Biorthogonal Wavelet Packet and Adaptive Filters for Noisy Speech Reduction

C. Shraddha
Department of Computer Science and Engineering, B.N.M. Institute of Technology (affiliated to Visvesvaraya Technological University), Bangalore, India | Department of Computer Science and Engineering, Vidyavardhaka College of Engineering (affiliated to Visvesvaraya Technological University), Mysuru, India
shraddha.sunilkumar@gmail.com (corresponding author)

M. L. Chayadevi
Department of Computer Science and Engineering, B.N.M. Institute of Technology (affiliated to Visvesvaraya Technological University), Bangalore, India
chayadevi1999@gmail.com

M. A. Anusuya
Department of Computer Science and Engineering, JSS Science and Technology University, Mysuru, India
anusuya_ma@sjce.ac.in

Received: 20 March 2025 | Revised: 4 July 2025 | Accepted: 6 July 2025

 DOI: https://doi.org/10.48084/etasr.11071

ABSTRACT

Minimizing noise in speech signals is crucial for applications such as speech recognition and enhancement. This paper proposes a hybrid technique that combines a biorthogonal wavelet packet with a Recursive Least Squares (RLS) adaptive filter to reduce environmental and colored noise during the preprocessing stage. Simulation results demonstrate a 2–10% improvement in speech signal strength under noisy conditions. The biorthogonal wavelet's vanishing moments and the length of the RLS filter play key roles in preserving speech characteristics while suppressing noise. Performance is evaluated using Signal-to-Noise Ratio (SNR), Mean Squared Error (MSE), and Peak Signal-to-Noise Ratio (PSNR) metrics, showing effective reduction of pink and babble noise across varying decibel levels, thereby ensuring enhancing the clarity of speech for recognition applications.

Keywords-noise reduction; biorthogonal wavelet packet; adaptive filtering; Recursive Least Squares (RLS); colored noise; environmental noise

I. INTRODUCTION

Speech recognition is the ability of a machine or software to accurately identify and interpret words and phrases in spoken language using acoustic and language modeling techniques [1]. The presence of unwanted noise in a speech signal significantly impacts the performance of recognition systems [2]. In speech-related applications, distinguishing noisy segments by retaining essential speech features poses a major challenge. Noise in speech signals can arise naturally or be introduced artificially, occurring in either additive or convolutive forms. These unwanted signals manifest as white, pink, babble, street noise, and many other environmental disturbances [3].

Noise removal from speech signals has been a key focus in speech processing research, as varying noise characteristics degrade communication quality and introduce errors. Moderate noise may affect comfort and naturalness, whereas excessive noise can render speech unintelligible. To ensure effective communication, digital signal processing techniques are applied to speech signals before storage, transmission, or playback, addressing these challenges through noise reduction methods. Removing noise from a degraded speech signal to improve perceived speech quality is addressed through the speech enhancement process. This enhances the speech by reducing the noise strength and increasing the speech signal strength [4].

This paper highlights the application and combination of wavelet packets and adaptive filters, which play a major role in moderately reducing colored and environmental noise present in speech signals. This process is implemented in the front-end processing of speech signals. The use of biorthogonal wavelets decomposes both low-pass and high-pass components, enabling noise reduction across frequency bands. Since biorthogonal wavelet packets offer symmetrical decomposition, they help reconstruct noisy speech using the weighted coefficients of Recursive Least Squares (RLS) filters.

Some related work on noise reduction using wavelet packets and the RLS adaptive filter, both in their independent and hybrid approaches, has been conducted by various researchers.

A hybrid technique that uses Biorthogonal Wavelet Transform-Butterworth (BWT-BW) filters to reduce street noise and babble noise at 5, 10, and 15 dB, achieving a significantly higher output Signal-to-Noise Ratio (SNR) and a Peak Signal-to-Noise Ratio (PSNR) with an average improvement of 8 dB, was presented by authors in [5]. Additionally, authors in [6] used a hybrid strategy to reduce additive white Gaussian noise and achieved 99% recognition accuracy at 10 dB by combining the Normalized Least Mean Squares (NLMS) filter with the Morlet wavelet. Authors in [7] demonstrated improved SNR of 4 dB using Wavelet Packet Transforms (WPT) for denoising speech signals corrupted by environmental noise. Authors in [8] combined WPT with adaptive thresholding techniques to reduce environmental noise. Authors in [9] investigated the optimal selection of biorthogonal wavelet bases, demonstrating that proper selection can further enhance noise reduction. Authors in [10] applied wavelet packets with adaptive filtering to reduce white, babble, and airport noise in speech signals. Authors in [11] used adaptive filtering in the wavelet transform domain, improving noise reduction in speech signals corrupted by white and pink noise with a Perceptual Evaluation of Speech Quality (PESQ) score of 3.308. As demonstrated by authors in [12], the RLS filter is effective in reducing background noise and improving an average SNR by 5 dB in speech signals [12].

It is evident from the literature mentioned above that not many authors have explored pink noise reduction. Additionally, negative decibel levels have also been largely overlooked. To enhance noise reduction in speech applications, authors in [13] suggested implementing the RLS filter. Hence, in this paper, the RLS filter is combined with the biorthogonal wavelet packet to analyze and evaluate its performance in reducing pink and babble noise.

II. PROPOSED NOISE REDUCTION APPROACH

This section explains how the Biorthogonal Wavelet Packet Transformation (BWPT) and RLS adaptive filtering techniques were adopted and combined for the noise reduction. The signal is decomposed up to 4 levels, and only detailed coefficients are extracted by setting $N = 4$ of the bior6.8 wavelet at the decomposition phase. The vanishing moment values are decided using the trial-and-error procedure such that the reconstructed signal has minimum error. Hence, the value of the vanishing parameter is set to 6. By setting the threshold parameter to the Stein's Unbiased Risk Estimator (SURE) function, the signal strength is recalculated during the reconstruction process. This step is performed iteratively until the difference between the original and the reconstructed signal is ≤0.5 dB. This threshold value provides clear perceptual quality between the pre- and post-reconstructed signal when the wavelet packet is used.

The signal is further processed by the adaptive RLS filter to improve noise reduction by computing the weighted vector and correlation matrix, setting the forgetting factor of the filter to 1. This factor is determined through a trial-and-error process by analyzing the Power Spectral Density (PSD) of the noisy speech signal.

The linear coupling between noise and the speech signal is assessed using the wavelet's high-frequency coefficients along with the covariance matrix coefficients of the RLS filter. The RLS filter ensures strong convergence between the noisy filtered coefficients and the clean speech coefficients by minimizing error through the recursive updating of filter coefficients. This effect is more noticeable for both types of noise, particularly pink and babble noise, as they have specific characteristics in common with the speech signal, including peak structure, duration, and PSD.

A. *Hybrid Approach Algorithm*

The following steps outline the proposed hybrid noise reduction algorithm.

```
Input : Synthesized noisy speech signal.
Output: Reduced noise signal.
Step 1: Read the synthesized noisy speech
      signal.
Step 2: Decompose the signal to level 4
      using biorthogonal wavelet packet.
Step 3: Compute the first 24 detailed
      coefficients for every frame of the
      speech signal.
Step 4: Compute lambda coefficients by
      setting Threshold=SURE.
Step 5: Initialize the gain vector and
      compute the error correlation matrix
      of the RLS filter:
      a) Compute the gain vector and error
         signal.
Step 6: Repeat the above process for the
      entire signal length using the RLS
      filter with the following settings:
      a step size of 0.0026, a forgetting
      factor of 1.043, and a filter length
      of 1 to obtain the final denoised
      signal.
Step 7: Reconstruct the signal using
      inverse WPT.
Step 8: Repeat steps 4-6 until the
      difference between the previous and
      current weighted vectors of the
      reconstructed noisy signal is <=0.5
      for the entire signal length.
Step 9: Compute the SNR, Mean Squared
      Error (MSE), and PSNR metrics of the
      reconstructed signal and compare the
      enhanced signal strength with the
      pre- and post-decibel values.
```

III. MATERIALS AND METHODS

A. Dataset

This section discusses the different types of datasets used for simulation purposes to verify the performance of the proposed approach. These datasets are utilized to create synthesized and derived datasets. Essentially, they consist of two types: the speech dataset and the noisy dataset. The standard datasets used to generate our derived dataset are discussed below:

- Speech dataset: Speech signals are sourced from the TIDIGITS speech corpus [14]. English read-digit utterances from five men and five women, each pronouncing a 10-digit sequence, are used in our experiment. These speech signals are recorded in a studio environment at a sampling frequency of 20 kHz.
- Noise dataset: This study makes use of two pre-existing real-world noise datasets, Aurora-2 and Noisex-92. There are eight distinct kinds of noise in the Noisex-92 dataset, of which we have selected four: white, pink, babble, and street, at a sampling frequency of 16 kHz [15]. Additionally, babble and street noise are sourced from the Aurora-2 dataset [16]. White and pink noises are obtained from Freesound [17], also at a sampling frequency of 16 kHz, for our experiment.
- Synthesized dataset: The synthesized noisy signals for our experiment are created by combining pre-scaled speech and noise files at various SNR levels. The SNR is varied in ±5 dB increments, ranging from -15 dB to 15 dB, including both positive and negative decibel values (-15, -10, -5, 0, 5, 10, and 15). The 'clean' speech signals from the TIDIGITS corpus and the noise recordings from NOISEX-92 and AURORA-2 are used to generate the synthesized datasets required for our experiment. The speech signal mixed with the NOISEX-92 dataset is referred to as Dataset-1; whereas the speech signal mixed with the AURORA-2 and Freesound datasets is referred to as Dataset-2.

B. Methodology

Figure 1 illustrates the proposed system architecture for noise reduction using a hybrid technique. The synthesized noisy speech signal is generated using the TIDIGITS speech corpus and noisy datasets from Aurora-2 and Noisex-92. To assess signal strength at the pre-processing stage, statistical metrics such as SNR, PSNR, and MSE are computed before applying the BWPT with the RLS filter. Research indicates that hybrid approaches combined with adaptive methods enhance noise reduction, supporting the choice of this technique. After pre-processing and post-processing, the reconstructed speech signal is evaluated, as shown in Figure 1. The simulation results demonstrate reduced noise strength and improved signal clarity for both colored and environmental noise at varying decibel levels. The underlying mathematical concepts of the noise reduction algorithm are also detailed.

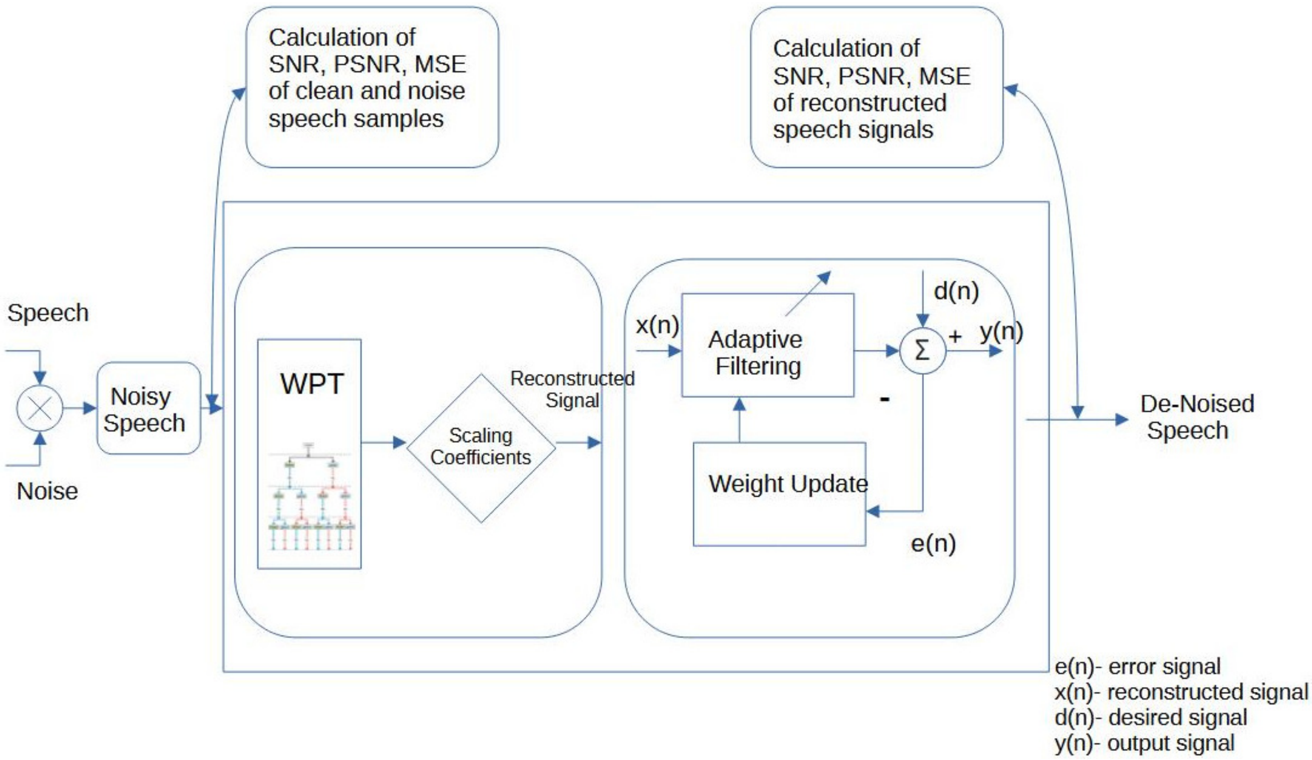


Fig. 1. System architecture of the proposed BWPT-RLS noise reduction method.

C. Adaptive Recursive Least Squares filter

The RLS filter is an adaptive filter commonly used in digital signal processing to minimize error and enhance signal quality [18]. It is designed to reduce the error between a desired signal and the actual output by recursively adjusting the filter coefficients to minimize a weighted linear least-squares cost function based on the input signals, using a step size of 0.0026, a forgetting factor of 1.043, and a filter length of 1 [19]. This filter exhibits a faster convergence rate compared to other adaptive filters, such as the Least Mean Squares (LMS) filter, making it more efficient in dynamic signal processing applications [20, 21]. The filter is mathematically defined by (1):

$$w(n) = w(n) + \mathrm{e}(n) \cdot k(n) \quad (1)$$

where $w(n)$ is the filter coefficient vector at time n, $\mathrm{e}(n)$ is the error signal, and $k(n)$ is the gain vector given by:

$$k(n) = \frac{p(n) \cdot u(n)}{m + u^T \cdot p(n)} \cdot u(n)$$

where $p(n)$ is the covariance matrix, $u(n)$ is the input vector, and m is a scalar.

D. Biorthogonal Wavelet Packet Transformation

BWPT is an extension of wavelet transforms. Biorthogonal wavelets use two different wavelet functions, one for decomposition and another for reconstruction [22, 23]. This helps to match the characteristics of the noisy speech signal being processed. The SURE thresholding function is essential for effective noise reduction during the reconstruction stage [24]. The use of the inverse BWPT reconstructs the signal efficiently with reduced noise [25].

Biorthogonal wavelets use two wavelet functions (ψ) and scaling functions (φ). The mother wavelet functions are derived from the scaling and wavelet filter coefficients defined in (2) and (3) [26, 27].

$$\varphi(t) = \sum_k h_k \varphi(2t - k) \quad (2)$$

$$\psi(t) = \sum_k g_k \varphi(2t - k) \quad (3)$$

where h_k and g_k are the low- and high-pass filter coefficients.

The BWPT operates in three main phases:

- Decomposition: The noisy speech signal is decomposed into 4 frequency bands using the biorthogonal wavelet packets defined in (4) and (5). This decomposition helps in isolating the noise components from the speech components. The functions used for the decomposition are:

$$W_{j+1}^{2n}(k) = \sum_m h(m - 2k) W_j^n(m) \quad (4)$$

$$W_{j+1}^{2n+i}(k) = \sum_m g(m - 2k) W_j^n(m) \quad (5)$$

- Thresholding: Once the signal is decomposed, the SURE thresholding technique is applied to the wavelet coefficients. This step helps in removing the noise components while preserving the important speech components [28, 29].
- Reconstruction: The reconstruction functions use the dual filters defined in (6):

$$W_j^n(k) = \sum_m (\tilde{h}(m - 2k) W_{j+1}^n(m) + \tilde{g}(m - 2k) W_{j+1}^{2n+1}(m)) \quad (6)$$

where $h(m)$ and $g(m)$ are low- and high-pass filter coefficients, $W_j^n(m)$ are the wavelet coefficients at level j, and k, m are the indices of the downsampled and input coefficients.

IV. PERFORMANCE EVALUATION

The performance of the hybrid algorithms is evaluated based on the following criteria to demonstrate noise reduction and speech signal enhancement.

A. Signal-to-Noise Ratio

SNR is a metric that calculates the ratio of the desired signal level to the noise level in a speech signal, as defined in (7) [30]. Table I presents the computed SNR values before and after applying the proposed filtering technique for initial signal SNR levels ranging from ±5 to ±15 db.

$$\mathrm{SNR} = 10 * \log_{10} \frac{\sum (s(n))^2}{\sum (s(n) - e(n))^2} \quad (7)$$

where $s(n)$ is the desired speech signal and $e(n)$ is error signal.

TABLE I. BWPT-RLS CALCULATION OF SNR FOR WHITE, PINK, BABBLE, AND STREET NOISE

Noise type	Noise level (dB)	Dataset-1		Dataset-2	
		Before filtering	After applying BWPT+RLS	Before filtering	After applying BWPT+RLS
White	-5	24.99	36.8034	25.02	33.5901
	-10	20.01	37.6437	20.06	35.7967
	-15	15.09	35.5755	15.16	35.2526
	5	6.085	37.3191	6.27	36.195
	10	2.88	42.5782	3.10	41.6773
	15	1.085	36.1497	1.27	35.6873
Pink	-5	25.01	37.4	25.02	33.4526
	-10	20.04	38.5422	20.06	36.765
	-15	15.13	35.9358	15.16	35.8673
	5	6.17	38.0101	6.26	37.1297
	10	2.98	44.2285	3.09	43.9706
	15	1.17	36.7187	1.26	36.0107
Babble	-5	24.99	35.9268	25.02	35.4326
	-10	20.01	36.8416	20.05	35.0653
	-15	15.08	37.6699	15.15	34.044
	5	6.07	37.9586	6.24	36.6558
	10	2.87	44.0505	3.06	43.4813
	15	1.07	36.8242	1.24	36.2657
Street	-5	25.02	34.6724	25.00	34.591
	-10	20.06	35.725	20.02	34.0753
	-15	15.09	36.4788	15.10	32.6612
	5	6.09	36.9305	6.10	35.7794
	10	2.89	41.623	2.90	40.7963
	15	1.09	35.7506	1.10	35.0608

As demonstrated in Table I, for both positive and negative decibel levels, the hybrid noise reduction method, BWPT with

RLS, demonstrates an average improvement of 38 dB in SNR across the two synthesized datasets. The adaptive nature of SURE ensures effective thresholding of the noise coefficients and minimizes the noise residue when the RLS filter is applied. An average SNR improvement of 40 dB is observed for pink and babble noise at 10dB.

B. Mean Square Error

MSE quantifies the difference between the input speech $s(n)$ and the reconstructed enhanced speech output $e(n)$, as defined in (8). Table II depicts the MSE values for Dataset-1 and Dataset-2.

$$\mathrm{MSE} = \mathrm{E}(s(n) - e(n))^2 \quad (8)$$

TABLE II. BWPT-RLS CALCULATION OF MSE FOR WHITE, PINK, BABBLE, AND STREET NOISE

Noise type	Noise level (dB)	Dataset-1		Dataset-2	
		Before filtering	After applying BWPT+RLS	Before filtering	After applying BWPT+RLS
White	-5	0.5421	2.09E-04	0.5527	2.15E-04
	-10	0.5166	1.72E-04	0.5239	1.83E-04
	-15	0.5073	2.77E-04	0.511	2.99E-04
	5	0.6644	1.85E-04	0.6717	1.96E-04
	10	0.7553	5.52E-05	0.7558	5.95E-05
	15	0.841	2.43E-04	0.8316	2.48E-04
Pink	-5	0.5356	1.82E-04	0.5556	1.88E-04
	-10	0.5099	1.40E-04	0.5236	1.46E-04
	-15	0.5027	2.55E-04	0.5101	3.08E-04
	5	0.6678	1.58E-04	0.6633	1.57E-04
	10	0.7529	3.78E-05	0.7416	3.44E-05
	15	0.8385	2.13E-04	0.8328	2.28E-04
Babble	-5	0.5414	2.55E-04	0.5275	2.74E-04
	-10	0.5148	2.07E-04	0.5065	2.11E-04
	-15	0.5049	1.71E-04	0.5008	1.93E-04
	5	0.6674	1.60E-04	0.6506	1.72E-04
	10	0.7556	3.94E-05	0.7393	3.83E-05
	15	0.8373	2.08E-04	0.8278	2.14E-04
Street	-5	0.5414	2.55E-04	0.5473	3.70E-04
	-10	0.5148	2.07E-04	0.5214	2.71E-04
	-15	0.5049	1.71E-04	0.5097	2.37E-04
	5	0.6674	1.60E-04	0.6678	2.14E-04
	10	0.7556	3.94E-05	0.7507	7.22E-05
	15	0.8373	2.08E-04	0.8306	2.84E-04

As demonstrated in Table II, the average noise reduction with respect to the MSE at 10 dB is 7.5E-05 dB for both Dataset-1 and Dataset-2. As the step size is optimized to 0.0026, the forgetting factor is set to 1.043, and the minimum filter length is maintained at 1, it is evident that the MSE decreases. An average improvement of 0.75 dB and 0.51 dB is observed for positive and negative decibels, respectively.

C. Peak Signal-to-Noise Ratio

PSNR is a metric used to evaluate the quality of the reconstructed speech, as defined in (9). Higher PSNR values are indicative of enhanced reconstruction quality. Table III tabulates the performance values of PSNR.

$$\mathrm{PSNR} = 10 * \log_{10}\left(\frac{1}{\mathrm{MSE}}\right) \quad (9)$$

TABLE III. BWPT-RLS CALCULATION OF PSNR FOR WHITE, PINK, BABBLE, AND STREET NOISE

Noise type	Noise level (dB)	Dataset-1		Dataset-2	
		Before filtering	After applying BWPT+RLS	Before filtering	After applying BWPT+RLS
White	-5	2.6592	35.5755	2.575	36.6675
	-10	2.8682	37.6437	2.8071	37.3709
	-15	2.9476	36.8034	2.9157	35.2403
	5	1.7759	37.3191	1.7284	37.0685
	10	1.219	42.5782	1.2159	42.2567
	15	0.7519	36.1497	0.8006	36.0586
Pink	-5	2.7114	35.9358	2.5521	37.2638
	-10	2.925	38.5422	2.8103	38.342
	-15	2.9869	37.4	2.9237	35.1093
	5	1.7539	38.0101	1.7826	38.0426
	10	1.2327	44.2285	1.2984	44.6312
	15	0.7651	36.7187	0.7944	36.4126
Babble	-5	2.665	37.6699	2.7781	35.6219
	-10	2.8835	36.8416	2.9541	36.7553
	-15	2.9681	35.9268	3.003	37.1513
	5	1.7561	37.9586	1.8666	37.6568
	10	1.2174	44.0505	1.3115	44.1731
	15	0.7713	36.8242	0.8205	36.6896
Street	-5	2.6174	36.4788	2.6174	34.3201
	-10	2.8281	35.725	2.8281	35.6663
	-15	2.9265	34.6724	2.9265	36.2498
	5	1.7534	36.9305	1.7534	36.6863
	10	1.2454	41.623	1.2454	41.4159
	15	0.8062	35.7506	0.8062	35.4595

As demonstrated in Table III, the proposed method significantly improves the PSNR, achieving values close to 40 dB across all noise types. Notably, shorter filters yield higher PSNR values compared to longer filters. The method achieves an average PSNR of 21 dB for positive decibels and 19 dB for negative decibels.

The simulation analysis from Tables I, II, and III indicates that the proposed hybrid method performs better for positive decibel levels compared to negative decibel levels. This is achieved due to the increase in the power of the signal, boosted due to the adoption of the bior6.8 wavelet packet, with its vanishing moment value set to 6 when decomposed at level 4. Among the above metrics, SNR and PSNR are highly related due to their approximated performance analysis of 40dB. Therefore, these two metrics are considered the most effective for evaluating the noise reduction performance in the proposed work [24, 31].

D. Comparison of the Proposed Method with Existing Techniques

1) Spectrogram Representation

Figures 2 and 3 present the spectrograms of the proposed method for Dataset-1 and Dataset-2, respectively. Figures 2(a) and 3(a) show the spectrogram of clean speech for the utterance of digit 1. Figures 2(b, c) and 3(b, c) depict the speech signals corrupted by pink and babble noise at 10 dB, respectively. Figures 2(d, e) and 3(d, e) show the first-level noise-reduced signals using BWPT, whereas Figures 2(f, g) and 3(f, g) illustrate the final-level noise-reduced signals.

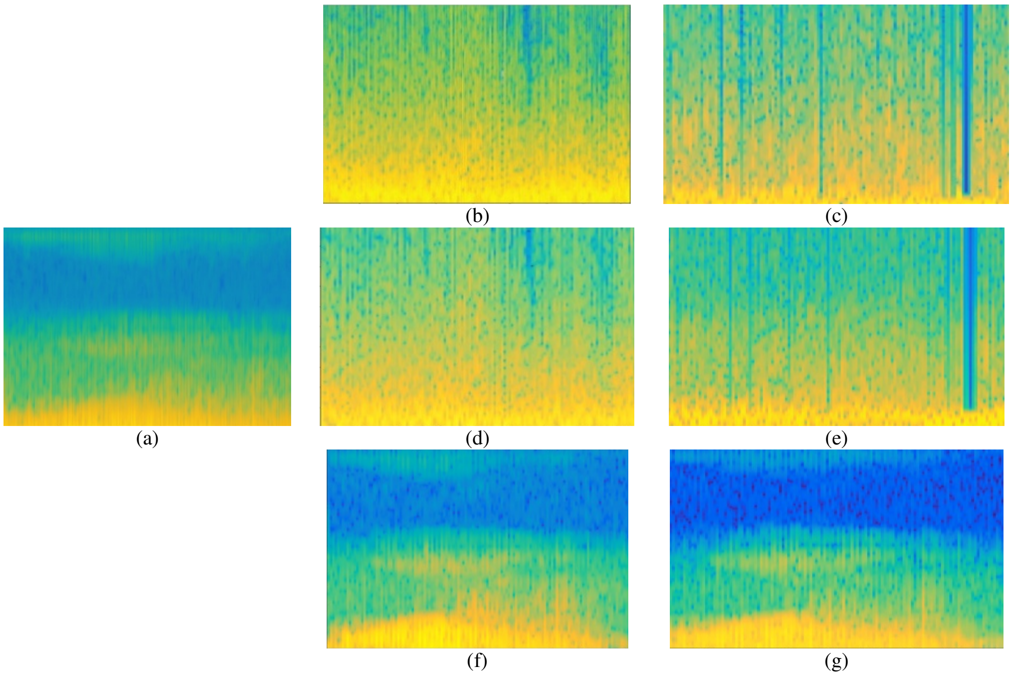


Fig. 2. Spectrogram representation of: (a) clean speech, (b) speech corrupted by pink noise at 10 dB, (c) speech corrupted by babble noise at 10 dB, (d) & (e) first-level denoised speech using BWPT, and (f) & (g) final denoised speech after applying BWPT with the RLS filter for Dataset-1.

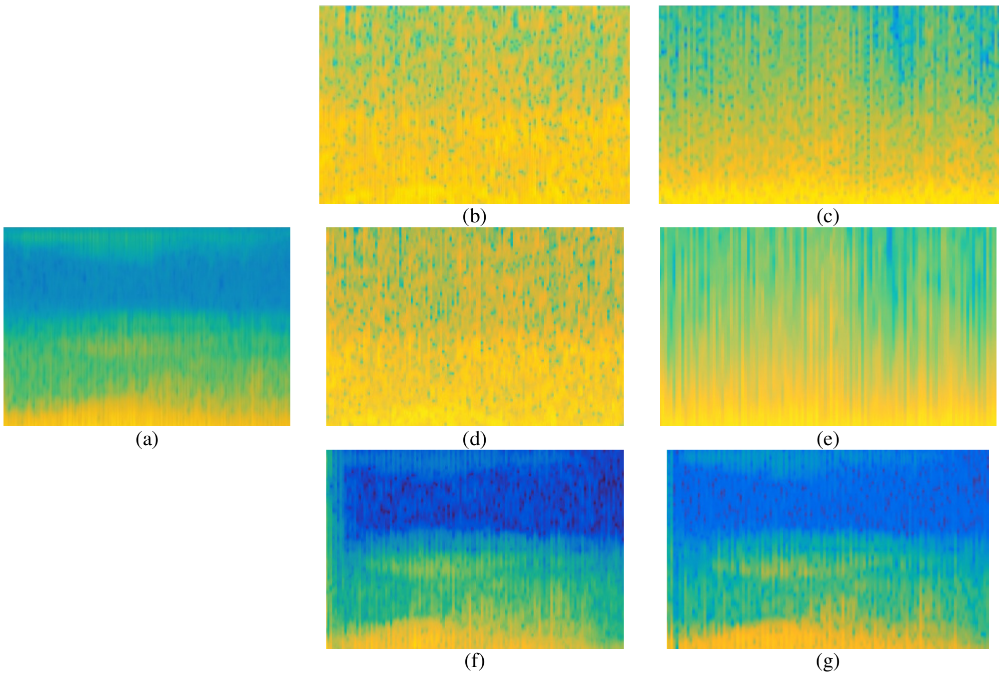


Fig. 3. Spectrogram representation of: (a) clean speech, (b) speech corrupted by pink noise at 10 dB, (c) speech corrupted by babble noise at 10 dB, (d) & (e) first-level denoised speech using BWPT, and (f) & (g) final denoised speech after applying BWPT with the RLS filter for Dataset-2.

Since the SNR metric demonstrates better noise reduction performance at 10 dB, it has been plotted for both positive and negative decibels of pink and babble noise. The proposed method achieves a higher SNR for both noise types, reducing noise by an average of 6 dB for pink noise and 7.12 dB for babble noise in Dataset-1. For Dataset-2, an average SNR improvement of 5 dB is observed for both noise types.

2) *Comparative Analysis with Existing Techniques*

The performance of the proposed method is compared with existing techniques from the literature. However, identifying applications of noise reduction specifically for pink and babble noise remains a significant challenge [32, 33]. Also, none of the existing works focus on noise reduction for negative decibel levels. The noise reduction techniques discussed in the literature are typically evaluated at 5dB and 10dB of noise levels for various speech datasets, such as microphone array recordings and speech recorded in a closed environment [34, 35].

Figure 4 presents a performance comparison of the proposed BWPT-RLS method with two baseline noise reduction techniques: BWT-BW and WT-NLMS. As shown in the figure, the proposed BWPT-RLS technique achieves superior SNR improvement across all four noise types: white, pink, babble, and street.

Table IV presents the results of the proposed noise reduction technique in comparison with existing works in the literature. It is evident from the literature that the RLS filter with a wavelet packet has not been adopted as a hybrid method, nor has its performance been explored by the research community. Additionally, pink, babble, and street noise reduction have been less frequently addressed by researchers for both decibel levels [36, 37]. The proposed method achieves an average improvement of 1.35 dB compared to wavelet and wavelet packet techniques. White and pink noise were applied to a microphone speech dataset and tested at 0 dB. Few experiments have been conducted on colored noises at this level.

3) Observations & Discussion

The simulations results indicate that the adaptive RLS filter with BWPT exhibits superior performance in reducing pink and babble noise at 10dB. An average improvement of 1.8 dB to 4 dB is observed, attributed to the PSD features and their similarity to natural speech sounds. Among environmental noises, babble noise is better identified and reduced due to its uniform PSD across frequencies, whereas street noise exhibits a variable PSD with higher power in lower frequencies. Among colored noises, white noise has a flat PSD with equal power across all frequencies, whereas in pink noise, the PSD decreases by 3 dB per octave, resulting in greater power in lower frequencies, which helps in identifying and reducing this noise. Clearly, pink noise (colored) and babble noise (environmental) exhibit better noise reduction at 10 dB. This is because the overlapping voices create a relatively constant noise level across different frequency bands. Hence, these two types of noise can achieve good noise reduction performance.

Figure 5 illustrates the gradual improvement in SNR performance when speech signals are corrupted by pink and babble noise on both datasets. This improvement is attributed to the convergence speed of the RLS algorithm, which heavily depends on the correlation and weighted vectors of the preprocessed signal. The process preserves the symmetry of both the speech and noise-modified signal structures, facilitating effective noise reduction by enhancing the adaptive filter's convergence speed.

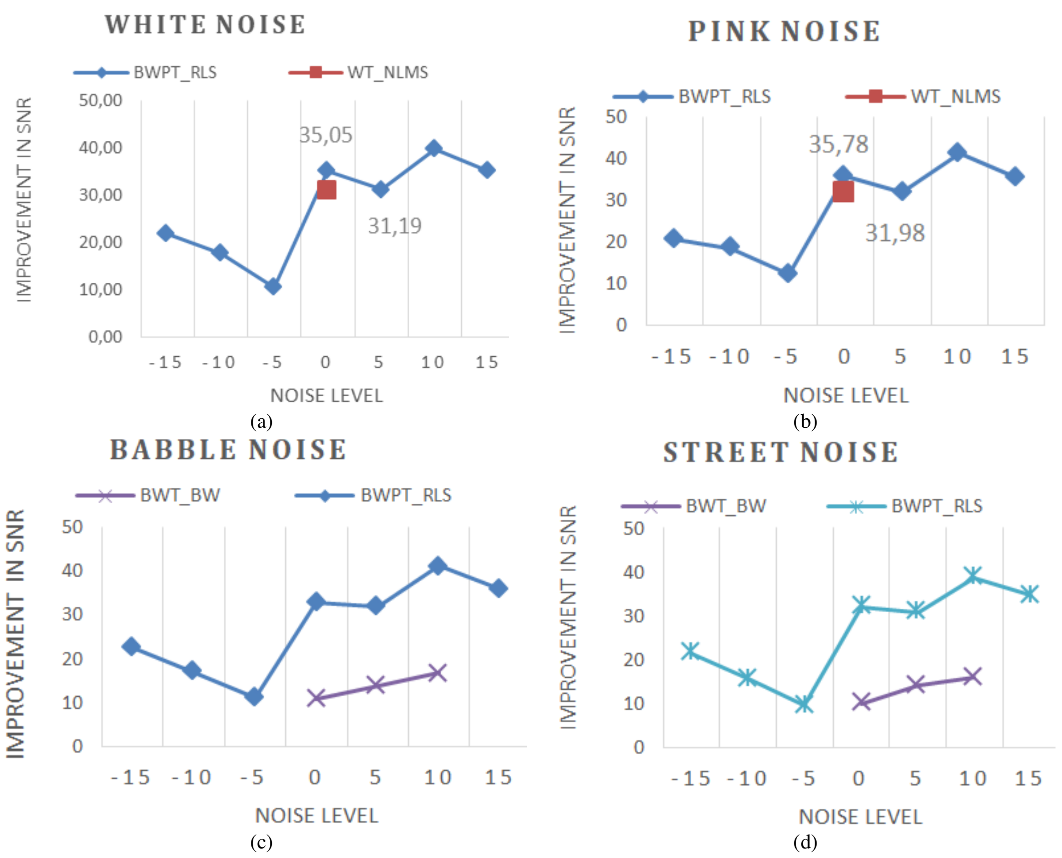
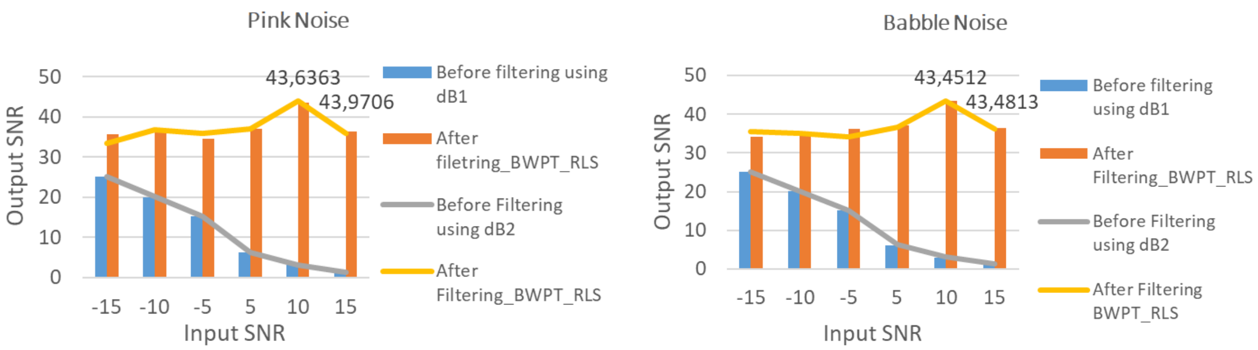


Fig. 4. SNR improvement comparison of the proposed BWPT-RLS method with BWT-BW and WT-NLMS for: (a) white, (b) pink, (c) babble, and (d) street noise.

TABLE IV. COMPARISON OF THE PROPOSED AND EXISTING HYBRID TECHNIQUES ON DATASET-1

Noise type	Noise level (dB)	BWT_BW [5] (dB)	WTD_NLMS [11] (dB)	MCBWT_RLS [18] (dB)	Proposed BPWT_RLS (dB)
White	-5	-	-	35.13	35.58
	-10	-	-	36.02	37.64
	-15	-	-	34.09	36.80
	0	-	31.19	-	35.05
	5	-	-	36.39	37.32
	10	-	-	41.97	42.58
	15	-	-	35.78	36.15
Pink	-5	-	-	35.71	35.94
	-10	-	-	36.91	38.54
	-15	-	-	34.47	37.40
	0	-	31.98	-	35.78
	5	-	-	37.14	38.01
	10	-	-	43.64	44.23
	15	-	-	36.34	36.72
Babble	-5	-	-	34.23	37.67
	-10	-	-	35.20	36.84
	-15	-	-	36.18	35.93
	0	-	-	-	32.68
	5	7.85	-	37.04	37.96
	10	3.93	-	43.45	44.05
	15	1.54	-	36.45	36.82
Street	-5	-	-	33.01	36.48
	-10	-	-	34.13	35.73
	-15	-	-	35.03	34.67
	0	-	-	-	32.01
	5	7.75	-	36.02	36.93
	10	3.98	-	41.00	41.62
	15	1.55	-	35.35	35.75



(a) (b)

Fig. 5. SNR improvement for the proposed BWPT-RLS noise reduction technique on: (a) Dataset-1, and (b) Dataset-2.

V. CONCLUSION AND FUTURE ENHANCEMENTS

In this paper, we propose a hybrid Biorthogonal Wavelet Packet Transformation with Recursive Least Squares (BWPT-RLS) filter to enhance noisy speech signals affected by white, pink, babble, and street noise. The TIDIGITS dataset and selected noise types were chosen for their relevance to real-world speech scenarios, providing a baseline before scaling to larger, more complex datasets. By combining BWPT with RLS filtering, and tuning parameters such as shorter filter lengths and suitable vanishing moments, the method leverages the strengths of both adaptive filtering and wavelet packet analysis, achieving significant noise reduction, particularly at 10 dB for pink and babble noise. Replacing the traditional wavelet with a wavelet packet yields an average Signal-to-Noise Ratio (SNR) improvement of 1.35 dB, attributed to Power Spectral Density (PSD) features that closely resemble natural speech when processed through wavelet packet decomposition and adaptive filtering.

Due to the limited comparable research in the existing literature on Dataset-2, which contains noisy speech signals where the speech signals are taken from the TIDIGITS corpus contaminated by two different types of noise—white and pink noises from Freesound, and babble and street noises from the AURORA-2 corpus—at various decibel levels, most discussion focuses on Dataset-1, which contains noisy speech signals where the speech signals are taken from the TIDIGITS corpus contaminated by four different types of noises—white, pink, babble, and street noises from the NOISEX-92 corpus. Based on available studies, an average of 2-10% improvement at 10 dB SNR is reported for pink and babble noise in Dataset-1, with SNR being the most impactful among evaluation metrics.

The proposed BWPT-RLS method shows a milder effect on Dataset-2, though the results still highlight the role of PSD in noise reduction. Given the convoluted nature of environmental and colored noise, which enhances the model's ability to distinguish between noise types, the BWPT-RLS approach may be better evaluated using Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR), as these metrics better reflect performance than SNR alone. Segmental SNR could also be explored to further enhance signal quality. Future work may extend this approach to feature extraction techniques such as Linear Predictive Cepstral Coefficients (LPCC), Mel-Frequency Cepstral Coefficients (MFCC), and multi-taper MFCC, as well as model-based systems, moving beyond preprocessing and traditional statistical metrics.

REFERENCES

[1] J. Kapoor, G. R. Mishra, A. Pathak, and M. Rai, "Analysis of BSC and AWGN Channel Distortion Effect on Sound Signal in Active Noise Cancellation Application," *Engineering Letters*, vol. 29, no. 3, pp. 926–930, Sept. 2021.

[2] S. C, C. M. L, A. M. A, and V. H. Y, "Enhancing Noise Reduction with Bionic Wavelet and Adaptive Filtering," *Inteligencia Artificial*, vol. 27, no. 74, pp. 214–226, Sept. 2024, https://doi.org/10.4114/intartif.vol27iss74pp214-226.

[3] I. P. Suryawanshi and C. A. Manjare, "Speech signal analysis and enhancement based on HNM synthesis," in *2016 International Conference on Computing Communication Control and automation*, Pune, India, 2016, pp. 1–5, https://doi.org/10.1109/ICCUBEA.2016.7860100.

[4] S. C, C. M L, and M. A. Anusuya, "Noise Cancellation and Noise Reduction Techniques: A Review," in *2019 1st International Conference on Advances in Information Technology*, Chikmagalur, India, 2019, pp. 159–166, https://doi.org/10.1109/ICAIT47043.2019.8987262.

[5] A. Garg and O. P. Sahu, "A hybrid approach for speech enhancement using Bionic wavelet transform and Butterworth filter," *International Journal of Computers and Applications*, vol. 42, no. 7, pp. 686–696, Oct. 2020, https://doi.org/10.1080/1206212X.2019.1614293.

[6] H. Y. Vani, M. A. Anusuya, and M. L. Chayadevi, "Morlet-Kernel Principal Component Analysis Features for Speech Recognition," *Journal of Computational and Theoretical Nanoscience*, vol. 17, no. 9–10, pp. 4482–4486, July 2020, https://doi.org/10.1166/jctn.2020.9102.

[7] G. Kumar, S. Kumar, and N. Kumar, "Comparative Study of Wavelet and Wavelet Packet Transform for Denoising Telephonic Speech Signal," *International Journal of Computer Applications*, vol. 110, no. 15, pp. 1–8, Jan. 2015, https://doi.org/10.5120/19389-9145.

[8] M. Shafieian and M. Rahmanian, "An unsupervised approach for improving speech enhancement using wavelet packet transform and adaptive thresholding," *Revista INGENIERÍA UC*, vol. 26, no. 3, pp. 319–336, Dec. 2019.

[9] E. Z. Engin and Ö. Arslan, "Selection of Optimum Mother Wavelet Function for Turkish Phonemes," *International Journal of Applied Mathematics Electronics and Computers*, vol. 7, no. 3, pp. 56–64, Sept. 2019, https://doi.org/10.18100/ijamec.556850.

[10] J. S. Jakati and S. S. Kuntoji, "A Novel Speech Enhancement Solution UsingHybrid Wavelet Transformation Least Means SquareMet hod," *International Journal of Engineering Trends and Technology*, vol. 69, no. 7, pp. 233–243, July 2021, https://doi.org/10.14445/22315381/IJETT-V69I7P230.

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

[12] C. B. Mbachu and S. A. Akaneme, "Noise Reduction in Speech Signals Using Recursive Least Square Adaptive Algorithm," *International Journal of Engineering and Advanced Technology Studies*, vol. 8, no. 2, pp. 1–17, June 2020.

[13] S. Kotte, "Performance Analysis of Adaptive Algorithms based on different parameters Implemented for Acoustic Echo Cancellation in Speech Signals," Ph.D. dissertation, Blekinge Institute of Technology, Karlskrona, Sweden, 2011.

[14] R. Gary Leonard and Doddington, George R., "TIDIGITS." Linguistic Data Consortium, 1993, https://doi.org/10.35111/72XZ-6X59.

[15] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, vol. 12, no. 3, pp. 247–251, July 1993, https://doi.org/10.1016/0167-6393(93)90095-3.

[16] D. Pearce and H.-G. Hirsch, "The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions," in *6th International Conference on Spoken Language Processing*, Beijing, China, 2000, pp. 29–32, https://doi.org/10.21437/ICSLP.2000-743.

[17] "Freesound." Freesound Labs. https://freesound.org/.

[18] S. C, C. M. L, M. A. Anusuya, and V. H. Y, "Noise Reduction using Modified Central Frequency BWT and RLS filter," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 11s, pp. 512–521, Jan. 2024.

[19] J. Kapoor, A. Pathak, M. Rai, and G. R. Mishra, "Speech Quality Enhancement through Noise Cancellation using an Adaptive Algorithm," *IAENG International Journal of Computer Science*, vol. 49, no. 3, pp. 653–665, Sept. 2022.

[20] M. Sugadev, M. Kaushik, V. Vijayakumar, and K. T. Ilayaraaja, "Performance Analysis of Adaptive filter Algorithms on Different Noise Sources," in *2022 International Conference on Computer Communication and Informatics*, Coimbatore, India, 2022, pp. 1–5, https://doi.org/10.1109/ICCCI54379.2022.9740807.

[21] S. C. Venkateswarlu, N. U. Kumar, and A. Karthik, "Speech Enhancement Using Recursive Least Square Based on Real-time adaptive filtering algorithm," in *2021 6th International Conference for Convergence in Technology*, Maharashtra, India, 2021, pp. 1–4, https://doi.org/10.1109/I2CT51068.2021.9417929.

[22] A. N. Shaker, "Comparison between Orthogonal and Bi-Orthogonal Wavelets," *Journal of Southwest Jiaotong University*, vol. 55, no. 2, Apr. 2020.

[23] K. Daqrouq, I. N. Abu-Isbeih, O. Daoud, and E. Khalaf, "An investigation of speech enhancement using wavelet filtering method," *International Journal of Speech Technology*, vol. 13, no. 2, pp. 101–115, June 2010, https://doi.org/10.1007/s10772-010-9073-1.

[24] I. N. Abu-Isbeih and K. Daqrouq, "The Use of Biorthogonal Wavelets in Speech Enhancement," in *7th International Conference on Electrical Engineering ICEENG 2010*, Cairo, Egypt, 2010, pp. 1–13, https://doi.org/10.21608/iceeng.2010.33042.

[25] D. P. Singh, D. Singh, and D. Sethi, "Reduction of Noise from Speech Signal using Haar and Biorthogonal Wavelet," *International Journal of Electronics & Communication Technology*, vol. 2, no. 3, pp. 263–269, Sept. 2011.

[26] A. A. Alasadi, T. H. Aldhayni, R. R. Deshmukh, A. H. Alahmadi, and A. S. Alshebami, "Efficient Feature Extraction Algorithms to Develop an Arabic Speech Recognition System," *Engineering, Technology & Applied Science Research*, vol. 10, no. 2, pp. 5547–5553, Apr. 2020, https://doi.org/10.48084/etasr.3465.

[27] A. Riadh, "New Speech Enhancement Method based on Wavelet Transform and Tracking of Non Stationary Noise Algorithm," in *Recent Advances on Electroscience and Computers*, Barcelona, Spain, 2015, pp. 45–52.

[28] R. Aggarwal, J. K. Singh, V. K. Gupta, S. Rathore, M. Tiwari, and A. Khare, "Noise Reduction of Speech Signal using Wavelet Transform with Modified Universal Threshold," *International Journal of Computer Applications*, vol. 20, no. 5, pp. 14–19, Apr. 2011, https://doi.org/10.5120/2431-3269.

[29] H. Kaur Ojhla and S. Patil, "A Wavelet Based Hybrid Threshold Transform Method for Speech Intelligibility and Quality in Noisy Speech Patterns of English Language," *Wireless Personal Communications*, vol. 112, no. 2, pp. 1113–1135, May 2020, https://doi.org/10.1007/s11277-020-07093-9.

[30] S. Singh, M. Tripathy, and R. S. Anand, "A wavelet based method for removal of highly non-stationary noises from single-channel hindi speech patterns of low input SNR," *International Journal of Speech Technology*, vol. 18, no. 2, pp. 157–166, June 2015, https://doi.org/10.1007/s10772-014-9255-3.

[31] T. Guo, T. Zhang, E. Lim, M. López-Benítez, F. Ma, and L. Yu, "A Review of Wavelet Analysis and Its Applications: Challenges and Opportunities," *IEEE Access*, vol. 10, pp. 58869–58903, 2022, https://doi.org/10.1109/ACCESS.2022.3179517.

[32] J. J. Chopade and N. P. Futane, "Implementation of optimised wavelet packet algorithms for audibility enhancement," *Australian Journal of Electrical and Electronics Engineering*, vol. 17, no. 1, pp. 7–12, Jan. 2020, https://doi.org/10.1080/1448837X.2019.1681849.

[33] L. Tan, Y. Chen, and F. Wu, "Research on Speech Signal Denoising Algorithm Based on Wavelet Analysis," *Journal of Physics: Conference Series*, vol. 1627, no. 1, Aug. 2020, Art. no. 012027, https://doi.org/10.1088/1742-6596/1627/1/012027.

[34] P. Saxena and V. Saxena, "Comparative Study of White Gaussian Noise Reduction for Different Signals using Wavelet," *International Journal of Research - GRANTHAALAYAH*, vol. 10, no. 7, pp. 112–123, Aug. 2022, https://doi.org/10.29121/granthaalayah.v10.i7.2022.4711.

[35] G. P. Singh and A. Pathak, "Biorthogonal Wavelet Packets in $H^S(K)$," *International Journal of Applied and Computational Mathematics*, vol. 8, no. 1, Dec. 2021, Art. no. 4, https://doi.org/10.1007/s40819-021-01154-8.

[36] R. R. Coifman and M. V. Wickerhauser, "Best adapted wavelet packets bases." Yale University, Feb. 1990.

[37] M. Gutiérrez-Muñoz and M. Coto-Jiménez, "An Experimental Study on Speech Enhancement Based on a Combination of Wavelets and Deep Learning," *Computation*, vol. 10, no. 6, June 2022, Art. no. 102, https://doi.org/10.3390/computation10060102.