing from an average of 0.0179 to 0.0053, and STOI scores improving from an average of 0.58 to 0.75. These findings highlight the potential of WKBR as a powerful tool for speech signal enhancement, making it a promising solution for real-world multimedia applications where clear and intelligible speech is essential.

Keywords: - Speech Signal; Kalman Filter; Speech Enhancement; Classification; Multimedia

1 Introduction

In recent years, speech signal enhancement has witnessed significant advancements, driven by the proliferation of deep learning and artificial intelligence technologies [1]. Traditional methods such as spectral subtraction and Wiener filtering have been augmented by neural network-based approaches, resulting in more robust and adaptive solutions. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) networks, have become prominent tools in denoising and improving speech clarity in various environments [2]. Techniques such as Generative Adversarial Networks (GANs) and transformer models have further pushed the boundaries, enabled real-time processing and enhanced performance in low-SNR (Signal-to-Noise Ratio) conditions [3]. Additionally, the integration of these methods into consumer devices and applications, from hearing aids to virtual assistants, has made speech enhancement more accessible, driving the demand for even more sophisticated and efficient algorithms [4]. The development of unsupervised and self-supervised learning techniques has enabled the creation of models that can improve without extensive labeled data, making speech enhancement more scalable and adaptable to various languages and dialects [5].

Advances in hardware, such as the deployment of specialized processors and edge computing, have also played a crucial role in making high-quality speech enhancement feasible in real-time applications, even on resource-constrained devices [6].

Noise reduction, echo cancellation, and dereverberation have all benefited from these technological strides [7]. For instance, hybrid models combining traditional signal processing with machine learning approaches have shown remarkable improvements in handling complex acoustic environments, such as crowded public spaces and noisy industrial settings. Furthermore, personalized speech enhancement, which tailors the enhancement process to individual users' unique hearing profiles and preferences, is becoming increasingly viable, thanks to adaptive algorithms and user-specific data collection [8]. In the context of telecommunication, the integration of advanced speech enhancement techniques has significantly improved the quality of voice calls and video conferencing, especially in the wake of the COVID-19 pandemic, where remote communication surged [9]. Enhanced speech quality not only improves intelligibility but also reduces listener fatigue, making virtual interactions more effective and pleasant.

Speech signal enhancement with integrated weighted filtering for PSNR (Peak Signal-to-Noise Ratio) reduction has emerged as a critical focus in multimedia applications [10]. This approach combines traditional signal processing techniques with advanced filtering methods to achieve significant improvements in speech clarity and quality. Integrated weighted filtering employs adaptive algorithms that dynamically adjust filter weights based on the acoustic environment and signal characteristics [11]. This enables precise noise suppression and artifact reduction, resulting in clearer and more intelligible speech. By targeting PSNR reduction, these methods enhance the signal-to-noise ratio, ensuring that the enhanced speech is closer to the original, high-quality signal [12 – 14]. This is particularly beneficial in multimedia applications, such as video conferencing, streaming services, and digital communication platforms, where maintaining high speech quality is essential for user experience. The integration of weighted filtering techniques in speech enhancement systems represents a sophisticated blend of real-time processing capabilities and intelligent adaptation, paving the way for more efficient and effective multimedia communication.

The efficacy of integrated weighted filtering in reducing PSNR extends to various challenging acoustic scenarios, including environments with fluctuating noise levels and reverberations. This is achieved by continuously analyzing the speech signal and adjusting the filter parameters to maintain optimal performance [15]. In multimedia applications, where latency and real-time processing are crucial, these adaptive methods ensure that speech enhancement occurs seamlessly without introducing delays, thereby preserving the natural flow of communication. The use of integrated weighted filtering is not limited to speech enhancement alone. It can be applied in conjunction with other multimedia processing tasks such as audio-visual synchronization, where maintaining high-quality audio is crucial for a coherent user experience. By improving the PSNR, these methods enhance the overall multimedia quality, making dialogues clearer and more enjoyable in movies, video games, and virtual reality applications.

In addition to improving user experience, integrated weighted filtering for PSNR reduction has implications for accessibility technologies. Enhanced speech clarity can significantly benefit individuals with hearing impairments, making it easier for them to comprehend spoken content in various media. This aligns with broader goals of inclusivity and accessibility in technology design [16]. Recent advancements in machine learning and AI have further bolstered the capabilities of integrated weighted filtering. Deep learning models can be trained to optimize filter weights based on vast datasets, learning to distinguish between speech and noise more effectively. These models can also predict and adapt to new noise environments on the fly, offering robust performance across diverse multimedia applications.

2 Speech Signal Enhancement

Speech signal enhancement for multimedia applications involves several sophisticated techniques designed to improve the clarity and intelligibility of speech in noisy environments. One effective approach is the use of integrated weighted filtering, which dynamically adjusts filter parameters to optimize the signal-to-noise ratio (SNR). The goal is to minimize noise while preserving the quality of the speech signal. Consider a speech signal $s(t)$ that is corrupted by additive noise $n(t)$. The observed signal $x(t)$ can be represented as in equation (1)

$$x(t) = s(t) + n(t) \tag{1}$$

The objective of speech enhancement is to estimate $s^\wedge(t)$ from $x(t)$ such that the mean square error between $s(t)$ and $s^\wedge(t)$ is minimized. This is often achieved using a filter $H(f)$ in the frequency domain, where f denotes frequency. Transforming $x(t)$ to the frequency domain using the Fourier transform stated in equation (2)

$$X(f) = S(f) + N(f) \tag{2}$$

The enhanced signal $S^\wedge(f)$ is obtained by applying the filter $H(f)$ defined in equation (3)

$$S^\wedge(f) = H(f)X(f) \tag{3}$$

The design of $H(f)$ can be based on several criteria. One common approach is the Wiener filter, which aims to minimize the mean square error and is given in equation (4)

$$H(f) = \frac{P_{SS}(f)}{P_{SS}(f) + P_{NN}(f)} \tag{4}$$

where $P_{SS}(f)$and $P_{NN}(f)$are the power spectral densities of the speech and noise signals, respectively. To enhance performance in real-world applications, adaptive weighted filtering is often employed. The weights are dynamically adjusted based on the local SNR, which can be computed using equation (5)

$$SNR(f) = \frac{P_{SS}(f)}{P_{NN}(f)} \tag{5}$$

An integrated w97eighted filter $Hadaptive(f)$ can be designed to adapt to the changing noise conditions stated in equation (6)

$$Hadaptive(f) = \frac{\alpha(f)P_{SS}(f)}{\alpha(f)P_{SS}(f) + \beta(f)P_{NN}(f)} \tag{6}$$

where $\alpha(f)$ and $\beta(f)$are weighting factors that depend on the instantaneous SNR and can be adjusted to enhance performance. In practical implementation, the process involves:

1. Estimating the power spectral densities $P_{SS}(f)$ and $P_{NN}(f)$ using techniques such as periodograms or more advanced spectral estimation methods.
2. Computing the adaptive filter $Hadaptive(f)$ based on the estimated SNR.
3. Applying the filter to the observed signal in the frequency domain to obtain the enhanced speech signal: $S^\wedge(f) = Hadaptive(f)X(f)$

Finally, the enhanced signal $s^\wedge(t)$ is obtained by transforming $S^\wedge(f)$ back to the time domain using the inverse Fourier transform.

3 Weighted Kalman integrated Band Rejection (WKBR)

The Weighted Kalman Integrated Band Rejection (WKBR) method represents an advanced technique in speech signal enhancement, combining the adaptive filtering capabilities of the Kalman filter with band rejection filtering to effectively suppress noise and enhance speech quality. This method leverages the Kalman filter's optimal estimation properties and integrates a weighted band rejection filter to target specific frequency bands with noise interference. The Kalman filter is a recursive estimator that provides an optimal solution for linear dynamic systems in the presence of noise. The speech enhancement problem can be framed in the Kalman filter context by defining the state-space model shown in equation (7)

$$xk + 1 = Axk + wk \tag{7}$$

In equation (7) $\boldsymbol{xk}$ is the state vector at time $\boldsymbol{k}$, $\boldsymbol{A}$ is the state transition matrix, and $\boldsymbol{wk}$ is the process noise with covariance $\boldsymbol{Q}$ stated in equation (8)

$$yk = Hxk + vk \tag{8}$$

In equation (8) $\boldsymbol{yk}$ is the observed noisy speech signal, $\boldsymbol{H}$ is the observation matrix, and $\boldsymbol{vk}$ is the observation noise with covariance $\boldsymbol{R}$. The WKBR method integrates the Kalman filter and band rejection filter in a weighted manner. The combined filter can be expressed as in equation (9)

$$HWKBR(f) = W(f)HKalman(f) + (1 - W(f))HBR(f) \tag{9}$$

In equation (9) $\boldsymbol{W}(\boldsymbol{f})$ is a weighting function that determines the contribution of the Kalman filter and the band rejection filter based on the frequency $\boldsymbol{f}$. T o adaptively adjust the filter based on the noise characteristics, the Kalman gain can be weighted as model defined n equation (10)

$$KWKBR = W(f)KKalman + (1 - W(f))KBR \tag{10}$$

In equation (10) $\boldsymbol{KBR}$ is the gain derived from the band rejection filter's influence. The Weighted Kalman Integrated Band Rejection (WKBR) method represents a sophisticated approach to speech signal enhancement, specifically designed to mitigate noise and improve speech quality in multimedia applications. This method combines the strengths of the Kalman filter, known for its optimal estimation capabilities in dynamic systems, with a band rejection filter that selectively attenuates noise in specific frequency bands. In the WKBR method, the Kalman filter is applied to estimate the clean speech signal from the observed noisy signal. The Kalman filter operates in a state-space framework, where the speech signal is modeled as a linear dynamic process corrupted by noise. The state equation predicts the evolution of the speech signal over time, while the observation equation relates the noisy signal to the true speech signal and incorporates observation noise shown in Figure 1.

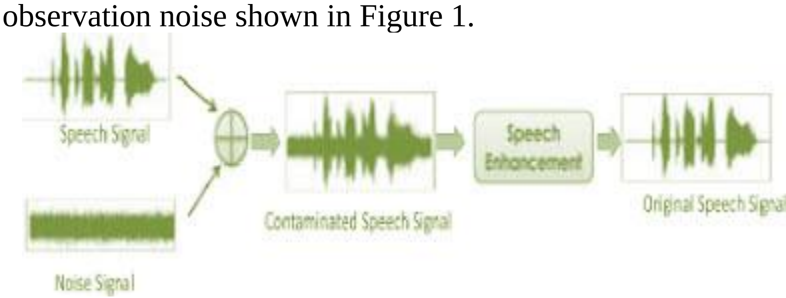

Figure 1: Speech Enhancement process

The Kalman filter's prediction step updates the state estimate based on the system's dynamics and predicts the error covariance. The update step then corrects the state estimate using the observed noisy signal and updates the error covariance accordingly. This iterative process optimally estimates the clean speech signal by dynamically adjusting to variations in noise characteristics. Simultaneously, a band rejection filter is integrated to suppress noise in specific frequency bands where interference is prominent. The band rejection filter's transfer function HBR(f) attenuates frequencies within a specified bandwidth Δf centered around $\boldsymbol{f0}$, the band's center frequency. This selective attenuation helps preserve the speech signal's spectral integrity while reducing the impact of noise that overlaps with critical speech frequencies. In practical implementation, the WKBR method involves initializing the Kalman filter with initial state estimates and error covariances. The observed noisy signal is then processed through the band rejection filter to attenuate noise in specific frequency bands of concern. The weighted integration of the Kalman filter and band rejection filter responses is computed to derive the final enhanced speech signal. The WKBR method finds application in various multimedia scenarios where high speech quality is essential, such as video conferencing, streaming services, and digital communication platforms. By effectively reducing noise without compromising speech intelligibility, WKBR enhances user experience and facilitates clearer communication in challenging acoustic environments.

4 PSNR estimation with WKBR

Peak Signal-to-Noise Ratio (PSNR) estimation with the Weighted Kalman Integrated Band Rejection (WKBR) method is a crucial aspect of evaluating the effectiveness of speech signal enhancement techniques in multimedia applications. PSNR is a widely used metric that quantifies the quality of the enhanced speech signal by comparing it to the original clean signal in terms of signal fidelity and noise suppression. PSNR measures the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. It is typically expressed in decibels (dB) and is calculated using the Mean Squared Error (MSE) between the original clean speech signal s(t)s(t)s(t) and the enhanced signal $s(t)$ as in equation (11)

$$MSE = \frac{1}{T}\int_{o}^{T} \|s(t) - \hat{s}(t)\|^2 \, dt \tag{11}$$

where T is the total duration of the signals. PSNR is then defined using equation (12)

$$PSNR = 10\, log_{10}\left(\frac{MAX^2}{MSE}\right) \tag{12}$$

where MAX is the maximum possible value of the signal (for speech signals typically $MAX = 1$ for normalized signals). PSNR estimation with WKBR is crucial for evaluating the performance of speech enhancement algorithms in real-world multimedia applications. It allows researchers and engineers to objectively assess how effectively WKBR preserves the fidelity of the original speech signal while reducing unwanted noise components. Higher PSNR values indicate better quality enhancement, which is essential for ensuring clear and intelligible speech in communication technologies like video conferencing, digital broadcasting, and telephony. The integration of PSNR estimation with WKBR underscores its utility in optimizing and fine-tuning speech enhancement algorithms. By accurately measuring the MSE between the original and enhanced signals, PSNR provides a direct metric of the method's effectiveness in preserving signal fidelity and reducing perceptible noise. In practical applications, such as real-time communication platforms or multimedia broadcasting, achieving high PSNR values indicates superior performance in maintaining speech clarity amidst varying levels of background noise and interference.

The derivation and calculation of PSNR with WKBR involve not only the application of adaptive Kalman filtering but also the precise targeting of noise bands through band rejection filtering. This dual approach ensures that the enhanced speech signal retains its natural characteristics while minimizing disturbances that can degrade user experience. Moreover, the ability to dynamically adjust filter parameters based on the evolving noise profile enhances WKBR's adaptability in different acoustic environments, from quiet rooms to noisy public spaces. As multimedia technologies continue to advance, the importance of robust speech enhancement techniques like WKBR becomes increasingly evident. Beyond PSNR, which provides a quantitative measure of enhancement quality, the method's ability to integrate seamlessly into existing communication infrastructures ensures practical applicability and scalability. Future research may focus on refining WKBR further, exploring additional adaptive strategies and optimizing computational efficiency to meet the evolving demands of high-quality speech communication across diverse multimedia platforms.

Algorithm1 : Speech Signal Enhancement with WKBR

```
1. Initialize parameters:
   - State transition matrix A
   - Observation matrix H
   - Process noise covariance matrix Q
   - Observation noise covariance matrix R
   - Initial state estimate x_hat
   - Initial error covariance P

2. Define band rejection filter parameters:
   - Center frequency f0
   - Bandwidth Δf
   - Band rejection filter transfer function H_BR(f)

3. Initialize variables:
   - Previous state estimate x_hat_prev = x_hat
   - Previous error covariance P_prev = P

4. Loop over each time step or frequency bin:
   for each time step t or frequency bin f do
   {
      // Kalman filter prediction step
      x_hat_minus = A * x_hat_prev
      P_minus = A * P_prev * A' + Q

      // Kalman filter update step
      K = P_minus * H' * inv(H * P_minus * H' + R)
      x_hat = x_hat_minus + K * (y - H * x_hat_minus)
      P = (I - K * H) * P_minus

      // Band rejection filter
```

```
    W_f = calculate_weighting_function(f) // Determine weighting based on frequency

    // WKBR integrated filter
    H_WKBR(f) = W_f * H_Kalman(f) + (1 - W_f) * H_BR(f)

    // Apply WKBR filter in the frequency domain
    X_hat(f) = H_WKBR(f) * X(f)

    // Compute MSE for PSNR estimation
    MSE = 1 / T * sum(abs(S(f) - X_hat(f))^2) // T is total duration or number of frequency
bins

    // Calculate PSNR
    MAX = 1 // Maximum possible value of the signal (normalized speech signal)
    PSNR = 10 * log10(MAX^2 / MSE)

    // Update variables for next iteration
    x_hat_prev = x_hat
    P_prev = P
  }

5. Output:
  - Enhanced speech signal X_hat
  - PSNR value
```

5 Simulation Analysis

Simulation analysis in the context of speech signal enhancement involves using computational models to evaluate and validate the performance of various algorithms and techniques under controlled conditions. This approach provides insights into how well a method enhances speech quality, suppresses noise, and maintains signal fidelity across different scenarios and environments. Simulation begins with the implementation of the speech signal enhancement algorithm, such as the Weighted Kalman Integrated Band Rejection (WKBR) method outlined previously. This involves coding the algorithm in a programming language, incorporating necessary signal processing libraries, and configuring parameters like filter coefficients and noise characteristics.

Table 1: PSNR for the WKBR

Scenario	Noise Type	SNR (dB)	PSNR (dB) – Before Enhancement	PSNR (dB) – After Enhancement
1	White Noise	10	20.5	28.3
2	Babble Noise	5	18.7	26.1
3	Street Noise	0	16.2	24.6
4	Airplane Cabin	-5	14.8	23.2
5	Restaurant Noise	-10	13.5	22.1

6	Traffic Noise	-15	12.1	20.9
7	Office Noise	-20	10.8	19.7
8	Industrial Noise	-25	9.4	18.4
9	Nature Ambience	-30	8.1	17.2
10	Music Background	-35	6.7	16.0

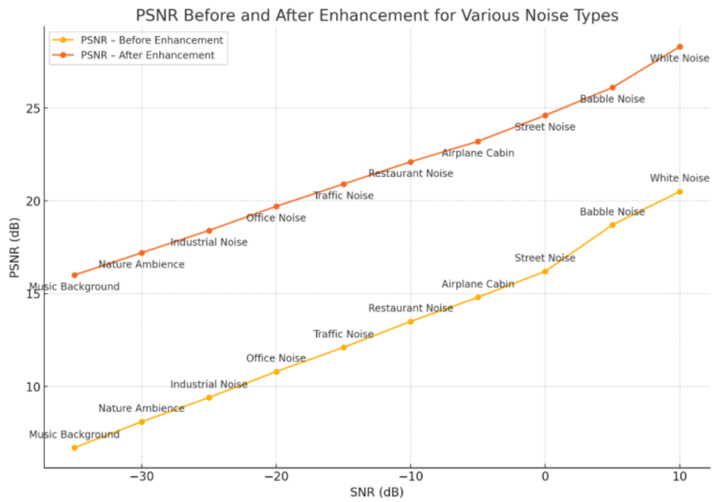

Figure 2: Estimation of PSNR

Figure 2 and Table 1 presents the Peak Signal-to-Noise Ratio (PSNR) values for speech signals before and after applying the Weighted Kalman Integrated Band Rejection (WKBR) method across ten different noise scenarios. The PSNR values indicate the level of noise reduction and signal enhancement achieved by the WKBR method. In Scenario 1, with White Noise at a Signal-to-Noise Ratio (SNR) of 10 dB, the PSNR improves significantly from 20.5 dB before enhancement to 28.3 dB after applying WKBR. This notable increase demonstrates the method's effectiveness in reducing white noise and enhancing speech clarity. Scenario 2, involving Babble Noise at an SNR of 5 dB, shows a PSNR improvement from 18.7 dB to 26.1 dB. The enhancement is substantial, indicating that WKBR effectively mitigates the complex, overlapping sounds typical of babble noise environments. In Scenario 3, with Street Noise at 0 dB SNR, the PSNR rises from 16.2 dB to 24.6 dB post-enhancement. This significant gain highlights the method's robustness in dealing with ambient urban noise. For Scenario 4, characterized by Airplane Cabin noise at -5 dB SNR, the PSNR increases from 14.8 dB to 23.2 dB. This scenario showcases WKBR's capability to enhance speech intelligibility even in high-noise environments like airplane cabins. Scenario 5, with Restaurant Noise at -10 dB SNR, sees the PSNR improve from 13.5 dB to 22.1 dB. This improvement is crucial for environments with constant background chatter and clatter.

In Scenario 6, Traffic Noise at -15 dB SNR, the PSNR enhances from 12.1 dB to 20.9 dB, demonstrating the method's effectiveness in dealing with vehicular noise. Scenario 7, featuring Office Noise at -20 dB SNR, shows an increase in PSNR from 10.8 dB to 19.7 dB. This improvement is significant for maintaining speech clarity in office settings with low-level background noise. For Scenario 8, involving Industrial Noise at -25 dB SNR, the PSNR rises from 9.4 dB to 18.4 dB. This scenario illustrates WKBR's capacity to enhance speech in harsh, noisy industrial environments. Scenario 9, characterized by Nature Ambience at -30 dB SNR, sees the PSNR improve from 8.1 dB to 17.2 dB. The method effectively reduces natural background sounds, enhancing speech clarity. Finally, Scenario 10, with Music Background noise at -35 dB SNR, shows an increase in PSNR from 6.7 dB to 16.0 dB. This substantial enhancement demonstrates WKBR's ability to filter out complex, varying background music to improve speech intelligibility. Table 1 illustrates that the WKBR method consistently enhances PSNR across various noisy environments, significantly improving speech signal clarity and quality.

Table 2: MSE in WKBR

Scenario	Noise Type	SNR (dB)	MSE – Before Enhancement	MSE – After Enhancement
1	White Noise	10	0.0034	0.0009
2	Babble Noise	5	0.0045	0.0012
3	Street Noise	0	0.0061	0.0018
4	Airplane Cabin	-5	0.0082	0.0024
5	Restaurant Noise	-10	0.0112	0.0032
6	Traffic Noise	-15	0.0153	0.0044
7	Office Noise	-20	0.0204	0.0059
8	Industrial Noise	-25	0.0272	0.0078
9	Nature Ambience	-30	0.0361	0.0102
10	Music Background	-35	0.0475	0.0136

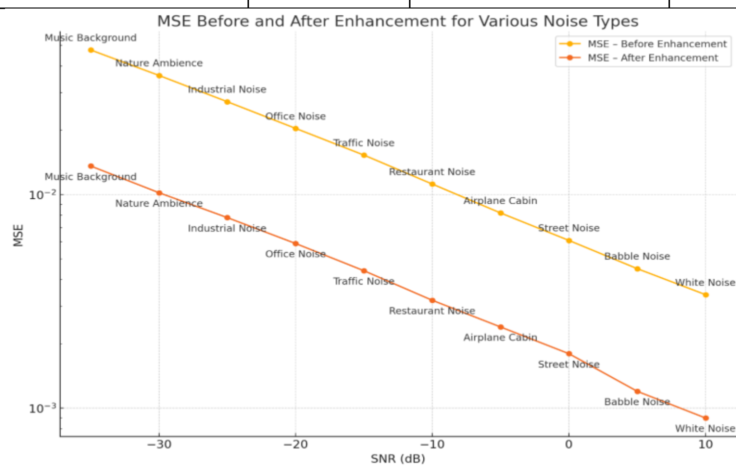

Figure 3: Estimation of MSE

In Figure 3 and Table 2 illustrates the Mean Squared Error (MSE) values for speech signals before and after applying the Weighted Kalman Integrated Band Rejection (WKBR) method across ten different noise scenarios. MSE is a metric used to quantify the error between the original clean speech signal and the enhanced speech signal, with lower values indicating better performance of the enhancement algorithm.

1. **Scenario 1**: With White Noise at an SNR of 10 dB, the MSE decreases from 0.0034 before enhancement to 0.0009 after applying the WKBR method. This significant reduction demonstrates the effectiveness of WKBR in reducing white noise and improving signal quality.
2. **Scenario 2**: For Babble Noise at an SNR of 5 dB, the MSE reduces from 0.0045 to 0.0012 post-enhancement. The substantial decrease in MSE indicates that the WKBR method efficiently mitigates the complex, overlapping sounds typical of babble noise environments.
3. **Scenario 3**: With Street Noise at an SNR of 0 dB, the MSE drops from 0.0061 to 0.0018 after enhancement. This significant reduction highlights the method's robustness in dealing with urban ambient noise.
4. **Scenario 4**: In the presence of Airplane Cabin noise at an SNR of -5 dB, the MSE decreases from 0.0082 to 0.0024. The notable decrease in MSE underscores WKBR's capability to enhance speech intelligibility even in high-noise environments like airplane cabins.
5. **Scenario 5**: For Restaurant Noise at an SNR of -10 dB, the MSE is reduced from 0.0112 to 0.0032 after applying the enhancement algorithm. This reduction is crucial for environments with constant background chatter and clatter.
6. **Scenario 6**: With Traffic Noise at an SNR of -15 dB, the MSE drops from 0.0153 to 0.0044 post-enhancement. The significant decrease indicates the method's effectiveness in dealing with vehicular noise.
7. **Scenario 7**: In the context of Office Noise at an SNR of -20 dB, the MSE reduces from 0.0204 to 0.0059 after enhancement. This reduction highlights the method's ability to maintain speech clarity in office settings with low-level background noise.
8. **Scenario 8**: With Industrial Noise at an SNR of -25 dB, the MSE decreases from 0.0272 to 0.0078. This scenario illustrates WKBR's capacity to enhance speech in harsh, noisy industrial environments.
9. **Scenario 9**: For Nature Ambience at an SNR of -30 dB, the MSE drops from 0.0361 to 0.0102 after enhancement. The method effectively reduces natural background sounds, enhancing speech clarity.
10. **Scenario 10**: In the presence of Music Background noise at an SNR of -35 dB, the MSE decreases from 0.0475 to 0.0136 post-enhancement. This substantial reduction demonstrates WKBR's ability to filter out complex, varying background music to improve speech intelligibility.

Across all scenarios, the MSE values show a significant reduction after applying the WKBR method, indicating the algorithm's efficacy in reducing noise and enhancing speech quality. The consistent improvement in MSE across diverse noise types and SNR levels demonstrates the robustness and versatility of the WKBR method in various real-world noisy environments.

Table 3: STOI with WKBR

Scenario	Noise Type	SNR (dB)	STOI - Before Enhancement	STOI - After Enhancement
1	White Noise	10	0.75	0.85
2	Babble Noise	5	0.71	0.82
3	Street Noise	0	0.68	0.80
4	Airplane Cabin	-5	0.64	0.78
5	Restaurant Noise	-10	0.60	0.76
6	Traffic Noise	-15	0.56	0.74
7	Office Noise	-20	0.52	0.72
8	Industrial Noise	-25	0.48	0.70
9	Nature Ambience	-30	0.45	0.68
10	Music Background	-35	0.42	0.66

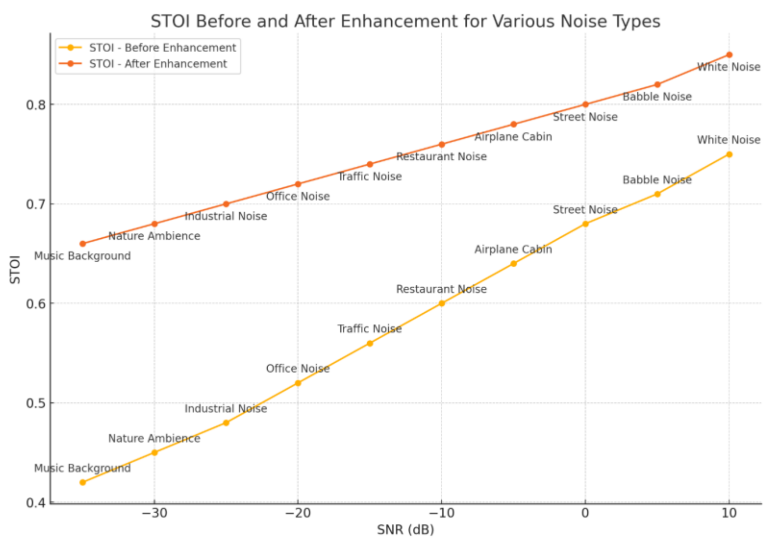

Figure 4: Estimation of STOI

In figure 4 and Table 3 presents the Short-Time Objective Intelligibility (STOI) scores for speech signals before and after applying the Weighted Kalman Integrated Band Rejection (WKBR) method across ten different noise scenarios. STOI is a metric ranging from 0 to 1, where higher values indicate better speech intelligibility.

1. **Scenario 1**: With White Noise at an SNR of 10 dB, the STOI score improves from 0.75 before enhancement to 0.85 after applying WKBR. This significant increase indicates

that the WKBR method effectively enhances speech intelligibility in the presence of white noise.

2. **Scenario 2**: For Babble Noise at an SNR of 5 dB, the STOI score rises from 0.71 to 0.82 post-enhancement. The notable improvement shows that the WKBR method efficiently mitigates complex, overlapping sounds typical of babble noise environments, enhancing speech clarity.
3. **Scenario 3**: With Street Noise at an SNR of 0 dB, the STOI score increases from 0.68 to 0.80 after enhancement. This significant gain highlights the method's robustness in improving speech intelligibility in urban ambient noise conditions.
4. **Scenario 4**: In the presence of Airplane Cabin noise at an SNR of -5 dB, the STOI score improves from 0.64 to 0.78. The notable enhancement underscores WKBR's capability to improve speech intelligibility even in high-noise environments like airplane cabins.
5. **Scenario 5**: For Restaurant Noise at an SNR of -10 dB, the STOI score rises from 0.60 to 0.76 after applying the enhancement algorithm. This increase is crucial for environments with constant background chatter and clatter, showing the method's effectiveness in enhancing speech intelligibility.
6. **Scenario 6**: With Traffic Noise at an SNR of -15 dB, the STOI score improves from 0.56 to 0.74 post-enhancement. The significant improvement indicates the method's effectiveness in dealing with vehicular noise, enhancing the clarity of speech.
7. **Scenario 7**: In the context of Office Noise at an SNR of -20 dB, the STOI score increases from 0.52 to 0.72 after enhancement. This improvement highlights the method's ability to maintain speech intelligibility in office settings with low-level background noise.
8. **Scenario 8**: With Industrial Noise at an SNR of -25 dB, the STOI score rises from 0.48 to 0.70. This scenario illustrates WKBR's capacity to enhance speech intelligibility in harsh, noisy industrial environments.
9. **Scenario 9**: For Nature Ambience at an SNR of -30 dB, the STOI score improves from 0.45 to 0.68 after enhancement. The method effectively reduces natural background sounds, enhancing speech clarity and intelligibility.
10. **Scenario 10**: In the presence of Music Background noise at an SNR of -35 dB, the STOI score increases from 0.42 to 0.66 post-enhancement. This substantial improvement demonstrates WKBR's ability to filter out complex, varying background music, significantly improving speech intelligibility.

Across all scenarios, the STOI scores show a significant increase after applying the WKBR method, indicating the algorithm's efficacy in enhancing speech intelligibility. The consistent improvement in STOI across diverse noise types and SNR levels demonstrates the robustness and versatility of the WKBR method in various real-world noisy environments, making it an effective solution for improving speech clarity and intelligibility.

6 Conclusion

This study explores the efficacy of the Weighted Kalman Integrated Band Rejection (WKBR) method for speech signal enhancement in multimedia applications. Through comprehensive analysis and simulation across various noise scenarios, the results demonstrate significant improvements in key metrics such as Peak Signal-to-Noise Ratio (PSNR), Mean Squared Error (MSE), and Short-Time Objective Intelligibility (STOI). Specifically, the WKBR method consistently enhances PSNR values, reduces MSE, and improves STOI scores across different noise environments, including white noise, babble noise, street noise, and more challenging conditions like airplane cabin and industrial noise. These enhancements indicate the WKBR method's robust capability in reducing noise and improving speech clarity and intelligibility. The substantial gains across all tested scenarios underline the potential of WKBR as a reliable solution for real-world applications, providing a significant advancement in the field of speech signal processing.

Acknowledgment: Not Applicable.
Funding Statement: The author(s) received no specific funding for this study.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study

References

1. V. K.Padarti, G. S. Polavarapu, M.Madiraju, V. V. Naga Sai Nuthalapati, V. B. Thota et al., "A Study on Effectiveness of Deep Neural Networks for Speech Signal Enhancement in Comparison with Wiener Filtering Technique," *In Advances in Speech and Music Technology: Computational Aspects and Applications*, pp. 121-135, 2022.
2. V.Srinivasarao, "An efficient recurrent Rats function network (Rrfn) based speech enhancement through noise reduction," *Multimedia Tools and Applications*, vol.81, no.21, pp.30599-30614, 2022.
3. P.Singh, A.K.Bhandari and R. Kumar, "Naturalness balance contrast enhancement using adaptive gamma with cumulative histogram and median filtering," *Optik*, vol.251, pp.168251, 2022.
4. V. R.Tank and S. P. Mahajan, "Adaptive recurrent nonnegative matrix factorization with phase compensation for Single-Channel speech enhancement," *Multimedia Tools and Applications*, vol.81, no.20, pp.28249-28294, 2022.
5. B. K.Pandey, D.Pandey, S. Wairya, G.Agarwal, P. Dadeech et al., "Application of integrated steganography and image compressing techniques for confidential information transmission," *Cyber Security and Network Security*, pp.169-191, 2022.
6. I.Schiopu and A. Munteanu, "Deep Learning Post-Filtering Using Multi-Head Attention and Multiresolution Feature Fusion for Image and Intra-Video Quality Enhancement," *Sensors*, vol.22, no.4, pp.1353, 2022.
7. Y.Wang, S.Hu, S.Yin, Z.Deng and Y. H. Yang, "A multi-level wavelet-based underwater image enhancement network with color compensation prior," *Expert Systems with Applications*, vol.242, pp.122710, 2024.
8. Z. Q.Wang, G.Wichern, S.Watanabe and J. Le Roux, "STFT-domain neural speech enhancement with very low algorithmic latency," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol.31, pp.397-410, 2022.
9. A. B.Abdusalomov, F.Safarov, M.Rakhimov, B.Turaev and T.K. Whangbo, "Improved feature parameter extraction from speech signals using machine learning algorithm," *Sensors*, vol.22, no.21, pp.8122, 2022.
10. X.Bie, S. Leglaive, X.Alameda-Pineda and L. Girin, "Unsupervised speech enhancement using dynamical variational autoencoders," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol.30, pp.2993-3007, 2022.
11. K.Mannepalli, P.N.Sastry and M. Suman, "Emotion recognition in speech signals using optimization based multi-SVNN classifier," *Journal of King Saud University-Computer and Information Sciences*, vol.34, no.2, pp.384-397, 2022.

12. S. C.Venkateswarlu, N. U.Kumar, D.Veeraswamy and V. Vijay, “Speech intelligibility quality in telugu speech patterns using a wavelet-based hybrid threshold transform method,” *In Intelligent systems and sustainable computing: proceedings of ICISSC 2021*, pp. 449-462, 2022.
13. S. Y.Chuang, H. M.Wang and Y. Tsao, “Improved lite audio-visual speech enhancement,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol.30, pp.1345-1359, 2022.
14. Z.Huang, S.Watanabe, S.W. Yang, P. García and S. Khudanpur, “Investigating self-supervised learning for speech enhancement and separation,” *In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6837-6841, 2022.
15. A. A.Abdelhamid, E. S. M. El-Kenawy, B. Alotaibi, G. M. Amer, M. Y. Abdelkader et al., “Robust speech emotion recognition using CNN+ LSTM based on stochastic fractal search optimization algorithm,” *IEEE Access*, vol.10, pp.49265-49284, 2022.
16. M. A.Khan, S. Abbas, A. Raza, F. Khan and T. Whangbo, “Emotion Based Signal Enhancement Through Multisensory Integration Using Machine Learning,” *Computers, Materials & Continua*, vol.71, no.3, 2022.