ISSN: 0067-2904

A Hybrid Method for Speech Noise Reduction Using Log-MMSE

Ruqaya Jamal Nasir*, Husam Ali Abdulmohsin
Department of Computer Science, College of Science, University of Baghdad, Baghdad, Iraq

Received: 25/9/2023 Accepted: 16/2/2024 Published: 28/2/2025

Abstract

Noise reduction is a significant field that appears in many aspects of life. There are many methods for speech noise reduction, especially for stationary noise. This paper talks about a new way to lower speech noise that combines the Log-MMSE (Logarithmic Minimum Mean Square Error) algorithm, which is used to improve speech signals that have been messed up by noise, with an adaptive Wiener filter with a decision-directed (DD) approach. This filter lowers musical noise and predicts the time-varying noise spectrum, which results in a better signal-to-noise ratio. The frame delay issue resulting from DD was resolved by utilizing the Two-Step Noise Reduction (TSNR) technique to reconstruct the harmonic structure of the voice signal that was distorted or missed during the processing and boost speech quality and intelligibility in loud circumstances. Harmonic Regeneration Noise Reduction (HRNR) was used. In this paper, we will investigate these methods in the field of stationary and non-stationary noise. The proposed method was evaluated using different techniques like SNR, perceptual evaluation of speech quality (PESQ), and short-time objective intelligibility (STOI). The proposed approach produced acceptable results in PESQ and STOI, with a considerable increase in the percentage of (SNR), where the percentage of our improvement reached 63.49% in the results of (SNR).

Keywords: Log-MMSE (Logarithmic Minimum Mean Square Error); Decision-Directed (DD); Two-Step Noise Reduction (TSNR); Harmonic Regeneration Noise Reduction (HRNR); MMSE (Minimum Mean Square Error).

طريقة هجينة لتقليل ضوضاء الكلام باستعمال*Log-MMSE*

رقية جمال ناصر*, حسام علي عبد المحسن
قسم علوم الحاسوب، كلية العلوم, جامعة بغداد, بغداد, العراق

الخلاصة

يعد تقليل الضوضاء مجالًا مهمًا يظهر في العديد من جوانب الحياة، هناك العديد من الطرق لتقليل ضوضاء الكلام، خاصة بالنسبة للضوضاء الثابتة. في هذه الورقة، سيتم تقديم نهج مختلط لتقليل ضوضاء الكلام باستعمال خوارزمية Log-MMSE (الحد الأدنى اللوغاريتمي لمتوسط الخطأ المربع)، والتي تُستعمل بشكل شائع لتحسين الكلام وتحاول تحسين جودة ووضوح إشارات الكلام التالفة بالضوضاء، ويتم استعمال مرشح Wiener التكيفي مع نهج موجه نحو القرار (DD)، لتقليل الضوضاء الموسيقية والتنبؤ بطيف الضوضاء المتغير بمرور الوقت، مما يؤدي إلى تحسين نسبة الإشارة إلى الضوضاء (SNR) وأداء الوضوح. تم استعمال

*Email: roqia.nasser2101m@sc.uobaghdad.edu.iq

خوارزمية تقليل الضوضاء بخطوتين (TSNR) لحل مشكلة تأخير الإطار الناتج عن DD، ولإعادة بناء البنية التوافقية للإشارة الصوتية التي تم تشويهها أو تفويتها أثناء المعالجة. وتعزيز جودة الكلام والوضوح في الظروف الصاخبة. يتم استعمال تقليل الضوضاء بالتجديد التوافقي (HRNR). في هذا البحث، سيتم دراسة هذه الطرق في مجال الضوضاء الثابتة وغير الثابتة. تم تقييم الطريقة المقترحة باستعمال تقنيات مختلفة مثل نسبة الإشارة إلى الضوضاء (SNR)، والتقييم الإدراكي لجودة الكلام (PESQ)، والوضوح الموضوعي في الوقت القصير (STOI). وقد أسفر النهج المقترح عن نتائج مقبولة، مع ارتفاع كبير في نسبة ال (SNR) حيث بلغت نسبة تحسيننا الى 63.49٪ في نتائج (SNR).

1. Introduction

Understanding the human voice production process is crucial for dealing with speech signal filters properly. The study of the sounds generated by human speech is known as phonetics. Pushing air from the lungs to the larynx (respiration) causes the vocal cords to expand to allow air passage or vibrate, generating sound (phonation). The articulators in the mouth and nose, which are responsible for articulation, will influence the airflow in the lungs [1]. Emotion can affect the voice, causing a difference in frequencies that causes a person's speech signal to change depending on whether they are happy, sad, angry, etc. [2], and noise is that thing that can affect speech while there are various noises in life, such as noises in the sonar images [3] or seismic data [4], which it is most important to remove. The same idea in speech noise reduction is now that it should be removed due to its importance in several fields.

Speech signal processing is a subset of digital signal processing that is used in a variety of applications, such as telecommunications, speech recognition, audio communication, multimedia, hearing aids, noise reduction, and more. since noise is part of life. It comes in many types, such as visual noise like in pictures, vibration noise in machinery, engines, and vehicles, or even environmental factors like earthquakes, and sound noise like unwanted sound in audio. As a result, most fields of life now use noise reduction. And background noise is the most prevalent cause of speech quality and intelligibility decline.

Due to the significance of telecommunications, speech noise reduction is an important approach in speech signal processing, and most researchers turned to speech noise reduction to improve the quality and intelligibility of speech signals affected by background noise. A wide categorization of speech noise reduction techniques is offered, like the wiener filter or adaptive filter algorithms. This paper will discuss different speech noise reduction techniques, such as the Log-MMSE (Logarithmic Minimum Mean Square Error) algorithm, which is an extension of the MMSE estimator. To better address the features of speech signals, the original MMSE, which was based on spectral amplitude estimation, was modified to operate in the log-spectral domain. The Log-MMSE filter might produce greater noise suppression and adapt to various noise environments by operating in the log domain. Over the years, the Log-MMSE filter has been extensively explored and modified, and it has become one of the essential strategies in speech enhancement research. First proposed in [5], Log-MMSE has been suggested in [6] to improve speech by removing impulsive noise in speech signals. And it will discuss the DD algorithm, which significantly reduces musical noise, but the estimated prior SNR skews because it is dependent on the assessment of the speech spectrum in the preceding frame. This causes a frame delay, which causes an irritating reverberation effect. For this problem, it will suggest a technique termed Two-Step Noise Reduction (TSNR), which was presented with the DD algorithm in 2006 in [7], for speech enhancement in noisy environments. It is used to refine the estimation of the a priori SNR, which eliminates the

disadvantages of the DD approach while retaining its advantage, namely a greatly decreased musical noise level. The main advantage of this strategy is that it suppresses the frame delay bias, which cancels out the irritating reverberation effect that is distinctive of the DD approach.

However, a significant drawback of traditional short-time suppression methods, such as the TSNR, is that some harmonics are suppressed throughout the noise reduction process because they are regarded as noise-only components [7]. Then, to solve this issue, we will present the Harmonic Regeneration Noise Reduction (HRNR) algorithm, which was proposed in 2013 for two separate loud environments in order to improve speech quality [8], which takes into account the harmonic nature of speech. In this method, the output signal of a standard noise reduction technique (with missing or degraded harmonics) is further processed to generate an artificial signal with automatically regenerated missing harmonics. The manufactured signal is then utilized to compute a suppression gain that preserves all harmonics. This fake signal contributes to the refinement of the a priori SNR used to compute a spectral gain capable of preserving the harmonics of the voice signal [9]. This paper is to produce an approach for high reduction of noise that will operate on different types of noise at different SNRs, but the limitation is that processing is time-consuming and complex. Finally, in this paper, it will operate on NOIZEUS data sets that represent different noise types at different SNRs, such as 5 dB, 10 dB, and 15 dB, and recorded at different frequencies. In the paper's outline, it will discuss the experimental work of the proposed approach in Section 2. Using log-MMSE, DD, TSNR, HRNR, and MMSE, discuss them, while Section 4 will address the MMSE methodologies, and Section 3 will give the results and go into depth about them. Additionally, Section 5 describes the measuring technique utilized in this study. Section 6 presents the results of our experimental work. Section 7 presents and discusses the results in detail. Section 8 presents the conclusion.

2. Related Work

Rapid growth in the use of the Internet in every field [10] has led to the development of many techniques for speech noise reduction. In 2014, researchers presented a study on speech improvement by removing impulsive noise from voice signals. To reduce impulsive disruptions in the voice stream, this research combines log spectral amplitude extraction with MMSE filtering. Regardless of the numerous background disturbances present in the voice, the suggested technique offers efficient outcomes in all applications, where the best result they reached in the SNR was almost equal to 20.4, and the best in the PESQ was equal to 2 in the car noise. The main advantage of this suggested strategy is that it is independent of the speaker [6]. Researchers presented studies in 2019 that attempt to reconcile these two disparate methods for speech improvement. In this work, deep learning techniques for MMSE approaches are explored with the goal of generating high-quality, comprehensible augmented speech. Here, the a priori SNR for the MMSE techniques is properly estimated using a causal ResLSTM and a non-causal ResBLSTM. Real-world, non-stationary, and colored noise sources at various SNR levels are included in the test settings. Compared to current masking and mapping-based deep learning algorithms, MMSE approaches that use the suggested a priori SNR estimator are able to produce better voice quality and intelligibility ratings. The results were tested using MOSLQO for objective quality and STOI for intelligibility of speech signals, where the best result of STOI was almost 0.9. The findings demonstrate that using deep learning considerably improves an MMSE approach's performance [11].

The updated findings in the last few years from other researchers, such as in 2023, when researchers suggested a single-channel speech enhancement framework that reduces speech

signal noise and improves intelligibility by combining particle swarm optimization (PSO), gravitational search algorithm (GSA), and harmonic regeneration noise reduction (HRNR), before employing a TSNR method with harmonic regeneration, used PSOGSA to find the degree of overlap between the noisy voice frames. The measurements used were SNRseg and PESQ. An improvement in input comprehension of speech is indicated by the increase in PESQ that this approach produces. the higher degree of PESQ improvement for babbling and exhibition noise, where the best in the PESQ was equal to 3.1 in the car noise, and in the TSNR, HRNR, and Log-MMSE, the result was in the range of 2.2 to 3.0. A rise in the segmental SNR value indicates that the increased speech quality has improved [12].

For DD, TSNR, and HRNR algorithms in 2017, researchers presented a study that improved the speech enhancement technique. They provide three algorithms for improving speech. The most often used techniques for determining the a priori SNR value are decision-directed (DD), two-step noise reduction (TSNR), and harmonic regeneration noise reduction (HRNR). An a priori estimate of the signal-to-noise ratio (SNR) determines how well a noisy speech augmentation technique performs. The measurement used in this work is SNRseg. The findings demonstrate that when comparing the segmented SNR ratio between improved noisy speech and clean speech, the HRNR methodology performs better than the TSNR method [13]. Researchers will provide a study in 2021. This work provides a unique hybrid speech enhancement strategy based on the combination of comb filters, harmonic regeneration noise reduction (HRNR), and two-step noise reduction (TSNR) for improving speech quality performance. The effectiveness of different enhancement strategies based on TSNR, HRNR, wavelet, and hybrid TSNRHRNR has been compared with the performance of the suggested methodology. The results of the performance study were compared using the following measurements: average segmental SNR (ASSNR), mean square error (MSE), mean opinion score (MOS), perceptual evaluation of speech quality (PESQ), and diagnostic rhyme test (DRT). These measurements demonstrate that the suggested method performs noticeably better in terms of spectrogram. where the best result of the proposed work in the PESQ was equal to 3.19 in 15dB SNR the airport noise, and in the methods TSNR with HRNR the result was reached to from 2.8 to 3 in 15dB SNR. The suggested scheme outperforms the other speech improvement methods taken into consideration in the performance comparison, according to the speech quality assessed in terms of average MOS and PESQ scores [14].

3. Wiener Filter

The Wiener filter was invented by Norbert Wiener in 1940. It was first published in 1949. Its objective is to minimize the amount of noise in a signal by comparing the received signal to an estimate of a desirable noiseless signal [15]. In the 1970s, researchers began investigating the Wiener filter's potential for voice improvement and noise reduction. Ephraim and Malah's (1984) publication [16] is an alternate approach for improving the voice signal to spectral subtraction. The Wiener filter is a linear filter that is used to recover the original speech signal from a noisy signal by reducing the mean square error (MSE) between the estimated and original signals [17].

Assume that a noisy speech signal $x(n)$ is created as a result of background noise that is additive $d(n)$ distorting a clear speech $s(n)$.It may be expressed mathematically, as shown below [18]:

$$x(n) = s(n) + d(n) \tag{1}$$

and the Short-Time Fourier Transform (STFT) of x(n) is,

$$X(p,k) = S(p,k) + D(p,k) \tag{2}$$

We construct SNR estimates using noisy features and utilize them to calculate the spectral gain $G(p,k)$. To estimate $S(p,k)$, we utilize this $G(p,k)$ to $X(p,k)$. The speech enhancement techniques used required the computation of two parameters: $a\ priori\ SNR$ and $a\ posteriori\ SNR$, which were specified as:

$$SNRprio(p,k) = \frac{E[|S(p,k)|2]}{E[|D(p,k)|2]} \quad (3)$$

For the $a\ posteriori\ SNR$

$$SNRpost(p,k) = \frac{|X(p,k)|^2}{E[|D(p,k)|2]} \quad (4)$$

There are several ways for measuring the coefficients of clean speech, one of which is the Wiener filter, which is based on MMSE estimation. The Wiener gain function is:

$$G(p,k) = \frac{E\{|S(p,k)|^2\}}{E\{|D(p,k)|^2\} + E\{|S(p,k)|^2\}} = \frac{\widehat{SNR}_{prio}(p,k)}{1 + \widehat{SNR}_{prio}(p,k)} \quad (5)$$

The calculation of the a $\widehat{SNR}_{prio}(p,k)$, is essential to calculate the $G(p,k)$, is taken into account. The DD method is commonly used to determine $\widehat{SNR}_{prio}(p,k)$. The estimator's behavior was investigated, and it noticed the $priori\ SNR$ of the current frame follows the a $posteriori\ SNR$ of the previous frame. Consequently, the intended behavior of spectral gain was not achieved. The a priori SNR estimate was improved using the $TSNR$ technique. Here, the second step ensures that the DD technique's annoying reverberation effect is eliminated while retaining its ability to reduce the level of musical noise [18].

3.1. Decision Directed (DD)

Ephra m and Malah proposed the (DD) technique in [16], Capp'e investigated this estimator's behavior in [19], and proved that the a priori SNR matches the form of the a posteriori SNR with a frame delay. As a result, because the spectral gain is dependent on the a priori SNR, it does not fit the present frame, and so the noise suppression system's performance suffers [7]. It was used to estimate the time-varying noise spectrum, resulting in improved intelligibility and less musical noise. However, the current frame's a priori signal-to-noise ratio (SNR) estimator is based on the previous frame's predicted speech spectrum. By using adaptive Wiener filtering and the DD technique, it may provide gain. However, the frame delay creates an annoying reverberation effect [20].

The following is the derivation of $SNRprio(p,k)$ based on its definition and relationship to $SNRpost(p,k)$ [18]:

$$SNRprio(p,k) = \frac{E[|S(p,k)|2]}{E[|D(p,k)|2]} \quad (6)$$

Where the $E[|S(p,k)|2] = [|\mathrm{X(p,k)}|^2] - E[|D(p,k)|^2]$

$$SNRprio(p,k) = \frac{[|X(p,k)|^2] - E[|D(p,k)|^2]}{E[|D(p,k)|^2]} = SNR_{post}(p,k) - 1 \quad (7)$$

It may be written by combining equations (7) and (8).

$$SNRprio(p,k) = E\left\{\frac{1}{2}\frac{[|\,\hat{s}\ (p\ -\ 1,k)|^2}{E[|D(p,k)|^2]} + \frac{1}{2}\left[SNR_{post}(p,k) - 1\right]\right\} \quad (8)$$

The suggested estimate $\widehat{SNR}_{prio}(p,k)$ comes from (9), and it is provided by:

$$SNR_{prio}^{DD}(p,k) = \alpha\frac{|\hat{s}\ (p\ -\ 1,k)|^2}{E[|D(p,k)|^2]} + (1-\alpha)P\left[SNR_{post}(p,k) - 1\right] \quad (9)$$
$$for 0 \le \alpha \le 1$$

Where $|\,\hat{s}\ (p\ -\ 1,k)|^2$ is the amplitude estimate of the $(\mathrm{p}-1)^{th}$ frame's k^{th} spectral component, and the function P [.] is defined as:

$$P[x] = \begin{cases} x & if\ x \geq 0 \\ 0 & otherwise \end{cases} \quad (10)$$

This $a\ priori\ SNR$ estimator as given in equation (10) corresponds to the DD approach and is denoted as $S\widehat{N}R_{prio}^{DD}(p,k)$ The behavior of $S\widehat{N}R_{prio}^{DD}(p,k)$ is controlled by α, a parameter whose typical value is 0.98. $G(p,k)$ in equation (6) was chosen to be the Wiener filter, which results in [18]:

$$G_{DD}(p,k) = \frac{S\widehat{N}R_{Prio}^{DD}(p,k)}{1+S\widehat{N}R_{Prio}^{DD}(p,k)} \quad (11)$$

3.2. Two-Step Noise Reduction (TSNR)

The Two-Step Noise Reduction (TSNR) approach used to eliminate delay and reverberation caused by DD approach while retaining the advantages of the decision-directed approach. Furthermore, one important limitation of traditional short-time suppression approaches, including the TSNR, is that some harmonics are treated as noise only components and are thus suppressed by the noise reduction process. This is due to the intrinsic inaccuracies induced by noise spectrum estimation, which is a tough problem for single channel noise reduction algorithms. It is worth noting that in most spoken languages, voiced sounds account for a high proportion of the pronounced sounds ($about$ 80%). Then it becomes quite intriguing to overcome this constraint [13]. Then implemented the $TSNR$ approach, which is a two-step method for calculating the a priori SNR. When the parameter α was set to 0.98 in the DD algorithm, musical noise was considerably decreased. Since we didn't want to interfere with the process of eliminating musical noise, one of the two procedures was exactly the same as the DD technique. The next step has to be able to removing the delay that created the issues mentioned as a drawback of the DD approach. As a result, the spectral gain estimated in the first step for the frame $(p+1)^{th}$ is applied to the p^{th} frame of noisy speech to obtain the enhanced p^{th} frame. The two stages are denoted mathematically as [18]:

$$S\widehat{N}R_{Prio}^{TSNR}(p,k) = S\widehat{N}R_{Prio}^{DD}(p+1,k)$$

$$S\widehat{N}R_{Prio}^{TSNR}(p,k) = \acute{\beta}\ \frac{|G_{DD}(p,k)X(p,k)|^2}{\hat{\gamma}_n(p,k)} + (1-\acute{\beta})\ P\ [S\widehat{N}R_{post}(p+1,k) - 1] \quad (12)$$

where the task of $\acute{\beta}$ is the same as that of α, but it might select a different value. it can see that in order to compute $SNRpost\ (p+1,k)$, information of $X\ (p+1,k)$, It is necessary to use the frame that follows, introducing an extra delay. As a result, it opted for $\acute{\beta} = 1$.
Equation (11) is now modified as follows:

$$S\widehat{N}R_{Prio}^{TSNR}(p,k) = \frac{|G_{DD}(p,k)X(p,k)|^2}{\hat{\gamma}_n(p,k)} \quad (13)$$

We will avoid wasting more processing time because the information about the next frame is not needed. Furthermore, the initial phase guarantees that the amount of musical noise is kept as low as possible using the DD technique. Finally, Wiener filtering was applied to determine the gain as follows:

$$G_{TSNR}(p,k) = \frac{S\widehat{N}R_{Prio}^{TSNR}(p,k)}{S\widehat{N}R_{Prio}^{TSNR}(p,k)} \quad (14)$$

To approximate the clean speech spectrum, the gain will be multiplied by the noisy speech spectrum.

$$\hat{S}(p,k) = G_{TSNR}(p,k)\ X(p,k) \quad (15)$$

3.3. Harmonic Regeneration Noise Reduction (HRNR)

The Harmonic Regeneration Noise Reduction (HRNR) approach considers the harmonic features of speech. In this method, the output signal of a standard noise reduction technique (with missing or degraded harmonics) is further processed to generate an artificial signal with automatically regenerated missing harmonics. The manufactured signal is then utilized to compute a suppression gain that preserves all harmonics [9]. The result of the TSNR technique is then employed in the HRNR method. The distortions are present in DD and TSNR. It was found that most of the aberrations were harmonic in character. In fact, numerous of the harmonics were suppressed by the algorithms since they were considered noise-only components. In order to prevent error, we analyzed the distorting signal and produced a fake signal that included the removed harmonics from the deformed signal with a response to frequency similar to a harmonic comb. Using the simulated signal, a spectral gain that might restore harmonics was calculated. [18].

Applying a nonlinear function to the time domain signal $\hat{s}$(t) makes it easy to complete this step, as demonstrated in:

$$S_{harmo}\,(t)\;=\;N\,L\,(\,\hat{s}(t)) \tag{16}$$

It is obvious that $S_{harmo}(t)$ harmonics will appear at the same locations as clear speech, but at prejudiced amplitudes. Therefore, it was only applied to raise the $priori\ SNR$:

$$\widehat{SNR}_{Prio}^{HRNR}(p,k) = \frac{\rho(p,k)\left|\hat{S}(p,k)\right|^2 + \big(1-\rho(p,k)\big)|S_{harmo}(p,k)|^2}{\hat{\gamma}_d(p,k)}, \rho(p,k) = G_{TSNR}(p,k) \tag{17}$$

$\widehat{SNR}_{Prio}^{HRNR}(p,k)$ was then used to compute a gain capable of preserving harmonics. Because the harmonics that were removed by the previous enhancing speech strategy are restored, the recreated speech after HRNR has all of the harmonics that were removed by the previous speech enhancement approach, so it sounds normal. The following formula is used to compute spectral gain:

$$G_{HRNR}(p,k)\;=\frac{\widehat{SNR}_{Prio}^{HRNR}(p,k)}{1+\widehat{SNR}_{Prio}^{HRNR}(p,k)} \tag{18}$$

and $\hat{S}(p,k)$ was calculated as:

$$\hat{S}\,(p,k) = G_{HRHR}\,(p,k)\,X(p,k) \tag{19}$$

4. Minimum Mean Square Error (MMSE)

MMSE can be used in a variety of ways, depending on the context. Here are a few typical types:

4.1. Linear Minimum Mean Square Error (MMSE)

By reducing the error between a linear model of the clean spectrum and a true spectrum, the Wiener estimator may be obtained. In terms of mean-square error, the Wiener estimator is thought to be the ideal complex spectrum estimator, although it's not the best one for estimating spectral magnitude. Several researchers have suggested the best techniques for extracting the spectral amplitudes from noisy data, acknowledging the significance of the short-time spectral amplitude (STSA) on speech intelligibility and quality. In contrast to the Wiener estimator, the MMSE estimator does not presuppose a linear relationship between the estimator and the observed data, but it does require an understanding of the probability distributions of the noise and speech DFT coefficients. assuming that the distributions of the speech and noise DFT coefficients were known to us previously.

MMSE is a fundamental concept in signal processing and estimation theory. The MMSE optimization criteria is frequently utilized in signal processing and estimation theory. It is a statistical approach for estimating an unknown signal based on a collection of observations,

with the goal of minimizing the mean square error between the real and estimated signals [21].
The MMSE gain function $G_{MMSE}(p,k)$ [22]:

$$G_{MMSE}(p,k) = \frac{\sqrt{\pi}}{2}\frac{\sqrt{v_k}}{\gamma_k}\exp\left(-\frac{v_k}{2}\right)\left[(1+v_k)I_0\left(\frac{v_k}{2}\right) + v_k I_1\left(\frac{v_k}{2}\right)\right] \quad (20)$$

Where the terms I_0, I_1 represent the Bassel function and the ξ_K and γ_k is are referred to as the $a\ prior\ SNR$ and $a\ posterior\ SNR$, respectively.

$$\lambda_k = \frac{\lambda_x(k)}{1+\xi_K}, v_k = \frac{\xi_K}{1+\xi_K}\gamma_k \quad (21)$$

$$\hat{S}(p,k) = G_{MMSE}(p,k)\,X(p,k) \quad (22)$$

4.2. Logarithmic of Minimum Mean Square Error (Log-MMSE)

The Logarithmic of Minimum Mean Square Error (Log-MMSE) method is a frequently used speech enhancement approach that tries to improve the quality and intelligibility of noise-corrupted audio signals. It is a more advanced variant of the MMSE estimator that operates in the log-spectral domain. It was proposed for the first time in 1984 by Ephraim, Y., and Malah, D., which was the first to propose the MMSE estimator for speech enhancement. It was subsequently extended to the log-spectral domain to address the limitations of the traditional MMSE estimator when dealing with speech signals [5]. The MMSE spectral amplitude estimator reduces the error in the spectral magnitude spectra. Although a measure based on the magnitude spectra's squared error is technically tractable, it may not be subjectively meaningful. A measure based on the squared error of the log-magnitude spectra has been proposed as better suited for voice processing. The following step is to develop an estimator that minimizes the mean square error of the log-magnitude spectra [22].

$$E\{(\log S - \log \hat{S})^2\} \quad (23)$$

The optimal log-MMSE estimator may be found by calculating the conditional mean of the log S, which is as follows:

$$\log \hat{S} = E\{\log S \mid X(\omega_k)\} \quad (24)$$

It allows us to calculate $\hat{S}$:

$$\hat{S} = exp\,(E\{\log S \mid X(\omega_k)\}) \quad (25)$$

The evaluation of $E\{\log S \mid X(\omega_k)\}$ is not simple; however, it may be made easier by using the moment-generating function of S conditioned on (ω_k). then $E\{\log S \mid X(\omega_k)\}$ will be as follows:

$$E\{\log S \mid X(\omega_k)\} = \frac{1}{2}\log\lambda_k + \frac{1}{2}\log v_k + \frac{1}{2}\int_{v_k}^{\infty}\frac{e^{-t}}{t}dt \quad (26)$$

To obtain the optimal log-MMSE:

$$\hat{S}(p,k) = \frac{\xi_K}{\xi_K+1}exp\left\{\frac{1}{2}\int_{v_k}^{\infty}\frac{e^{-t}}{t}dt\right\} Y_k \triangleq G_{LSA}(p,k)\,X(p,k) \quad (27)$$

Where $G_{LSA}(p,k)$ is the gain function of the log-MMSE estimator, and (LSA) is log-spectral amplitude [18].

5. Signal-To-Noise Ratio (SNR) Estimator

The Short-Time Silence of Speech Signal as Signal-to-Noise Ratio Estimator (STS-SNR) has been used in this work. It was proposed in 2016 by [23], and they approved that it is the

best estimator for SNR among different types of estimators. This approach will only evaluate the first 30 ms of the audio stream to predict the SNR by processing a limited number of samples from the audio signal. This estimator assumes that the first 30 milliseconds of the tested audio are silent rather than speech. This study also assumes that the SNR does not change during the time of interest.

Firstly, take the first 30 ms, which is referred to as the noise frame N_{Frame}here.

Use the Fast-Fourier Transform (FFT) of 512 points to estimate the power spectral density (PSD) of the N_{Frame} while only accounting for the 0–8 kHz band. Where the PSD for the NPSD of 30 ms is [23]:

$$N_{PSD} = |N_{Frame}(\omega)|^2 \tag{28}$$

Where $N_{Frame}$is the spectrum of the audio frame.

Reformat the PSD using the following steps to create a white-like PSD:

$$N_{Reformed} = N_{PSD} - N_{PSD}^{T} \tag{29}$$

Where N_{PSD}^{T} is the flipped version of the noise power spectral density vector N_{PSD}.

In decibels, the estimated SNR is:

$$SNR_{db} = offset - 10\, log_{10}\,(\widehat{N}_{PSD}) \tag{30}$$

6. Experimental work

Noise reduction is an appealing subject for researchers to investigate. There is also a vast category of speech noise reduction techniques available, and many articles have worked to enhance noise reduction approaches and find a new method. In 2006, Cyril Plapous, Claude Marro, and Pascal Scalart produced an article about how to improve SNR by utilizing the DD algorithm, TSNR, and HRNR, and the results demonstrate the good performance of these approaches [7]. And in 2016, Siddala V. and others produced articles that utilized spectral subtraction and the Wiener filter with DD, TSNR, and HRNR algorithms, whose results were respectable and reached a good SNR [18]. This paper will be an improvement of this method (DD, TSNR, and HRNR) to get more noise reduction and improve SNR in terms of stationary noise at different noise levels and different types of noise. In this paper, we choose the best arrangement of filters that gives us the highest SNR. The DD approach for estimating the prior SNR was important for the MMSE-type (linear (MMSE) and log (MMSE)) algorithms. The origin of the DD approach is the Wiener filter, which assumes that there is a linear relationship between the filtered signal and the distorted signal, while the MMSE expresses that the filtered signal is the expectation value respected by the joint pdf. Based on those ideas, the MMES filter was chosen to improve the filtering of DD, TSNR, and HRNR.

The block diagram shows an overview of the concept employed in this study:

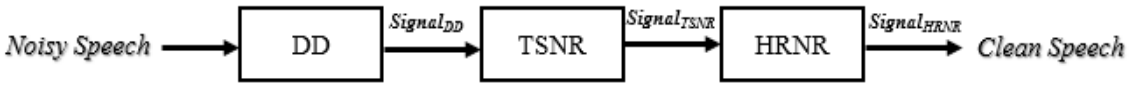

Figure -1 First arrangement

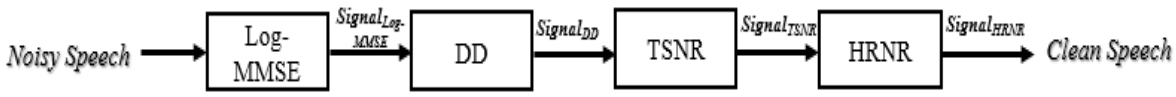

Figure -2 Second arrangemen

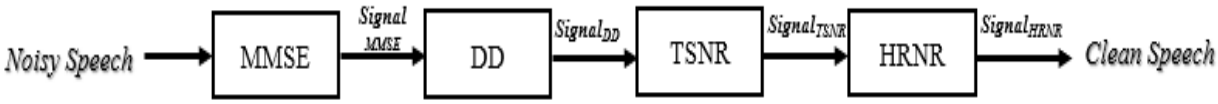

Figure -3 :Third arrangement

7. Results and Discussion

In this study, we will operate on NOIZEUS [24], data sets representing various noise types at various SNRs, such as 5 dB, 10 dB, and 15 dB. And captured at various frequencies is a publicly accessible noisy voice corpus used to compare speech improvement methods. Three male and three female speakers deliver 30 phonetically balanced IEEE English statements. The phrases are contaminated with one of six typically occurring real-world noises: babbling, automobile, street, train, restaurant, and airport. The sounds are from the AURORA database. The sentences were recorded at 25 kHz and then down-sampled to 8 kHz. Each statement lasts 3 seconds on average. WAV files are used to save all sample files [25]. To demonstrate the results and performance of the developed algorithms, wave signals and spectrograms are displayed after the proposed method, as shown in Figures 4–8. Figure 9 presents a compression of the used measurements in different SNRs.

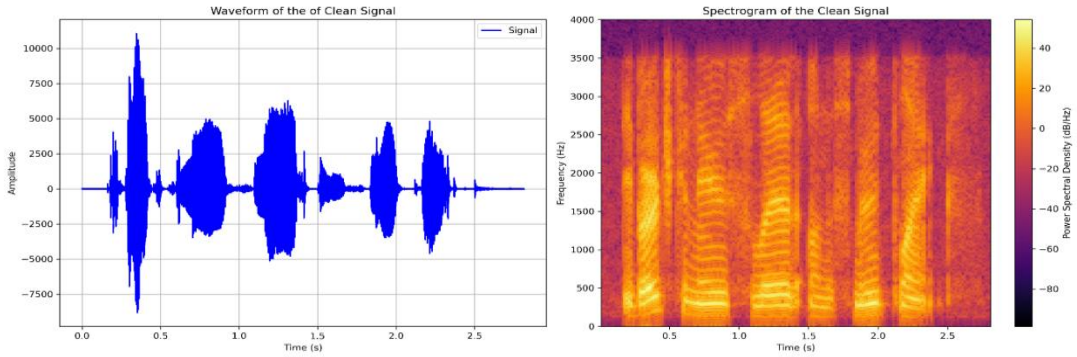

Figure -4 The time history and spectrogram of the true clean speech signal

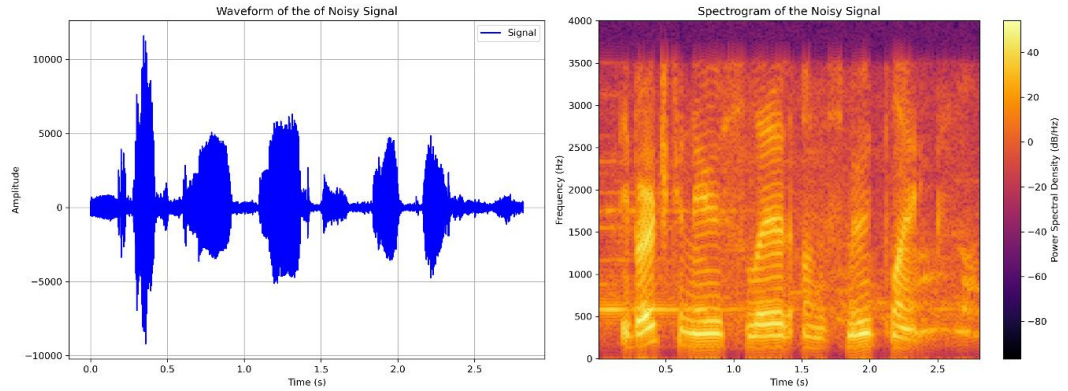

Figure -5 The time history and spectrogram of the noisy speech signal

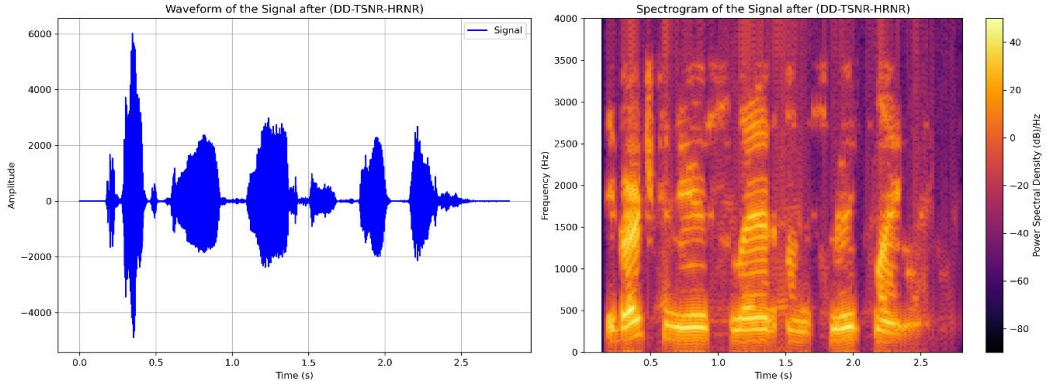

Figure -6 : The time history and spectrogram of the clean speech signal after filtering by (DD+TSNR+HRNR)

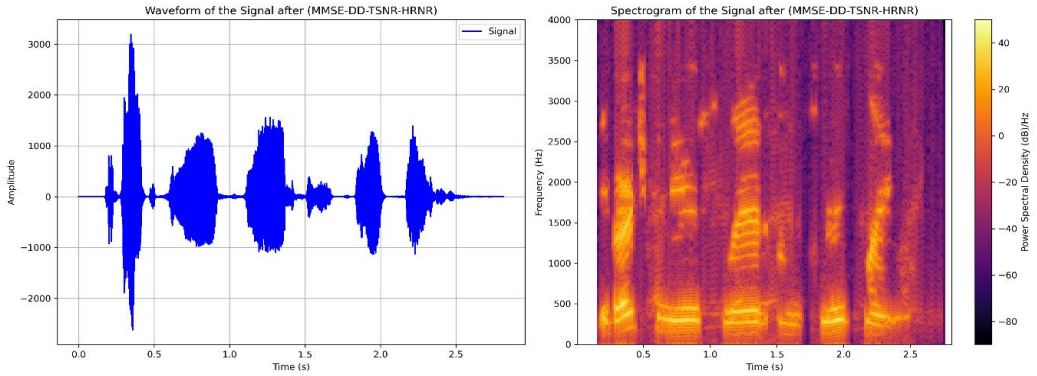

Figure -7 : The time history and spectrogram of the clean speech signal after filtering by (MMSE+DD+TSNR+HRNR)

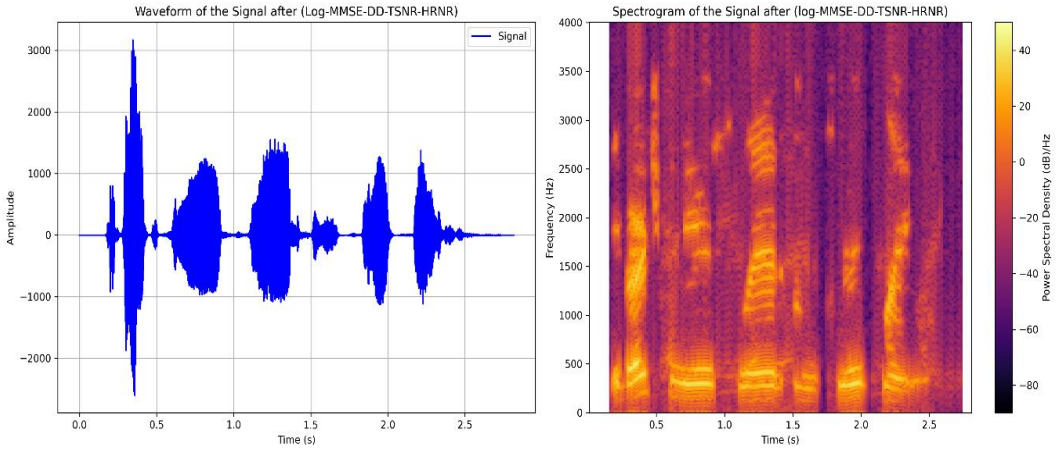

Figure -8 The time history and spectrogram of the clean speech signal after filtering by (Log-MMSE+DD+TSNR+HRNR)

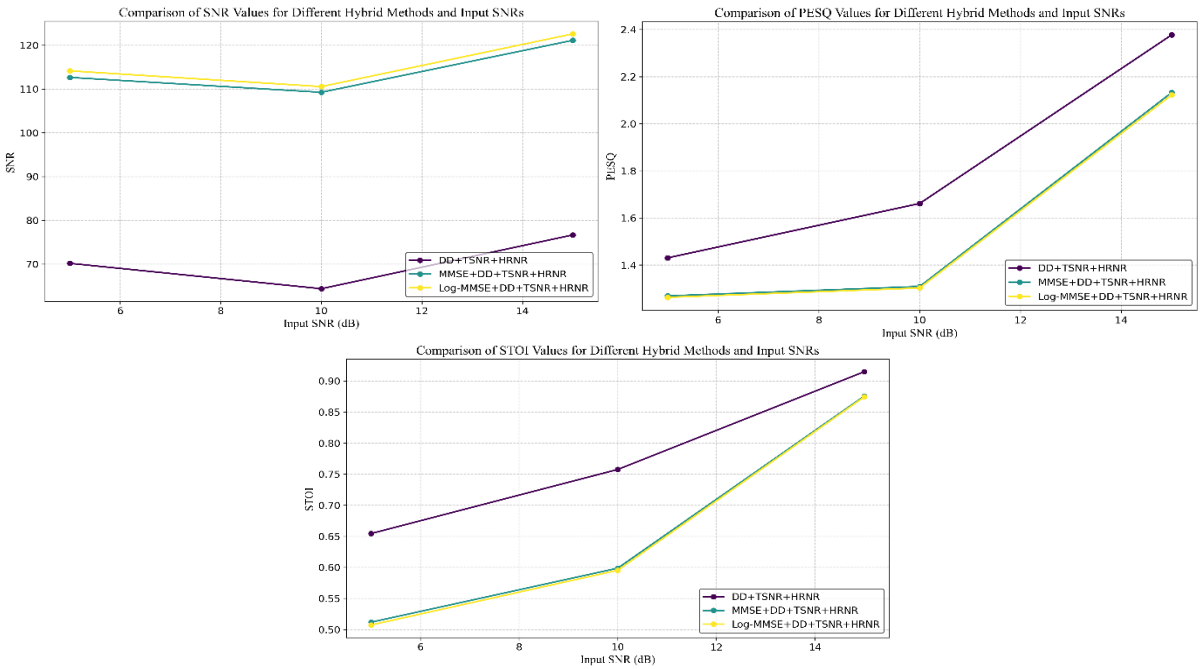

Figure -9 : Comparison of SNR, PESQ, and STOI Values for Different Hybrid Methods and Input SNRs

In this section, the time history and spectrogram of the true clean speech signal and the noisy speech signal are presented in Figures 4 and 5. After applying the different arrangements of filters on the noisy speech signal in Figure 6, it was noticed that the filtering arrangement of Figure 1 (DD+TSNR+HRNR) made a good filtering on the noisy speech signal. It was observed that the noise has been significantly reduced, especially in the silent time segments at the beginning, between speech intervals, and at the end of the speech record. In Figure 7, it was noticed that the filtering of the arrangement of Figure 2 (MMSE+DD+TSNR+HRNR) made better filtering on the noisy speech signal from the previous arrangement. It was observed that the noise had been significantly reduced. Lastly, in Figure 8, it was noticed that the filtering of the arrangement of Figure 3 (Log-MMSE+DD+TSNR+HRNR) made excellent filtering on the noisy speech signal. From the whole last arrangement in Figure 7 and Figure 6, it was observed that the noise has been significantly reduced, almost to zero noise intensity, especially in the silent time segments beginning, between speech intervals, and at the end of the speech record. Also, to demonstrate the results and performance of the developed algorithms, tables have been made for three measurements (SNR, PESQ, and STOI) for each arrangement of filters for five types of noise with three different levels of input SNR, as shown in Tables 1, 2, and 3.

Table 1: SNR for each arrangement of filters for five types of noise with three different level of input SNR

No.	Type	Input SNR(dB)	SNR(dB) DD+TSNR+HRNR	SNR(dB) MMSE+DD+TSNR+HRNR	SNR(dB) Log-MMSE+ DD +TSNR+HRNR
1	Airport	5	56.2481	91.1375	92.2163
		10	64.4950	106.8280	108.2783
		15	48.3529	78.9705	79.3962
2	Babble	5	52.0281	89.0521	105.7713
		10	62.8655	101.1710	102.5076
		15	64.7819	103.7125	104.6406
3	Car	5	70.1645	112.6568	114.1564
		10	64.3672	109.2560	110.5603
		15	76.6575	121.1384	122.6040
4	Restaurant	5	49.1617	82.6512	83.5945
		10	62.4669	100.4130	101.5889
		15	61.2998	97.4559	98.4092
5	Street	5	44.1863	73.0636	73.5617
		10	70.4265	104.0540	105.0860
		15	71.7137	109.7828	111.0962

In Table 1, it is observed that the SNR increases significantly through the arrangements (DD+TSNR+HRNR), (MMSE+DD+TSNR+HRNR), and (Log-MMSE+DD+TSNR+HRNR). We did not rely just on the numbers in the results; we also examined them through hearing, by listening to the results of each filtered file, observing the clarity and understanding of the output voice, and comparing them with each other. and we noticed that each of the presented algorithms improved the SNR and the quality of voice, but when they were combined, the best result was obtained, which is the proposed method. The best arrangement is the last one, which collects the four filters (Log-MMSE+DD+TSNR+HRNR) in this paper. The most improvement in SNR was in car noise, followed by street noise, and it can be noticed that the SNR of the output signal is improving concurrently with the value of the input SNR. The best performance for each arrangement is with the input 15 SNR for car and street noise, where for other noises it is different in each input SNR, where 10 SNR is best for airport, babble, and restaurant noise.

Table 2: PESQ for each arrangement of filters for five types of noise with three different levels of input PESQ

No.	Type	Input SNR(dB)	PESQ DD+TSNR+HRNR	PESQ MMSE+DD+TSNR+ HRNR	PESQ Log-MMSE+ DD +TSNR+HRNR
1	Airport	5	1.4308	1.2692	1.2633
		10	1.6616	1.3093	1.3031
		15	2.3772	2.1325	2.1225
2	Babble	5	1.4814	1.2776	1.2383
		10	1.7780	1.3807	1.3726
		15	2.2915	1.7353	1.7239
3	Car	5	1.3432	1.2242	1.2290
		10	1.7471	1.3528	1.3501
		15	2.0439	1.5367	1.5280
4	Restaurant	5	1.5505	1.3388	1.3320
		10	1.7459	1.3741	1.3682
		15	2.2873	1.7966	1.7852
5	Street	5	1.5761	1.3322	1.3267
		10	1.4445	1.2526	1.2490
		15	2.0939	1.5886	1.5787

Table 3: STOI for each arrangement of filters for five types of noise with three different levels of input STOI

No.	*Type*	*Input SNR(dB)*	*STOI DD+TSNR+HRNR*	*STOI MMSE+TSNR+HRNR*	*STOI Log-MMSE+ DD +TSNR+HRNR*
1	Airport	5	0.6544	0.5118	0.5073
		10	0.7575	0.5987	0.5953
		15	0.9148	0.8756	0.8743
2	Babble	5	0.6499	0.5244	0.4888
		10	0.7684	0.6093	0.6058
		15	0.8840	0.7704	0.7678
3	Car	5	0.5929	0.4471	0.4413
		10	0.7799	0.6256	0.6218
		15	0.8622	0.7238	0.7209
4	Restaurant	5	0.7410	0.6073	0.6039
		10	0.7915	0.6387	0.6357
		15	0.8949	0.7813	0.7787
5	Street	5	0.7351	0.6077	0.6045
		10	0.6415	0.5103	0.5066
		15	0.8655	0.7308	0.7278

In Table 2 and Table 3, it is observed that the PESQ and STOI do not increase significantly through the arrangements, but there is no huge difference between airport noise results. The best results for the PESQ extract from the (DD+TSNR+HRNR) arrangement were 2.3772 airport noise in 15 SNR, followed by (MMSE+DD+TSNR+HRNR) and (Log-MMSE+DD+TSNR+HRNR), where the PESQ was very close. Lastly, the best results for the STOI extract from the (DD+TSNR+HRNR) arrangement were 0.9148 in airport noise at 15 SNR, followed by (MMSE+DD+TSNR+HRNR) and (Log-MMSE+DD+TSNR+HRNR), where the STOI was very close to 0.8756 and 0.8743 in airport noise at 15 SNR.

8. Conclusion

This paper aims to improve this method (DD, TSNR, and HRNR) to get more noise reduction and improve SNR in terms of stationary noise at different noise levels (5, 10, and 15 dB) and different types of noise (airport, babble, car, restaurant, and street). the proposed work in terms of complexity using a hybrid method like DD, TSNR, and HRNR with Log-MMSE to improve the SNR It was a complex procedure to implement and made a limitation like the time of processing, but in terms of performance, our proposed work result gives the best result compared to each method if it was used alone and compared to the other hybrid methods that it presented in this work. The arrangement of filters that gives the highest SNR was chosen as the best arrangement. This article introduces a hybrid approach for speech noise reduction using the Log-MMSE (Logarithmic Minimum Mean Square Error) algorithm compared to the MMSE (Minimum Mean Square Error) algorithm. It can be concluded from the previous results:

Using Log-MMSE in conjunction with the DD, TSNR, and HRNR algorithms considerably improves performance.

1- The arrangement (DD+TSNR+HRNR) gives a good increase in the SNR measurement.
2- The arrangement (MMSE+DD+TSNR+HRNR) gives a very good increase in the SNR measurement.
3- The arrangement (Log-MMSE+DD+TSNR+HRNR) compared with other arrangements gives the best increase in the SNR measurement.

4- The best PESQ value can be obtained from the (DD+TSNR+HRNR) arrangement.
5- The best STOI value can be obtained from the (DD+TSNR+HRNR) arrangement.
6- Considerable rise in the ratio of the SNR, where our improved parentage reached 63.49% in the SNR results compared with the (DD+TSNR+HRNR) arrangement.
7- The output SNR is increasing concurrently with the SNR of the input signal, and the best performance of the filter's arrangement was when the SNR was equal to 15 in (Car, Street), where in other noises the best result was 10 SNR.

Furthermore, when compared to existing approaches, the improved speech produced by the suggested method is more effective than other methods. And as a future work, it may be used with deep learning or transfer learning to improve intelligibility.

9. Acknowledgements

We extend our sincere gratitude to all researchers in the field of noise reduction for the huge efforts they have given in this field.

References

[1] H. A. Abdulmohsin, B. Al-Khateeb, S. S. Hasan, and R. Dwivedi, "Automatic illness prediction system through speech," *Computers and Electrical Engineering*, vol. 102, p. 108224, July. 2022, DOI: https://doi.org/10.1016/j.compeleceng.2022.108224.

[2] H. A. Abdulmohsin, "A new proposed statistical feature extraction method in speech emotion recognition," *Computers & Electrical Engineering*, vol. 93, p. 107172, July. 2021, DOI: https://doi.org/10.1016/j.compeleceng.2021.107172.

[3] N. H. Resham, H. K. Abbas, H. J. Mohamad, and A. H. Al-Saleh, "Noise reduction, enhancement and classification for sonar images," *Iraqi Journal of Science*, vol. 62, no. 11, pp. 4439-4452, 2021, DOI: 10.24996/ijs.2021.62.11(SI).25.

[4] A. H. Ali and A. M. Al-Rahim, "Linear Noise Removal Using Tau-P Transformation on 3D Seismic Data of Al-Samawah Area-South West of Iraq," *Iraqi Journal of Science*, vol. 60, no. 12, pp. 2664-2671, 2019, DOI: 10.24996/ijs.2019.60.12.16.

[5] Y. Ephraim and M. David, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE transactions on acoustics, speech, and signal processing*, vol. 33, no. 2, pp. 443-445, April. 1985, DOI: 10.1109/TASSP.1985.1164550.

[6] D.A. Suganthi, "Improved speech enhancement by removal of impulsive noise," in *2014 IEEE 8th International Conference on Intelligent Systems and Control (ISCO)*, Jan. 2014: IEEE, pp. 145-148, DOI: 10.1109/ISCO.2014.7103934.

[7] C. Papous, 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, pp. 2098-2108, November. 2006, DOI: 10.1109/TASL.2006.872621.

[8] P. Gael, M. Chandra, P. Saxena, and V. K. Gupta, "Comparative analysis of speech enhancement methods," in *2013 Tenth International Conference on Wireless and Optical Communications Networks (WOCN)*, Bhopal, India, July. 2013: IEEE, pp. 1-5, DOI: 10.1109/WOCN.2013.6616238.

[9] C. Plapous, C. Marro, and P. Scalart, "Speech enhancement using harmonic regeneration," in *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing.*, Philadelphia, PA, USA, March. 2005, vol. 1: IEEE, pp. 157-160 DOI: 10.1109/ICASSP.2005.1415074.

[10] N. Abdullah and M. Abduljaleel, "Adaptive medical image watermarking technique based on wavelet transform," *Iraqi journal of science*, vol. 55, no. 2A, pp. 548-555, 2014.

[11] A. Nicolson and K. K. Paliwal, "Deep learning for minimum mean-square error approaches to speech enhancement," *Speech Communication*, vol. 111, pp. 44-55, 2019, DOI: https://doi.org/10.1016/j.specom.2019.06.002.

[12] K. Ghorpade and A. Khaparde, "Single-channel speech enhancement by PSO-GSA with harmonic regeneration noise reduction," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 5, pp. 2895-2902, 2023, DOI: 10.11591/eei.v12i5.5373.

[13] A. Akash and K. Rajesh, "Adaptive Weiner filter for speech enhancement under various noisy conditions," *International Journal of Computer Applications*, vol. 170, no. 7, pp. 9-11, July. 2017, DOI: 10.5120/ijca2017914912.
[14] I. Bahadur, S. Kumar, and P. Agarwal, "Performance measurement of a hybrid speech enhancement technique," *International Journal of Speech Technology*, vol. 24, pp. 665-677, 2021, DOI: https://doi.org/10.1007/s10772-021-09830-2.
[15] N. Wiener, *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. MIT Press Direct, 1949, p. 163
[16] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 32, no. 6, pp. 1109-1121, December. 1984, DOI: 10.1109/TASSP.1984.1164453.
[17] M. A. Abd El-Fattah, M I. Dessouky, A M. Abbas, S M. Diab, E M. El-Rabaie, W. Al-Nuaimy, S A. Alshebeili and F E .Abd El-samie. , "Speech enhancement with an adaptive Wiener filter," *International Journal of Speech Technology*, vol. 17, no. 1, pp. 53-64, 2014, DOI: 10.1007/s10772-013-9205-5.
[18] S. Vihari, A. S. Murthy, P. Soni, and D. Naik, "Comparison of speech enhancement algorithms," *Procedia computer science*, vol. 89, pp. 666-676, 2016, DOI: https://doi.org/10.1016/j.procs.2016.06.032.
[19] O. Cappé, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE transactions on Speech and Audio Processing*, vol. 2, no. 2, pp. 345-349, April. 1994, DOI: 10.1109/89.279283.
[20] A. Ouardia and F. Merazka, "Denoising of Speech Signal Using Decision Directed Approach," *International Journal of Informatics and Applied Mathematics*, vol. 3, no. 1, pp. 70-83, June. 2020.
[21] T. Trainer. "MMSE minimum mean square error." Telecom Trainer. https://www.telecomtrainer.com/mmse-minimum-mean-square-error-4/ (accessed May 4, 2023).
[22] P. C. Loizou, *Speech enhancement: theory and practice*, 1st Edition ed. Boca Raton: CRC press, 2007, p. 632.
[23] A. S. Abdulaziz and V. Z. Këpuska, "The short-time silence of speech signal as signal-to-noise ratio estimator," *International Journal of Engineering Research and Applications (IJERA)*, vol. 6, no. 8, pp. 99-103, 2016.
[24] Y. Hu and P. C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," *Speech communication*, vol. 49, no. 7-8, pp. 588-601, August. 2007, DOI: https://doi.org/10.1016/j.specom.2006.12.006.
[25] R. a. D. R. Jaiswal, "Implicit wiener filtering for speech enhancement in non-stationary noise," in *2021 11th International Conference on Information Science and Technology (ICIST)*, Chengdu, China, May. 2021: IEEE, pp. 39-47, DOI: 10.1109/ICIST52614.2021.9440639.