International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Speech Noise Reduction via Intelligent Spectral Gain Selection and Modification

Huiting Yu[1], Anton Louise De Ocampo*[2], Rowell Hernandez[3]

Submitted:13/03/2024 **Revised**: 28/04/2024 **Accepted**: 05/05/2024

Abstract*:* Speech information is distributed across the frequency band of 300 Hz to about 3000-4000 Hz. Noise affects speech information distinctively between different spectral components of the speech signal. This means that noise infusion to individual spectral components is variable when measured in signal-to-noise ratio. This variation can be picked up and efficiently discerned by the workings of the human brain. The human brain can perform spectral selection and gain modification to address the various effects of noise in a speech signal. When faced with noise, the brain combines top-down and bottom-up processing to improve the signal-to-noise ratio of the speech signal. The brain's ability to select spectra with more information is not utilized in existing or traditional methods. This paper proposes a novel deep neural network-based method for voice noise reduction. The formulations are determined by attention-based neural networks using spectral gain adjustments as the basis for the proposed technique. The intelligent spectral gain selection and modification using attention mechanisms are introduced after speech signal preprocessing. The noisy signal passes through spectral decomposition where each component has assigned weights based on the proposed attention network. The work shows how the addition of spectral gain adjustments affects the suppression of noise in voice signals by an attention-network-based algorithm. The attention network can focus on the spectral components that are most informative by suppressing bands where noise is excessive. Based on the identified weights, the spectral component is attenuated or amplified to extract the most information content possible. The proposed framework obtained an SNR and segmented SNR of 7.1138 and 1.9950 respectively, higher than existing methods.

Keywords: *attention mechanism, deep learning, noise suppression, speech processing*

1. Introduction

Background or ambient noise and other factors can all make it difficult to discern the information content of a speech signal. But still, the human brain can extract snippets of information from a noise-infused speech signal. Noise often affects the whole spectrum of the speech signal. The objective is to mimic the human brain's ability in speech noise reduction to computers and machines. In contrast with the traditional filtering techniques, brain-inspired speech processing algorithms may be significantly more effective at suppressing noise. One of the most dominant phenomena is the auditory masking process attributed to the human brain's capability to reduce the effect of noise in speech signals.

The auditory masking phenomenon happens when a sound partially or completely obscures another sound. Wang & Xu demonstrated the effect of this phenomenon by studying speech perception in noisy environments, such as the "cocktail party problem", where they synthesized existing findings on the factors that release speech perception from masking in both English and Mandarin Chinese [1]. The authors compare and contrast the mechanisms that allow listeners to "unmask" and focus on a target speech signal despite competing auditory information. The "other sound" masking another is either of the two languages. The study highlights the potential language-specific influences on the auditory process. Auditory models mimic the human capability to mask sounds to suppress noise in speech signals. A trade-off between noise reduction and speech distortion characterizes the subtraction process in auditory masking models.

In another paper, [2] proposes a hypothesis of harmonic cancellation as a fundamental mechanism underlying auditory scene analysis. According to this hypothesis, an interfering sound is suppressed or canceled based on its harmonicity (or periodicity in the time domain) to focus on a target sound. Another approach to noise reduction in the speech signal is the Perceptual Wavelet Packet Decomposition or Transform which combines wavelet packet decomposition with a perceptual filter bank [3]. The basis of this method is the human auditory system's ability to discriminate between unwanted noise and signal according to the characteristics of the former. To distinguish the information from the noise speech utterances are converted into wavelet entropy characteristics but not into distinct components. Non-negative matrix factorization (NMF), on the other hand, can introduce this separation by dividing a signal into components suitable for non-

[1] College of Engineering, Batangas State University, Batangas City, 4200, PHILIPPINES.
ORCID ID : 0009-0001-1334-1758
[2] College of Engineering, Batangas State University, Batangas City, 4200, PHILIPPINES.
ORCID ID : 0000-0002-6280-6259
[3] College of Informatics and Computer Sciences, Batangas State University, Batangas City, 4200, PHILIPPINES.
ORCID ID : 0000-0002-8748-6271
** Corresponding Author Email: antonlouise.deocampo@ieee.org*

stationary noise which is frequently present in speech signals [4]. This technique can then be used in voice processing to separate noise from speech signals like how the human auditory system separates speech from background noise. Another interesting approach, inspired by the human auditory system, uses Artificial Neural Networks to detect information-carrying parts of the speech signal from the noise-infused components. When sufficiently trained, ANN can recognize patterns in speech signals and differentiate speech from noise [5].

When these approaches are combined with parametric optimization using evolutionary algorithms, such as genetic algorithms, the noise suppression capabilities of the mentioned approaches are improved. Positive optimization results of feed-forward ANNs, i.e. Extreme learning machines – ELM, show successful noise reduction in speech signals when parameters are optimized. Evolutionary techniques in optimization perform well when applied to parameterized Active Noise Reduction techniques, which are frequently used in Automatic Speech Recognition (ASR) systems. The process of natural selection, which may be used to identify the best solution to a problem, served as the model for this approach [6].

In speech applications, conventional filtering and noise suppression techniques also performed satisfactorily. For instance, the signal processing technique known as Weiner Harmonic Regeneration Noise Reduction can enhance the quality of noisy signals, especially in audio and speech applications [7]. It combines harmonic regeneration which seeks to return the signals harmonic structure with the Wiener filter a statistical method for estimating the original signal from a noisy observation. The Wiener filter minimizes the mean square error between the estimated and actual signals thereby reducing noise [8]. By reconstructing lost harmonics harmonic regeneration improves the signal's perceptual quality and produces a clearer more natural-sounding voice or audio. The use of digital biquad filters for the noise reduction method is an additional strategy that uses a kind of second-order filter made up of the ratio of two quadratic functions to reduce signal noise [9]. Such approaches can target specific noise frequencies while maintaining the desired signal by implementing a variety of frequency responses including low-pass high-pass band-pass and notch filters. Selectively filtering out noise from a speech signal remains a challenge leading to the use of adaptive techniques. Some of the well-known adaptive filters are the Least Mean Squares (LMS) algorithm and its diverse variations (Normalized LMS [10], FIR-LMS [11], Cascaded X-filter LMS [12], and LMS with spectral subtraction [13]).

The LMS adapts dynamically to the changes in the characteristic nature of noise and the speech signal. The coefficients in the LMS adaptive filter are adjusted with a target of minimizing the square error's mean between the output of the filter and the desired signal. Another powerful adaptive filtering method for speech noise reduction is based on using the Recursive Least Square algorithm [14]. Contrary to the LMS algorithm, which uses the gradient descent method or steepest-descent procedure to update the filter coefficient, the RLS algorithm minimizes the least-square error between the desired and actual signals in a recursive way [15]. Chebyshev speech noise reduction utilizes Chebyshev filters, which have acquired wide popularity because of their sharp cutoff and high efficiency in separating the noise and signal frequencies [16]. It is possible to design these filters to attenuate specific noise frequencies in speech noise reduction while keeping the integrity of the speech signal. Another approach is the Savitzky-Golay filter, which reduces noise and does not alter the basic structure of the speech signal [17]. It becomes particularly useful in applications where the characteristics of the signals need to be preserved, such as speech recognition audio augmentation and communication. The results show that this filter reduces noise but still retains essential details of the speech signal by fitting a polynomial to the data within the window and then replacing the central point with the value from the fitted polynomial.

The brain's ability to select spectra with more information is not utilized in any of the above-mentioned methods. This paper proposes a novel deep neural network-based method for voice noise reduction. The formulations are determined by attention-based neural networks using spectral gain adjustments as the basis for the proposed technique. The contributions of this paper are summed up as follows:

1. The work shows how the addition of spectral gain adjustments affects the suppression of noise in voice signals by an attention-network-based algorithm.

2. The study proposes the use of attention networks to detect and modify spectral components in order to reduce noise in a speech signal.

2. METHODS

Fig. 1 shows the proposed system takes in the pure speech signal. Spectral decomposition allows analysis of the frequency contents of the recorded signal. Fourier transform of the signal provides insights on which frequency beds most of the speech energies are distributed. Taking note of these slices in the frequency domain allows the separation of the original signal into its constituent spectral components. Simultaneously, the same original signal is infused with Gaussian noise to mimic real-world conditions. The next step involves utilizing the input of an attention network to perform spectral gain computations which identify the critical spectral components that hold speech information. Simultaneously relevant features are extracted from the noisy speech stream to enable additional processing. The attention network can focus on the spectral components that are most informative by suppressing bands where noise is excessive.

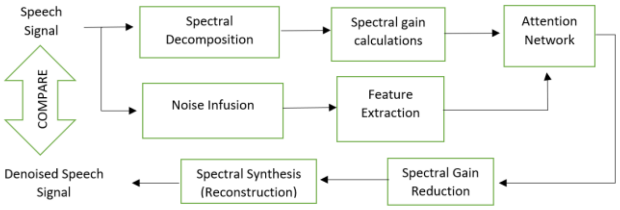

Fig. 1. Overview of the Proposed approach to speech noise reduction

Information-related speech components experience an increase in spectral gain after attention-based modulation whereas noise-dominated components see a decrease in spectral gain attenuating noise in the spectrum domain. These gain-modified spectral components are then used to reconstruct the denoised speech signal. This process produces a denoised speech signal which should ideally have a significant reduction in noise while preserving the original speech quality and comprehensibility. To assess how effectively the algorithm suppresses noise the denoised speech signal is carefully compared to the original noisy speech signal. Combining quantitative measurements with perceptual evaluations the capacity of the algorithm to improve speech quality and intelligibility in noisy environments is appraised.

2.1. Collecting and Preprocessing a Dataset for Speech Samples

The reference data used to calculate the noise reduction technique included noise-free speech samples. These speech samples were taken from specific English audiobooks and were recorded in a quiet studio environment with high-fidelity instruments. Following the selection of the audiobooks, the files will be stored at the standard sampling frequency of 8 kHz and split into segments lasting five to twenty seconds. Each segment's actual transcript will then be graded. One can get the written transcript of the speech samples using a speech-to-text tool such as Baidu Translator or Google Translate. After the audiobooks are divided into speech samples and annotated with the text transcript noise signals are added to each speech sample. Measured and documented in the metadata will be the SNR of each segment. Assessing the noise reduction technique will make it easier to consult the references.

To prepare the dataset for use in the noise reduction system for voice recognition format conversion and volume normalization are necessary. This will ensure that sound samples are compatible with one another and that loudness levels remain constant. The collected and preprocessed speech samples are divided into three subsets: testing validation and training. The machine learning industry standard which is 70 percent training 20 percent validation and 10 percent testing can be used to calculate the proportion of each subset.

2.2. Designing and Implementing an Attention-Network-Based Algorithm for Speech Noise Suppression with Spectral Gain Modifications

The intended attention mechanism for choosing the spectral components for gain modification is shown in Fig. 2. A speech signal in the time domain that has had its spectral components separated using filter banks serves as the input of the encoder. The query (Q) keys (K) and values (V) are generated from the separated spectral components which are then transformed into input embeddings.

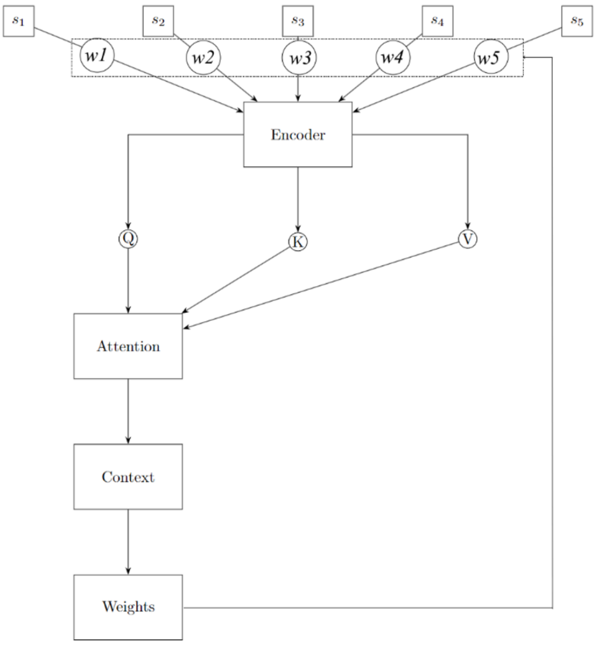

Fig. 2. Attention network for the proposed speech noise reduction system

The attention network's focus in the input sequence is determined by the query. The input embeddings that are going to be compared to the corresponding query are represented by the keys. The values in these embeddings represent the real information. Two sub-modules make up the attention block developed in this study: the feature attention module (FAM) and the temporal attention module (TAM) as shown in Fig. 3. The required data from each of the individual time segments and the whole speech signal are combined by the temporal attention module (TAM). The spectral component of the signal that requires amplification (information) or attenuation (noise) is then identified by the feature attention module by gathering the feature maps. The input embeddings are represented more contextually and accurately capturing relationships and dependencies between spectral components of the input speech signal. The attention block creates the attention weights required to compute a weighted sum of the value vectors.

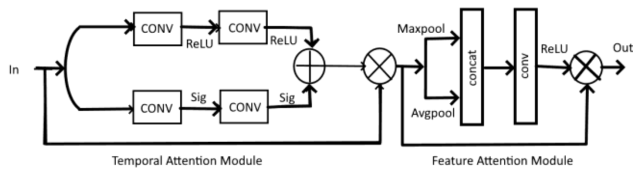

Fig. 3. Sub-modules of the attention block.

The final weights of the attention mechanism are generated by two feed-forward neural networks that are part of the post-processing module called the context block. More specifically by feeding the attention mechanisms output through two feed-forward neural networks, the context block creates the final attention weights. The attention mechanism itself determines the relevance of each input element to the current output. It does this by comparing the current output with each input piece and weighing each input based on its significance. It is important to note that the attention mechanism might not directly produce the final weights to be used in determining the weighted sum of the inputs. The produced attention weights from the mechanism need further post-processing in the context block to yield the final weights. Two feed-forward neural networks are contained in the context block. The output from the attention mechanism is fed as input to the first neural network, which develops an intermediate representation. After that, the second neural network generates the final attention weights depending on the intermediate representation. The context blocks are designed in a way to help the attention mechanism learn the ever-increasing complexities in relationships between the inputs and outputs. These attention weights can then be manipulated based on the general context of the input and output sequences using a context block that is trained with two different neural networks for source and target tokens.

2.3. Evaluating the Performance of the Attention-Network-Based Speech Noise Suppression Algorithm Using Objective Metrics

These noise reduction and filtering techniques should be evaluated by important metrics for assessing speech quality, such as signal-to-noise ratio, segmented signal-to-noise ratio, log-spectral distortion, perceptual evaluation of speech quality, and short-time objective intelligibility. Higher values on the SNR scale show more efficient noise suppression. The improvement in the overall intended signal over the background noise is represented by SNR. The segSNR, by which the uniformity of the noise reduction throughout the signal is represented, makes a wholesale evaluation over short fixed-length segments. A higher segSNR would therefore mean that the performance is more consistent, which in most real-time applications is very important because noise characteristics can change rapidly. The LSD quantifies the amount of distortion brought on by the noise reduction process and is a measure of the average log-spectral distance between the processed and clean signals. Lower LSD levels are recommended because they exhibit less distortion and better preservation of the original speech characteristics. A higher score corresponds to a higher degree of user satisfaction. The perceptual quality of speech post-processing is assessed using the PESQ. This step is crucial for high-quality audio-dependent applications such as telecommunications and hearing aids. The STOI evaluates the intelligibility of the processed speech and predicts listeners' comprehension. A greater STOI value indicates improved intelligibility which is crucial for effective communication in noisy environments.

3. RESULTS AND DISCUSSION

To illustrate how effective the recommended noise reduction technique is, Figure 4 shows a time-domain representation of the original filtered and noisy signals. Figure 5 shows the original signals' spectral representation where the principal information-carrying components are represented by discrete peaks at 15, 180, 310, 500, and 750 Hz. Figure 6 shows the spectra of the three signals (filtered noise-infused and original). Iteratively the suggested method locates spectral components to attenuate or boost. Figure 7 illustrates how the spectral component of a noisy signal can be enhanced. Table 1 presents the performance metrics of various filtering and noise reduction methods based on five critical metrics: signal-to-noise ratio (SNR), segmented SNR (segSNR), short-time objective intelligibility (STOI), log-spectral distance (LSD), perceptual evaluation of speech quality (PESQ).

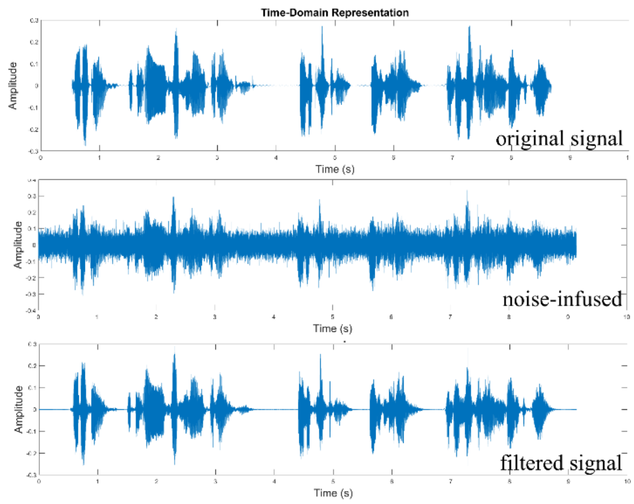

Fig. 4. Time-domain representations of the original, noise-infused, and filtered speech signals.

The human brain can perform spectral selection and gain modification to address the various effects of noise in a speech signal. When a speech signal is infused with noise, the brain combines top-down and bottom-up processing to improve the signal-to-noise ratio of the speech signal. The brain's ability to select spectra with more information is not utilized in existing or traditional methods.

The original signals used in this study are studio recorded narratives which were consolidated as audiobooks. These recordings are clean, well-defined signals of which the waveforms can serve as the benchmark for comparison. It can be noticed that during quiet intervals, negligible signal amplitude can be recorded. However, the noisy signal shows additional oscillations and abnormalities due to background noise which obscures the speech components. A filtered signal that ideally looks very similar to the original signal with less noise-induced fluctuations and smoother variations is an indicator of effective noise suppression. Fine details like distinct transients and crisp peaks should be preserved to maintain the voice signals intelligibility and clarity. It is imperative to minimize distortion and avoid artifacts such as echoes and delay. The performance of the noise reduction algorithm in real-world applications is thoroughly examined using both quantitative data and visual evaluations (Figure 4).

The original signal half-power bandwidth is limited to a frequency range of 0-370 Hz where the amplitudes of the Fourier coefficients suggest the information-carrying components. Outside of this region, signal energy is extremely low which suggests that less speech information can be extracted from those regions. Components of broadband noise are frequently scattered across a wide frequency range. When noise is added more frequency components that are not within the bandwidth of the original signal are introduced. When noise is added the overall power spectrum of the combined signal (original + noise) widens (Fig. 5).

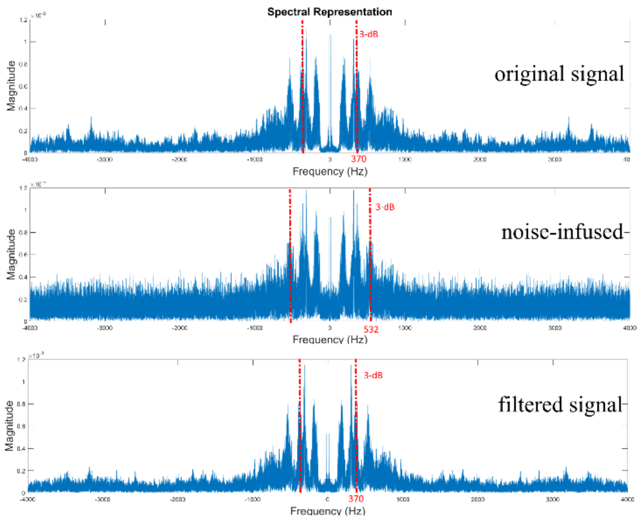

Fig. 1. Spectral representation with half-power bandwidth markings

The half-power bandwidth rises from 740 Hz to 1064 Hz upon the introduction of noise suggesting that the noise has expanded the signals frequency range. By adding more frequency components noise broadens the signal and increases the bandwidth where the power is within 3 dB of the peak power. Noise spreads its energy over a wider frequency range and weakens the signal's frequency domain distinctness which lowers the signals spectral purity. The proposed technique involves identifying the spectrum components and adjusting the weight of the attention network for each component to modify its gain. Due to the noise adding power at frequencies where the original signal had little to none the overall power over a wider frequency range increases causing this broadening. The extra noise has caused the signal strength to become more evenly distributed across a larger frequency range. The half-power bandwidth increases as more of the spectrum components of the noise are now covered by frequencies that are 3 dB below the peak power. When some of the signal components are obscured by noise it could be harder to distinguish the signal from the noise. As a result, the signal's power falls within 3 dB of its peak power within an extended effective range reflecting the noise contribution. The attention network proposed in this study focuses on the spectral components where information are obscured by noise. Once determined, spectral gain modification is performed to amplify those spectra that contain speech information while attenuating those that have no or less information content.

Specific frequency ranges and time intervals are the focus of a selective gain modification. As a demonstration, a single spectrum is amplified at a specific time frame. For the first part of the signal duration, the frequencies around 1000 Hz are only amplified by a factor of two (Fig. 6 & Fig. 7). This selective gain modification amplifies or attenuates specific spectra over time which can manipulate the targeted signal characteristics to enhance speech information-rich or reduce noise-infused components.

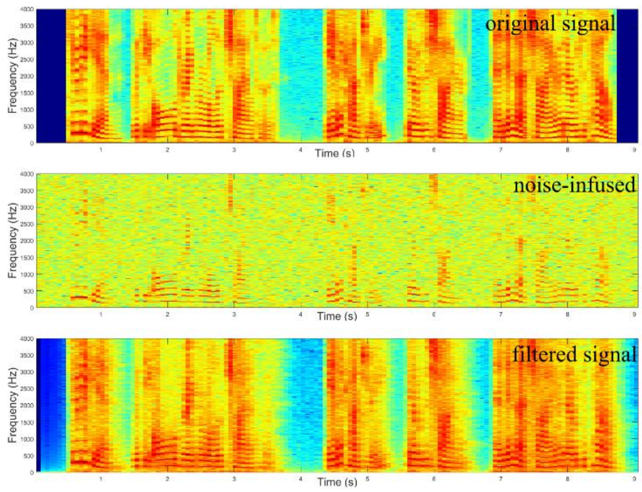

Fig. 6. Spectrogram of the original, noise-infused, and filtered signal

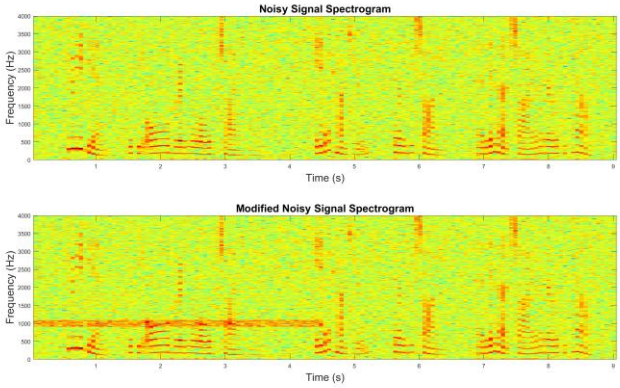

Fig. 7. Boosting noisy signal's spectral component

The best-performing filters and their effects for each parameter are highlighted in this section analysis of the data (Table 1). Our proposed method shows the highest SNR and segSNR when compared with previous approaches indicating its superior ability to increase the overall signal-to-noise ratio. It is apparent from this that it is effective in reducing ambient noise and improving the clarity of the output signal. Given its ability to maintain consistent signal quality across a range of audio channel segments, the segSNR is a dependable choice for applications requiring noise reduction. However, for the other metrics, our proposed method did not achieve the best scores but still performed acceptable. For example, if the speech processing will focus on the filter with the least distortion, then the Recursive Least Squares (RLS) filter is the best choice in this cohort because it achieves the lowest spectral distortion or LSD score. Lower LSD values are better because they suggest that the filter more accurately preserves the original signal properties by modifying the spectral content of the signal with less distortion. RLS is therefore particularly well-suited for applications where maintaining the spectral integrity of the original signal is crucial.

Another criterion is the perceived quality of the filtered signal. The Least Mean Square (LMS) algorithm works remarkably well on the perceptual evaluation of speech quality (PESQ) metric that closely resembles how people perceive speech quality. A higher PESQ score denotes better perceived audio quality. The results show that the LMS algorithm successfully enhanced speech intelligibility and naturalness. The algorithm is therefore suitable for usage in voice communication systems. Our proposed algorithm performed not too far from the LMS since it is second among the other techniques evaluated. In addition, LMS also obtained the highest score in short-time objective intelligence (STOI) although our proposed algorithm is not too far behind gaining 3rd place. This statistic is important to evaluate the quality of the speech signal understanding in short bursts for real-time communication applications. The high STOI score confirms that LMS can keep speech understandable even when noise levels are reduced.

Table 1. The performance of different noise reduction methods

Methods	SNR	segSNR	LSD	PESQ	STOI
Weiner+HRNR [7]	5.5564	-0.5826	8.6322	1.9553	0.7840
Weiner [8]	4.3992	-0.2622	6.7307	1.4648	0.7611
LMS [15]	5.5617	0.1458	5.4341	2.8005	0.9454
RLS [14]	6.7711	1.6528	4.9316	2.2960	0.8761
Chebyshev [16]	3.3057	-1.7722	10.396	2.0125	0.7044
Savitzky-Golay [17]	5.6493	-0.7370	7.9836	1.9573	0.7808
Normalized LMS [10]	2.8624	-1.9260	8.1491	1.4847	0.7507
FIR+LMS [11]	6.0250	1.3868	5.1120	2.2132	0.8371
Cascaded [12]	4.0442	-1.6908	9.1313	1.9272	0.7894
LMS + Spectral Subtraction [13]	6.2662	0.0970	7.0714	2.0035	0.7865
Ours*	7.1138[a]	1.9950[b]	5.0052	2.3149	0.8541

4. Conclusion

A strategy such as the brains' ability to recognize spectra with more information is not utilized by most noise reduction techniques. The proposed method can identify which spectral components of a voice signal require modification to remove noise by using attention networks. This study has shown that noise in voice signals can be effectively reduced by adjusting the spectral component gain which is determined by the attention network. The recommended approach performed better in terms of SNR and segSNR than a number of the available noise reduction methods.

In future work, the parameters of the attention network can be optimized to provide more focus on the information-carrying spectra of a signal. Future studies can also include the evaluation of overlapping spectra, as well as temporal window, in performing selective gain modification. Such approach can provide insight on how to reduce distortion in the signal brought by selectively varying spectral attributes. The overlap can reduce the drastic variation caused by the difference in gain factor for adjacent spectra in a signal.

Acknowledgments

We thank our colleagues from Batangas State University who provided insight and expertise that greatly assisted the research.

Author contributions

Huiting Yu: Conceptualization, Methodology, Data curation, **Anton Louise De Ocampo:** Validation, Writing-Reviewing, and Editing, **Rowell Hernandez:** Writing-

Original draft preparation, and Software.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] X. Wang and L. Xu, "Speech perception in noise: Masking and unmasking," Journal of Otology, vol. 16, no. 2, pp. 109–119, 2021.

[2] A. de Cheveigné, "Harmonic cancellation—A fundamental of auditory scene analysis," Trends in Hearing, vol. 25, p. 23312165211041424, 2021.

[3] Mahadevaswamy and D. Ravi, "Robust perceptual wavelet packet features for recognition of continuous Kannada speech," Wireless Personal Communications, vol. 121, no. 3, pp. 1781–1804, 2021.

[4] M. S. Islam, Y. Zhu, M. I. Hossain, R. Ullah, and Z. Ye, "Supervised single channel dual domains speech enhancement using sparse non-negative matrix factorization," Digital Signal Processing, vol. 100, p. 102697, 2020.

[5] M. Iqbal, S. A. Raza, M. Abid, F. Majeed, and A. A. Hussain, "Artificial neural network based emotion classification and recognition from speech," International Journal of Advanced Computer Science and Applications, vol. 11, no. 12, 2020.

[6] M. A. A. Albadr, S. Tiun, M. Ayob, F. T. AL-Dhief, K. Omar, and M. K. Maen, "Speech emotion recognition using optimized genetic algorithm-extreme learning machine," Multimedia Tools and Applications, vol. 81, no. 17, pp. 23963–23989, 2022.

[7] M. Une and R. Miyazaki, "Evaluation of sound quality and speech recognition performance using harmonic regeneration for various noise reduction techniques," in RISP Int. Workshop Nonlinear Circuits, Commun Signal Process.(NCSP), 2017, pp. 377–380.

[8] H. H. Nuha and A. A. Absa, "Noise reduction and speech enhancement using wiener filter," in 2022 International Conference on Data Science and Its Applications (ICoDSA), 2022, pp. 177–180.

[9] C. I. Muresan, I. R. Birs, E. H. Dulf, D. Copot, and L. Miclea, "A review of recent advances in fractional-order sensing and filtering techniques," Sensors, vol. 21, no. 17, p. 5920, 2021.

[10] J. S. Jakati and S. S. Kuntoji, "A noise reduction method based on modified LMS algorithm of real time speech signals," WSEAS Trans Environ Dev, vol. 16, no. 13, 2021.

[11] T. P. Zieliński and T. P. Zieliński, "FIR Adaptive Filters," Starting Digital Signal Processing in Telecommunication Engineering: A Laboratory-based Course, pp. 317–343, 2021.

[12] X. H. Xie, L. N. Zhou, and Y. J. Xie, "Design and Simulation of Active Noise Cancelling Earphone System Based on FXLMS Algorithm," in 2022 4th International Conference on Natural Language Processing (ICNLP), 2022, pp. 626–630.

[13] O. Barkovska, V. Kholiev, and V. Lytvynenko, "Study of noise reduction methods in the sound sequence when solving the speech-to-text problem," Advanced Information Systems, vol. 6, no. 1, pp. 48–54, 2022.

[14] S. C. Venkateswarlu, N. U. Kumar, and A. Karthik, "Speech enhancement using recursive least square based on real-time adaptive filtering algorithm," in 2021 6th International Conference for Convergence in Technology (I2CT), 2021, pp. 1–4.

[15] A. Chiheb and H. Khelladi, "Performance Comparison of LMS and RLS Algorithms for Ambient Noise Attenuation," International Journal of Electrical and Computer Engineering Research, vol. 4, no. 1, pp. 14–19, 2024.

[16] S. Olika and A. Rajani, "Adaptive noise cancellation for speech signal," Int J Sci Res Publ, vol. 10, no. 9, 2020.

[17] J. Blessy and C. S. Christopher, "A Survey on Filtering Methods Used to Remove Noise in Speech and Music Signal," in 2022 6th International Conference on Electronics, Communication and Aerospace Technology, 2022, pp. 140–145.