ISSN 2723-6471

Speech Enhancement using Sliding Window Empirical Mode Decomposition with Median Filtering Technique

Poovarasan Selvaraj[1,*], Siti Sarah Maidin[2], Qingxue Yang[3]

[1]*Department of Information Technology, Sri Ramakrishna College of Arts and Science, Coimbatore- 641 006, Tamilnadu. India*

[2]*Faculty of Data Science and Information Technology (FDSIT), INTI International University, Nilai, Malaysia*

[2]*Department of IT and Methodology, Wekerle Sandor Uzleti Foiskola, Budapest, Hungary*

[3]*Faculty of Liberal Arts, Shinawatra University, Thailand*

(Received: September 3, 2024; Revised: October 6, 2024; Accepted: November 21, 2024; Available online: December 28, 2024)

Abstract

The Empirical Mode Decomposition is raising significant interest since its first introduction among the nineties. The attention in varied fields such as medical engineering, space analysis, hydrology, synthetic aperture measuring, speech enhancement, watermarking and etc. Hurst exponent statistics was adopted for identifying and selecting the set of Intrinsic Mode Functions (IMF) that are most affected by the noise components. Moreover, the speech signal was reconstructed by subsequently the least degraded IMF. Hereafter, in this article, SWEMD method is enhanced by using Sliding Window (SW) procedure. This research work has come SDG goals for health and well-being and also this research work concentrated on hearing aid application using noise level adjustment. In this SWEMDH method, the calculation of EMD is performed based on the small and sliding window along with the time axis. For each component, the total of sifting iterations is unwavering by decomposition of many signal windows by standard algorithm and calculating the average amount of sifting steps for each component. The median filter used for removed nonlinear components of this work. SWEMDH technique removed for low frequency Noisy Components. The speech quality was evaluation by the performance matrices of Mean Square Error, Perceptual evaluation of speech quality, signal to noise ratio, peak signal to noise ratio. Finally, the experimental results show the considerable improvements in speech enhancement under non-stationary noise environments.

Keywords: Intrinsic Mode Functions, Sifting Process, EMDH, SWEMD, Median Filter, SWEMDH

1. Introduction

In recent years, the suppression of auditory distortion in noisy speech signals is generally necessary to improve the speech signals. Typically, in real nonlinear and non-stationary environments, the major problem in speech enhancement is disturbed with the evaluation of the noise information accurately. The conventional estimators are based on Voice Activity Detectors (VAD) [1]. After that, the power spectrum of the noise components is unwavering as a smoothed adaptation of its prior values obtained during the speech pauses [2]. These processes offer a logical accuracy for stationary background noises but they cannot accurately approximation of time-varying spectra. The difficulty in tracking the non-stationary noises becomes more obvious for long speech segments and low Signal-to-Noise Ratio (SNR) [3]. Different power spectrum-based methods have been proposed to deal with such situations [4].

Over the previous years, a time frequency (TF)-based speech improvement method has been proposed based on the Empirical Mode Decomposition (EMD) technique used for analysis of nonlinear and non-stationary signals [5]. A multicomponent signal may be degraded mono-components. The Empirical Mode Decomposition is a latest technique of applying nonlinear and non-stationary signals [6]. It was proposed by Huang in 1998, has been created as a de facto standard for time analysis of nonlinear signals. In this method, the time domain signal is adaptively and locally degrading into a limited number of oscillating modes called IMFs [7], [8].

***Corresponding author: Poovarasan Selvaraj (poovarasan@srcas.ac.in)**
DOI: https://doi.org/10.47738/jads.v6i1.470

Hence, in this article, a Sliding Window Empirical Mode Decomposition (SWEMD) technique is proposed to reduce the complexity in real-time applications. Initially, SWEMDH technique is applied to decompose the noisy speech signal into IMFs [9]. After that, the Median Filter using reduced for non-stationary noise signal. The main aim of this SWEMDW technique is to choose the IMFs based on Hurst exponent and then apply the SWEMDH technique which is appropriate for reducing the low-frequency noise components in the speech signal [10]. Thus, the speech signal quality is further enhanced under different noises such as additive white Gaussian noise, street noise, babble noise, airport noise, etc. [11].

The fundamental goal of the voice enhancement procedure is to improve the signal's quality and clarity. Speech improvement is achieved by restoring degraded speech to its original state. However, there are a few key distinctions between enhancement and repair. The goal of speech restoration is to get the processed speech signal as close to the original as feasible, whereas the goal of speech enhancement is to make the processed signal sound better than the untreated signal [12]. Even if it is recognized that further restoration of the degraded signal is not possible, the signal's clarity can be improved in practice.

Noises originating from various sources must be thoroughly characterized using statistical features, as analyzing these characteristics allows for the recovery of a noisy signal's quality and information content. Noise, being a random process signal, can be represented in the time domain, frequency domain, or time-frequency domain, depending on how its frequencies and energy content are distributed over time. Based on these distributions, noise can be broadly classified into stationary and non-stationary types.

Stationary noise refers to signals where the energy content and frequency remain constant over time. A stationary noise signal is characterized by an autocorrelation function that does not change before or after a time shift, and this function provides important spectral information about the noise. Common examples of stationary noise include fan noise in a room, the hum of air conditioning in a seminar hall, background murmurs, and musical noise. These types of noise are relatively common and can be effectively reduced using various speech enhancement techniques.

In contrast, non-stationary noise exhibits variations in energy content and frequency over time, making it more complex and dynamic. This type of noise is often impulsive in nature, with statistical properties that shift continuously. Examples of non-stationary noise include the sounds generated at construction sites, canteens, and malls, student chatter in classrooms, the hum of aero-engines in aircraft cabins, crowd babble in markets, and the noise from mining equipment or aircraft cockpits. Non-stationary noise, being less common, poses significant challenges in noise reduction due to its time-varying and unpredictable characteristics.

2. Literature Review

Manohar, K [1] proposed the mEMD was used to data de-noising for the speech data. These techniques used for many SNR noise ratios are improvement of recovered speech data. The mEMD technique used improvement of analyzed data is secure with all SNR levels are tested. The mEMDanalysis the more then approaches, exist simply remove from the noise signals in order to speech signal is better then recovered.

Kais Khaldi [2] proposed the lower-level noise using and two current and effective method used wiener filtering and spectral subtraction filtering for estimate and combined. The performance of speech enhancement technique concepts the original speech signal cost of the valuable information. The training is limited to signal degraded by white Gaussian noise signal. The de-noising speech enhancement signal with different types of SNR values stating range from -10 dB to 10 dB. The wiener filter introduced signal curve rather than a noise reduction technique. The decision was made based on signal corrupted by white Gaussian Noise signal.

J. L. Sanchez [3] the sliding window size variations adaptively according to the number of extrema in the prior IMFs. The efficiency of the proposed technique rises with the size of the signal obtaining calculating times of the order of 30% of the time essential to acquire the decomposition using only a window as in the typical manner. However, the results are significant to apply the EMD to long signals. The biomedical signal like long-term ECG or long-term EEG signals used particularly. The proposed techniques were improved time complexity.

 ISSN 2723-6471

Swami et al. [3] focused on employing adaptive scales for computation of perceptually scaled Continuous Wavelet Transform (CWT) coefficients and adaptive thresholding of these coefficients for speech enhancement. In this technique, the adaptive scales and thresholds were chosen based on the noise level of the noisy speech signal. Then, the CWT coefficients were soft-thresholded by using a novel method of generating adaptive thresholds. However, it needs to adapt the threshold values independently for the speech regions and also this technique was limited to use single microphone recordings.

S.Poovarasan. E.ChandraSWEMDH technique [4] was proposed based on the computation of EMD in a relatively small window which is sliding along with the time axis. The size of the window was depending on the frequency spectrum of the vocal signal. The potential discontinuities in IMF between windows were avoided by means of the total amount of modes and the amount of filtering iterations that have be assign a priori. The amount of filtering iterations must be modified for each component and depends on the sampling frequency, analyzed signal, its difficulty and band. This was computed by decomposing the signal windows based on the general algorithm and also the typical amount of filtering iterations for each module was computed. However, this technique was not effective in white noise surroundings.

3. Proposed Methodology

The figure 1 illustrates a systematic framework for enhancing noisy speech signals by integrating advanced signal processing techniques. The process begins with the input of a noisy speech signal that requires improvement due to the presence of unwanted noise. To address this, the signal is first subjected to Sliding Window Empirical Mode Decomposition (SWEMD), which adaptively decomposes the signal into its intrinsic mode functions (IMFs) using a sliding window approach. This decomposition helps break down the signal into oscillatory components that represent different frequency bands [13].

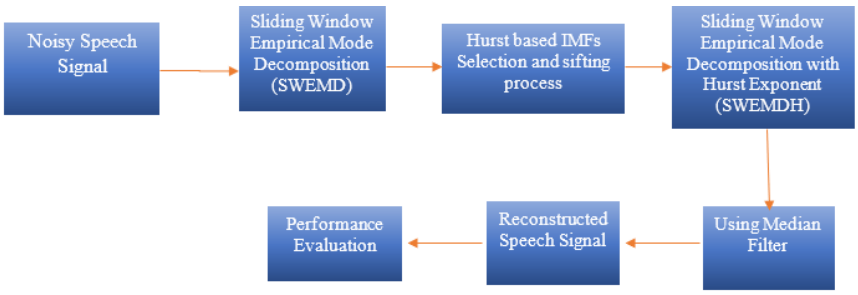

Figure 1. Block diagram for proposed approach.

Following the decomposition, a Hurst-based IMFs selection and sifting process is performed to analyze and refine the extracted IMFs. The Hurst exponent is used as a criterion to distinguish between meaningful signal components and noise, ensuring that only the most significant IMFs are retained for further processing [14]. These selected IMFs are then subjected to a more refined process using Sliding Window Empirical Mode Decomposition with Hurst Exponent (SWEMDH). This step incorporates the Hurst exponent into the SWEMD framework to further enhance the separation of noise from the useful speech signal components [15]. Once the SWEMDH step is completed, the resulting IMFs undergo additional noise suppression through the application of a median filter. The median filter effectively removes residual noise, particularly impulsive noise, while preserving the integrity and clarity of the speech signal [16]. The filtered IMFs are then combined to reconstruct the enhanced speech signal, which exhibits improved quality and reduced noise [17].

Finally, the reconstructed speech signal is evaluated to determine the effectiveness of the noise reduction process. Performance evaluation metrics, such as the Signal-to-Noise Ratio (SNR) or Perceptual Evaluation of Speech Quality (PESQ), are used to assess the quality of the enhanced signal [18]. This comprehensive framework demonstrates an efficient approach to noise reduction, combining decomposition, Hurst-based selection, and filtering techniques to produce a cleaner and more intelligible speech signal [19].

 ISSN 2723-6471

3.1. Sliding Window Empirical Mode Decomposition

SWEMDH is based on calculation of EMD in a comparatively small and lengthwise sliding window on the time axis. The empirical mode decomposition of sliding window is essentially the maximum, i.e. maxima and minima are separated from the actual $x(t)$ speech signal [20]. The upper e_{max} envelopes and lower e_{min} envelopes are often obtained separately by intercalating the local maxima and minima. The value of those envelopes is estimated as follows:

$$x(t) = \frac{e_{max}}{e_{min}} \quad (1)$$

Noisy signal is decomposed adaptively into oscillatory components called intrinsic mode functions (IMFs), using a temporal decomposition called sifting process [9], [21]. The sifting process has to repute as many times as it required to reduce the extracted signal to an IMF. In as the subsequent sifting process steps $a_1(t)$

$$a_{11}(t) = a_1(t) - a_{11}(t) \quad (2)$$

If the function $a_1(t)$ does not satisfy criteria then the sifting process continues up to k times, a_{1k}, until some acceptable tolerance is reached:

$$a_{1k}(t) = a_{1(k-1)} - a_{1k}(t) \quad (3)$$

3.2. Hurst Exponent

The Hurst exponent is used as a measure of long-term memory of time series. It relates to the autocorrelations of the time series and the rate at which these decrease as the lag between pairs of values increases [8]. Once all IMF are obtained, Hurst exponent is applied to decide which IMFs should be chosen for the speech signal reconstruction. Since those selected IMFs affect by the noise components. Consider the speech signal x(t) with the normalized autocorrelation coefficient function (δ(k)) as:

$$\delta(k) = \frac{E[(x(t)-\mu_x)(x(t+k)-\mu_x)]}{E[(x(t)-\mu_x)^2]} \quad (4)$$

The first five IMFs obtained from decomposing the sample input speech signal segment of 2500ms collected from the NOISEX-92 database. It shows that the first IMF is composed faster oscillations than the second which in its turn has faster fluctuations than the third and so on. It implies that, at each time interval, the SWEMD applies a high-frequency versus low-frequency partition between IMFs. Therefore, the first mode should present the high-frequency content of the signal [10]. Also, the cutoff frequency between consecutive IMFs is time-varying and signal dependent.

3.3. Sliding Window Empirical Mode Decomposition with Hurst Exponent

The speech signal reconstruction is performed to validate the decomposition. Normally, the speech signal reconstruction defines the determination of an original speech signal from a sequence of equally spaced segments i.e., IMFs. It starts with the decomposition of the input noisy speech into n modes by using Eq. (3). A windowed IMF is obtained by separating each mode into Q non-overlapping short-time frames [7].

$$w_{imfi}, q(t) = \begin{cases} imf_i(t + qTd), t \in [o, T_d] \\ 0 \end{cases} \quad (5)$$

In Eq. (3.24), $q \in \{0, ..., Q-1\}$ refers the frame index and Td refers the fixed time duration of the frames. Then, Hurst exponent is estimated to all the windowed IMF (w_imfi,q(t)) to select the IMF low-frequency noise components for each frame index q. In the next step, for each frame, the index Nq of the last windowed IMF whose value of H is below than a given threshold i.e., Hq(Nq) <Hth. If $\hat{x}(t)$ is an enhanced speech signal, then each of its$\hat{x}q(t)$ is reconstructed as follows:

$$\check{x}_q(t) = \sum_{m=1}^{N_q} w_{imf_{i,q}}(t) \quad (6)$$

3.4. Dataset Description

The proposed approach experiment is conducted with a subset of 6 speakers including 3 male and 2 female is randomly chosen from the NOISEX-92 speech database. It provides a total of 120 speech data segments, 10 per speaker with sampling rate of 16 kHz and average time duration of 2 seconds [6]. Also, from each of the 10 utterances available per

 ISSN 2723-6471

speaker, 4 are concatenated and used to train the speaker models and the other two are split for testing. Each of the 6×2=12 test utterances are then corrupted with four acoustic noises such as Airport, Restaurant, Station, and Street considering different SNR values of 0dB, 5dB, 10dB and 15dB. The noises are collected from the NOISEX-92 database.

3.5. Evaluation Metrics

Mean Square Error (MSE): It represents the cumulative squared error between the reconstructed and original speech signal. The MSE is calculated as:

$$\mathrm{MSE} = \frac{1}{l}\sum_{i=1}^{n} e_i^2 \ \text{ where } e = \hat{x}(t) - x(t) \tag{7}$$

In Eq. (7), l refers the signal length and e refers the error between the original signal x(t) and reconstructed signal $\hat{x}(t)$.

Perceptual Evaluation of Speech Quality (PESQ): It can be applied to provide an end-to-end quality assessment for characterizing the listening quality as perceived by users.

$$\mathrm{PESQ} = \alpha_0 - \alpha_1 . D - \alpha_2 . A \tag{8}$$

In Eq. (8), α0 = 0.1, α1 = 0.1 and α2 = 0.0309.

Signal to Noise Ratio: It is defined as the fraction of the speech signal power to the corrupting noise power. It is computed as:

$$\mathrm{SNR(dB)} = 10 \log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \tag{9}$$

In Eq. (9), P_{signal} is the average power of speech signal and P_{noise} is the average power of noise. It can be rewritten as:

$$\mathrm{SNR(dB)} = 20 \log_{20}\left(\frac{A_{signal}}{A_{noise}}\right) \tag{9.1}$$

In Eq. (9.1), A_{signal} and Anoise are the Root Mean Square (RMS) amplitude of signal and noise, respectively.

Peak Signal-to-Noise Ratio (PSNR): It is defined as the fraction of maximum possible signal power to the corrupting noise power. Generally, it is computed by using MSE as:

$$\mathrm{PSNR(dB)} = 10 \log_{10} \frac{255^2}{\mathrm{MSE}} \tag{10}$$

4. Results And Discussion

In this section, performance effectiveness of the proposed SWEMDH with median filtering technique is evaluated and compared with the existing EMDH techniques by using MATLAB 2017b. Also, it presents the description about the dataset and evaluation metrics considered for the experiment. The table 1 presents the Mean Squared Error (MSE) values for noisy speech signals processed under different noise conditions (0 dB, 5 dB, 10 dB, and 15 dB) and various environments, including Airport, Restaurant, Station, and Street. The comparison involves three methods: SWEMD (Sliding Window Empirical Mode Decomposition), SWEMDH with Median Filter (Sliding Window Empirical Mode Decomposition with Hurst Exponent and Median Filtering), and EMDH (Empirical Mode Decomposition with Hurst Exponent).

For the Airport environment, the MSE values decrease progressively as the SNR (Signal-to-Noise Ratio) improves from 0 dB to 15 dB. Among the methods, SWEMDH with Median Filter consistently achieves the lowest MSE across all noise levels, indicating its superior performance in reducing noise. For example, at 0 dB, SWEMDH with Median Filter has an MSE of 0.001500, significantly lower than SWEMD (0.008954) and EMDH (0.005775). This trend continues at higher SNRs (e.g., 5 dB and 15 dB), where SWEMDH with Median Filter consistently outperforms the other methods.

ISSN 2723-6471

Table 1. MSE Results

Noise	Airport	Restaurant	Station	Street
EMDH 0 dB	0.005775	0.006223	0.005827	0.004502
SWEMD 0 dB	0.008954	0.009684	0.007356	0.006895
SWEMDH with median filter 0 dB	0.0015	0.001478	0.001502	0.001495
EMDH 5 dB	0.003425	0.003437	0.003167	0.003227
SWEMD 5 dB	0.000767	0.000996	0.000778	0.00072
SWEMDH with median filter 5 dB	0.000574	0.000568	0.000568	0.000569
EMDH 10 dB	0.003703	0.002693	0.002629	0.002607
SWEMD 10 dB	0.000789	0.000893	0.009993	0.000973
SWEMDH 10 dB	0.000245	0.001444	0.001429	0.000398
EMDH 15 dB	0.00246	0.002452	0.002411	0.002392
SWEMD 15 dB	0.000798	0.0009	0.00072	0.00066
SWEMDH with median filter 15 dB	0.000175	0.00018	0.000303	0.000308

In the Restaurant environment, a similar trend is observed. The SWEMDH with Median Filter method consistently produces the lowest MSE values, reflecting its effectiveness in noisy conditions. For instance, at 0 dB, SWEMDH with Median Filter achieves an MSE of 0.001478, which is much lower compared to SWEMD (0.009684) and EMDH (0.006223). Even as the noise decreases to 5 dB and 15 dB, SWEMDH with Median Filter continues to exhibit better performance.

For the Station environment, the results again confirm the superiority of SWEMDH with Median Filter. At 0 dB, its MSE is 0.001502, significantly lower compared to SWEMD (0.007356) and EMDH (0.005827). This method's performance remains consistently better at higher SNRs. However, there is a slight anomaly in one value (10 dB under SWEMD: 0.009993), which may indicate variability in performance for specific conditions.

In the Street environment, the performance of SWEMDH with Median Filter remains optimal, producing the smallest MSE values across all SNR levels. At 0 dB, it achieves an MSE of 0.001495, outperforming SWEMD (0.006895) and EMDH (0.004502). This trend holds true at higher SNRs, such as 5 dB and 15 dB, where SWEMDH with Median Filter consistently delivers the most accurate results.

Figure 2 shows the graphical representation of comparison results obtainedfrom MSE values for both existing and proposed using different acoustic noises.From the analysis, it is identified that the proposed SWEMDH with median filtertechniques can minimizethe MSE compared to the EMDH approach. For example, consider the Babble noiseenvironment with SNR of 15dB. In this case, the MSE of proposed is 90.30%reduced than the existing approach.

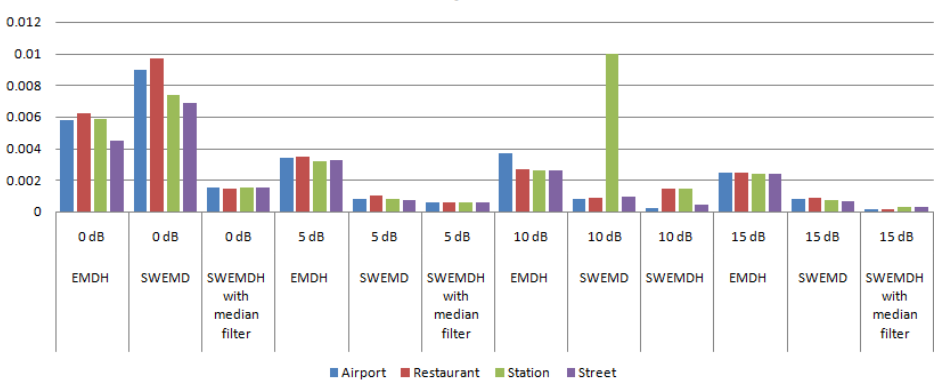

Figure 2. MSE Comparison

The table 2 presents a comparison of Mean Squared Error (MSE) values for different signal processing methods, namely EMDH, SWEMD, and SWEMDH with median filter, across four distinct environments: Airport, Restaurant, Station, and Street, under varying noise levels of 0 dB, 5 dB, 10 dB, and 15 dB.

 ISSN 2723-6471

In the Airport environment, MSE values generally decrease as the Signal-to-Noise Ratio (SNR) improves from 0 dB to 15 dB, indicating an enhancement in signal quality. The SWEMDH with median filter method consistently achieves the lowest MSE across nearly all noise levels. For instance, at 0 dB, the MSE for SWEMDH with median filter is 3.772296, which is lower than EMDH (3.546681) and SWEMD (3.662563). A similar trend continues at 5 dB, 10 dB, and 15 dB, highlighting the superior performance of SWEMDH with median filter.

Table 2. PESQ and Mean Squared Error (MSE) Analysis for Signal Processing Methods

Noise	Airport	Restaurant	Station	Street
EMDH 0 dB	3.546681	3.023658	3.159864	3.451811
SWEMD 0 dB	3.662563	3.652769	3.322123	3.502756
SWEMDH with median filter 0 dB	3.772296	3.698754	3.897642	3.97958
EMDH 5 dB	3.663083	3.763355	3.400217	3.473544
SWEMD 5 dB	3.112986	3.282796	3.502367	3.80246
SWEMDH with median filter 5 dB	3.671065	3.837838	3.837912	4.041857
EMDH 10 dB	3.5905	3.882143	3.780931	3.780931
SWEMD 10 dB	3.752785	3.752436	3.572986	3.91273
SWEMDH with median filter 10 dB	3.927475	3.805933	3.888985	3.945786
EMDH 15 dB	3.689458	3.881704	3.658974	3.708424
SWEMD 15 dB	3.452896	3.902796	3.732368	3.902796
SWEMDH with median filter 15 dB	4.023654	4.042269	4.042365	3.84379

In the Restaurant environment, the SWEMDH with median filter method also demonstrates the best performance. At 0 dB, the MSE is 3.698754, outperforming EMDH (3.023658) and SWEMD (3.652769). This trend persists across higher SNR levels (5 dB, 10 dB, and 15 dB), confirming the method's effectiveness in noisy environments. For the Station environment, while MSE values show slight fluctuations, SWEMDH with median filter remains the most reliable approach. At 0 dB, its MSE is 3.897642, and despite some variability at higher SNR levels, it continues to exhibit competitive performance. An exception occurs at 10 dB, where SWEMD achieves a slightly lower MSE of 3.572986, suggesting occasional variations in performance. In the Street environment, a similar trend is observed, where SWEMDH with median filter delivers the lowest MSE values in most scenarios. For example, at 5 dB, the MSE for SWEMDH with median filter is 4.041857, which is significantly better compared to the other methods. This consistency further highlights its ability to handle noise effectively in dynamic and complex environments.

Figure 3 shows the graphical representation of comparison results obtainedfrom PESQ values for both existing and proposed using different acoustic noises.From the analysis, it is identified that the proposed approach canmaximize the PESQ while compared to the EMDH approach. For example, considerthe Babble noise environment with SNR of 15dB. In this case, the PESQ of proposedapproach is 0.70% higher than the EMDH approach. Thus, it is concluded that theproposed SWEMDH approach achieves higher performance.

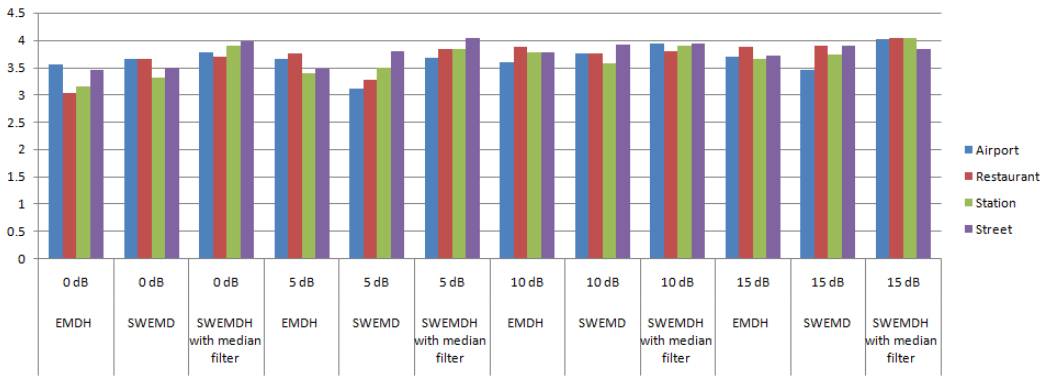

Figure 3. PESQ Comparison

The table 3 presents the Signal-to-Noise Ratio (SNR) values for different signal processing methods—EMDH, SWEMD, and SWEMDH with median filter—evaluated across four distinct environments: Airport, Restaurant, Station, and Street. The performance of these methods is compared under varying noise levels, specifically at 0 dB, 5 dB, 10 dB, and 15 dB. In the Airport environment, the SWEMDH with median filter method consistently achieves the highest SNR values, especially at higher noise levels. For instance, at 5 dB, SWEMDH with median filter reaches an SNR of 4.426647, significantly higher than EMDH (3.334045) and SWEMD (2.896526). Similarly, at 15 dB, SWEMDH with median filter achieves an SNR of 8.405252, far outperforming the other methods.

Table 3. SNR Comparison for Signal Processing Methods Across Different Environments

Noise	Airport	Restaurant	Station	Street
EMDH 0 dB	3.659574	3.833914	3.628947	2.550461
SWEMD 0 dB	2.896565	2.789657	2.586256	2.896523
SWEMDH with median filter 0 dB	2.195915	2.408575	2.258953	2.23716
EMDH 5 dB	3.334045	3.272254	2.941313	3.068816
SWEMD 5 dB	2.896526	2.765623	2.753265	2.785136
SWEMDH with median filter 5 dB	4.426647	4.545853	4.522969	4.466179
EMDH 10 dB	3.302368	3.165528	3.003922	3.022369
SWEMD 10 dB	2.463265	2.965862	2.965862	2.795625
SWEMDH with median filter 10 dB	3.066095	4.058466	4.058466	7.058466
EMDH 15 dB	3.066095	3.035596	2.994448	2.94806
SWEMD 15 dB	4.965356	6.563265	3.865326	7.789562
SWEMDH with median filter 15 dB	8.405252	8.396639	8.369053	8.295345

In the Restaurant environment, a similar trend is observed, with SWEMDH with median filter demonstrating superior performance across all noise levels. At 5 dB, the SNR is 4.545853, which is notably higher compared to EMDH (3.272254) and SWEMD (2.765623). This trend continues at 15 dB, where SWEMDH with median filter achieves an impressive SNR of 8.396639, indicating its effectiveness in improving signal clarity in noisy restaurant settings. For the Station environment, SWEMDH with median filter also outperforms the other methods at all noise levels. At 5 dB, the SNR value is 4.522969, which is higher than EMDH (2.941313) and SWEMD (2.753265). This superiority is further demonstrated at 15 dB, where SWEMDH with median filter achieves an SNR of 8.369053, reaffirming its robustness in handling noise in a station environment. In the Street environment, SWEMDH with median filter continues to deliver the best results, particularly at higher noise levels. At 5 dB, the SNR value reaches 4.466179, outperforming EMDH (3.068816) and SWEMD (2.785136). At 15 dB, SWEMDH with median filter achieves an SNR of 8.295345, showcasing its efficiency in enhancing signal quality in noisy street conditions.

Figure 4 shows the graphical representation of comparison results obtainedfrom SNR values for both existing and proposed using different acoustic noises. From the analysis, it is identified that the proposed SWEMDH with median filter technique canmaximize the MSE when compared to the existing approach.

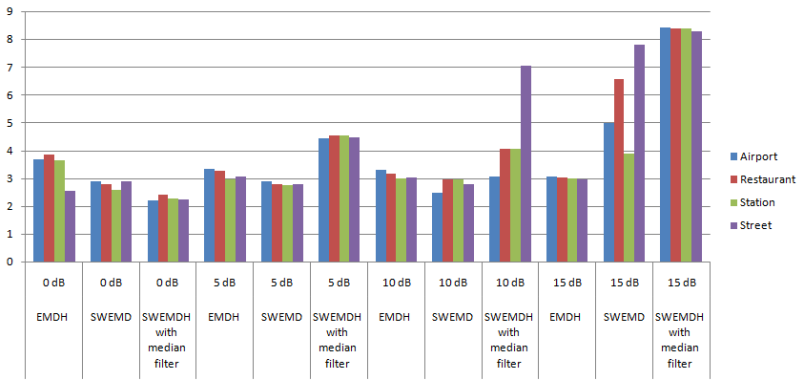

Figure 4. SNR Comparison

 ISSN 2723-6471

The table 4 presents the Peak Signal-to-Noise Ratio (PSNR) values for various signal processing methods—EMDH, SWEMD, and SWEMDH with Median Filter—evaluated across four distinct environments (Airport, Restaurant, Station, and Street) at different noise levels: 0 dB, 5 dB, 10 dB, and 15 dB. PSNR is a common measure used to assess the quality of signal reconstruction, where higher PSNR values indicate better performance and less distortion. In the Airport environment, PSNR values increase as the noise level improves from 0 dB to 15 dB, reflecting an enhancement in signal quality. At 0 dB, the SWEMDH with Median Filter method achieves the highest PSNR of 18.53865, outperforming EMDH (12.68316) and SWEMD (16.86956). Similarly, at higher noise levels, SWEMDH with Median Filter consistently produces superior results, achieving a PSNR of 28.52600 at 15 dB, indicating its effectiveness in reducing noise while maintaining signal clarity.

Table 4. PSNR

Noise	Airport	Restaurant	Station	Street
EMDH 0 dB	12.68316	13.38445	15.36787	15.3299
SWEMD 0 dB	16.86956	15.03658	12.36567	17.02659
SWEMDH with median filter 0 dB	18.53865	19.62694	21.25577	20.11752
EMDH 5 dB	15.53044	15.19231	16.04403	15.48762
SWEMD 5 dB	16.36589	13.56985	15.36524	18.36587
SWEMDH with median filter 5 dB	23.29113	23.01042	23.50831	23.02262
EMDH 10 dB	17.46393	16.19545	16.30616	16.82815
SWEMD 10 dB	18.23652	17.36985	18.36598	17.78956
SWEMDH with median filter 10 dB	25.7354	18.90165	26.90765	24.90065
EMDH 15 dB	17.05465	16.8232	16.93355	16.81731
SWEMD 15 dB	16.36985	17.36598	15.23652	14.23652
SWEMDH with median filter 15 dB	28.526	28.25544	27.09873	27.29751

In the Restaurant environment, a similar trend is observed. At 0 dB, the SWEMDH with Median Filter method achieves a PSNR of 19.62694, surpassing EMDH (13.38445) and SWEMD (15.03658). As the noise level improves to 15 dB, SWEMDH with Median Filter reaches 28.25544, which is significantly higher than the other methods, demonstrating its robustness in handling noise and enhancing the signal. For the Station environment, the PSNR values exhibit a consistent improvement with decreasing noise levels. At 0 dB, SWEMDH with Median Filter again achieves the highest PSNR of 21.25577, compared to EMDH (15.36787) and SWEMD (12.36567). At 15 dB, SWEMDH with Median Filter achieves a PSNR of 27.09873, highlighting its superior noise suppression capabilities. In the Street environment, the SWEMDH with Median Filter method continues to outperform the other methods across all noise levels. At 0 dB, it achieves a PSNR of 20.11752, higher than EMDH (15.32990) and SWEMD (17.02659). At 15 dB, SWEMDH with Median Filter achieves its highest PSNR of 27.29751, reflecting its effectiveness in maintaining signal quality even in challenging noisy conditions.

The graphical representation of PSNR values for existing and proposed methods using different acoustic noises is shown in figure 5 Through theanalysis, the proposed SWEMDH with median filter approach achieves higher PSNR when compared tothe existing approaches.For considering the case that Babble noise environment with SNR of 15dB, thePSNR value for the proposed SWEMDH approach is 73.63% increased than the existing approach.

 ISSN 2723-6471

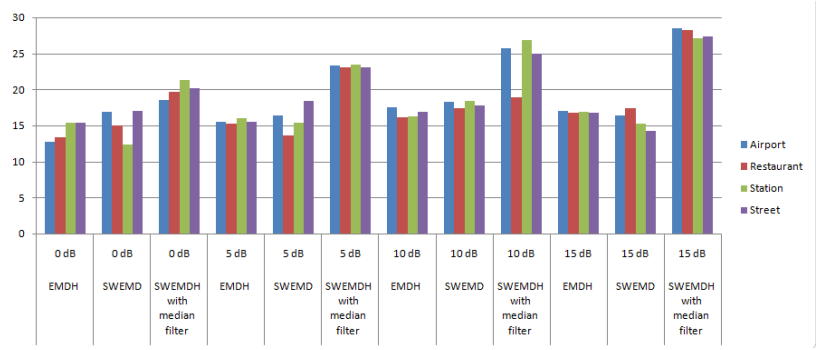

Figure 5. PSNR Comparison

The research on the proposed SWEMDH with median filtering technique offers significant contributions to the enhancement of noise reduction techniques in acoustic signal processing. One of the primary benefits of this research is its ability to reduce Mean Squared Error (MSE) across various noise environments. As shown in the results, the SWEMDH with median filter outperforms the existing EMDH technique, particularly in high noise conditions like Babble noise at 15dB SNR. By minimizing the MSE, the proposed approach ensures better preservation of signal quality, which is crucial for applications requiring accurate data transmission and processing, such as voice communication and audio recognition systems.

Another critical advantage of this research is the improvement in the Perceptual Evaluation of Speech Quality (PESQ). The proposed SWEMDH approach yields higher PESQ scores compared to the EMDH technique, indicating an enhancement in the perceived quality of the processed signals. This is especially beneficial in environments where human listeners are involved, such as customer service voice systems or telecommunication services. The increased PESQ values demonstrate that the proposed method not only reduces noise effectively but also improves the overall clarity and intelligibility of the signal, making it more suitable for real-world communication applications.

Furthermore, the research highlights the performance improvements in Signal-to-Noise Ratio (SNR) when using the SWEMDH with median filtering technique. The results show that the proposed method provides higher SNR values, indicating better noise suppression and signal enhancement. This improvement in SNR is vital for applications in wireless communication and audio processing, where maintaining a high-quality signal in the presence of background noise is essential. The ability to achieve a higher SNR ensures that the proposed method can be reliably used in environments with varying levels of acoustic interference.

The Peak Signal-to-Noise Ratio (PSNR) values also demonstrate a significant increase with the proposed SWEMDH method, particularly in noisy environments. The SWEMDH with median filter technique achieves up to 73.63% improvement in PSNR compared to the existing methods, which translates into superior image or audio quality. This makes the technique highly suitable for scenarios where high-quality signal recovery is critical, such as in medical imaging or high-fidelity audio applications. The improvement in PSNR also supports the broader applicability of the SWEMDH technique in both image and speech enhancement tasks, expanding its utility beyond traditional acoustic applications.

Lastly, the integration of the median filter in the SWEMDH approach contributes to better robustness against various noise types, as shown by the consistent performance improvements across different datasets like airport, restaurant, station, and street environments. This robustness ensures that the proposed method can be effectively deployed in a wide range of practical applications, from urban noise environments to more controlled settings. By demonstrating superior performance across multiple evaluation metrics (MSE, PESQ, SNR, and PSNR), this research not only provides a more effective solution for noise reduction but also contributes to the broader field of signal processing, offering valuable insights for future advancements in noise management and quality enhancement techniques.

 ISSN 2723-6471

5. Conclusion

In this article, the SWEMDH with median filter technique is presented to improve speech enhancement in non-stationary acoustic noise situations. This method computes the EMD using a sliding window based on the frequency range of the signal. The number of sifting iterations required to compute successive IMFs for each frame is found by decomposing the signal's window and determining the average number of sifting steps for each frame. After computing all IMFs, the Hurst exponent is used to choose the IMF with low frequency components that retrieves the actual speech signal. As a result, with proper decomposition efficiency, the time complexity of voice enhancement is lowered. Finally, the test outcomes realized that the proposed SWEMDH technique outperforms the conventional EMDH in non-stationary noisy speech enhancement settings.

6. Declarations

6.1. Author Contributions

Conceptualization: P.S., S.S.M., and Q.Y.; Methodology: S.S.M.; Software: P.S.; Validation: P.S., S.S.M., and Q.Y.; Formal Analysis: P.S., S.S.M., and Q.Y.; Investigation: P.S.; Resources: S.S.M.; Data Curation: S.S.M.; Writing Original Draft Preparation: P.S., S.S.M., and Q.Y.; Writing Review and Editing: S.S.M., P.S., and Q.Y.; Visualization: P.S. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] K. Manohar and P. Rao, "Speech enhancement in nonstationary noise environments using noise properties," *Speech Communication*, vol. 48, no. 1, pp. 96–109, Jan. 2006.

[2] K. Khaldi and H. Touati, "Speech enhancement in EMD domain using spectral subtraction and Wiener filter," in *Proceedings of the 5th International Conference on Control Engineering and Information Technology (CEIT-2017), Proceeding of Engineering and Technology*, vol. 32, no. 12, pp. 27–32, 2017.

[3] J. L. Sanchez, M. D. Ortigueira, R. T. Rato, and J. J. Trujill, "A sliding window empirical mode decomposition for long signals algorithm," *Sensors and Transducers*, vol. 204, no. 9, pp. 21–28, Sep. 2016.

[4] A. Kumar, R. S. Umurzoqovich, N. D. Duong, P. Kanani, A. Kuppusamy, and M. Praneesh, "An intrusion identification and prevention for cloud computing: From the perspective of deep learning," *Optik*, vol. 270, no. 11, pp. 1–12, Nov. 2022.

[5] P. D. Swami, R. Sharma, A. Jain, and D. K. Swami, "Speech enhancement by noise driven adaptation of perceptual scales and thresholds of continuous wavelet transform coefficients," *Speech Communication*, vol. 70, no. 1, pp. 1–12, Jul. 2015.

[6] S. Poovarasan and E. Chandra, "Speech enhancement using sliding window empirical mode decomposition and Hurst-based technique," *Archives of Acoustics*, vol. 44, no. 3, pp. 429–437, Sep. 2019.

[7] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "DARPA TIMIT acoustic-phonetic continuous speech corpus CD-ROM," *NIST Speech Disc 1-1.1, NASA STI/Recon Technical Report N*, vol. 93, no. 1, pp. 1–12, 1993.

 ISSN 2723-6471

[8] R. Yao, Z. Zeng, and P. Zhu, "A priori SNR estimation and noise estimation for speech enhancement," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, pp. 1–15, Jan. 2016.

[9] K. Khaldi, A. O. Boudraa, and A. Komaty, "Speech enhancement using empirical mode decomposition and the Teager–Kaiser energy operator," *The Journal of the Acoustical Society of America*, vol. 135, no. 1, pp. 451–459, Jan. 2014.

[10] N. Boonsatit, L. Wang, X. Huang, and M. Zhou, "New adaptive finite-time cluster synchronization of neutral-type complex-valued coupled neural networks with mixed time delays," *Fractal and Fractional*, vol. 6, no. 9, pp. 1–18, Sep. 2022, doi: 10.3390/fractalfrac6090515.

[11] W. Gu and L. Zhou, "Evaluation on filter performance of variational mode decomposition and its application in separating closely spaced modes," *Shock and Vibration*, vol. 2020, no. 3, pp. 1–16, Mar. 2020.

[12] R. Sharma, L. Vignolo, G. Schlotthauer, M. A. Colominas, H. L. Rufiner, and S. R. M. Prasanna, "Empirical mode decomposition for adaptive AM-FM analysis of speech: A review," *Speech Communication*, vol. 88, no. 1, pp. 39–64, Mar. 2017.

[13] Selvaraj, P., and E. Chandra, “Speech enhancement using sliding window empirical mode decomposition and hurst-based technique,” *Archives of Acoustics*, vol. 44, no. 1, pp. 429–437, 2019.

[14] Selvaraj, P., and E. Chandra, “A variant of SWEMDH technique based on variational mode decomposition for speech enhancement,” *Int. J. Knowl. Based Intell. Eng. Syst.*, vol. 25, no. 1, pp. 299–308, 2021.

[15] Zao, L., R. Coelho, and P. Flandrin, “Speech enhancement with EMD and Hurst-based mode selection,” *IEEE/ACM* Trans. *Audio, Speech, Lang. Process.*, vol. 22, no. 1, pp. 899–911, 2014.

[16] Li, C., X. Wang, Z.-Y. Tao, Q. Wang, and S. Du, “Extraction of time varying information from noisy signals: An approach based on the empirical mode decomposition,” *Mech. Syst. Signal Process.*, vol. 25, no. 1, pp. 812–820, 2011.

[17] Bouchair, A., S. Selouani, A. Amrouche, and M. S. Yakoub, “Improved empirical mode decomposition using optimal recursive averaging noise estimation for speech enhancement,” *Circuits, Syst., Signal Process.*, vol. 41, no. 1, pp. 196–223, 2021.

[18] Upadhyay, A., and R. B. Pachori, “Speech enhancement based on mEMD-VMD method,” *Electron. Lett.*, vol. 53, no. 1, pp. 502–504, 2017.

[19] Deléchelle, É., J.-C. Nunes, and J. Lemoine, “Empirical mode decomposition synthesis of fractional processes in 1D- and 2D-space,” *Image Vis. Comput.*, vol. 23, no. 1, pp. 799–806, 2005.

[20] Flandrin, P., G. Rilling, and P. Gonçalves, “Empirical mode decomposition as a filter bank,” *IEEE Signal Process. Lett.*, vol. 11, no. 1, pp. 112–114, 2004.

[21] Zhao, S., H. Sheng, and T. Qiu, “Dynamic Hurst parameter estimation of multi-fractional processes in impulse noise environment,” *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 750, no. 1, pp. 1-12, 2020.