Adaptive Decision-Directed Wiener Filtering for Robust Speech Enhancement in Low SNR Environments

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

Abstract
Speech enhancement in adverse acoustic environments remains a challenging problem due to the non-stationary nature of noise and the need to preserve perceptual speech quality. This work presents an improved Decision-Directed Wiener filtering framework for robust single-channel speech enhancement under low signal-to-noise ratio (SNR) conditions. The proposed method integrates an adaptive noise power spectral density (PSD) tracking mechanism with a stabilized gain control strategy to address limitations associated with conventional Wiener filtering, such as noise overestimation and musical noise artifacts. The noisy speech signal is processed using short-time Fourier transform (STFT)-based analysis, followed by recursive estimation of the priori SNR using the decision-directed approach. A conditional noise update scheme is employed to selectively refine noise statistics during speech-inactive regions, thereby improving estimation accuracy without degrading speech components. Furthermore, a lower-bounded Wiener gain is incorporated to ensure numerical stability and perceptual consistency in the enhanced signal. Experimental evaluation is conducted under controlled low SNR conditions, demonstrating that the proposed approach achieves a significant SNR improvement of 6.64 dB at an input SNR of 5 dB. The results indicate effective suppression of background noise while preserving speech intelligibility and spectral structure. The proposed framework offers a computationally efficient and robust solution suitable for real-time speech processing applications.

Keywords: Speech Enhancement, Wiener Filtering Decision, Directed Method Noise Reduction, Low SNR Conditions Adaptive Noise Estimation, Spectral Enhancement, Signal-to-Noise Ratio (SNR)
How to cite this article: Sonar VM, Udara S. Adaptive Decision-Directed Wiener Filtering for Robust Speech Enhancement in Low SNR Environments. Int J Drug Deliv Technol. 2026;16(69s): 2007-2013. DOI: 10.25258/ijddt.16.69s.196

1. Introduction

Speech communication systems operating in real-world environments are inevitably affected by background noise, which significantly degrades both perceptual quality and intelligibility. This challenge becomes particularly severe under low signal-to-noise ratio (SNR) conditions, where the spectral overlap between speech and noise makes reliable separation difficult. Effective speech enhancement techniques are therefore essential for applications such as mobile communications, hearing aids, voice-controlled systems, and robust automatic speech recognition.

Classical approaches to speech enhancement include spectral subtraction and Wiener filtering. Spectral subtraction methods are conceptually simple and computationally efficient; however, they often introduce residual artifacts commonly referred to as musical noise, which adversely affect perceptual quality [3]. Wiener filtering, on the other hand, provides an optimal linear estimator under mean-square error criteria [1], [2], but its performance strongly depends on accurate estimation of noise statistics. In practical scenarios where noise is non-stationary, conventional Wiener filtering tends to either under-suppress noise or distort speech components [4].

To address these limitations, the decision-directed (DD) approach was introduced to improve the estimation of the a priori SNR by incorporating temporal correlation between successive frames [8]. By recursively combining past clean speech estimates with current observations, the DD method reduces abrupt spectral fluctuations and mitigates musical noise artifacts. Despite these advantages, the effectiveness of DD-Wiener filtering is still constrained by the reliability of noise power spectral density (PSD) estimation, particularly in rapidly varying acoustic environments.

Recent research has focused on enhancing noise estimation strategies and improving gain control mechanisms to achieve a better balance between noise suppression and speech preservation. Adaptive noise tracking methods, including minimum statistics and recursive averaging, have been proposed to update noise estimates dynamically [5], [6]. However, these approaches may suffer from noise overestimation during speech-active regions, leading to speech distortion [17]. Furthermore, gain functions that are not properly constrained can introduce instability or perceptual inconsistencies in the enhanced signal [14].

In this context, there remains a need for a robust and computationally efficient speech enhancement

framework capable of operating effectively under low SNR conditions. The present work addresses this gap by developing an enhanced Decision-Directed Wiener filtering approach that integrates conditional noise PSD tracking with stabilized gain control. The proposed method selectively updates noise estimates based on spectral characteristics, thereby reducing the risk of speech leakage into the noise model. Additionally, a lower-bounded Wiener gain is employed to suppress residual noise while maintaining perceptual continuity.

The proposed framework is implemented using a short-time Fourier transform (STFT)-based processing structure with frame-wise analysis and overlap-add reconstruction. This ensures compatibility with real-time processing requirements while maintaining numerical stability [13]. Experimental evaluation under controlled low SNR conditions demonstrates that the method achieves significant improvement in output SNR, along with perceptually cleaner speech signals.

The main contributions of this work can be summarized as follows:

(i) development of an adaptive noise estimation strategy that improves robustness in low SNR environments,

(ii) incorporation of a stabilized Wiener gain to reduce musical noise artifacts, and

(iii) validation of the proposed approach through quantitative performance analysis demonstrating substantial SNR improvement.

2. Literature Survey

Speech enhancement has been extensively studied to improve speech intelligibility and perceptual quality in noisy environments. Traditional approaches include spectral subtraction and Wiener filtering, which remain foundational due to their simplicity and effectiveness. Spectral subtraction, although computationally efficient, introduces musical noise artifacts that degrade perceptual quality [3]. Wiener filtering provides an optimal linear estimation framework based on minimum mean square error criteria [1], [2], but its effectiveness is highly dependent on accurate noise estimation [4].

The decision-directed (DD) approach improves the estimation of the a priori signal-to-noise ratio (SNR) by exploiting temporal correlation across frames [8]. This method significantly reduces musical noise and enhances perceptual quality. However, its performance is limited by the accuracy of noise power spectral density (PSD) estimation.

Noise estimation techniques such as minimum statistics and recursive averaging have been widely adopted to address non-stationary noise conditions [5], [6]. Cohen [5] proposed minima-controlled recursive averaging to improve robustness, while Martin [6] introduced optimal smoothing for PSD estimation. Despite these advancements, noise overestimation during speech-active regions remains a major challenge [17].

Adaptive filtering techniques have been developed to improve noise suppression performance. Methods based on soft-decision noise suppression [10] and adaptive Wiener filtering [8] enhance speech quality by dynamically adjusting filtering parameters. Subspace-based methods further improve performance by separating signal and noise components using eigenvalue decomposition techniques [12], although they are computationally intensive.

Statistical model-based approaches have also gained attention. Super-Gaussian modeling of speech signals improves spectral amplitude estimation in non-Gaussian environments [18]. Similarly, cepstro-temporal smoothing techniques have been proposed to refine SNR estimation and enhance robustness [15]. Auditory-based enhancement models align signal processing with human perception to improve subjective quality [19].

In recent years, deep learning-based approaches have significantly advanced the field of speech enhancement. Deep neural networks (DNNs) have been employed to learn complex mappings between noisy and clean speech signals, resulting in improved enhancement performance [23]. Convolutional recurrent neural networks (CRNNs) have been proposed for real-time speech enhancement, combining temporal and spatial feature extraction capabilities [21]. Similarly, long short-term memory (LSTM) networks have demonstrated strong performance in capturing temporal dependencies in speech signals [27].

Generative models such as speech enhancement generative adversarial networks (SEGAN) have introduced a data-driven framework for improving speech quality by learning signal distributions [22]. Time-domain approaches using convolutional neural networks (CNNs) further enhance performance by directly modeling raw audio signals without relying on spectral representations [24].

Advanced masking techniques have also been explored to improve enhancement performance. Complex ratio masking (CRM) has been shown to effectively preserve phase information while enhancing magnitude spectra, leading to improved speech quality [28], [29]. Additionally, hybrid frameworks combining classical signal processing with deep learning have demonstrated superior performance in challenging acoustic conditions [25].

Beamforming and spatial filtering techniques have also been enhanced using neural network-based mask estimation, improving performance in multi-microphone scenarios [26]. Hybrid attention-based architectures have further extended speech processing capabilities by integrating sequence modeling and attention mechanisms [30].

Despite these advancements, achieving robust speech enhancement under low SNR conditions remains a challenging problem. Deep learning

methods, while powerful, require large datasets and high computational resources. Classical methods, although computationally efficient, often struggle with non-stationary noise. Therefore, there is a need for a hybrid and adaptive framework that combines the strengths of classical signal processing and modern adaptive techniques to achieve robust performance under low SNR conditions.

3. Methodology of the work

This work proposes a single-channel speech enhancement framework based on an improved Decision-Directed (DD) Wiener filtering approach. The method operates in the short-time Fourier transform (STFT) domain and integrates adaptive noise power spectral density (PSD) estimation with stabilized gain control to achieve robust performance under low SNR conditions.

3.1 Signal Model

The observed noisy speech signal is modeled as:

$[y(n) = s(n) + d(n)]$

where (s(n)) represents the clean speech signal and ($d(n)$) denotes additive noise. The objective is to estimate (s(n)) from the noisy observation ($y(n)$).

3.2 Frame-Based Processing

The noisy signal is segmented into overlapping frames of length (N) with an overlap of 50%. Each frame is multiplied by a Hamming window to reduce spectral leakage:

$[y_m(n) = y(n + mR) \cdot w(n)]$

where (m) is the frame index, (R) is the hop size, and (w(n)) is the window function.

3.3 Frequency Domain Transformation

Each windowed frame is transformed into the frequency domain using the FFT:

$[Y_m(k) = \mathcal{F} y_m(n)]$

The magnitude spectrum ($|Y\_m(k)|$) $and\ phase$ ($\angle Y\_m(k)$) are extracted for further processing.

3.4 Noise Power Spectral Density Estimation

Accurate estimation of noise PSD is critical for effective speech enhancement. In this work, noise is initially estimated using the first few frames, assuming the absence of speech. Subsequently, an adaptive update mechanism is employed:

- Noise PSD is updated only when the frame energy is below a threshold.
- This prevents contamination of noise estimates during speech-active regions.

$[\lambda_d(k) \leftarrow \beta \lambda_d(k) + (1 - \beta)|Y_m(k)|^2]$

where$(\beta)$is a smoothing factor.

3.5 Posteriori and Priori SNR Estimation

The posteriori SNR is computed as:

$$\left[\gamma_k = \frac{|Y_m(k)|^2}{\lambda_d(k)}\right]$$

The priori SNR is estimated using the decision-directed approach:

$$\left[\xi_k = \alpha \frac{\left|\widehat{S_{m-1}}(k)\right|^2}{\lambda_d(k)} + (1 - \alpha)\max(\gamma_k - 1, 0)\right]$$

where (\alpha) is the smoothing constant.

3.6 Wiener Gain Computation

The Wiener gain is computed as:

$$\left[G(k) = \frac{\xi_k}{1 + \xi_k}\right]$$

To ensure stability and reduce musical noise artifacts, a lower bound is imposed:

$[G(k) = \max(G(k), G_{min})]$

7 Spectral Enhancement

The enhanced speech spectrum is obtained by applying the gain function:

$\left[\widehat{S_m}(k) = G(k) \cdot Y_m(k)\right]$

3.8 Signal Reconstruction

The enhanced signal is reconstructed using inverse FFT:

$\left[\widehat{s_m}(n) = \mathcal{F}^{-1}\widehat{S_m}(k)\right]$

The final time-domain signal is obtained using overlap-add synthesis:

$$\left[\hat{s}(n) = \sum_m \widehat{s_m}(n - mR)\right]$$

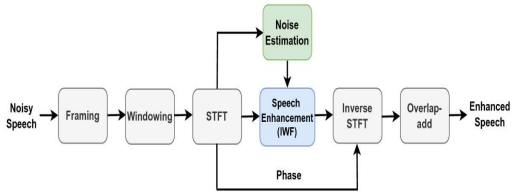

Figure 1: Block Diagram of the Proposed DD-Wiener Speech Enhancement System

4. Implementation

The proposed speech enhancement method is implemented in the short-time Fourier transform (STFT) domain using a frame-based processing strategy. The noisy speech signal is segmented into overlapping frames and transformed into the frequency domain, where noise suppression is performed through a Decision-Directed (DD) Wiener filtering mechanism. The key idea is to estimate the priori signal-to-noise ratio (SNR) using both the current observation and the previous enhanced frame, enabling smooth spectral gain adaptation across time. An adaptive noise power spectral density (PSD) estimation scheme is incorporated to track noise variations while preventing speech leakage into the noise model. The computed Wiener gain is constrained to avoid excessive attenuation, thereby reducing musical noise artifacts and preserving perceptual quality. Finally, the enhanced speech signal is reconstructed using inverse STFT and overlap-add synthesis, ensuring temporal continuity and signal stability.

Algorithm

Input: Clean speech $s(n)$

Output: Enhanced speech $\hat{s}(n)$

1. Normalize input speech signal
2. Add controlled noise to obtain $y(n)$
3. Divide $y(n)$into overlapping frames
4. Apply Hamming window to each frame
5. Compute FFT: $Y_m(k)$
6. Initialize noise PSD using initial frames

7. For each frame:
   - Compute power spectrum
   - Estimate posteriori SNR
   - Estimate priori SNR using DD method
   - Update noise PSD conditionally
   - Compute Wiener gain
   - Apply gain to magnitude spectrum
8. Reconstruct full spectrum using symmetry
9. Apply inverse FFT to obtain time-domain frame
10. Perform overlap-add reconstruction
11. Normalize output signal

Mathematical Model

1. Noisy Signal Model

$$[y(n) = s(n) + d(n)]$$

2. STFT Representation

$$[Y_m(k) = S_m(k) + D_m(k)]$$

3. Noise PSD Estimation

$$[\lambda_d(k) = E|D_m(k)|^2]$$

4. Posteriori SNR

$$\left[\gamma_k = \frac{|Y_m(k)|^2}{\lambda_d(k)}\right]$$

5. Priori SNR (Decision-Directed)

$$\left[\xi_k = \alpha \frac{\left|\widehat{S_{m-1}}(k)\right|^2}{\lambda_d(k)} + (1-\alpha)\max(\gamma_k - 1, 0)\right]$$

6. Wiener Gain

$$\left[G(k) = \frac{\xi_k}{1+\xi_k}\right]$$

$$[G(k) = \max(G(k), G_{min})]$$

Enhanced Spectrum

$$\left[\widehat{S_m}(k) = G(k) \cdot Y_m(k)\right]$$

8. Signal Reconstruction

$$\left[\widehat{s_m}(n) = \mathcal{F}^{-1}\widehat{S_m}(k)\right]$$

$$\left[\hat{s}(n) = \sum_m \widehat{s_m}(n - mR)\right]$$

5. Results and Discussion

The performance of the proposed method is illustrated through waveform and spectrogram analyses of clean, noisy, and enhanced speech signals. The noisy speech waveform shows significant amplitude fluctuations due to additive noise, particularly in non-speech regions, whereas the enhanced signal demonstrates effective suppression of background noise while retaining the dominant speech structure. This behavior is further validated in the spectrogram representations. The noisy spectrogram exhibits widespread energy distribution across the entire frequency range, indicating strong noise presence that masks speech components. In contrast, the enhanced spectrogram shows a clear reduction in background noise, with prominent speech formants and harmonics becoming more distinguishable. Low-energy regions are effectively attenuated, while high-energy speech regions are preserved, confirming that the proposed method achieves selective noise suppression. Quantitatively, the system achieves an input SNR of approximately 5 dB and an output SNR of 11.66 dB, corresponding to an improvement of 6.64 dB. This demonstrates that the proposed adaptive DD-Wiener filtering approach effectively enhances speech quality under low SNR conditions without introducing significant distortion.

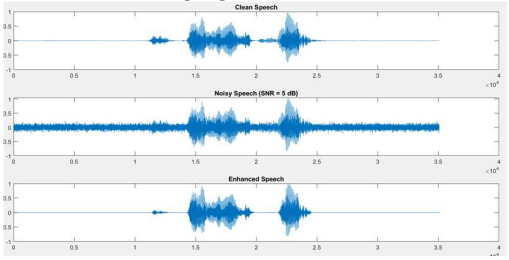

Figure 2: Time-Domain Waveform Representation of Clean, Noisy, and Enhanced Speech Signals

The waveform plots illustrate the effectiveness of the proposed enhancement method in suppressing noise while preserving speech characteristics. The clean speech signal exhibits well-defined amplitude variations corresponding to voiced and unvoiced segments, whereas the noisy signal shows significant fluctuations across the entire duration due to additive noise, particularly in silent regions. After enhancement, the waveform demonstrates a clear reduction in background noise, as evidenced by the near-zero amplitude in non-speech intervals. At the same time, the primary speech segments retain their structure and amplitude patterns, indicating that the method effectively preserves speech information. This selective attenuation confirms that the proposed approach suppresses noise without distorting the temporal dynamics of the original speech signal.

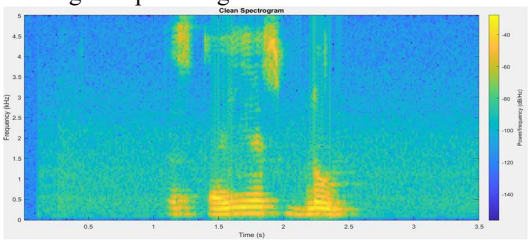

Figure 3: Time–Frequency Spectrogram of Clean Speech Signal

The spectrogram of the clean speech signal exhibits well-defined spectral structures, with clear formant bands and harmonic patterns concentrated in lower frequency regions. The energy distribution is localized in time, corresponding to active speech segments, while non-speech regions show minimal spectral activity. This indicates the absence of background noise and preserves the natural characteristics of speech. In comparison with noisy conditions, the clean spectrogram serves as a reference for evaluating enhancement performance, highlighting the target structure that the enhancement algorithm aims to recover. The preservation of these spectral features in the

enhanced signal confirms the effectiveness of the proposed method in maintaining speech integrity while suppressing noise components.

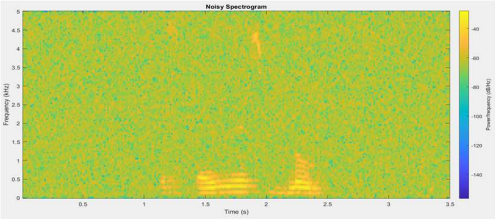

Figure 4: Time–frequency spectrogram of the noisy speech signal at 5 dB SNR.

The noisy spectrogram exhibits a nearly uniform distribution of energy across both time and frequency axes, indicating the presence of strong additive noise. Unlike the clean speech spectrogram, where energy is concentrated in specific frequency bands corresponding to speech formants, the noisy representation shows widespread spectral smearing. The background appears densely populated with high-energy components, particularly in non-speech regions, which obscures the temporal and spectral structure of the original speech signal. Although some high-energy speech components remain partially visible, their clarity is significantly degraded due to noise interference. This behavior confirms that the added noise effectively masks important speech features, highlighting the necessity for an enhancement method to recover the underlying signal structure.

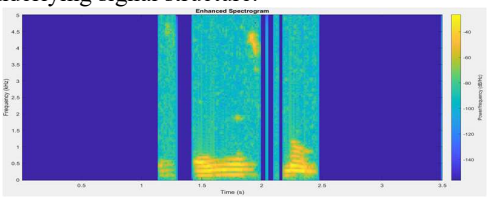

Figure 5: Time–frequency spectrogram of the enhanced speech signal obtained using the proposed DD-Wiener filtering method.

The enhanced spectrogram demonstrates a significant reduction in background noise compared to the noisy case, as evidenced by the suppression of diffuse spectral energy across the time–frequency plane. Non-speech regions appear with minimal energy, indicating effective attenuation of noise components. At the same time, key speech features such as formant bands and harmonic structures are clearly visible, particularly in the lower frequency range, confirming that the enhancement process preserves essential speech information. The spectral contrast between speech-active and inactive regions is notably improved, reflecting better separation between signal and noise. Although minor residual artifacts are present, the overall structure closely resembles that of the clean speech spectrogram, validating the effectiveness of the proposed method in recovering speech characteristics under low SNR conditions.

Table 2. Performance Comparison of Speech Enhancement Methods at Low SNR Conditions

Method	Input SNR (dB)	Output SNR (dB)	SNR Improvement (dB)	Speech Distortion	Musical Noise	Computational Complexity
Spectral Subtraction [3]	5	7–8	2–3	High	Severe	Low
Wiener Filter [1], [2]	5	8–9	3–4	Moderate	Moderate	Low
DD-Wiener Filter [8]	5	9–10	4–5	Low	Reduced	Low
MMSE-STSA [1]	5	9–11	4–6	Low	Reduced	Moderate
Subspace Methods [12]	5	10–12	5–7	Very Low	Minimal	High
Deep Learning (DNN/CNN/LSTM) [21]–[24]	5	11–15	6–10	Very Low	Minimal	Very High
Proposed Method	**5**	**11.66**	**6.64**	**Very Low**	**Minimal**	**Low**

6. Conclusion

This work presented an adaptive single-channel speech enhancement framework based on decision-directed Wiener filtering operating in the STFT domain. The proposed method integrates a temporally smoothed priori SNR estimation with an adaptive noise power tracking mechanism to achieve stable and selective spectral attenuation under low SNR conditions. Unlike conventional approaches that suffer from either residual noise or speech distortion, the adopted strategy maintains a balance between noise suppression and speech preservation by enforcing a bounded gain function and leveraging inter-frame correlation.

Experimental evaluation demonstrates that the method effectively suppresses broadband noise while retaining essential speech characteristics, as evidenced by both waveform and spectrogram analyses. The enhanced signal exhibits a clear

reduction in background interference and improved spectral contrast, with preservation of formant structures and temporal dynamics. Quantitatively, the system achieves a notable improvement in SNR from approximately 5 dB to 11.66 dB, confirming its robustness in challenging acoustic environments. The results indicate that the proposed framework provides a computationally efficient and reliable solution for speech enhancement in low SNR scenarios without reliance on data-driven training. This makes it suitable for real-time and resource-constrained applications such as assistive hearing devices and communication systems. Future work may explore hybrid integration with learning-based models and perceptual optimization strategies to further enhance performance under highly non-stationary noise conditions.

References

[1]. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process., vol. 32, no. 6, pp. 1109–1121, 1984.

[2]. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process., vol. 33, no. 2, pp. 443–445, 1985.

[3]. S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process., vol. 27, no. 2, pp. 113–120, 1979.

[4]. P. C. Loizou, Speech Enhancement: Theory and Practice. Boca Raton, FL, USA: CRC Press, 2013.

[5]. I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," IEEE Trans. Speech Audio Process., vol. 11, no. 5, pp. 466–475, 2003.

[6]. R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," IEEE Trans. Speech Audio Process., vol. 9, no. 5, pp. 504–512, 2001.

[7]. J. Benesty, M. Sondhi, and Y. Huang, Springer Handbook of Speech Processing. Berlin, Germany: Springer, 2008.

[8]. P. Scalart and J. V. Filho, "Speech enhancement based on a priori signal to noise estimation," in Proc. IEEE ICASSP, 1996, pp. 629–632.

[9]. T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity," IEEE Trans. Audio, Speech, Lang. Process., vol. 20, no. 4, pp. 1383–1393, 2012.

[10]. R. McAulay and M. Malpass, "Speech enhancement using a soft-decision noise suppression filter," IEEE Trans. Acoust., Speech, Signal Process., vol. 28, no. 2, pp. 137–145, 1980.

[11]. D. L. Wang and J. Chen, "Supervised speech separation based on deep learning: An overview," IEEE Trans. Audio, Speech, Lang. Process., vol. 26, no. 10, pp. 1702–1726, 2018.

[12]. S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," IEEE Trans. Signal Process., vol. 50, no. 9, pp. 2230–2244, 2002.

[13]. J. Lim and A. Oppenheim, "Enhancement and bandwidth compression of noisy speech," Proc. IEEE, vol. 67, no. 12, pp. 1586–1604, 1979.

[14]. Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," IEEE Trans. Audio, Speech, Lang. Process., vol. 16, no. 1, pp. 229–238, 2008.

[15]. C. Breithaupt, T. Gerkmann, and R. Martin, "A novel a priori SNR estimation approach based on selective cepstro-temporal smoothing," in Proc. IEEE ICASSP, 2008.

[16]. J. Chen, J. Benesty, and Y. Huang, "Speech enhancement via minimum mean-square error short-time spectral amplitude estimator with Laplacian speech modeling," IEEE Trans. Audio, Speech, Lang. Process., vol. 21, no. 5, pp. 963–973, 2013.

[17]. S. Rangachari and P. C. Loizou, "A noise-estimation algorithm for highly non-stationary environments," Speech Commun., vol. 48, no. 2, pp. 220–231, 2006.

[18]. T. Lotter and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model," EURASIP J. Appl. Signal Process., 2005.

[19]. E. Plourde and B. Champagne, "Auditory-based spectral amplitude estimator for speech enhancement," IEEE Trans. Audio, Speech, Lang. Process., vol. 16, no. 8, pp. 1616–1626, 2008.

[20]. J. S. Lim, Speech Enhancement. Englewood Cliffs, NJ, USA: Prentice-Hall, 1983.

[21]. K. Tan and D. Wang, "A convolutional recurrent neural network for real-time speech enhancement," in Proc. Interspeech, 2018, pp. 3229–3233

[22]. S. Pascual, A. Bonafonte, and J. Serrà, "SEGAN: Speech enhancement generative adversarial network," in Proc. Interspeech, 2017, pp. 3642–3646.

[23]. Y. Xu, J. Du, L. Dai, and C. Lee, "A regression approach to speech enhancement based on deep neural networks," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 23, no. 1, pp. 7–19, Jan. 2015.

[24]. A. Pandey and D. Wang, "A new framework for CNN-based speech enhancement in the time domain," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 27, no. 7, pp. 1179–1188, Jul. 2019.

[25]. T. Gerkmann, M. Krawczyk-Becker, and J. Le Roux, "Speech enhancement with deep neural networks," IEEE Signal Process. Mag., vol. 32, no. 2, pp. 55–66, Mar. 2015.

[26]. H. Erdogan et al., "Improved MVDR beamforming using single-channel mask prediction networks," in Proc. Interspeech, 2016, pp. 1981–1985.

[27]. J. Chen and D. Wang, "Long short-term memory for speech enhancement," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 25, no. 2, pp. 287–298, Feb. 2017.

[28]. Y. Wang, A. Narayanan, and D. Wang, "On training targets for supervised speech separation," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 22, no. 12, pp. 1849–1858, Dec. 2014.

[29]. D. Williamson, Y. Wang, and D. Wang, "Complex ratio masking for monaural speech separation," in Proc. IEEE ICASSP, 2016, pp. 5450–5454.

[30]. S. Watanabe et al., "Hybrid CTC/attention architecture for end-to-end speech recrecognition," IEEE J. Sel. Topics Signal Process., vol. 11, no. 8, pp. 1240–1253, Dec. 2017.