IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 December 2024, accepted 21 February 2025, date of publication 27 February 2025, date of current version 7 March 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3546434

RESEARCH ARTICLE

A Hybrid Speech Enhancement Technique Based on Discrete Wavelet Transform and Spectral Subtraction

YASIR IQBAL[1], TAO ZHANG[1], TEDDY SURYA GUNAWAN[2], (Senior Member, IEEE), AGUS PRATONDO[3], (Senior Member, IEEE), XIN ZHAO[1], YANZHANG GENG[1], MIRA KARTIWI[4], (Member, IEEE), NASIR SALEEM[5], AND SAMI BOUROUIS[6]

[1]School of Electrical and Information Engineering, Tianjin University, Nankai, Tianjin 300072, China
[2]Department of Electrical and Computer Engineering, International Islamic University Malaysia, Kuala Lumpur 53100, Malaysia
[3]Department of Multimedia Engineering Technology, School of Applied Science, Telkom University, Bandung 40257, Indonesia
[4]Information Systems Department, International Islamic University Malaysia, Kuala Lumpur 53100, Malaysia
[5]Department of Electrical Engineering, Faculty of Engineering and Technology (FET), Gomal University, Dera Ismai Khan 29050, Pakistan
[6]Department of Information Technology, College of Computers and Information Technology, Taif University, Taif 21944, Saudi Arabia

Corresponding authors: Teddy Surya Gunawan (tsgunawan@iium.edu.my) and Yanzhang Geng (gregory@tju.edu.cn)

This work was supported in part by Telkom University under Grant PU.007/AKD2.4/PPM/2022, and in part by the National Natural Science Foundation of China under Grant 62271344. The authors extend their appreciation to the Taif University, Saudi Arabia, for supporting this work through project number (TU-DSPP-2024-60).

ABSTRACT Speech quality and intelligibility are often severely degraded by background noise in communication systems such as hearing aid (HA) and speech recognition technologies, compromising their effective use. In low Signal-to-Noise Ratio (SNR) conditions, various approaches and algorithms are applied to improve speech quality and intelligibility. This study introduces a novel hybrid speech enhancement framework that synergistically integrates Spectral Subtraction (SS) and Discrete Wavelet Transform (DWT) to address limitations of traditional noise reduction techniques. Traditional SS methods generate musical noise artifacts due to static noise estimation, while standard DWT approaches struggle with selective thresholding and static coefficient processing. To overcome these challenges, the proposed SS method incorporates iterative noise estimation, Voice Activity Detection (VAD), minimum statistics for dynamic noise adaptation, Spectral Smoothing and phase-aware spectral reconstruction. Concurrently, in the enhanced DWT method adaptive noise refinement with phase-aware soft thresholding is employed to detail coefficients, and the Spatial and Intensity filter is adapted to the approximation coefficients to improve low-frequency features and retain structural integrity while reducing distortion. The integrated SS-DWT framework significantly improves noise suppression, reduces musical noise artifacts, and enhances signal clarity as it leverages the strengths of both phase-aware spectral reconstruction in improved SS and phase-aware soft thresholding in DWT, particularly in adaptive noise refinement and thresholding. Proposed speech enhancement network evaluated and experimental results show that the hybrid SS-DWT method outperforms existing systems, achieving up to 34.15 dB in SDR, 0.98 in STOI, and 3.84 in PESQ, demonstrating significant improvements in speech quality under various noisy conditions.

INDEX TERMS Speech enhancement, adaptive noise refinement, DWT, spectral subtraction, phase-aware construction, music noise.

I. INTRODUCTION

Noise-induced intelligibility is an unavoidable problem with communication technologies like hearing aids. Hearing clear speech in noisy situations is a challenge for many hearing aids users. During a hearing sequence, these devices pick up excess or unnecessary speech impulses and amplify each one of them. In that scenario, the candidate may grow weary and attempt to take off the hearing aid, choosing not to hear his or her surroundings rather than hear everything [1]. Thus, improving speech that has been distorted by ambient noise presents a problem for hearing aid applications [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Manuel Rosa-Zurera.

One of the main concerns of HA users is noise-induced hearing impairment, which speech enhancement (SE) may help to solve [3], [4]. Hearing aid technologies need to amplify sounds and eliminate background noise in order to improve speech intelligibility [5]. Several SE techniques were invented to enhance the clarity of distorted speech signals in communication platforms such as HA devices, where the desired signal is received without distorting the informative signal [6], [7]. Trade-off between noise reduction and voice distortion, however, limits its performance [8]. The most typical situation is a single channel framework, in which the voice and noise originate from different sources and are difficult for the microphone to record and control. A correlation between the recorded signal, noise, and speech gives rise to such a circumstance. Computing complexity and implementation cost are essential considerations for putting recommended speech enhancement algorithms into practice in real-time applications such as mobile communications, hearing aids, and intelligent hearing protection. Most popular technique for improving speech is the wavelet transform [9]. The wavelet thresholding technique exclusively splits signals into low-frequency ranges and works on the assumption that the processed data are independent, uniformly distributed Gaussian noise inputs [10]. Among commonly used methods for enhancing speech is Discrete Wavelet Transform (DWT), which works incredibly well for processing non-stationary signals [11], [12]. To be more precise, DWT is used in conjunction with thresholding and spectrum subtraction techniques for noise suppression applications. In particular, DWT is used in combination with spectrum subtraction and thresholding techniques to improve speech in many applications. Choosing appropriate threshold values is one of the challenges in DWT-thresholding approaches. In recent applications, several decision-making technologies are used for the process of estimating noise within signal's sub-bands, and either thresholding or spectral subtraction is used to clean the voice signal [13], [14], [15]. Furthermore, spectral subtraction has been the subject of a tremendous deal of research in recent times due to its ease of use and abstraction for portable devices like mobile phones and hearing aids [16], [17].

The remaining paper is arranged in sections as follows:
Related work is discussed in Section II. The methodology for the proposed speech enhancement approaches is explained in Section III. Details on Speech Signal Collection and Pre-processing are discussed in Section IV. The Analysis of Experimental Results is shown in Section V. In Section VI Comparisons with recent methods based on SS-DWT combination are evaluated. The conclusions are contained in Section VII.

II. RELATED WORK

Spectral Subtraction (SS) and wavelet transform (WT) are filtering techniques that have been used for Background noise reduction in audio signals in many recent works. Several technologies were developed to reduce the various categories of non-stationary noise present in original signals and attempts were made to improve understandability of the audio signal. However, because traditional methods are unable to remove the noise associated with audio frequency, maintaining the unique characteristics of the original audio remains difficult. This section provides a summary of some important research in this field that has been done by various teams.

A. WAVELET TRANSFORM

When analyzing non-stationary noisy signals, the discrete wavelet transform (DWT) is utilized because it produces a signal with a higher frequency resolution and a time-frequency representation [18]. A noisy signal is broken down by the DWT into detail and approximation coefficients at each level, with level of decomposition determined by an input signal. Moreover, choosing the appropriate mother wavelet [19] and decomposition level [20] determines how well the DWT denoising performs. The formulation of the DWT [21] takes into account both the temporal and frequency aspects of the signal to be studied, in contrast to Fourier transform (FT), that only considers the frequency portions. Wavelet transform is used in Wavelet Thresholding Denoising (WTD) [22] to divide time-domain data into sub-bands. Subsequently, the generated wavelet coefficients (subbands) undergo thresholding. In [23], data reduction and noise robustness in recognition were achieved concurrently by applying the DWT [24] to audio signals and merely conserving generated approximation fraction. The authors of [25] produced a clear signal by removing noise from the signal using DWT semi-soft thresholding. Recently, a few hybrid approaches have been devised to enhance Wavelet-based speech enhancement performance [26], [27], [28].

B. SPECTRAL SUBTRACTION

Adaptive filtering, wavelet transform, spectral subtraction, Kalman filtering, Wiener filtering, etc. are some of the currently popular speech improvement techniques [29], [30], [31]. Due to its ease of calculation and strong real-time performance, spectral subtraction is frequently employed in speech noise reduction processing research [36]. However, at low SDRs, it tends to produce music noise and it directly decreases intelligibility of speech signal [32], [33]. Researchers have suggested enhancements to the conventional spectrum subtraction method in response to this issue. These include the adaptive gain average spectral subtraction proposed by Gustasson [35], the multiband spectral subtraction offered by Berouti [34], and others. Spectral subtraction filters have been employed by numerous studies recently to enhance the quality of speech that is deteriorated in real time [37], [38]. This is why using DWT and SS approaches to obtain effective Speech Enhancement methods with an enhanced SDR value for loud speech signals in hearing aids is motivated by the above-mentioned factors.

C. PROBLEM STATEMENT

In communication systems like speech recognition and hearing aids in low SNR situations, noise reduction frameworks such as Spectral Subtraction (SS) and DWT approaches are used to minimize background noise. However, the intrinsic issue with the creation of musical noise involves Spectral Subtraction. Nevertheless, SE employing DWT-based thresholds might not be suitable for all kinds of non-stationary noise because of different frequencies and time-based scales. Such concerns may result in insufficient noise reduction in real-time communication systems which could impact speech signal de-noising evaluation metrics.

D. AIMS OF CURRENT STUDY

In order to get improved speech quality and intelligibility in terms of SNR, SDR, MSE, PESQ, and STOI and reduce noise for audio signals in communication networks like wireless communication and hearing aid systems, this investigation provides speech enhancement methods based on a combination of SS and DWT algorithms. Unlike traditional methods, in proposed SS filter tools like iterative Spectral Subtraction procedure, Voice Activity Detection (VAD), minimum statistics for adaptive noise tracking, Spectral Smoothing and phase-aware spectral reconstruction are used whereas in DWT phase-aware soft thresholding applied to the detail coefficients, along wit spatial and intensity filters adapted to the approximation coefficients are integrated to enhance speech quality and intelligibility in terms of SNR, SDR, MSE, PESQ and STOI. This removes the inefficiency of DWT while handling noise throughout varying frequencies and time scales, as well as the music noise in SS.

E. OUR CONTRIBUTION

This paper introduces a novel hybrid approach that combines Spectral Subtraction (SS) with Discrete Wavelet Transform (DWT), addressing key limitations of traditional methods like musical noise artifacts and insufficient noise reduction across varying frequencies and time scales. Initially, an iterative Spectral Subtraction procedure with dynamic noise estimation, Voice Activity Detection (VAD), minimum statistics for adaptive noise tracking, Spectral Smoothing and phase-aware spectral reconstruction is applied to reduce the noise spectrum and minimize musical noise artifacts. After applying SS, DWT is employed for adaptive noise refinement with phase-aware soft thresholding applied to the detail coefficients, along with spatial and intensity filters adapted to the approximation coefficients to enhance low-frequency features and retain structural integrity while reducing distortion. The proposed hybrid method significantly improves speech enhancement performance, as evidenced by experimental results showing superior metrics in SDR (up to 34.15 dB), STOI (0.98), and PESQ (3.84), compared to existing systems. We assessed each filter's de-noising performance, including the wavelet types, proposed hybrid noise reduction techniques, and the Spectral Subtraction filter.

III. METHODOLOGY

A. DISCRETE WAVELET TRANSFORM (DWT)

One of most important temporal-frequency analysis methods that has a big impact on Speech Enhancement is Discrete Wavelet Transform (DWT). Because DWT can break down an input signal into its component sub-signals over a wide range of scales and frequency ranges, it can effectively extract and amplify relevant information from the signal. During decomposition phase, a signal is defined as a collection of orthonormal wavelet functions that comprise a wavelet basis [39]. It was reported that the better performances were achieved by employing mother-wavelet functions such as Symlets while addressing problems related to enhancing voice signals [40].

1) DISCRETE WAVELET TRANSFORM (DWT) WITH PHASE-AWARE SOFT THRESHOLDING

The proposed framework uses the Discrete Wavelet Transform (DWT) to decompose noisy signals into corresponding components such as approximation and detail coefficients, to reduce noise and enhance speech quality and intelligibility. Unlike the fixed thresholding and simple filtering that is commonly employed in traditional approaches, the proposed DWT framework incorporates adaptive noise refinement together with efficient complex phase-aware soft thresholding techniques to enhance the detail coefficients. The spatial and intensity are also applied on the approximation coefficients for enhancing the low-frequency detailed and reducing the artifacts of noise, and it makes the design framework more flexible and effective for noise reduction. The proposed DWT incorporates two different methods for DWT thresholds which are as follows:
This framework comprises two different base threshold methodologies:
a. Discrete Wavelet Transform - Constant Threshold (DWT-CT)
b. Discrete Wavelet Transform - Adaptive Threshold (DWT-AT)

Both of the methods discussed here apply wavelet decomposition first and then process the detail coefficients employing their specific thresholding method. An innovative adaptive phase-aware soft thresholding procedure is then used to further optimize these thresholds by modifying threshold values concerning the magnitude and phase of the coefficients to further enhance signal quality. In addition, Spatial and Intensity filtering is used for the approximation coefficients, as well as a residual refinement and fusion process, which is overlooked to some degree in standard denoising methods. The detailed methodology incorporated in this paper for the proposed DWT SE method is shown below in Figure 1, which consists of the following steps:

2) WAVELET DECOMPOSITION

The noisy signal $x_{\text{noisy}}(t)$ is decomposed using the Discrete Wavelet Transform (DWT) into: Approximation coefficients

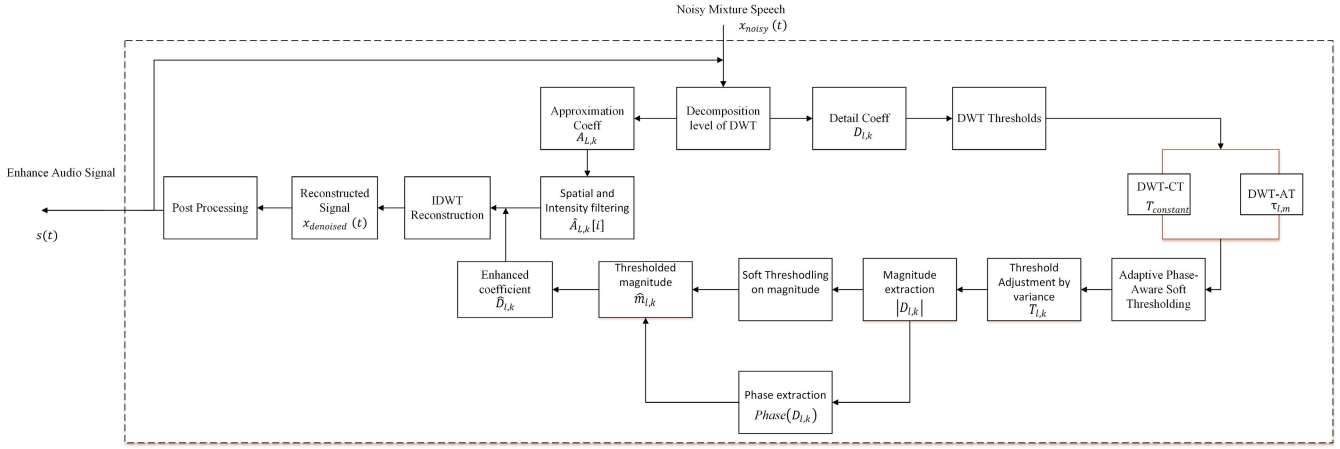


FIGURE 1. Illustration of proposed Discrete Wavelet Transform (DWT) speech enhancement scheme.

$A_{(L,k)}$ which represents low-frequency, large-scale features of the signal and Detail coefficients $D_{(l,k)}$ which represents high-frequency, fine-scale components where noise is typically concentrated. The decomposition can be expressed as:

$$x_{\text{noisy}}(t) = \sum_{k} A_{(L,k)}\phi_{(L,k)}(t) + \sum_{l=1}^{L}\sum_{k} D_{(l,k)}\psi_{(l,k)}(t) \quad (1)$$

where in the decomposition phase, $\phi_{(L,k)}(t)$ is scaling functions representing low-frequency components and $\psi_{(l,k)}(t)$ is the wavelet basis functions representing high-frequency components respectively, both depend on the selected mother wavelet. In this study, Symlet-5 (Sym5), Symlet-10 (Sym10), and Biorthogonal-1.1 (Bior1.1) wavelets are employed for decomposition due to their distinctive characteristics in noise suppression and signal preservation. The decomposition level L is set to 2 for both proposed DWT-CT and DWT-AT methodologies. The choice of wavelet family influences both the decomposition and the denoising process, as each wavelet has unique properties of smoothness, symmetry, and localization. This decomposition enables multi-resolution analysis of the signal, separating noise from signal components across scales. Unlike traditional wavelet-based methods, this framework refines both low-frequency and high-frequency components for superior noise suppression.

3) BASE THRESHOLDS

a: DISCRETE WAVELET TRANSFORM - CONSTANT THRESHOLD (DWT-CT)

This methodology employs a constant global base threshold named DWT-CT for noise suppression, followed by adaptive phase-aware soft thresholding to refine detail coefficients. The DWT-Constant Threshold (DWT-CT)T_{constant} is calculated using statistical properties of the detail coefficients $D_{(l,k)}$ at first decomposition level:

$$T_{\text{constant}} = \sqrt{2\log N} \cdot \text{MAD}_{\text{detail}} \quad (2)$$

where N represents length of the signal (or the number of coefficients) and $\text{MAD}_{\text{detail}}$ represents Median Absolute Deviation of detail coefficients denoted by $D_{(l,k)}$ which is used to estimate level for the noise in the noisy mixture signal. The Median Absolute Deviation is computed as:

$$\text{MAD}_{\text{detail}} = \text{median}\left(\left|D_{(l,k)} - \text{median}(D_{(l,k)})\right|\right) \quad (3)$$

This measure provides an estimate of the noise energy in the signal which is then used to determine the constant threshold. This threshold is chosen by determining the noise level of the signal from this measure. Unlike other methods that use rigid thresholding or simple noise elimination techniques, this threshold provides a uniform noise elimination method across all levels of decomposition. This is however done by employing phase-aware adaptive soft thresholding which learns from the local noise characteristics and signal structures. By this combination, the framework obtains the merits of the existing approach to achieve both global and local contexts at the same time.

b: DISCRETE WAVELET TRANSFORM - ADAPTIVE THRESHOLD (DWT-AT)

This work employs a base threshold named DWT-Adaptive Threshold (DWT-AT) with L1-L2 norms to suppress noise and achieve sparse and smooth results. An adaptive base threshold $\tau_{(l,k)}$is derived to well balance the sparseness and smoothness of the high-frequency detail coefficients $D_{(l,k)}$ with the help of L1-L2 regularization. The expression for regularizing coefficients is as follows:

$$\hat{D}_{(l,m)} = \frac{\max\left(|D_{(l,m)}| - \alpha\tau_{(l,m)}, 0\right)}{1+\beta} \quad (4)$$

In above eq. (4), α is L1 penalty for sparsity control, which controls signal to shrink toward zero, where larger value increases sparsity and small values retains more coefficients. The second parameter is β L2 smoothness penalty: It provides a balance between removing noise from the signal and retaining the signal, with higher values leading to smoother signals but a potential loss of important details. So an empirical tuning of alpha (α) and beta (β) or cross validation

is applied to balance sparsity and smoothness. In proposed DWT-AT the Adaptive threshold is computed as:

$$\tau_{(l,m)} = \sigma.\sqrt{2\log N} \tag{5}$$

where Noise standard deviation is computed as:

$$\sigma = \frac{\text{median}\left(|D_{(l,m)}|\right)}{0.6745} \tag{6}$$

In this approach, some modifications have been applied such as through adaptive thresholding which adjusts itself adaptively to the local noise variance to enhance noise suppression while phase-aware adaptive soft thresholding is used in order to preserve both magnitude and phase for enhanced structural details. L1-L2 regularization also allows for a proper balance between noise elimination and information retention. It can handle non-stationary noise as compared to other techniques while at the same time maintaining adaptively as well as signal quality.

4) ADAPTIVE PHASE-AWARE SOFT THRESHOLDING FOR DETAIL COEFFICIENTS

In the proposed DWT SE method the adaptive phase aware soft thresholding is subsequently applied to the detail coefficients $D_{(l,k)}$ at each decomposition level l after base thresholding (adaptive or constant). In this soft thresholding step, the threshold gets set depending on the local rate of noise, and the signal's structural information is maintained through using the phase along with magnitude of the coefficients. Adaptive phase aware soft thresholding is then applied to the detail coefficients $D_{(l,k)}$ at each decomposition level l after base thresholding (constant or adaptive). In this step, the threshold is adjusted to the local noise characteristics, and the structural integrity of the signal is preserved by taking into account both magnitude and phase of the coefficients. It undergoes the following steps:

a: THRESHOLD ADJUSTMENT BY VARIANCE

The base threshold T_{constant} or $\tau_{(l,m)}$ is adjusted adaptively at each decomposition level according to the variance of the coefficients:

$$T_{(l,k)} = T_{\text{baseline}}\sqrt{\text{Var}(D_{(l,k)})} \tag{7}$$

where T_{baseline} is the initial base threshold (either constant or adaptive) and $\text{Var}(D_{(l,k)})$ is the variance of the detail coefficients $D_{(l,k)}$ at each level l, capturing the noise characteristics in that region of the signal.

b: MAGNITUDE AND PHASE EXTRACTION

The next step involves extracting the magnitude and phase of each detail coefficient $D_{(l,k)}$. These coefficients are expressed in polar form:

$$|D_{(l,k)}| = \sqrt{R(D_{(l,k)})^2 + I(D_{(l,k)})^2} \tag{8}$$

$$Phase(D_{(l,k)}) = \tan^{-1}\left(\frac{I(D_{(l,k)})}{R(D_{(l,k)})}\right) \tag{9}$$

Here $|D_{(l,k)}|$ is the Magnitude of the coefficient, $\text{Phase}(D_{(l,k)})$ is the Phase of the coefficient, and $R(D_{(l,k)})$ and $I(D_{(l,k)})$ represent the real and imaginary parts of the coefficient, respectively.

c: SOFT THRESHOLDING ON MAGNITUDE

Now apply soft thresholding to the magnitude of the coefficients to suppress noise. Soft thresholding is a form of shrinkage that reduces the magnitude of coefficients that are smaller than the threshold, and leaves larger coefficients unchanged (or shrunk proportionally). The soft thresholding rule is:

$$\hat{m}_{(l,k)} = \max\left(|D_{(l,k)}| - T_{(l,k)}, 0\right) \tag{10}$$

where $\hat{m}_{(l,k)}$ is the enhanced magnitude of the coefficient after thresholding, $|D_{(l,k)}|$ is the original magnitude and $T_{(l,k)}$ is the threshold applied at each decomposition level, adjusted by the variance.

d: RECOMBINE MAGNITUDE WITH ORIGINAL PHASE

Once the magnitude has been thresholded, we recombine the adjusted magnitude with the original phase to obtain the final enhanced coefficient. The enhanced coefficient $\hat{D}_{(l,k)}$ is:

$$\hat{D}_{(l,k)} = \hat{m}_{(l,k)} \cdot e^{i\cdot\text{Phase}(D_{(l,k)})} \tag{11}$$

Here $\hat{m}_{(l,k)}$ is the thresholded magnitude. This process ensures both magnitude and phase are preserved, maintaining the temporal and structural alignment of the signal while suppressing noise effectively.

5) SPATIAL AND INTENSITY FILTERING FOR APPROXIMATION COEFFICIENTS

The approximation coefficients $A_{(L,k)}$, often overlooked in traditional methods, are refined using Spatial and Intensity filtering. This process enhances the signal by removing some of the low-frequency noise but still retains some important features. The refinement for each coefficient $\hat{A}_{(L,k)}$is computed as:

$$\hat{A}_{(L,k)}[i] = \frac{\sum_{j\in\Omega(i)} \omega_s(j-i)\cdot\omega_i\left(A_{(L,k)}[j] - A_{(L,k)}[i]\cdot A_{(L,k)}[j]\right)}{\sum_{j\in\Omega(i)} \omega_s(j-i)\cdot\omega_i\left(A_{(L,k)}[j] - A_{(L,k)}[i]\right)} \tag{12}$$

where $\omega_s(j-i)$ represents spatial proximity weights, which ensure that closer coefficients have a higher influence and $\omega_i\left(A_{(L,k)}[j] - A_{(L,k)}[i]\right)$ represents Intensity similarity weights. These weights make it possible to filter out noise whilst avoiding over-smoothing and distortion of edges and transitions in the signal. The coefficients $\hat{A}_{(L,k)}$ which are obtained after the refinement process help to improve the structure of the denoised signal, particularly in the low frequency features of the signal.

6) RECONSTRUCTION OF DENOISED SIGNAL USING IDWT

The denoised signal $x_{\text{denoised}}(t)$ is then reconstructed from the refined approximation coefficients $\hat{A}_{(L,k)}$ and detail coefficients $\hat{D}_{(l,k)}$ using the process of Inverse Discrete Wavelet Transform (IDWT):

$$x_{\text{denoised}}(t) = \sum_{k} \hat{A}_{(L,k)} \phi_{(L,k)}(t) + \sum_{l=1}^{L} \sum_{k} \hat{D}_{(l,k)} \psi_{(l,k)}(t) \tag{13}$$

This step aims at reconstructing the denoised signal from the refined wavelet coefficients.

7) RESIDUAL PROCESSING AND FUSION

Residual noise R is calculated as the difference between the noisy signal and the initially denoised signal:

$$R = x_{\text{noisy}} - x_{\text{denoised}} \tag{14}$$

This residual contains high-frequency noise components that were not eliminated in the first step of the de-noising process. The residual R undergoes wavelet decomposition and adaptive soft thresholding as follows:

$$D_{(l,k)}^{\text{residual}} = soft_{\text{denoised}} \left(D_{(l,k)}^{\text{residual}}, T \right) \tag{15}$$

The threshold T could be either a constant or adaptive. The refined residual $R_refined$ is then reconstructed and this involves both approximation (low-frequency) and detail (high-frequency) coefficients:

$$R_{\text{refined}} = \sum_{k} A_{(L,K)} \phi_{(L,K)}(t) + \sum_{l=1}^{L} \sum_{k} D_{(l,k)}^{\text{refined}} \psi_{(l,k)}(t) \tag{16}$$

Finally, the refined residual R_{refined} is then fused with the initially denoised signal x_{denoised} to get the final signal s(t):

$$s(t) = x_{\text{denoised}} + R_{\text{refined}} \tag{17}$$

This step ensures that the fine details (those that are not well captured in the initial denoising) are restored by adding the residual back to the denoised signal. This fusion helps preserve signal integrity while effectively suppressing noise. The scheme with the proposed DWT-based framework is capable of eliminating noise while retaining signal quality with phase-aware adaptive soft thresholding, Spatial and Intensity filtering for low-frequency enhancement, and Residual fusion for fine details enhancement. In this way, the method produces high-quality denoising without requiring any additional normalization, which shows the stability of the proposed approach for further tasks.

B. PROPOSED SPECTRAL SUBTRACTION (SS) FILTER

The proposed Spectral Subtraction (SS) filter intends to improve the quality of noisy audio signals through noise estimation and accurate phase reconstruction. Contrary to many techniques based on standard spectral subtraction for noise attenuation that works with a fixed estimate of noise and very simple reconstruction, the developed approach deals with non-stationary noise, takes into account the phase information and employs multiple iterations to improve speech estimation. Thus, overcoming these limitations allows the framework to provide higher signal fidelity, better intelligibility, and a decrease in the degree of perceptive artifacts. The detail methodology for proposed SS filter is shown in below Figure 2 which involves the following steps:

1) FRAME DIVISION AND SHORT-TIME FOURIER TRANSFORM (STFT)

Being inherently non-stationary, speech signals require localized processing in the time-frequency domain. To specify a noisy signal x_{noisy}, the signal is divided into overlapping frames of size N with a frameshift M. Each frame i is extracted as:

$$x_i(t) = x_{\text{noisy}}[t + i \cdot M], \quad t = 0, 1, \ldots, N-1 \tag{18}$$

where i represents the frame index. Each frame is decomposed into a signal splitting into its magnitude and phase components using the Short-Time Fourier Transform (STFT):

$$X_i(f) = |X_i(f)| \cdot e^{j\phi_i(f)} \tag{19}$$

Here $|X_i(f)|$ is the i_{th} frame, the magnitude spectrum, corresponds to the energy distribution of frequencies, $\phi_i(f)$ is the temporal alignment of signal components encodes in the phase spectrum of i_{th} frame and $X_i(f)$ is the Complex spectrum at frequency f of the i_{th} frame. Frequency domain manipulation is performed using the STFT, and this allows specific noise to be suppressed, without modifying the phase of the underlying signal.

2) NOISE ESTIMATION

Accurate noise estimation is critical for effective spectral subtraction. To identify noise-dominant regions, Voice Activity Detection (VAD) is first applied, followed by a Minimum Statistics approach.

a: VOICE ACTIVITY DETECTION (VAD)

VAD distinguishes speech and non-speech frames based on their energy levels. The energy for each frame E_i is computed as:

$$E_i = \sum_{f} |X_i(f)|^2 \tag{20}$$

If the energy E_i is below a certain threshold E_{th}, the frame is classified as non-speech (noise), and its corresponding magnitude spectrum $|X_i(f)|$ is used for the noise estimate. The VAD output is a binary decision:

$$\text{VAD}(i) = \begin{cases} 1 & \text{if } E_i > E_{\text{th}} \text{ (Speech)} \\ 0 & \text{if } E_i \leq E_{\text{th}} \text{ (Non-Speech)} \end{cases} \tag{21}$$

Frames classified as non-speech (noise) (i.e., $VAD(i) = 0$are used to compute the noise spectrum.

b: PRELIMINARY NOISE ESTIMATED SPECTRUM

The noise spectrum $\hat{N}(f)$ is computed as the average of the magnitude spectra from the non-speech frames $VAD(i) = 0$:

$$\hat{N}(f) = \frac{1}{K} \sum_{k=1}^{K} |X_k(f)| \cdot [\text{VAD}(k) = 0] \tag{22}$$

where $\hat{N}(f)$ is preliminary Estimated noise spectrum at frequency f, K is the Number of non-speech frames, $(X_k(f)$ is the k_{th} frame non-speech frame and VAD(k) is the Indicator function that selects only non-speech frames for noise estimation.

c: MINIMUM STATISTICS

To refine this noise estimate over time, a minimum statistics approach is used. This approach tracks the minimum value of the noise spectrum across a sliding time window, capturing low-energy variations and avoiding overestimations of noise. The refined noise spectrum is calculated as:

$$\hat{N}_{\min}(f) = \min\left(\hat{N}_{\min}(f), \hat{N}(f)\right) \tag{23}$$

where $\hat{N}_{\min}(f)$ is the Minimum noise spectrum estimate at frequency f and $\hat{N}(f)$ is the Preliminary noise spectrum estimate at frequency f, computed from non-speech frames identified by VAD. This equation updates $\hat{N}_{\min}(f)$ by retaining the smaller of its current value and the newly estimated $\hat{N}(f)$. This dynamic refinement ensures the framework adapts to varying noise profiles, maintaining accuracy across diverse conditions.

3) SPECTRAL SUBTRACTION WITH PHASE-AWARE RECONSTRUCTION

a: PRELIMINARY NOISE SUPPRESSION

After estimating the Preliminary Estimated Spectrum $\hat{N}(f)$ and refining it using the minimum statistics approach $\hat{N}_{\min}(f)$ the actual spectral subtraction is applied. To enhance the performance of spectral subtraction, this process is repeated iteratively. In each iteration, the residual noise is computed and further refined by applying spectral subtraction to the residuals, progressively improving noise suppression. For the i_{th} frame of noisy signal's magnitude spectrum $|X_i(f)|$ is processed by subtracting the minimum noise spectrum $\hat{N}_{\min}(f)$ to reduce noise. The formula for spectral subtraction for i_{th} frame is:

$$|X_{\text{prelim-denoised},i}(f)| = \max\left(|X_i(f)| - \alpha \cdot \hat{N}_{\min}(f), \epsilon\right) \tag{24}$$

where $|X_{\text{prelim-denoised},i}(f)|$ is the preliminary denoised magnitude spectrum for i_{th} frame, $|X_i(f)|$ is magnitude spectrum of the noisy signal for i_{th} frame, α is the Subtraction factor controlling the level of noise suppression and ϵ is the small constant to prevent negative or zero magnitudes. This subtraction ensures that the noise is attenuated while preventing over-subtraction, which could distort speech components.

b: PHASE-AWARE RECONSTRUCTION

Once the magnitude spectrum has been denoised, it is combined with the original phase spectrum $\phi_i(f)$from the noisy signal for accurate signal reconstruction. For the i_{th} frame, the phase-aware reconstruction is given by:

$$|\tilde{X}_{\text{reconstructed},i}(f)| = |X_{\text{prelim-denoised},i}(f)| \cdot e^{j\phi_i(f)} \tag{25}$$

where $\tilde{X}_{\text{reconstructed},i}(f)$ is the reconstructed complex spectrum for the i_{th} frame (combining the preliminarily denoised magnitude and the original phase).

4) ITERATIVE REFINEMENT

a: ADAPTIVE SUBTRACTION FACTOR

The subtraction factor α_t is updated adaptively with each iteration to improve noise suppression. The update is given by:

$$\alpha_{t+1} = \alpha_t \cdot \gamma, \quad \gamma > 1 \tag{26}$$

where γ is the increment rate that controls how the subtraction factor evolves in each iteration.

b: RESIDUAL NOISE CALCULATION

After the initial spectral subtraction, the residual noise $R_i(f)$ is calculated as the difference between the noisy signal $X_i(f)$ and the denoised signal $\tilde{X}_{\text{reconstructed},i}(f)$:

$$R_i(f) = X_i(f) - \tilde{X}_{\text{reconstructed},i}(f). \tag{27}$$

This residual noise represents the part of the signal that was not effectively suppressed by the initial subtraction process.

c: RESIDUAL NOISE REFINEMENT

In each iteration, the refined residual noise spectrum $R_{\text{refined},i}(f)$ is obtained by applying spectral subtraction to the residual noise spectrum. This iterative process progressively refines the residual noise in each step, ensuring that noise components that were missed in the earlier iterations are suppressed more effectively. The updated spectrum is given by:

$$R_{\text{refined},i}(f) = \max\left(|R_i(f)| - \alpha_t \cdot \hat{N}_{\min}(f), \epsilon\right) \tag{28}$$

where α_t is the updated subtraction factor. The residual spectrum is refined using the same subtraction process, and the refined residual is integrated with the denoised signal:

$$X_{\text{final},i}(f) = \tilde{X}_{\text{reconstructed},i}(f) + R_{\text{refined},i}(f) \tag{29}$$

5) SPECTRAL SMOOTHING

Artifacts such as "musical noise" can result from abrupt changes in the spectrum across frames. To mitigate this, spectral smoothing is applied:

$$X_{\text{smoothed},i}(f) = \frac{1}{2W+1} \sum_{w=-W}^{W} X_{\text{final},i+w}(f) \tag{30}$$

where W is the Smoothing window size. This step ensures temporal and spectral coherence, enhancing the perceptual quality of the output.

6) INVERSE SHORT-TIME FOURIER TRANSFORM (ISTFT)

The smoothed speech signal in the time domain is obtained using the Inverse Fourier Transform (IFFT) after performing spectral smoothing given by:

$$x_{\text{smoothed},i}(t) = \text{iSTFT}\left[X_{\text{smoothed},i}(f)\right] \tag{31}$$

7) RECONSTRUCTION USING OVERLAP-ADD SYNTHESIS

The final enhanced signal is synthesized in the time domain using overlap-add synthesis, a technique that combines overlapping frames while preserving continuity:

$$S(t) = \sum_i x_{\text{smoothed},i}(t) \cdot w(t - iM) \tag{32}$$

where S(t) is the enhanced synthesized signal (after overlap-add in the time domain), the i_{th} smoothed frame is $x_{\text{smoothed},i}(t)$ and M is the frameshift. In above eq. (32) component w(t-iM) which is windowing function is utilized in order to do smooth synthesis frame wise. This process is used to blend the overlapping regions more smooth to avoid artifacts in combining the enhanced frames achieved from proposed SS filter.

C. PROPOSED HYBRID SS-DWT SE APPROACHES

The SS-DWT frame work is proposed in this paper to address the music noise challenge Spectral Subtraction (SS) filter and in addition to it, this method provides solution for DWT thresholds to reduce background noises of varying frequencies with respect to time scale. In this paper, Hybrid Speech Enhancement methods are proposed which combine Spectral Subtraction (SS) and Discrete Wavelet Transform (DWT) to improve the quality and the intelligibility of noisy mixture speech signals corrupted with non-stationary background noises. Iterative noise estimation, Voice Activity Detection (VAD), minimum statistics for dynamic noise adaption, spectral smoothing, and phase-aware spectral reconstruction are all used in the suggested SS method, in contrast to the conventional approach. In the suggested DWT methods Spatial and Intensity filter is adjusted to the approximation coefficients to enhance low-frequency features and maintain structural integrity while lowering distortion and adaptive noise refinement with phase-aware soft thresholding is used to detail coefficients. In proposed SS-DWT framework to remove any remaining noise from the speech signals, SS filter is applied to lower the average noise intensity. Afterward, we use proposed DWT with its threshold to enable the reduction of musical noise and enhance the speech quality and intelligibility.

Our approach is depicted in Figure 3 below, where we introduce two primary threshold approaches: SS-DWTAT (Adaptive thresholding) and SS-DWTCT (Constant Thresholding). These approaches result in more reliable and consistent noise reduction, yielding improved speech quality in noisy environments.

IV. SPEECH SIGNAL COLLECTION AND EVALUATION METRICS

A. SPEECH SIGNAL COLLECTION

A 7-second speech signal recorded in a quiet setting makes up the clean audio signal. Additionally, this speech signal was distorted for various conditions using noise signals captured in the actual environment. Aircraft engine noise, white noise, pink noise, siren noise, café ambience noise, and engine idle noise are the noise signals that are used. All these audio signals, including the pink noise were sourced from the website ''www.freesound.com''. This website provides audio signals captured in natural settings, without copyright issues, specifically for scientific research and application development [41]. In other experiments to compare proposed SE methods with other baseline methods clean speech signals are sourced from TIMIT dataset [42] and CASIA dataset, whereas Gaussian noise is sourced from Noisex-92 database and babble noise from AURORA-2 dataset.

B. INSTRUMENTAL EVALUATION METRICS

In this study assessing the efficiency of proposed SE methods employing SS and DWT, a number of commonly used metrics are calculated including Signal-to-distortion ratio (SDR) [43], Mean Square Error (MSE) [44] b45, perceptual evaluation of speech quality (PESQ) [46] and Short-time objective intelligibility (STOI) [47]. Although a lower Mean Square Error (MSE) score indicates higher levels of resemblance in the original and compressed audio signals, indicating superior compressed quality, SDR is employed to demonstrate the effects of the algorithms on noise isolation and reduction. During the experiments, we conduct and employ perceptual evaluation of speech quality (PESQ) [46] to generate a mean opinion score for listening quality objective (MOS-LQO).

For real-time speech quality improvement, the recommended framework uses PESQ as a metric to assess performance and provide approximation values to MOS, a human listening subjective measure. The enhanced speech is evaluated for readability using the Short-time Objective Intelligibility (STOI) measure. With a value within a range of [0, 1], the STOI measure is specifically designed to evaluate noise suppression strategies; high scores are strongly correlated with superior intelligibility. Additionally to compare results with baselines the speech quality metrics such as segmental SNR (segSNR) [48] and Signal-to-Noise Ratio (SNR) [43] of the enhanced speech are measured. To evaluate the Speech Quality MOS predictions were evaluated by 20 male listeners and 20 female listeners of age between 20 to 40 years. Listeners rated the enhanced speech signals on scale of 1 to 5 used for poor and excellent speech signal quality.

Here are the results of the investigation and experimentation with real-world noise reduction algorithms for communication systems.

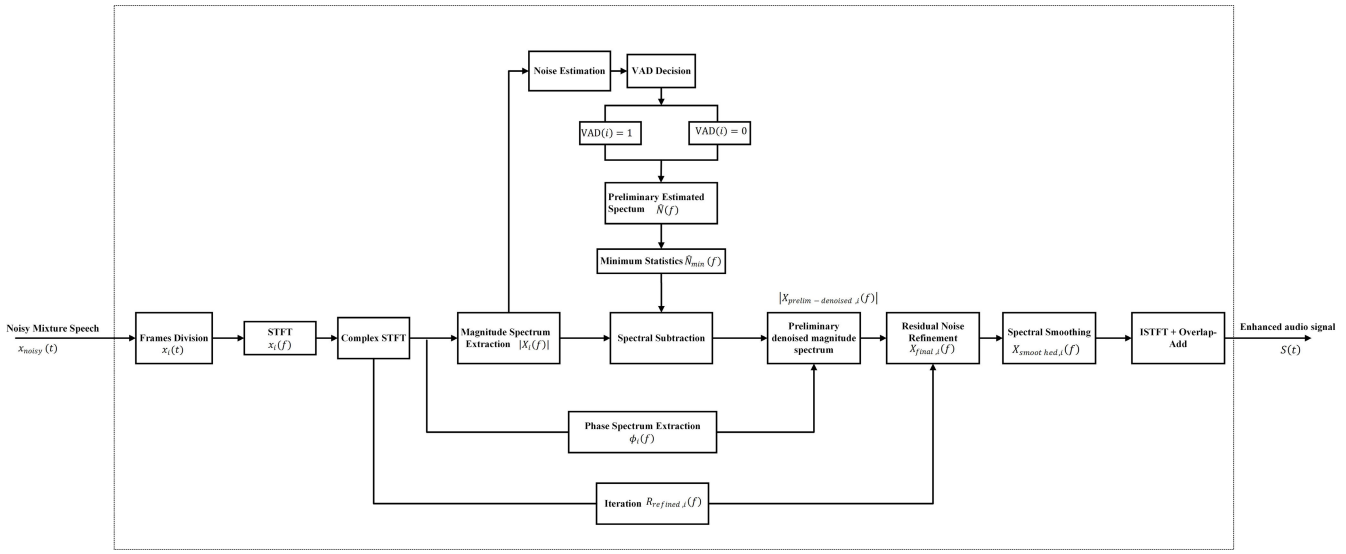


FIGURE 2. Methodology of spectral subtraction speech enhancement.

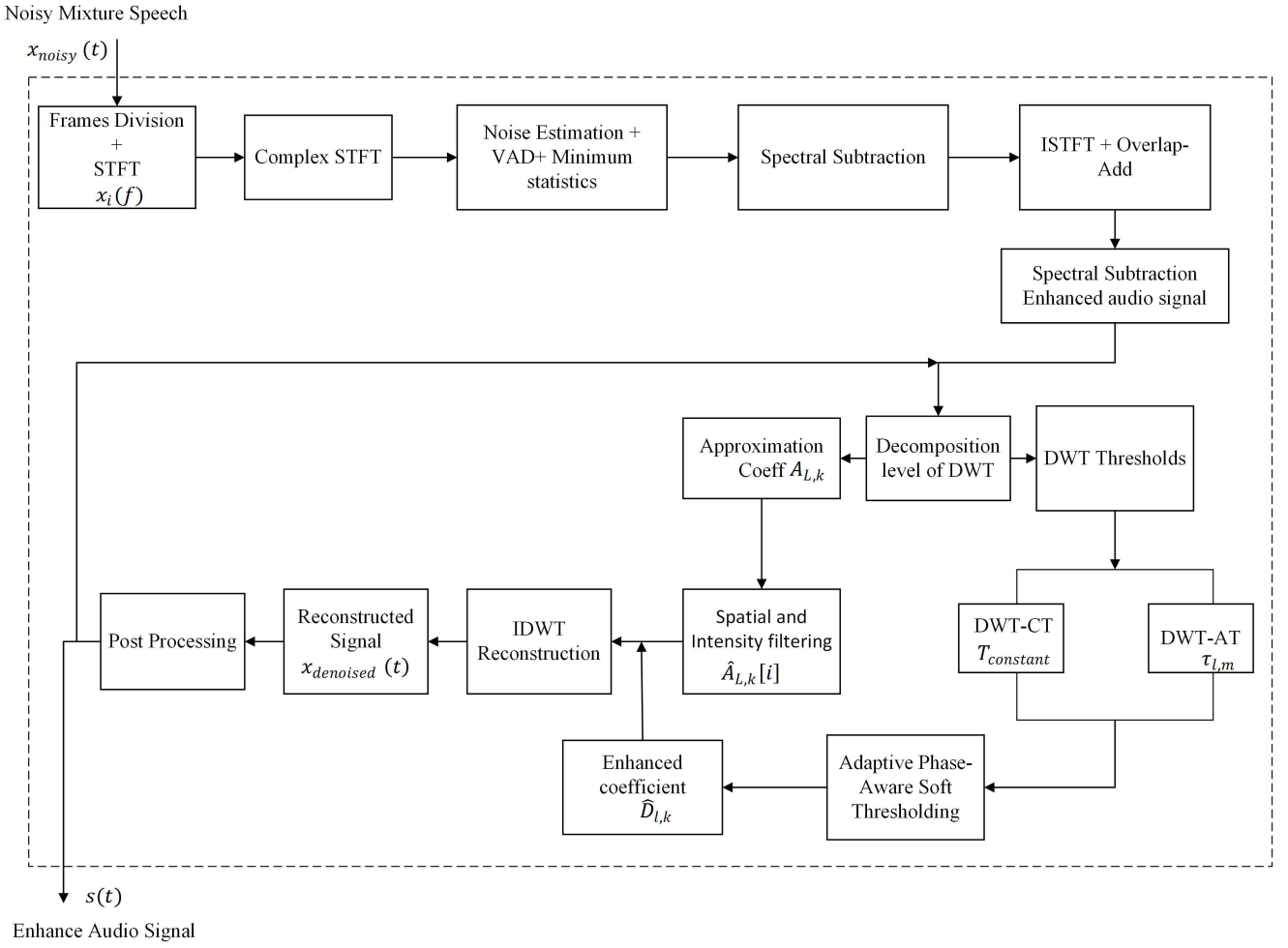


FIGURE 3. Data flow diagram for the hybrid SS-DWT speech restoration scheme.

V. ANALYSIS OF EXPERIMENTAL RESULTS

In this section, we present the results of the investigation and experimentation with real-world noise reduction algorithms. We investigated the effects of employing Discrete Wavelet Transform (DWT) and Spectral Subtraction (SS) noise reduction methods to make promising quality and clarity

TABLE 1. Proposed speech enhancement methods assessment using MSE and SDR.

Method	Noise	Wavelets	Ini-MSE	Final-MSE	Ini-SDR (dB)	Final-SDR (dB)
DWT-AT	Pink	sym5	8.26×10^{-4}	2.90×10^{-5}	$8.8 \times 10^{-15} - 0$	25.42
		sym10	8.26×10^{-4}	2.81×10^{-5}	$8.8 \times 10^{-15} - 0$	28.35
		bior1.1	8.26×10^{-4}	2.95×10^{-5}	$8.8 \times 10^{-15} - 0$	25.10
	Siren	sym5	8.26×10^{-4}	2.59×10^{-5}	$2.22.8 \times 10^{-15} - 0$	28.14
		sym10	8.26×10^{-4}	2.64×10^{-5}	$2.22.8 \times 10^{-15} - 0$	29.77
		bior1.1	8.26×10^{-4}	2.55×10^{-5}	$2.22.8 \times 10^{-15} - 0$	28.10
DWT-CT	Pink	sym5	8.26×10^{-4}	3.11×10^{-5}	$8.8 \times 10^{-15} - 0$	24.19
		sym10	8.26×10^{-4}	2.91×10^{-5}	$8.8 \times 10^{-15} - 0$	27.01
		bior1.1	8.26×10^{-4}	3.25×10^{-5}	$8.8 \times 10^{-15} - 0$	23.00
	Siren	sym5	8.26×10^{-4}	2.95×10^{-5}	$2.22.8 \times 10^{-15} - 0$	26.04
		sym10	8.26×10^{-4}	2.88×10^{-5}	$2.22.8 \times 10^{-15} - 0$	29.11
		bior1.1	8.26×10^{-4}	3.02×10^{-5}	$2.22.8 \times 10^{-15} - 0$	25.90
SS	Pink	-	8.26×10^{-4}	2.80×10^{-5}	$8.8 \times 10^{-15} - 0$	30.01
	Siren	-	8.26×10^{-4}	2.64×10^{-5}	$2.22.8 \times 10^{-15} - 0$	31.11
SS-DWTAT	Pink	sym5	8.26×10^{-4}	2.77×10^{-5}	$8.8 \times 10^{-15} - 0$	28.27
		sym10	8.26×10^{-4}	2.59×10^{-5}	$8.8 \times 10^{-15} - 0$	**32.14**
		bior1.1	8.26×10^{-4}	2.81×10^{-5}	$8.8 \times 10^{-15} - 0$	28.01
	Siren	sym5	8.26×10^{-4}	2.36×10^{-5}	$2.22.8 \times 10^{-15} - 0$	32.71
		sym10	8.26×10^{-4}	2.11×10^{-5}	$2.22.8 \times 10^{-15} - 0$	**34.15**
		bior1.1	8.26×10^{-4}	2.43×10^{-5}	$2.22.8 \times 10^{-15} - 0$	31.88
SS-DWTCT	Pink	sym5	8.26×10^{-4}	2.57×10^{-5}	$8.8 \times 10^{-15} - 0$	29.85
		sym10	8.26×10^{-4}	2.50×10^{-5}	$8.8 \times 10^{-15} - 0$	32.10
		bior1.1	8.26×10^{-4}	2.70×10^{-5}	$8.8 \times 10^{-15} - 0$	28.91
	Siren	sym5	8.26×10^{-4}	2.47×10^{-5}	$2.22.8 \times 10^{-15} - 0$	32.65
		sym10	8.26×10^{-4}	2.38×10^{-5}	$2.22.8 \times 10^{-15} - 0$	34.07
		bior1.1	8.26×10^{-4}	2.51×10^{-5}	$2.22.8 \times 10^{-15} - 0$	30.04

of speech samples affected by background noise. Following provides an explanation of outcomes of using these speech enhancement algorithms.

A. MSE AND SDR MEASURES FOR PROPOSED SE METHODS

Considering non-stationary background noises like Pink and Siren noise types the proposed SE methods are evaluated under same noise conditions as in [11] and the Mean Square Error (MSE) and Signal-to-Distortion Ratio (SDR) are measured in Table 1. The experimental results show that wavelet family Symlet-10 (Sym10) outperforms Symlet-5 (Sym5) and Biorthogonal-1.1 (Bior1.1) in the DWT-AT method, achieving the lowest MSE 2.81×10^{-5} and highest SDR (28.35 dB), compared to MSE of 2.90×10^{-5} and SDR of 25.42 dB for Sym5, and MSE of 2.95×10^{-5} and SDR of 25.10 dB for Bior1.1. For the Sym10 wavelet family, the proposed DWT-AT SE method achieves SDR improvements of 28.35 dB and 29.77 dB for Pink and Siren noise respectively, and surpassing the proposed DWT-CT which achieves SDR improvements of 27.01 and 29.11 respectively. The proposed Spectral Subtraction (SS) filter outperforms both DWT-AT and DWT-CT in MSE reduction achieving 2.80×10^{-5} and 2.64×10^{-5} for noisy mixture audios degraded with Pink and Siren background noises, respectively. In Table 1 it is observed that proposed SS-DWT hybrid methods (SS-DWTAT and SS-DWTCT) yield higher performance in terms of MSE reduction and SDR improvement compared to other methods. For Sym10 SS-DWTAT and SS-DWTCT achieve the highest MSE reduction of 2.11×10^{-5} and 2.38×10^{-5} respectively, and also achieve the highest improvement in SDR of 34.15 dB and 34.07 dB respectively against the non-stationary siren noise type. This suggests that integrating spectral subtraction with wavelet thresholding effectively enhances both speech quality and noise suppression, addressing the issue of musical noise associated with SS and non-stationary variations across frequencies and time scales encountered in DWT.

B. STOI AND PESQ MEASURES FOR PROPOSED SE METHODS

To evaluate the proposed methods SE performance STOI and PESQ scores are measured for noisy signals degraded with non-stationary background noises like Pink and Siren noise types as shown in Table 2. Using the sym10 wavelet family Proposed DWT-AT SE method improves STOI score from 0.86 to 0.96 and 0.68 to 0.93 for Pink and Siren noise respectively and surpassing the proposed DWT-CT which achieves STOI scores of 0.93 and 0.94 respectively. Similarly DWT-AT SE approach PESQ improvement of 3.41 for speech signal degraded with Pink noise and 3.61 for Siren noise sample. Furthermore, proposed Spectral Subtraction (SS)

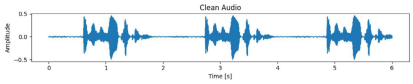
TABLE 2. Proposed speech enhancement methods assessment using STOI and PESQ.

Method	Noise	Wavelets	Initial STOI	Final STOI	Initial PESQ	Final PESQ
DWT-AT	Pink	sym5	0.68	0.91	1.92	3.22
		sym10	0.68	0.93	1.92	3.41
		bior1.1	0.68	0.91	1.92	3.15
	Siren	sym5	0.86	0.95	2.26	3.58
		sym10	0.86	0.96	2.26	3.61
		bior1.1	0.86	0.94	2.26	3.53
DWT-CT	Pink	sym5	0.68	0.92	1.92	3.27
		sym10	0.68	0.92	1.92	3.30
		bior1.1	0.68	0.91	1.92	3.09
	Siren	sym5	0.86	0.92	2.26	3.47
		sym10	0.86	0.96	2.26	3.53
		bior1.1	0.86	0.92	2.26	3.30
SS	Pink	-	0.68	0.95	1.92	3.59
	Siren	-	0.86	0.96	2.26	3.61
SS-DWTAT	Pink	sym5	0.68	0.96	1.92	3.50
		sym10	0.68	0.97	1.92	3.67
		bior1.1	0.68	0.94	1.92	3.42
	Siren	sym5	0.86	0.97	2.26	3.68
		sym10	0.86	**0.98**	2.26	**3.84**
		bior1.1	0.86	0.95	2.26	3.61
SS-DWTCT	Pink	sym5	0.68	0.95	1.92	3.38
		sym10	0.68	0.96	1.92	3.53
		bior1.1	0.68	0.91	1.92	3.11
	Siren	sym5	0.86	0.97	2.26	3.49
		sym10	0.86	0.98	2.26	3.71
		bior1.1	0.86	0.95	2.26	3.42

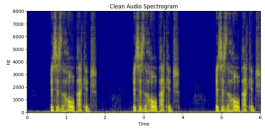
filter outperforms both DWT-AT and DWT-CT in STOI improvements achieving 0.95 and 0.96 for noisy mixture audios degraded with Pink and Siren background noises, respectively. In Table 2 it is observed that proposed SS-DWT hybrid methods (SS-DWTAT and SS-DWTCT) yield higher performance in terms of STOI and PESQ improvement compared to other methods. Compared to other wavelets sym5 and bior1.1, the Sym10 for both SS-DWTAT and SS-DWTCT achieved highest STOI of 0.98, and also achieved highest improvement in PESQ of 3.84 and 3.71 respectively against the non-stationary siren noise type. This suggests that integrating spectral subtraction with wavelet thresholding effectively enhances speech quality and intelligibility, addressing the issue of musical noise associated with SS and non-stationary variations across frequencies and time scales encountered in DWT.

C. TIME-AMPLITUDE AND TIME-FREQUENCY GRAPHICAL ANALYSIS

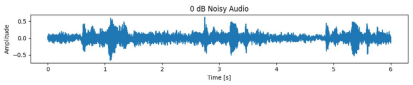
Improved speech achieved from the proposed Spectral Subtraction (SS) and Discrete Wavelet Transform (DWT) methods were investigated using Time-Amplitude Graphs and spectrograms as shown in Figure 4. To better understand the fundamental reasons for the quality improvements observed with the proposed Hybrid SS-DWT, an investigation of a clean audio signal corrupted by Pink noise considering the 0 dB SNR level was conducted. Using Improved Spectral Subtraction and Discrete Wavelet Transform, Figure 4 displays the spectrograms for denoised audio signals and clean audio signals along alongside their corresponding noisy audio signals. Spectrogram enhancement shows that considerably more noise suppression is feasible with the suggested approaches when looking at the outputs. Due to sophisticated processing of audio, enhanced audios are achieved by having both phase and magnitude information of the speech, and it helped to reconstruct better audio signals. When DWT-CT and DWT-AT SE approaches are applied to 0 dB SNR noisy audio signal the average noise level across speech signal decreased, however, residual noise remains as shown in graphs in Figure 4 (e - h). Similarly, when Spectral Subtraction (SS) is applied to a noisy mixture signal, a reduction in background noise is observed but some music noise persists in the signal as shown in Figure 4 (k - l). To address these limitations in proposed hybrid SS-DWT, SS is first applied to noisy speech signals and then followed by the application of proposed DWT-AT and DWT-CT approaches. The time-amplitude and time-frequency plots in Figure 4 (k - n) show a significant reduction in noise addresses the problem of musical noise within SS filter as well as the lower frequency noises in DWT method.



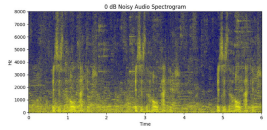
(a) Time-Amplitude Graph of Clean Speech



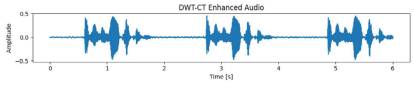
(b) Time-Frequency Graph of Clean Speech



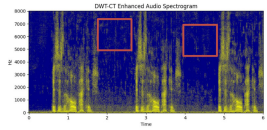
(c) Time-Amplitude Graph of 0dB Noisy Mixture Speech



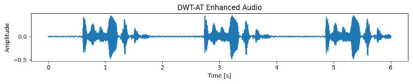
(d) Time-Frequency Graph of Noisy Mixture Speech



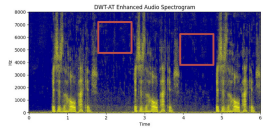
(e) Time-Amplitude Graph of DWT-CT



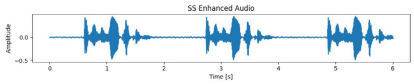
(f) Time-Frequency Graph of DWT-CT



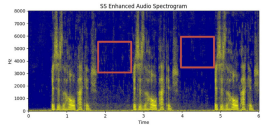
(g) Time-Amplitude Graph of DWT-AT



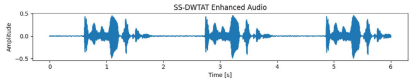
(h) Time-Frequency Graph of DWT-AT



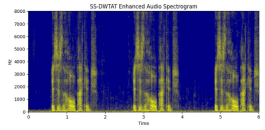
(i) Time-Amplitude Graph of Spectral Subtraction (SS)



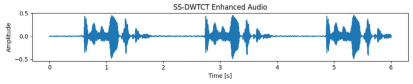
(j) Time-Frequency Graph of Spectral Subtraction (SS)



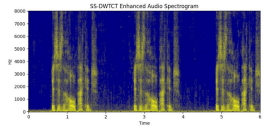
(k) Time-Amplitude Graph of Hybrid SS-DWTAT



(l) Time-Frequency Graph of Hybrid SS-DWTAT



(m) Time-Amplitude Graph of Hybrid SS-DWTCT



(n) Time-Frequency Graph of Hybrid SS-DWTCT

FIGURE 4. Time-Amplitude and Time-Frequency graph comparison of proposed speech enhancement methods.

D. DISCUSSION ON PERFORMANCE OF SPEECH ENHANCEMENT METHODS

Table 3 shows our solutions for speech signals affected by different background noises correlating the proposed Speech Enhancement algorithms for AWGN noise types with previous approaches. SS-DWTAT (Sym10), our suggested technique, outperforms all other Speech Enhancement Techniques, with a maximum SNR improvement of 34.15 for 0 dB SNR speech signals. From SNR scores for estimated clean signals, it is suggested that the suggested noise reduction performance provides high SNR values for assistive technology. In addition, all of our suggested

TABLE 3. Quality measures analysis of speech improvement procedures exploiting different background noise types.

Methods	Noise Type	Frequency of Audio Signals	Initial-SNR (dB)	Final-SNR (dB)
DWT-thresholds [14]	AWGN	8K	5.0	10.190
			5.0	7.510
			5.0	8.440
WTD-LMS [28]	Aircraft Engine	8K	0.0	19.930
			5.0	21.360
			15	27.01
			30	34.03
WTD-NLMS [11]	AWGN	8K	Pink Noise → 0	32
			Siren → 0	34.03
Proposed DWT-AT (sym10)	AWGN	8K	Pink Noise → 0	28.35
			Siren → 0	29.77
Proposed DWT-CT (sym10)	AWGN	8K	Pink Noise → 0	27.01
			Siren → 0	29.11
Proposed SS	AWGN	8K	Pink Noise → 0	30.01
			Siren → 0	31.11
Proposed SS-DWTAT (Sym10)	AWGN	8K	Pink Noise → 0	**32.14**
			Siren → 0	**34.15**
Proposed SS-DWTCT (Sym10)	AWGN	8K	Pink Noise → 0	32.10
			Siren → 0	34.07

methods—DWT-AT (sym10), DWTCT (sym10), SS-DWTAT (sym10), SS-DWTCT (sym10), and SS—show significant gains in SNR over previous approaches.

VI. COMPARISON WITH RECENT METHODS BASED ON SS-DWT COMBINATION

To evaluate the Speech Enhancement performance proposed Hybrid SS-DWT methods are compared with traditional methods and wavelet family selected in these experiments is Sym10, the experimental results are as follows:

A. SPEECH ENHANCEMENT EVALUATION USING TIMIT AND GAUSSIAN NOISE

The proposed hybrid SS-DWT methods namely SS-DWTAT and SS-DWTCT speech enhancement algorithms have been tested on the spoken English sentence chosen from TIMIT database against a baseline method proposed in [49]. The sentence used is ''Please shorten this skirt for Joyce'' with a sampling rate of 16 kHz and spoken by a male and female speaker. The clean speech is corrupted with white Gaussian noise resulting in global SNR levels ranging from -10db to 20db. Compared to baseline method proposed methods achieve highest SNR improvement as shown in Table 4. The SS-DWTAT surpasses all its counterparts and improves the noisy signal up to 11.42 dB in challenging SNR like -10 dB for female speakers and 11.64 dB for male speakers. It is observed that proposed SS-DWTAT has better Speech enhancement performance for all noisy conditions.

B. SPEECH ENHANCEMENT EVALUATION USING PURE RECORDED VOICE SIGNAL AND GAUSSIAN NOISE

The proposed speech enhancement (SE) methods are further evaluated with a pure recorded speech signal against baseline methods [33]. The experiments are conducted with the sampling frequency of 8 kHz, length of frame set to 1024, and 50% overlap between frames. The pure speech signals are degraded with Gaussian white noise at different SNRs (-5, 0, 5, and 10 dB) to obtain noisy audio signals for processing. As given in Table 5 that in comparison to baseline spectral subtraction (SS), the output SNR of proposed spectral subtraction and DWT are enhanced keeping varied input noisy circumstances. Proposed SS-DWTAT method achieves highest SNR improvement of 16.06 dB 17.84 dB, 21.1 dB, and 22.05 dB for SNR levels of -5 dB, 0 dB, 5 dB, and 10 dB, respectively. The rationale behind this improved performance is due to the suggested SS-DWTAT approach further improves the output SNR by processing the speech through wavelet transform after it has been processed by improved spectral subtraction. Whenever the given input SNR is lower, the speech enhancement impact is more noticeable and all suggested methods perform better than traditional methods.

C. SPEECH ENHANCEMENT EVALUATION USING CASIA DATABASE

The Hybrid SS-DWTAT and SS-DWTCT methods are evaluated against methods recently developed in [50] using Clean audio signals sourced from CASIA dataset which is based on Chinese Mandarin speech data produced by Institute of Automation of Chinese Academy of Science. The Chinese female speaker leverages 8 kHz sampling and 8-bit coding to pronounce the chosen utterance as ''xíng zhèng qū yù''. The Noisex-92 database is the source of the white Gaussian noise. At the condition of 0 dB SNR, the noise is added to the clean speech signals to create the noisy speech signals. As shown in Table 6 the proposed SS-DWTAT obtains improvements up

TABLE 4. SNR score comparison between proposed and referenced SE methods using TIMIT dataset and gaussian noise.

Method	Female							Male						
Noisy SNR (dB)	-10	-5	0	5	10	15	20	-10	-5	0	5	10	15	20
Spectral Subtraction in Wavelet Domain [49]	2.3	4.9	7.5	10.7	14	18.1	22.3	3.2	5.5	8.7	11.7	14.9	18.3	22.4
DWT-AT	9.55	13.6	16.83	21.82	22.4	23.11	26.91	10.35	14.16	17.2	21.16	23.03	24.18	27.05
DWT-CT	9.22	13.1	16.01	19.17	21.3	22.04	25.11	10.01	13.94	16.84	20.05	22.51	23.9	26.42
SS	10.01	13.73	17.11	22.51	22.51	23.84	27.13	10.77	14.84	17.57	21.66	23.78	24.44	27.30
SS-DWTAT	**11.42**	**14.59**	**18.61**	**23.75**	**24.02**	**25.17**	**28.43**	**11.64**	**15.34**	**18.37**	**22.65**	**24.11**	**25.38**	**28.19**
SS-DWTCT	11.05	14.25	18.33	23.18	23.96	24.87	27.86	11.18	15.02	18.14	22.24	23.89	25.17	27.95

TABLE 5. SNR comparison considering recoded pure voice signal and degraded with gaussian white noise.

Method	SNR (dB)			
Noisy SNR (dB)	-5	0	5	10
Traditional spectral subtraction [33]	2.355	4.314	8.390	11.770
Improved spectrum subtraction [33]	3.690	5.252	9.250	12.482
Improved speech enhancement algorithm [33]	5.569	6.553	9.925	12.936
DWT-AT	14.40	16.17	20.73	21.53
DWT-CT	13.53	16.04	20.41	21.38
SS	14.66	16.52	21.10	22.05
SS-DWTAT	**16.14**	**18.06**	**22.62**	**24.58**
SS-DWTCT	16.06	17.84	22.13	24.27

TABLE 6. SegSNR comparison of various methods.

Methods	SegSNR
heursure [50]	9.8442
rigrsure [50]	9.8442
sqtwolog [50]	2.5925
minimaxi [50]	4.4048
SS-DWTAT	**15.2801**
SS-DWTCT	15.2106

TABLE 7. Evaluation of proposed methods with TIMIT dataset and babble noise against baseline SE methods.

Methods	PESQ (MOS points)				
SNR (dB)	-5	0	5	10	15
Noisy	1.35	1.52	1.77	2.11	2.53
MMSE-LSA [51]	1.39	1.64	1.97	2.36	2.76
SG-JMAP [51]	1.38	1.63	1.98	2.41	2.87
DCCNN [52]	2.87	2.95	3.01	3.27	3.65
SS-DWTAT	**2.91**	**2.97**	**3.15**	**3.31**	**3.74**
SS-DWTCT	2.90	2.95	3.13	3.28	3.69

to 15.2801 dB segSNR, the highest improvement compared to best baseline method. The proposed SS-DWTCT method achieves a SegSNR of 15.2106, indicating its superior performance compared to the baseline methods, such as Heursure (9.8442), Rigrsure (9.8442), Sqtwolog (2.5925), and Minimax (4.4048).

D. SPEECH ENHANCEMENT EVALUATION USING MOS MEASURES

To further evaluate proposed Hybrid Spectral Subtraction-Discrete Wavelet Transform (SS-DWT) SE methods clean speech data from TIMIT is mixed with babble noise from AURORA-2 dataset and its PESQ in terms of MOS points is compared with baselines MMSE-LSA and SG-JMAP [51] and DCCNN [52] method in Table 7. The results show that proposed SS-DWTCT network is able to improve quality of noisy mixture by 0.03 MOS points (2.90) for lower SNR conditions such as -5 over the best baseline method DCCNN. The proposed SS-DWTAT approaches improve quality of noisy mixture by 1.56 MOS points (2.91) over noisy speech and surpass all the other SE methods. The proposed SS-DWT consistently outperforms the minimum setting of PESQ with 1.34 MOS points (2.97) as well as 1.33 MOS points (2.95) for SS-DWTAT and SS-DWTCT, correspondingly, whereas conventional MMSE-LSA and SG-jMAP only slightly improve PESQ across unprocessed noisy speech for the challenging 0 dB condition. Across all noisy conditions with improved Spectral Subtraction (SS) filter and Discrete Wavelet Transform (DWT), the proposed hybrid SS-DWTAT and SS-DWTCT methods highly suppress the background noise compared to baseline methods and achieve highest improvements in PESQ in terms of MOS points.

E. EVALUATION OF COMPUTATIONAL COMPLEXITY

To evaluate the computational complexity the proposed SS-DWT methods are evaluated with a testing speech signal sp26.wav (female) corrupted by babble noise at 5 dB taken from the NOIZEUS speech corpus (NOIZEUS) having a sampling frequency of 8 kHz. In Table 8, the scores for evaluation metrics like processing time and SNR achieved

TABLE 8. Computational efficiency in terms of processing time and SNR.

Methods	Processing Time (s)		SNR (dB)	
Noisy	5.00	10.00	5.00	10.00
DFT [53]	138.60	115.09	5.45	6.32
DCT [53]	59.56	57.49	7.03	7.97
DWT (db1) [53]	10.10	9.80	10.89	11.37
Proposed SS-DWTAT	7.83	7.52	**16.59**	**17.41**
Proposed SS-DWTCT	**7.71**	**7.39**	15.35	16.86

from proposed SE methods are compared against baselines DFT, DCT, and DWT [53]. It is observed that proposed SS-DWTAT achieved processing times of 7.83 seconds and 7.52 seconds under 5 dB and 10 dB SNR conditions, respectively, resulting in SNR improvements of 16.59 dB and 17.41 dB. These performance metrics, coupled with significantly reduced computational power requirements compared to traditional methods like DFT (138.60 s and 115.09 s) and DCT (59.56 s and 57.49 s), highlight its efficiency and effectiveness for real-time applications. The processing times for proposed SS DWTCT were 7.71 sec and 7.39 sec in 5 dB and 10 dB SNR conditions, with SNR improvements of 15.35 dB and 16.86 dB. With less computational demand than traditional approaches (DFT, DCT, and DWT (db1)), it achieves better efficiency and noise reduction efficacy than other methods.

VII. CONCLUSION

In this study, a hybrid speech enhancement framework was proposed which integrates both the Spectral Subtraction (SS) and Discrete Wavelet Transform (DWT) effectively to reduce nonstationary background noise in low Signal-to-Noise Ratio (SNR) environments. The proposed SS DWT methods overcome limitations of traditional SS and DWT methods by integrating iterative noise estimation, Voice Activity Detection (VAD), dynamic noise adaptation, and phase aware spectral reconstruction within the SS component and, adaptive noise refinement and phase aware soft thresholding within the DWT module. Experimental results indicate that the presented hybrid SS DWT techniques dramatically improve noise suppression and reduce musical noise artifacts resulting in the subjective noticeable improvement of speech clarity and intelligibility. The SS-DWT model is consistently shown to produce superior performance, according to metrics including Mean Square Error (MSE), Signal-to-Distortion Ratio (SDR), Short-Time Objective Intelligibility (STOI), and Perceptual Evaluation of Speech Quality (PESQ) in different noise types and datasets. These results confirm the applicability of this integrated approach for real-time applications in communication systems that require hearing aids and speech recognition systems. However, the proposed method has certain limitations: it depends heavily on detailed ambient noise knowledge and needs a dual-channel setup (which could lead to a cost increase and added complexity). Future work will explore integrating machine learning algorithms to enhance noise estimation without prior noise information and developing robust denoising techniques for single-microphone systems. Additionally, implementing the hybrid SS-DWT framework in multi-channel systems and communication technologies such as mobile devices and hearing aids will be pursued to enhance versatility and cost-effectiveness.

REFERENCES

[1] A. Keshavarz and M. Divandari, "Improving speech quality in hearing aids using fuzzy complex and wavelet," in *Proc. 8th Int. Conf. Control, Instrum. Autom. (ICCIA)*, Mar. 2022, pp. 1–6, doi: 10.1109/ICCIA54998.2022.9737161.

[2] H. B. Vanjari and M. T. Kolte, "Comparative analysis of speech enhancement techniques in perceptive of hearing aid design," in *Proc. 3rd Int. Conf. Inf. Manage. Mach. Intell.*, Singapore, D. Goyal, A. Kumar, V. Piuri, and M. Paprzycki, Eds., Cham, Switzerland: Springer, Aug. 2022, pp. 117–125.

[3] I. Fedorov, M. Stamenovic, C. Jensen, L.-C. Yang, A. Mandell, Y. Gan, M. Mattina, and P. N. Whatmough, "TinyLSTMs: Efficient neural speech enhancement for hearing aids," 2020, *arXiv:2005.11138*.

[4] H. Vanjari and M. Kolte, "Comparative analysis of compressive sensing methods for speech enhancement in hearing aid applications," in *Proc. 7th Int. Conf. Signal Process. Commun. (ICSC)*, Nov. 2021, pp. 137–141.

[5] G. Park, W.-H. Cho, K. Kim, and S. Lee, "Speech enhancement for hearing aids with deep learning on environmental noises," *Appl. Sci.*, vol. 10, no. 17, p. 6077, Sep. 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/17/6077

[6] M. Brahim, "Denoising and enhancement speech signal using wavelet," *J. Inf. Syst. Telecommun.*, vol. 9, no. 33, pp. 37–44, Apr. 2021.

[7] P. G. Patil, T. H. Jaware, S. P. Patil, R. D. Badgujar, F. Albu, I. Mahariq, B. Al-Sheikh, and C. Nayak, "Marathi speech intelligibility enhancement using I-AMS based neuro-fuzzy classifier approach for hearing aid users," *IEEE Access*, vol. 10, pp. 123028–123042, 2022.

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

[9] S. Sahu and N. Rayavarapu, "Compressive speech enhancement using semi-soft thresholding and improved threshold estimation," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 13, no. 3, p. 2788, Jan. 2023.

[10] C. Venkatesan, P. Karthigaikumar, and R. Varatharajan, "A novel LMS algorithm for ECG signal preprocessing and KNN classifier based abnormality detection," *Multimedia Tools Appl.*, vol. 77, no. 8, pp. 10365–10374, Apr. 2018.

[11] E. Özen Acarbay and N. Özkurt, "Performance analysis of the speech enhancement application with wavelet transform domain adaptive filters," *Int. J. Speech Technol.*, vol. 26, no. 1, pp. 245–258, Mar. 2023, doi: 10.1007/s10772-023-10022-3.

[12] M. Talbi and M. S. Bouhlel, "A new speech enhancement technique based on stationary bionic wavelet transform and MMSE estimate of spectral amplitude," *Secur. Commun. Netw.*, vol. 2021, pp. 1–11, Dec. 2021.

[13] S. R. Chiluveru and M. Tripathy, "Speech enhancement using a variable level decomposition DWT," *Nat. Acad. Sci. Lett.*, vol. 44, no. 3, pp. 239–242, Aug. 2020.

[14] S. Özaydın and I. K. Alak, "Speech enhancement using maximal overlap discrete wavelet transform," *Gazi Univ. J. Sci. A, Eng. Innov.*, vol. 5, no. 4, pp. 159–171, Dec. 2018.

[15] M. S. E. Abadi, H. Mesgarani, and S. M. Khademiyan, "The wavelet transform-domain LMS adaptive filter employing dynamic selection of subband-coefficients," *Digit. Signal Process.*, vol. 69, pp. 94–105, Oct. 2017.

[16] M. Gupta, R. K. Singh, and S. Singh, "Analysis of optimized spectral subtraction method for single channel speech enhancement," *Wireless Pers. Commun.*, vol. 128, no. 3, pp. 2203–2215, Feb. 2023, doi: 10.1007/s11277-022-10039-y.

[17] R. Kumar, M. Tripathy, and R. S. Anand, "Iterative thresholding-based spectral subtraction algorithm for speech enhancement," in *Advances in VLSI, Signal Processing, Power Electronics, IoT, Communication and Embedded Systems*. Cham, Switzerland: Springer, 2021, pp. 221–232.

[18] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[19] J. Rafiee, P. W. Tse, A. Harifi, and M. H. Sadeghi, "A novel technique for selecting mother wavelet function using an intelli gent fault diagnosis system," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 4862–4875, Apr. 2009.
[20] M. Srivastava, C. L. Anderson, and J. H. Freed, "A new wavelet denoising method for selecting decomposition levels and noise thresholds," *IEEE Access*, vol. 4, pp. 3862–3877, 2016.
[21] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Process. Mag.*, vol. 8, no. 4, pp. 14–38, Oct. 1991.
[22] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1532–1546, Sep. 2000.
[23] S.-S. Wang, P. Lin, Y. Tsao, J.-W. Hung, and B. Su, "Suppression by selecting wavelets for feature compression in distributed speech recognition," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 3, pp. 564–579, Mar. 2018.
[24] D. Huang, L. Ke, B. Mi, G. Wei, J. Wang, and S. Wan, "A cooperative denoising algorithm with interactive dynamic adjustment function for security of stacker in industrial Internet of Things," *Secur. Commun. Netw.*, vol. 2019, pp. 1–16, Feb. 2019.
[25] S. Pradeep Kumar, A. Daripelly, S. M. Rampelli, S. K. R. Nagireddy, A. Badishe, and A. Attanthi, "Noise reduction algorithm for speech enhancement," in *Proc. Int. Conf. Signal Process., Comput., Electron., Power Telecommun. (IConSCEPT)*, May 2023, pp. 1–5, doi: 10.1109/IConSCEPT57958.2023.10170204.
[26] G. Tejaswini, "Speech enhancement using discrete wavelet transform with long short-term memory algorithm," *Nanotechnol. Perceptions*, vol. 20, pp. 18–32, May 2024.
[27] P. Cherukuru and M. B. Mustafa, "CNN-based noise reduction for multi-channel speech enhancement system with discrete wavelet transform (DWT) preprocessing," *PeerJ Comput. Sci.*, vol. 10, p. e1901, Feb. 2024.
[28] E. Ozen and N. Özkurt, "Speech noise reduction with wavelet transform domain adaptive filters," in *Proc. Global Congr. Electr. Eng. (GC-ElecEng)*, Dec. 2021, pp. 15–20, doi: 10.1109/gc-eleceng52322.2021.9788190.
[29] M. Parchami, W.-P. Zhu, B. Champagne, and E. Plourde, "Recent developments in speech enhancement in the short-time Fourier transform domain," *IEEE Circuits Syst. Mag.*, vol. 16, no. 3, pp. 45–77, 3rd Quart., 2016, doi: 10.1109/MCAS.2016.2583681.
[30] H. Pardede, K. Ramli, Y. Suryanto, N. Hayati, and A. Presekal, "Speech enhancement for secure communication using coupled spectral subtraction and Wiener filter," *Electronics*, vol. 8, no. 8, p. 897, Aug. 2019.
[31] M. Balasubrahmanyam, R. S. Valarmathi, and C. H. M. S. Kumar, "A comprehensive review of conventional to modern algorithms of speech enhancement," in *Innovations in Electrical and Electronic Engineering*. Singapore: Springer, 2024, pp. 633–648.
[32] N. Upadhyay, "Iterative-processed multiband speech enhancement for suppressing musical sounds," *Multimedia Tools Appl.*, vol. 83, no. 15, pp. 45423–45441, Oct. 2023, doi: 10.1007/s11042-023-17336-z.
[33] Y. Yang, P. Liu, H. Zhou, and Y. Tian, "A speech enhancement algorithm combining spectral subtraction and wavelet transform," in *Proc. IEEE 4th Int. Conf. Autom., Electron. Electr. Eng. (AUTEEE)*, Nov. 2021, pp. 268–273, doi: 10.1109/AUTEEE52864.2021.9668622.
[34] T. Yadava, B. Nagaraja, and H. Jayanna, "A spatial procedure to spectral subtraction for speech enhancement," *Multimedia Tools Appl.*, vol. 81, no. 17, pp. 23633–23647, Jul. 2022.
[35] H. Gustafsson, S. E. Nordholm, and I. Claesson, "Spectral subtraction using reduced delay convolution and adaptive averaging," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 8, pp. 799–807, Nov. 2001.
[36] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 1979, pp. 208–211.
[37] G. Vinothkumar and M. Kumar, "Speech enhancement with background noise suppression in various data corpus using bi-LSTM algorithm," *Int. J. Electr. Electron. Res.*, vol. 12, no. 1, pp. 322–328, Mar. 2024.
[38] G. Ioannides and V. Rallis, "Real-time speech enhancement using spectral subtraction with minimum statistics and spectral floor," 2023, *arXiv:2302.10313*.
[39] W. Huang, "Wavelet transform adaptive signal detection," Dept. Comput. Eng., North Carolina State Univ., Raleigh, NC, USA, Tech. Rep., 1999.
[40] R. Bendoumia and M. Djendi, "Two-channel variable-step-size forward-and-backward adaptive algorithms for acoustic noise reduction and speech enhancement," *Signal Process.*, vol. 108, pp. 226–244, Mar. 2015.
[41] T. L. Kumar and K. Rajan, "Noise suppression in speech signals using adaptive algorithms," *Int. J. Eng. Res. Appl.*, vol. 2, no. 1, pp. 718–721, 2012.
[42] J. W. Lyons, "DARPA TIMIT acoustic-phonetic continuous speech corpus," Linguistic Data Consortium, Nat. Inst. Standards Technol. (NIST), Gaithersburg, MD, USA, Tech. Rep. LDC93S1, 1993. [Online]. Available: https://catalog.ldc.upenn.edu/LDC93S1
[43] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, Jul. 2006.
[44] U. Ijaz, F. Gillani, A. Iqbal, M. S. Sharif, M. F. Anwar, and A. Ijaz, "Fine-tuning audio compression: Algorithmic implementation and performance metrics," *Int. J. Innov. Sci. Technol.*, vol. 6, no. 1, pp. 220–236, 2024.
[45] S. S. Haykin, *Adaptive Filter Theory*. London, U.K.: Pearson, 2002.
[46] *Perceptual Evaluation of Speech Quality (PESQ): an Objective Method for End-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs*, Standard ITU-T P.862, ITU, 2001.
[47] C. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *Proc. ICASSP*, Mar. 2010, pp. 4214–4217.
[48] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)–A new method for speech quality assessment of telephone networks and codecs," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, May 2001, pp. 749–752.
[49] M. Hassani and M. K. Mollaei, "Speech enhancement based on spectral subtraction in wavelet domain," in *Proc. IEEE 7th Int. Colloq. Signal Process. Appl.*, 2011, pp. 366–370.
[50] G.-Y. Wang, X.-Q. Zhao, and X. Wang, "Speech enhancement based on the combination of spectral subtraction and wavelet thresholding," in *Proc. Int. Conf. Apperceiving Comput. Intell. Anal.*, Oct. 2009, pp. 136–139.
[51] H. Yu, Dept. Post-filter optimization for multichannel automotive speech enhancement," Ph.D. dissertation, Technische Universität Braunschweig, Braunschweig, Germany, 2013.
[52] Y. Iqbal, T. Zhang, M. Fahad, S. U. Rahman, A. Iqbal, Y. Geng, and X. Zhao, "Speech enhancement using deep complex convolutional neural network (DCCNN) model," *Signal, Image Video Process.*, vol. 18, no. 12, pp. 8675–8692, Dec. 2024.
[53] S. Sahu and N. Rayavarapu, "Performance comparison of sparsifying basis functions for compressive speech enhancement," *Int. J. Speech Technol.*, vol. 22, no. 3, pp. 769–783, Sep. 2019.

YASIR IQBAL received the master's degree in electrical engineering from Bahria University, Islamabad, Pakistan, in 2020. He is currently pursuing the Ph.D. degree with the School of Electrical and Information Engineering, Tianjin University, China. His research interests include digital signal and image processing and machine and deep learning networks.

TAO ZHANG received the M.S. degree from the School of Electronic Information Engineering, Tianjin University, Tianjin, China, in 2001, and the Ph.D. degree from Tianjin University, in 2004. He is currently an Associate Professor with Texas Instruments DSP Joint Laboratory, School of Electrical and Information Engineering, Tianjin University. His current interests include intelligent audio and video processing and intelligent computing.

TEDDY SURYA GUNAWAN (Senior Member, IEEE) received the B.Eng. degree (cum laude) in electrical engineering from Institut Teknologi Bandung (ITB), Indonesia, in 1998, the M.Eng. degree from Nanyang Technological University, Singapore, in 2001, and the Ph.D. degree from the University of New South Wales (UNSW), Australia, in 2007. He is a Professor with the Department of Electrical and Computer Engineering, International Islamic University Malaysia (IIUM). He has held esteemed roles, including Visiting Research Fellow at UNSW (2010–2021) and an Adjunct Professor at Telkom University (2022–2023), previously chairing the IEEE Instrumentation and Measurement Society – Malaysia Section. Recognized for his contributions to speech and audio processing, biomedical signal processing, image and video processing, and parallel computing, he received IIUM's Best Researcher Award in 2018 and is listed among the World's Top 2% Scientists in Artificial Intelligence and Image Processing by Elsevier for 2023 and 2024. In addition to his academic and research accomplishments, he holds multiple professional engineering certifications, including CEng (IET, U.K., 2016), Insinyur Profesional Utama (PII, Indonesia, 2019), ASEAN Engineer (2018), ASEAN Chartered Professional Engineer (2020), APEC Engineer (2023), CPEng (Australia, 2024), and PEng (Malaysia, 2025), reflecting his commitment to professional excellence. Within IIUM, he has also served as Head of Department (2015–2016) and Head of Programme Accreditation and Quality Assurance (2017–2018) at the Faculty of Engineering, reinforcing his leadership and expertise in the field.

AGUS PRATONDO (Senior Member, IEEE) received the bachelor's degree in informatics engineering and the master's degree in electrical engineering from the Institut Teknologi Bandung, and the Ph.D. degree in electrical and computer engineering from the National University of Singapore. He is currently a Professor with the Department of Multimedia Engineering, Telkom University, was recognized in 2024 among Elsevier's World Top 2% Scientists in Artificial Intelligence and Image Processing. His research interests span artificial intelligence, machine learning, computer vision, data analytics, and multimedia applications. His notable contributions and expertise make him a distinguished AI research and innovation figure.

XIN ZHAO received the Ph.D. degree from Tianjin University, Tianjin, China, in 2020. He is currently an Associate Professor with the School of Electrical and Information Engineering, Tianjin University. His current research interests include evolutionary computation, machine learning, optimization, and image processing.

YANZHANG GENG received the M.S. degree from the School of Mechanical and Power Engineering, North University of China, China, in 2017, and the Ph.D. degree from the School of Electrical and Information Engineering, Tianjin University, China, in 2023. He is currently a Research Fellow at Tianjin University. His research interests include speech signal processing and microphone array signal processing.

MIRA KARTIWI (Member, IEEE) is currently a Professor with the Department of Information Systems, Kulliyyah of Information and Communication Technology, and the Deputy Director of e-learning with the Centre for Professional Development, International Islamic University Malaysia (IIUM). She is an experienced consultant specializing in the health, financial, and manufacturing sectors. Her current research interests include health informatics, e-commerce, data mining, information systems strategy, business process improvement, product development, marketing, delivery strategy, workshop facilitation, training, and communications. She was one of the recipients of Australia Postgraduate Award (APA), in 2004. For her achievement in research, she was awarded the Higher Degree Research Award for Excellence, in 2007. She has also been appointed as an editorial board member in local and international journals to acknowledge her expertise.

NASIR SALEEM received the B.S. degree in telecommunication engineering from the University of Engineering and Technology, Peshawar, Pakistan, in 2008, the M.S. degree in electrical engineering from CECOS University, Peshawar, in 2012, and the Ph.D. degree in electrical engineering with a specialization in digital speech processing and deep learning from the University of Engineering and Technology, in 2021. Following the Ph.D. degree, he was a Postdoctoral Fellow with Islamic International University Malaysia (IIUM), where he researched modern artificial intelligence-based speech processing algorithms. From 2008 to 2012, he was a Lecturer with the Institute of Engineering Technology (IET), Gomal University, engaging in both teaching and research. Currently, he is an Assistant Professor with the Department of Electrical Engineering, Faculty of Engineering and Technology (FET), Gomal University. He also holds the position of the Deputy Director of the Quality Assurance Directorate with Gomal University. He has published several research papers in renowned journals and conferences, including those by Elsevier, Springer, and IEEE. In addition to his research, he actively participates in academic activities, such as guest editing and paper reviewing. His research interests include human–machine interaction, speech enhancement, speech recognition, speech and video processing, and machine learning applications.

SAMI BOUROUIS received the Engineering, M.Sc., and Ph.D. degrees in computer science from the University of Tunis, Tunisia, in 2003, 2005, and 2011, respectively. He is currently a Professor at the College of Computers and Information Technology, Taif University, Saudi Arabia. His research interests include data mining, image processing, statistical machine learning, cybersecurity, and pattern recognition applied to several real-life applications.