

Review Article

A New Speech Enhancement Technique Based on Stationary Bionic Wavelet Transform and MMSE Estimate of Spectral Amplitude

Mourad Talbi ¹ and Med Salim Bouhlel²

¹Laboratory of Nanomaterials and Systems for Renewable Energies (LaNSER),
Center of Researches and Technologies of Energy of Borj Cedria, Tunis 952050, Tunisia

²Sciences and Technologies of Images and Telecommunications, Higher Institute of Electronics and Telecommunication of Sfax,
Sfax University, Sfax, Tunisia

Correspondence should be addressed to Mourad Talbi; talbi1969@yahoo.fr

Received 2 April 2021; Revised 26 April 2021; Accepted 10 October 2021; Published 24 December 2021

Academic Editor: Manjit Kaur

Copyright © 2021 Mourad Talbi and Med Salim Bouhlel. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Speech enhancement has gained considerable attention in the employment of speech transmission via the communication channel, speaker identification, speech-based biometric systems, video conference, hearing aids, mobile phones, voice conversion, microphones, and so on. The background noise processing is needed for designing a successful speech enhancement system. In this work, a new speech enhancement technique based on Stationary Bionic Wavelet Transform (SBWT) and Minimum Mean Square Error (MMSE) Estimate of Spectral Amplitude is proposed. This technique consists at the first step in applying the SBWT to the noisy speech signal, in order to obtain eight noisy wavelet coefficients. The denoising of each of those coefficients is performed through the application of the denoising method based on MMSE Estimate of Spectral Amplitude. The SBWT inverse, $SBWT^{-1}$, is applied to the obtained denoised stationary wavelet coefficients for finally obtaining the enhanced speech signal. The proposed technique's performance is proved by the calculation of the Signal to Noise Ratio (SNR), the Segmental SNR (SSNR), and the Perceptual Evaluation of Speech Quality (PESQ).

1. Introduction

In many speech-related applications, an input speech signal is frequently corrupted by environmental noise and needs further processing using a speech enhancement technique for ameliorating the associated quality before being employed [1]. Generally, speech enhancement techniques can be grouped into two groups which are supervised and unsupervised. Unsupervised techniques include spectral subtraction (SS) [2–4], Wiener filtering [5, 6], short-time spectral amplitude (STSA) estimation [7], and short-time log-spectral amplitude estimation (logSTSA) [8]. Concerning the supervised speech enhancement techniques, they employ a training set for learning diverse models for noisy and clean speech signals, and examples include

codebook-based methods [9] and Hidden Markov Model (HMM)-based techniques [10]. Classical speech enhancement techniques are frequently processing a noisy utterance in a frame-wise way, that is, for enhancing each short-time period of the utterance nearly in independent manner. Some research works showed that considering the inter-frame variation over a relatively long span of time can contribute to superior performance in enhancing speech [1]. Famous approaches along this direction include modulation-domain spectral subtraction [11], Kalman filtering, and modulation-domain Wiener filtering [12, 13]. Moreover, when we compare the discrete wavelet transform (DWT) to the Fourier transform (FT) where only the frequency parts are taken into consideration, though, in the expression of the DWT [14], both temporal and frequency

characteristics of the signal to be analyzed are taken into consideration. The DWT has become a well-known method in speech analysis. In Wavelet Thresholding Denoising (WTD) [15], the wavelet transform is applied for splitting the time-domain signal into sub-bands. After that, thresholding of the obtained wavelet coefficients (sub-bands) is performed. In [16], the DWT [17, 18] was applied to the speech signal to simply conserve the obtained approximation portion, which simultaneously attains data compression and noise robustness in recognition. In [1], the DWT was employed for analyzing the spectrogram of a noisy utterance along the temporal axis, and then the resulting detail portion was devalued with an expect of reducing noise effect in order to promote speech quality. Despite the ease of its implementation, the preliminary evaluation results indicate that the technique proposed in [1] permits to have input signals with better perceptual quality. It was proved that this technique [1] can be paired with many well-known speech enhancement approaches for achieving even better performance [1]. In this work, a novel speech enhancement technique based on the Stationary Bionic Wavelet Transform (SBWT) [19–21] and Minimum Mean Square Error (MMSE) Estimate of Spectral Amplitude [22] is proposed. In this paper, this approach is evaluated and compared to four other speech enhancement approaches which are as follows:

- (i) Unsupervised speech denoising via perceptually motivated robust principal component analysis [23].
- (ii) The speech enhancement technique based on MSS-SMPO [24, 25].
- (iii) The denoising technique based on MMSE Estimate of Spectral Amplitude [22].
- (iv) Our previous speech enhancement technique based on LWT and Artificial Neural Network (ANN) and using MMSE Estimate of Spectral Amplitude [26].

The fourth technique which is based on LWT and ANN [27–29] and uses MMSE Estimate of Spectral Amplitude [26] can be summarized by the following steps:

- (i) **First step:** applying the LWT to the noisy speech signal for obtaining two noisy details coefficients, $cD1$ and $cD2$, and one approximation coefficient, $cA2$.
- (ii) **Second step:** denoising $cD1$ and $cD2$ by soft thresholding, and for their thresholding, suitable thresholds, $thr_j, 1 \leq j \leq 2$, have to be used. Those thresholds are determined by using an Artificial Neural Network (ANN). This soft thresholding is performed for having two denoised coefficients, $cDd1$ and $cDd2$.
- (iii) **Third step:** applying the denoising approach based on MMSE Estimate of Spectral Amplitude [22] to $cA2$ for obtaining a denoised coefficient, $cAd2$.
- (iv) **Fourth step:** applying the inverse of LWT, LWT^{-1} , to $cDd1$, $cDd2$, and $cAd2$, for finally obtaining the enhanced signal.

As a future work, we will develop a novel speech enhancement approach using ANN [30–36] or deep learning [37, 38] for thresholding the noisy stationary bionic wavelet coefficients. Those coefficients are obtained by applying the BWT to the noisy speech signal.

In Section 2 of this paper, materials and methods are presented. Section 2.4 describes the speech enhancement technique proposed in this work. In Section 3, results and discussion are presented. Finally, Section 4 concludes the paper.

2. Materials and Methods

2.1. The Stationary Bionic Wavelet Transform (SBWT). In [19], the SBWT has been proposed as a novel wavelet transform. This transform was initially introduced for solving the problem of perfect reconstruction that exists with the Bionic Wavelet Transform (BWT). Its application was performed for speech enhancement [19, 20] and also for ECG denoising [21].

2.2. The MMSE Estimate of Spectral Amplitude. In the literature, it was proposed to estimate the noise power spectral density employing MMSE (Minimum Mean Square Error) optimal estimation [22]. It was proved that the obtained estimator can be considered as a VAD (Voice Activity Detector)-based noise power estimator, and the noise power is updated alone if speech absence is detected, compensated with a required bias compensation [22]. It was proved that the bias compensation is not needed if the VAD is substituted by a soft SPP (Speech Presence Probability) with fixed priors [22]. When choosing fixed priors, this has the benefit of decoupling the noise power estimator from subsequent steps in a speech enhancement algorithm, such as the estimation of the speech power and that of the clean speech [22]. Gerkmann and Richard [22] proved that the proposed SPP approach permits to maintain the quick noise tracking performance of the bias-compensated MMSE-based technique while exhibiting less overestimation of the spectral noise power and an even lower complexity of calculation.

2.3. Signal Model. In [22], Gerkmann and Richard considered frame-by-frame processing of time-domain signals where the Discrete Fourier Transform (DFT) is applied to these frames. Let the complex spectral noise and speech coefficients be given, respectively, by $N_k(l)$ and $S_k(l)$, where l is the time frame index and k is the frequency bin index [22]. In [22], it was assumed that in the short-time Fourier domain, both noise and speech signals tend to be additive. Therefore, the complex spectral noisy observation has the following expression:

$$Y_k(l) = S_k(l) + N_k(l). \quad (1)$$

In [22], it was supposed that the noise and speech signals own zero mean and are independent so that

$$E(|Y|^2) = E(|S|^2) + E(|N|^2), \quad (2)$$

where $E(\bullet)$ denotes the statistical expectation operator.

The spectral noise and speech power are expressed as follows:

$$\begin{aligned} E(|N|^2) &= \sigma_N^2, \\ E(|S|^2) &= \sigma_S^2. \end{aligned} \quad (3)$$

Then, both a posteriori SNR and a priori SNR are expressed as follows:

$$\begin{aligned} \gamma &= \frac{(|Y|^2)}{\sigma_N^2} \text{ (a posteriori SNR),} \\ \xi &= \frac{\sigma_S^2}{\sigma_N^2} \text{ (a priori SNR).} \end{aligned} \quad (4)$$

All details about MMSE-based noise power estimation are given in [22].

2.4. The Proposed Speech Enhancement Technique. The speech enhancement technique introduced in this work is based on the SBWT [19–21] and the MMSE Estimate of Spectral Amplitude [22]. The novelty of this approach consists in applying the speech enhancement method based on MMSE Estimate of Spectral Amplitude [1, 22] in the SBWT domain. In fact, this technique [22] is applied to each noisy stationary bionic wavelet coefficient for its denoising. Those noisy coefficients are obtained by applying the SBWT to the noisy speech signal. Then, the inverse of SBWT ($SBWT^{-1}$) is applied to the obtained denoised coefficients in order to obtain finally the enhanced speech signal. Figure 1 illustrates the flowchart of this proposed technique.

According to Figure 1, the first step of the proposed approach is to apply the SBWT to the noisy speech signal for obtaining eight noisy stationary bionic wavelet coefficients. Those coefficients are named wb_i , $1 \leq i \leq 8$, and each of them is denoised by the speech enhancement technique based on MMSE Estimate of Spectral Amplitude [1, 22]. and we obtain eight denoised coefficients, wd_i , $1 \leq i \leq 8$ (Figure 1). In those coefficients, wd_i , $1 \leq i \leq 8$ inverse is applied for SBWT ($SBWT^{-1}$) in order to obtain the enhanced signal finally.

2.5. Minimum Mean Square Error (MMSE) Estimate of Spectral Amplitude in the SBWT Domain. In general, classical speech enhancement approaches based on thresholding in the wavelet transform domain can introduce some distortions to the original speech signal. This particularly occurs for the unvoiced sounds. Consequently, a great number of speech enhancement techniques based on wavelet transforms are employing other tools such as spectral subtraction (SS), Wiener filtering, and MMSE-STSA estimation [39, 40]. This is the reason why we apply the Minimum Mean Square Error (MMSE) Estimate of Spectral Amplitude in the SBWT domain in our speech enhancement system. The application of the SBWT permits to solve the problem of the perfect reconstruction existing when we apply the BWT [19].

Furthermore, the SBWT among all wavelet transforms [41, 42] tends to uncorrelated data [43] and facilitates the noise suppression. The fact that the Minimum Mean Square Error (MMSE) Estimate of Spectral Amplitude [22] is applied to each noisy stationary bionic coefficient permits to have a better adaptation for speech and noise estimations compared to the application of this technique [22] to the whole noisy speech signal.

2.6. Unsupervised Speech Denoising via Perceptually Motivated Robust Principal Component Analysis [23]. To overcome the shortcomings in the existing sparse and low-rank speech denoising technique that the auditory perceptual properties are not fully exploited and the speech degradation is simply perceived, a perceptually motivated robust principal component analysis (ISNRPCA) technique was presented. In order to reflect the non-linear property for frequency perception of the basilar membrane, cochleagram is employed as inputs of ISNRPCA. The latter employs the perceptually meaningful Itakura–Saito measure as its optimization objective function. Furthermore, non-negative constraints are also compulsory for regularizing the decomposed terms with respect to their physical meaning [23]. In [23], Min et al. proposed an alternating direction technique of multipliers (ADMM) for solving the optimization problem of ISNRPCA. The latter is completely unsupervised, and neither the noise nor the speech model requires to be trained beforehand. Experimental results under diverse kinds of noise and different SNRs prove that the ISNRPCA is showing promising results for speech denoising [23].

2.7. The Speech Enhancement Technique Based on MSS-SMPO [25]. In [25], a two-step enhancement technique based on spectral subtraction and phase spectrum compensation was presented for noisy speeches in diverse environments requiring non-stationary noise and medium to low levels of SNR. In the first step of the technique proposed in [25], the magnitude of the noisy speech spectrum is modified by a spectral subtraction technique, where a noise estimation approach was introduced. The latter is based on the low-frequency information of the noisy speech. This noise estimation technique is able to estimate precisely the non-stationary noise. In the second step, the phase spectrum of the noisy speech is modified consisting of phase spectrum compensation, where an SNR-dependent technique is incorporated for determining the amount of compensation to be compulsory on the phase spectrum [25]. A modified complex spectrum is obtained by aggregating the magnitude from the step of spectral subtraction and the modified phase spectrum from the step of phase compensation, which is found to be a better representation of enhanced speech spectrum.

3. Results and Discussion

In this work, the evaluation of the proposed technique is performed by its application to ten Arabic speech sentences pronounced by a male speaker and ten others by a female speaker (Table 1). Those speech signals are degraded in

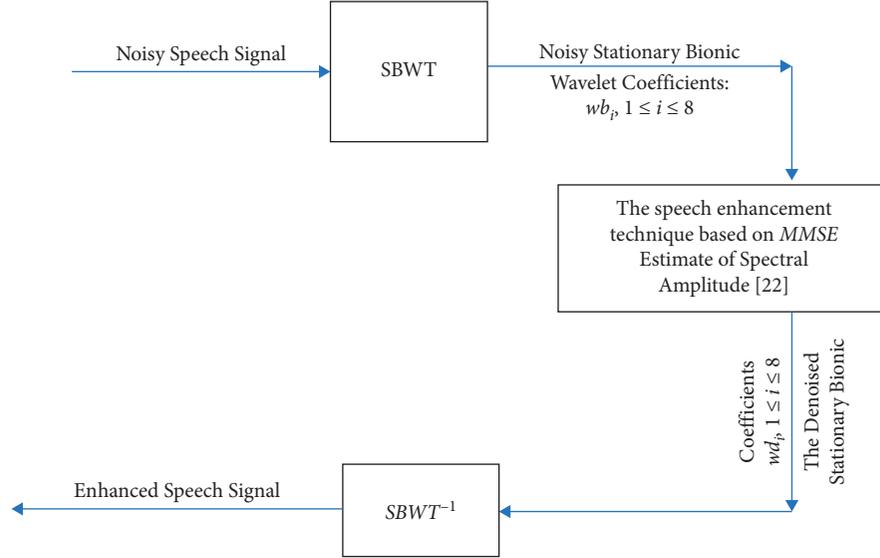


FIGURE 1: The flowchart of the proposed speech enhancement technique.

TABLE 1: The list of the used Arabic speech sentences.

Female speaker	Male speaker
أحفظ من الأرض: signal 1	يذيع الخبر نل ال: signal 1
أين المسافرين: signal 2	الكميل بالاسلام رسالتك: signal 2
مستمتع بتمهارة ال: signal 3	قطت إبرتس: signal 3
يؤديهم زمانناس: signal 4	من لم يبتفع: signal 4
بنت قذوة لهم: signal 5	فل عن ضحكاته اغ: signal 5
ئم اص رازا: signal 6	و لماذا نشف مالهم: signal 6
بكال و غبط الكبش: signal 7	أين زوايانا و قانوننا: signal 7
هل لذعته بقول: signal 8	د الموروث مدلع اص: signal 8
عرف واليا و قايذا: signal 9	به آبائكم من: signal 9
خال بالنا منكم: signal 10	أظهره و قم: signal 10

artificial manner by an additive noise at different values of SNR_i (before denoising). In order to corrupt those speech signals (Table 1), we have chosen four kinds of noise which are white Gaussian, car, F16, and tank noises. Those twenty speech signals are sampled at 16kHz and are listed in Table 1.

Also, for evaluating the proposed technique, it is compared with other three speech enhancement approaches which are as follows:

- (i) The denoising approach based on MMSE Estimate of Spectral Amplitude [22].
- (ii) The unsupervised speech denoising technique via perceptually motivated robust principal component analysis [23].
- (iii) The speech enhancement approach based on MSS-SMPO [24].

This evaluation is performed through the computations of the SNR (Signal to Noise Ratio), the Segmental SNR (SSNR), and the PESQ (Perceptual Evaluation of Speech Quality). The results obtained from these computations are presented in Tables 2–16.

According to these tables, the best results are the values in italics and they are practically obtained from the application of the proposed technique. Therefore, this technique outperforms the other speech enhancement approaches [22–25] applied for this evaluation.

Figure 2 illustrates an example of speech enhancement applying the proposed technique to the clean speech signal (Figure 2(a)) corrupted in additive manner by a car noise (Volvo) with $SNR = 0dB$ (Figure 2(b)). According to this figure, this technique permits to considerably reduce noise and to obtain an enhanced speech signal (Figure 2(c)) with little distortions despite the fact that the value of the SNR is low (0 dB). Figure 3 illustrates the spectrograms of the clean, noisy, and enhanced speech signals.

The spectrogram in Figure 3(b) shows that the type of noise corrupting the speech signal is localized in low-frequency parts. The spectrogram in Figure 3(c) shows that the car noise is considerably reduced by using the proposed speech enhancement technique. Moreover, this technique permits to have an enhanced speech signal with low distortions compared to the clean speech signal (Figure 2(a)).

In the following, we will compare the proposed technique with our previous speech enhancement approach which is based on LWT and ANN and uses MMSE [26]. The first difference between the speech enhancement technique proposed in this work and our previous approach is that they use two completely different wavelet transforms which are the SBWT for the technique proposed in this paper and the LWT for our previous approach proposed in [26]. The second difference between these two techniques is that the denoising approach based on MMSE Estimate of Spectral Amplitude is applied [22] to all stationary bionic wavelet coefficients for the technique proposed in this paper. However, we apply this approach [22] only to the approximation coefficient for our previous speech enhancement technique proposed in [26]. The latter also uses an

TABLE 2: Results obtained from the computation of SNR (signal 7 (female voice) corrupted by Gaussian white noise).

SNR_i (dB)	SNR_f (dB)			
	The speech enhancement technique			
	Unsupervised speech denoising via perceptually motivated robust principal component analysis [23]	The speech enhancement technique based on <i>MSS-SMPO</i> [24]	The proposed speech enhancement technique	The denoising technique based on <i>MMSE</i> Estimate of Spectral Amplitude [22]
-5	2.7870	8.1682	8.4026	6.3331
0	6.9200	11.5437	12.4447	10.4737
5	11.0291	14.7845	15.9887	14.2200
10	14.1329	18.2255	19.3911	17.6035
15	16.7798	21.3456	22.4836	20.8019

The bold values show the values obtained from the application of the proposed speech enhancement technique, and they are the best values.

TABLE 3: Results obtained from the computation of SSNR (signal 7 (female voice) corrupted by Gaussian white noise).

SNR_i (dB)	$SSNR$ (dB)			
	The speech enhancement technique			
	Unsupervised speech denoising via perceptually motivated robust principal component analysis [23]	The speech enhancement technique based on <i>MSS-SMPO</i> [24, 25]	The proposed speech enhancement technique	The denoising technique based on <i>MMSE</i> Estimate of Spectral Amplitude [22]
-5	-2.4187	0.8504	1.3313	1.2531
0	-2.4610e - 04	3.1099	3.9607	2.7508
5	2.7044	5.4567	6.2536	5.1757
10	5.0003	8.6660	8.9350	7.4637
15	7.7230	11.7670	12.1193	10.4470

The bold values show the values obtained from the application of the proposed speech enhancement technique, and they are the best values.

TABLE 4: Results obtained from the computation of PESQ (signal 7 (female voice) corrupted by Gaussian white noise).

SNR_i (dB)	$PESQ$			
	The speech enhancement technique			
	Unsupervised speech denoising via perceptually motivated robust principal component analysis [23]	The speech enhancement technique based on <i>MSS-SMPO</i> [24, 25]	The proposed speech enhancement technique	The denoising technique based on <i>MMSE</i> Estimate of Spectral Amplitude [22]
-5	1.0636	1.4235	1.4884	1.2531
0	1.4606	1.8776	1.9593	1.7421
5	1.9352	2.2374	2.3747	2.1773
10	2.3212	2.5908	2.7116	2.5503
15	2.7461	2.9835	3.0695	2.8944

The bold values show the values obtained from the application of the proposed speech enhancement technique, and they are the best values.

TABLE 5: Results obtained from the computation of SNR (signal 5 (male voice) corrupted by F16 noise).

SNR_i (dB)	SNR_f (dB)			
	The speech enhancement technique			
	Unsupervised speech denoising via perceptually motivated robust principal component analysis [23]	The speech enhancement technique based on <i>MSS-SMPO</i> [24, 25]	The proposed speech enhancement technique	The denoising technique based on <i>MMSE</i> Estimate of Spectral Amplitude [22]
-5	1.2722	2.5071	3.7283	2.3837
0	5.1589	6.2402	8.2589	7.1434
5	8.9032	9.5351	11.9891	11.1435
10	12.9517	14.0785	15.5901	14.6596
15	15.9880	17.9726	19.6030	18.3975

The bold values show the values obtained from the application of the proposed speech enhancement technique, and they are the best values.

Artificial Neural Network (ANN), and this fact differentiates this technique [26] from our technique proposed in this paper. The comparison of these two techniques is also in

terms of SNR, SSNR, and PESQ. These two techniques are applied to a speech signal degraded by a car noise with diverse values of SNR before denoising (SNR_i). Tables 17–19

TABLE 6: Results obtained from the computation of SSNR (signal 5 (male voice) corrupted by F16 noise).

SNR_i (dB)	SSNR (dB)			
	The speech enhancement technique			
	Unsupervised speech denoising via perceptually motivated robust principal component analysis [23]	The speech enhancement technique based on MSS-SMPO [24, 25]	The proposed speech enhancement technique	The denoising technique based on MMSE Estimate of Spectral Amplitude [22]
-5	-4.0510	-3.7447	-3.0533	-3.6560
0	-2.2335	-1.9010	-0.7847	-1.3701
5	-0.4451	-0.1214	1.0887	0.6802
10	1.7251	2.7250	3.1337	2.6493
15	3.7474	6.0977	5.9488	5.0968

The bold values show the values obtained from the application of the proposed speech enhancement technique, and they are the best values.

TABLE 7: Results obtained from the computation of PESQ (signal 5 (male voice) corrupted by F16 noise).

SNR_i (dB)	PESQ			
	The speech enhancement technique			
	Unsupervised speech denoising via perceptually motivated robust principal component analysis [23]	The speech enhancement technique based on MSS-SMPO [24, 25]	The proposed speech enhancement technique	The denoising technique based on MMSE Estimate of Spectral Amplitude [22]
-5	1.2831	1.2227	1.2951	1.2593
0	1.6378	1.5769	1.8125	1.7017
5	2.0750	2.1242	2.2982	2.1889
10	2.4313	2.5673	2.7291	2.6444
15	2.8416	3.0733	3.1164	3.0182

The bold values show the values obtained from the application of the proposed speech enhancement technique, and they are the best values.

TABLE 8: Results obtained from the computation of SNR (signal 3 (male voice) corrupted by tank noise).

SNR_i (dB)	SNRf (dB)			
	The speech enhancement technique			
	Unsupervised speech denoising via perceptually motivated robust principal component analysis [23]	The speech enhancement technique based on MSS-SMPO [24, 25]	The proposed speech enhancement technique	The denoising technique based on MMSE Estimate of Spectral Amplitude [22]
-5	1.8657	2.8506	4.5261	3.0084
0	5.2569	6.0328	8.2533	6.6513
5	8.8912	9.7367	12.5241	10.7510
10	12.5670	13.6433	16.6318	14.7634
15	15.9158	18.2296	21.0643	19.0401

The bold values show the values obtained from the application of the proposed speech enhancement technique, and they are the best values.

TABLE 9: Results obtained from the computation of SSNR (signal 3 (male voice) corrupted by tank noise).

SNR_i (dB)	SSNR (dB)			
	The speech enhancement technique			
	Unsupervised speech denoising via perceptually motivated robust principal component analysis [23]	The speech enhancement technique based on MSS-SMPO [24, 25]	The proposed speech enhancement technique	The denoising technique based on MMSE Estimate of Spectral Amplitude [22]
-5	-3.0742	-4.1830	-3.6883	-4.2203
0	-1.2104	-2.2805	-1.7476	-2.6546
5	1.0990	-0.0078	0.9546	-0.1952
10	3.6778	2.6210	3.8260	2.5252
15	6.2864	6.0870	7.2508	5.7392

The bold values show the values obtained from the application of the proposed speech enhancement technique, and they are the best values.

TABLE 10: Results obtained from the computation of PESQ (signal 3 (male voice) corrupted by tank noise).

<i>PESQ</i>				
The speech enhancement technique				
<i>SNR_i</i> (dB)	Unsupervised speech denoising via perceptually motivated robust principal component analysis [23]	The speech enhancement technique based on <i>MSS-SMPO</i> [24, 25]	The proposed speech enhancement technique	The denoising technique based on <i>MMSE</i> Estimate of Spectral Amplitude [22]
-5	0.9941	1.3792	1.5538	1.3848
0	1.3492	1.8875	2.0085	1.8400
5	1.7720	2.3089	2.3781	2.2503
10	2.2027	2.5906	2.6361	2.5094
15	2.6084	2.7785	2.8201	2.7240

The bold values show the values obtained from the application of the proposed speech enhancement technique, and they are the best values.

TABLE 11: Results obtained from the computation of SNR (signal 8 (female voice) corrupted by factory noise).

<i>SNR_f</i> (dB)				
The speech enhancement technique				
<i>SNR_i</i> (dB)	Unsupervised speech denoising via perceptually motivated robust principal component analysis [23]	The speech enhancement technique based on <i>MSS-SMPO</i> [24, 25]	The proposed speech enhancement technique	The denoising technique based on <i>MMSE</i> Estimate of Spectral Amplitude [22]
-5	3.0176	3.7705	4.9624	3.4404
0	6.6901	7.8983	9.0631	7.2378
5	10.4536	11.7369	12.5889	11.2022
10	13.5863	14.9194	15.7971	14.4791
15	16.0742	19.2111	20.1601	18.6555

The bold values show the values obtained from the application of the proposed speech enhancement technique, and they are the best values.

TABLE 12: Results obtained from the computation of SSNR (signal 8 (female voice) corrupted by factory noise).

<i>SSNR</i> (dB)				
The speech enhancement technique				
<i>SNR_i</i> (dB)	Unsupervised speech denoising via perceptually motivated robust principal component analysis [23]	The speech enhancement technique based on <i>MSS-SMPO</i> [24, 25]	The proposed speech enhancement technique	The denoising technique based on <i>MMSE</i> Estimate of Spectral Amplitude [22]
-5	-2.6512	-2.3588	-1.4115	-2.1701
0	-0.5142	0.0580	0.8899	-0.0307
5	1.6661	2.4913	2.9617	2.1659
10	3.5242	5.0128	5.3739	4.4255
15	5.5958	8.6024	8.9245	7.7369

The bold values show the values obtained from the application of the proposed speech enhancement technique, and they are the best values.

TABLE 13: Results obtained from the computation of PESQ (signal 8 (female voice) corrupted by factory noise).

<i>PESQ</i>				
The speech enhancement technique				
<i>SNR_i</i> (dB)	Unsupervised speech denoising via perceptually motivated robust principal component analysis [23]	The speech enhancement technique based on <i>MSS-SMPO</i> [24, 25]	The proposed speech enhancement technique	The denoising technique based on <i>MMSE</i> Estimate of Spectral Amplitude [22]
-5	0.6882	0.5311	0.8818	0.7038
0	0.9916	0.9724	1.1770	1.0327
5	1.4558	1.6078	1.7891	1.5358
10	1.9024	2.1297	2.3493	2.1484
15	2.4498	2.6051	2.7664	2.6077

The bold values show the values obtained from the application of the proposed speech enhancement technique, and they are the best values.

TABLE 14: Results obtained from the computation of SNR (signal 2 (male voice) corrupted by Volvo noise).

SNR_i (dB)	SNR_f (dB)			
	The speech enhancement technique			
	Unsupervised speech denoising via perceptually motivated robust principal component analysis [23]	The speech enhancement technique based on MSS-SMPO [24, 25]	The proposed speech enhancement technique	The denoising technique based on MMSE Estimate of Spectral Amplitude [22]
-5	4.5435	3.3737	5.4782	4.2192
0	8.3623	7.4176	9.9016	8.3451
5	12.4591	12.8324	14.4917	12.6024
10	16.1163	17.5726	18.9803	17.4120
15	18.1761	20.6149	22.9204	21.4578

The bold values show the values obtained from the application of the proposed speech enhancement technique, and they are the best values.

TABLE 15: Results obtained from the computation of SSNR (signal 2 (male voice) corrupted by Volvo noise).

SNR_i (dB)	$SSNR$ (dB)			
	The speech enhancement technique			
	Unsupervised speech denoising via perceptually motivated robust principal component analysis [23]	The speech enhancement technique based on MSS-SMPO [24, 25]	The proposed speech enhancement technique	The denoising technique based on MMSE Estimate of Spectral Amplitude [22]
-5	-1.1159	-1.2951	-0.1793	-1.1347
0	1.2097	1.3262	2.9410	1.7861
5	4.0138	5.1028	5.9979	4.7166
10	6.8442	8.3857	9.1966	7.8228
15	9.5287	10.9227	12.4662	10.9850

The bold values show the values obtained from the application of the proposed speech enhancement technique, and they are the best values.

TABLE 16: Results obtained from the computation of PESQ (signal 2 (male voice) corrupted by Volvo noise).

SNR_i (dB)	$PESQ$			
	The speech enhancement technique			
	Unsupervised speech denoising via perceptually motivated robust principal component analysis [23]	The speech enhancement technique based on MSS-SMPO [24, 25]	The proposed speech enhancement technique	The denoising technique based on MMSE Estimate of Spectral Amplitude [22]
-5	2.3027	2.2235	2.5406	2.4021
0	2.6867	2.5464	2.8408	2.7163
5	3.0848	2.8339	3.1719	3.0184
10	3.3970	2.9781	3.4203	3.2461
15	3.5439	3.0602	3.6435	3.4789

The bold values show the values obtained from the application of the proposed speech enhancement technique, and they are the best values.

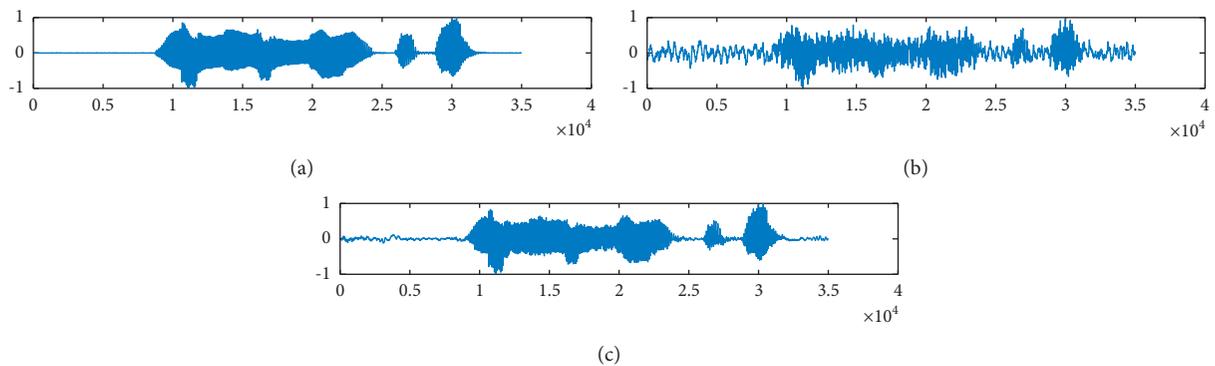


FIGURE 2: An example of speech enhancement applying the proposed speech enhancement approach. (a) Clean speech signal (male voice (signal 4)). (b) Noisy speech signal (clean signal degraded by additive car noise with $SNR_i = 0$ dB). (c) Enhanced speech signal with $SNR_f = 13.2999$, $SSNR = 4.3802$, and $PESQ = 3.0888$.

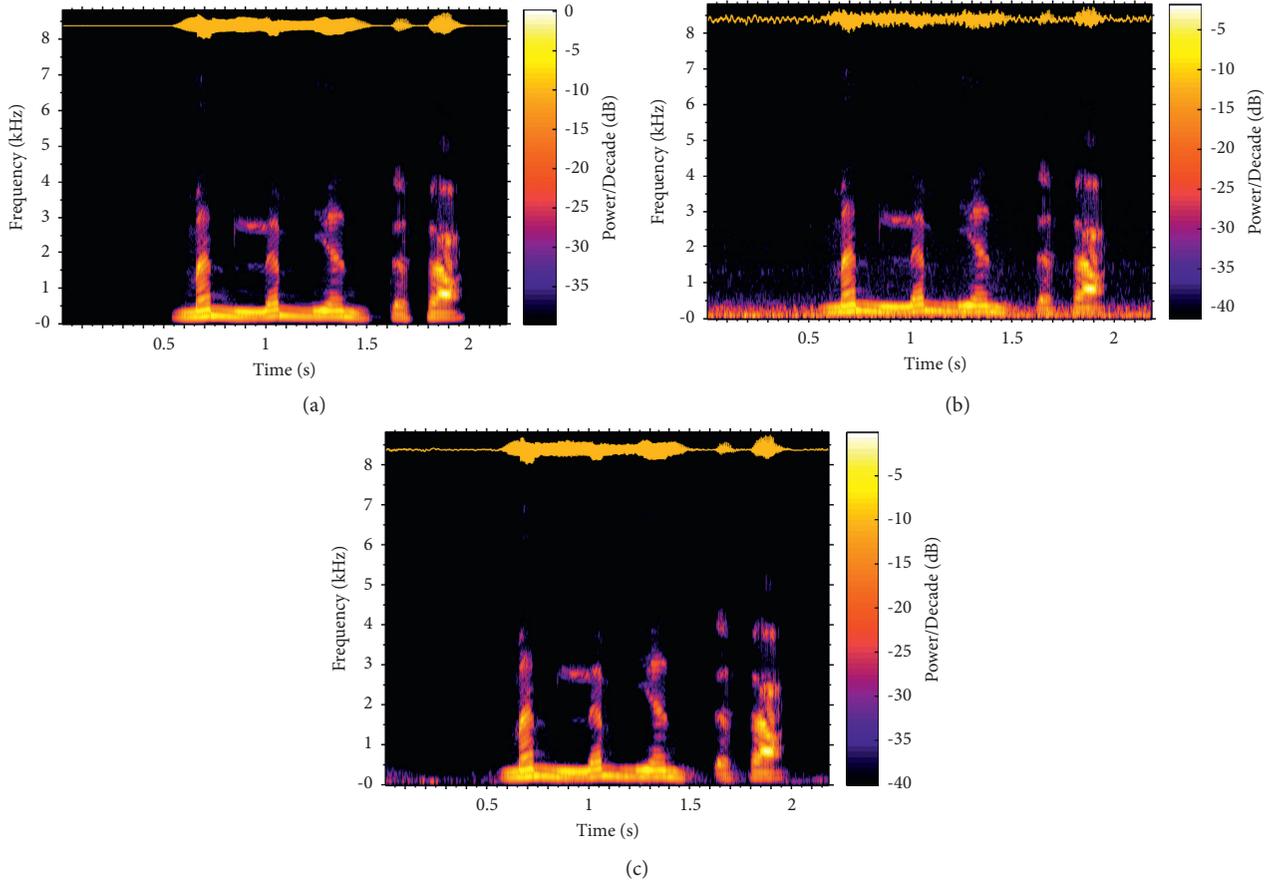


FIGURE 3: (a) The spectrogram of the clean speech signal (Figure 2(a)). (b) The spectrogram of the noisy speech signal (Figure 2(b)). (c) The spectrogram of the enhanced speech signal (Figure 2(c)).

TABLE 17: Results obtained from the computation of SNR (signal 2 (male voice) corrupted by Volvo noise).

SNR_i (dB)	SNR_f (dB)	
	The speech enhancement technique	
	Speech enhancement based on <i>LWT</i> and <i>ANN</i> and using <i>MMSE</i> Estimate of Spectral Amplitude [26]	The proposed speech enhancement technique
-5	5.8737	5.4782
0	9.8414	9.9016
5	14.1647	14.4917
10	18.5308	18.9803
15	22.5102	22.9204

The bold values show the values obtained from the application of the proposed speech enhancement technique, and they are the best values.

TABLE 18: Results obtained from the computation of SSNR (signal 2 (male voice) corrupted by Volvo noise).

SNR_i (dB)	$SSNR$ (dB)	
	The speech enhancement technique	
	Speech enhancement based on <i>LWT</i> and <i>ANN</i> and using <i>MMSE</i> Estimate of Spectral Amplitude [26]	The proposed speech enhancement technique
-5	0.2145	-0.1793
0	2.7478	2.9410
5	5.6644	5.9979
10	8.8942	9.1966
15	11.9663	12.4662

The bold values are the best values: for the SNR_i values, 0, 5, 10, and 15 dB, these best values are obtained from the application of the proposed speech enhancement technique.

TABLE 19: Results obtained from the computation of PESQ (signal 2 (male voice) corrupted by Volvo noise).

SNR _i (dB)	PESQ	
	Speech enhancement based on LWT and ANN and using MMSE Estimate of Spectral Amplitude [26]	The proposed speech enhancement technique
-5	2.2837	2.5406
0	2.5999	2.8408
5	2.8709	3.1719
10	3.1190	3.4203
15	3.3590	3.6435

present the results obtained from the computation of SNR, SSNR, and PESQ for the two techniques.

According to these tables, the best results are the values in italics and they are obtained from the application of the proposed technique. Therefore, this technique outperforms the other speech enhancement approach proposed in [26].

4. Conclusion

In this paper, we propose a new speech enhancement technique based on SBWT and MMSE Estimate of Spectral Amplitude. In the first step of this technique, the SBWT is applied to the noisy speech signal for obtaining eight noisy stationary bionic wavelet coefficients. The denoising of each of those coefficients is performed through the application of the denoising approach based on MMSE Estimate of Spectral Amplitude. Finally, the inverse of SBWT ($SBWT^{-1}$) is applied to the obtained stationary wavelet coefficients, for obtaining the enhanced speech signal. An evaluation of this technique is performed by its comparison with four other speech enhancement approaches where the first one is the denoising technique based on MMSE Estimate of Spectral Amplitude. The second one is the speech enhancement technique based on MSS-SMPO. The third one is the unsupervised speech denoising approach through perceptually motivated robust principal component analysis. The fourth one is the speech enhancement technique based on LWT and ANN and using MMSE Estimate of spectral amplitude. This evaluation is performed through the computations of Signal to Noise Ratio (SNR), the Segmental SNR (SSNR), and the Perceptual Evaluation of Speech Quality (PESQ). The results obtained from these computations show that the proposed technique outperforms the other previously mentioned techniques. Furthermore, the technique proposed in this work permits to considerably reduce the noises corrupting the clean speech signal and to have an enhanced speech signal with good perceptual quality.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] S.-K. Lee, S.-S. Wang, T. Yu, and J.-W. Hung, "Speech enhancement based on reducing the detail portion of speech spectrograms in modulation domain via Discrete wavelet transform," in *Proceedings of the 2018 11th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, Taipei City, Taiwan, November 2018.
- [2] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech, & Signal Processing*, vol. 27, no. 2, Article ID 113120, 1979.
- [3] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 208–211, Washington, D. C., USA, April 1979.
- [4] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Orlando, Florida, USA, May 2002.
- [5] C. Plapous, C. Marro, and P. Scalart, "Improved signal-to-noise ratio estimation for speech enhancement," *IEEE Transactions on Audio Speech and Language Processing*, vol. 14, no. 6, Article ID 20982108, 2006.
- [6] P. Scalart and J. V. Filho, "Speech enhancement based on a priori signal to noise estimation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 629–632, Atlanta, GA, USA, June 1996.
- [7] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, & Signal Processing*, vol. 32, no. 6, Article ID 11091121, 1984.
- [8] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, & Signal Processing*, vol. 33, no. 2, pp. 443–445, 1985.
- [9] S. Srinivasan, J. Samuelsson, and W. Kleijn, "Codebook driven short-term predictor parameter estimation for speech enhancement," *IEEE Transactions on Audio Speech and Language Processing*, vol. 14, no. 1, Article ID 163176, 2006.
- [10] D. Y. Zhao and W. B. Kleijn, "HMM-based gain modeling for enhancement of speech in noise," *IEEE Transactions on Audio Speech and Language Processing*, vol. 15, no. 3, Article ID 882892, 2007.
- [11] K. K. Paliwal, K. K. Wojcicki, and B. Schwerin, "Single-channel speech enhancement using spectral subtraction in the short-time modulation domain," *Speech Communication*, vol. 52, no. 5, pp. 450–475, 2010.
- [12] C.-C. Hsu, K.-M. Cheong, J.-T. Chien, and T.-S. Chi, "Modulation Wiener filter for improving speech intelligibility," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 370–374, Queensland, Australia, April 2015.
- [13] S. So and K. K. Paliwal, "Modulation-domain Kalman filtering for single-channel speech enhancement," *Speech Communication*, vol. 53, no. 6, pp. 818–829, 2011.

- [14] O. Rioul and M. Vetterli, *Wavelets and Signal Processing*, Springer, Berlin Heidelberg, Germany, 1991.
- [15] S. G. Chang, B. Bin Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, 2000.
- [16] 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. on Audio, Speech, and Language Processing*, vol. 26, no. 3, pp. 564–579, 2018.
- [17] D. Huang, K. Lanyan, 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," *Hindawi, Security and Communication Networks*, vol. 2019, Article ID 4049765, 16 pages, 2019.
- [18] M. Ali Nematollahi, C. Vorakulpipat, and H. G. Rosales, "Optimization of a blind speech watermarking technique against amplitude scaling," *Hindawi, Security and Communication Networks*, vol. 2017, Article ID 5454768, 13 pages, 2017.
- [19] T. Mourad, "Speech enhancement based on stationary bionic wavelet transform and maximum a posterior estimator of magnitude-squared spectrum," *International Journal of Speech Technology*, vol. 20, no. 1, pp. 75–88, 2017.
- [20] M. Talbi and M. S. Bouhleh, "A novel approach of speech enhancement based on SBWT and MMSE estimate of spectral amplitude," in *Proceedings of the 2020 4th International Conference on Advanced Systems and Emergent Technologies (IC_ASET)*, Hammamet, Tunisia, March 2020.
- [21] M. Talbi, "New approach of ECG denoising based on 1-D double-density complex DWT and SBWT," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 8, no. 6, pp. 608–620, 2020.
- [22] T. Gerkmann and C. H. Richard, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Transactions on Audio Speech and Language Processing*, vol. 20, no. 4, pp. 1383–1393, 2012.
- [23] G. Min, X. Zou, W. Han, X. Zhang, and W. Tan, "Unsupervised speech denoising via perceptually motivated robust principal component analysis," *Shengxue Xuebao/Acta Acustica*, vol. 42, no. 2, pp. 246–256, 2017.
- [24] Y. Lu and P. C. Loizou, "Estimators of the magnitude-squared spectrum and methods for incorporating SNR uncertainty," *IEEE Transactions on Audio Speech and Language Processing*, vol. 19, no. 5, pp. 1123–1137, 2011.
- [25] M. T. Islam, A. Asaduzzaman, C. Shahnaz, W. P. Zhu, and M. O. Ahmad, "Speech enhancement in adverse environments based on non-stationary noise-driven Spectral Subtraction and SNR-dependent phase compensation," arXiv preprint <https://arxiv.org/abs/1803.00396>, 2018.
- [26] M. Talbi, R. Baazaoui, and M. Salim Bouhleh, "Speech enhancement based on LWT and artificial neural Network and using MMSE estimate of spectral amplitude," *Deep Learning Applications*, 2021.
- [27] T. Chen, N. Kapron, and J. C.-Y. Chen, "Using evolving ANN-based algorithm models for accurate meteorological forecasting applications in vietnam," *Hindawi, Mathematical Problems in Engineering*, vol. 2020, Article ID 8179652, 8 pages, 2020.
- [28] E. Vilavicencio-Arcadia, S. G. Navarro, S. G. Navarro et al., "Application of artificial neural networks for the automatic spectral classification," *Hindawi Mathematical Problems in Engineering*, vol. 2020, Article ID 1751932, 15 pages, 2020.
- [29] K.-C. Yang, C. Yang, P.-Y. Chao, and Po-H. Shih, "Applying artificial neural Network to predict semiconductor machine outliers," *Hindawi Publishing Corporation Mathematical Problems in Engineering*, vol. 2013, Article ID 210740, 10 pages, 2013.
- [30] B. Ramesh Murlidhar, R. K. Sinha, E. T. Mohamad, R. Sonkar, and M. Khorami, "The effects of particle swarm optimisation and genetic algorithm on ANN results in predicting pile bearing capacity," *International Journal of Hydro-mechanics*, vol. 3, no. 1, p. 69, 2020.
- [31] M. Safa, M. Ahmadi, J. Mehrmashadi et al., "Selection of the most influential parameters on vectorial crystal growth of highly oriented vertically aligned carbon nanotubes by adaptive neuro-fuzzy technique," *International Journal of Hydromechanics*, vol. 3, no. 3, p. 238, 2020.
- [32] C. Zhu, W. Yan, X. Cai, S. Liu, T. H. Li, and G. Li, "Neural saliency algorithm guide bi-directional visual perception style transfer," *CAAI Transactions on Intelligence Technology*, vol. 5, no. 1, pp. 1–8, 2020.
- [33] T. Sangeetha and G. Mary Amalanathan, "Outlier detection in neutrosophic sets by using rough entropy based weighted density method," *CAAI Transactions on Intelligence Technology*, vol. 5, no. 2, pp. 121–127, 2020.
- [34] Z. Ali and T. Mahmood, "Complex neutrosophic generalised dice similarity measures and their application to decision making," *CAAI Transactions on Intelligence Technology*, vol. 5, no. 2, pp. 78–87, 2020.
- [35] T. Goehring, F. Bolner, J. J. M. Monaghan et al., "Speech enhancement based on neural networks improves speech intelligibility in noise for cochlear implant users," *Hearing Research*, vol. 344, pp. 183–194, 2017.
- [36] R. Birok, R. Kapoor, and M. Singh Choudhry, "ECG denoising using artificial neural networks and complete ensemble empirical mode decomposition," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 2, pp. 2382–2389, 2021.
- [37] J. Llombart, D. Ribas, A. Miguel, L. Vicente, A. Ortega, and E. Lleida, "Progressive loss functions for speech enhancement with deep neural networks," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2021, no. 1, 2021.
- [38] P. Karjol, M. Ajay Kumar, and P. K. Ghosh, "Speech Enhancement Using Multiple Deep Neural Networks," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Canada, April 2018.
- [39] H. Tasmaz and E. Erc, elebi, "Speech enhancement based on undecimated wavelet packet-perceptual filterbanks and MMSE- STSA estimation in various noise environments," *Digital Signal Processing*, vol. 18, no. 5, pp. 797–812, 2008.
- [40] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, & Signal Processing*, vol. 32, no. 6, pp. 1109–1121, 1984.
- [41] A. Biswas, P. K. Sahu, A. Bhowmick, and M. Chandra, "Feature extraction technique using ERB like wavelet sub-band periodic and aperiodic decomposition for TIMIT phoneme recognition," *International Journal of Speech Technology*, vol. 17, no. 4, pp. 389–399, 2014.
- [42] S. Singh and A. M. Mutawa, "A wavelet-based transform method for quality improvement in noisy speech patterns of Arabic language," *International Journal of Speech Technology*, vol. 19, no. 4, pp. 677–685, 2016.
- [43] M. Bahoura and J. Rouat, "Wavelet speech enhancement based on time-scale adaptation," *Speech Communication*, vol. 48, no. 12, pp. 1620–1637, 2006.