Machine learning based noise suppression in narrow-band speech communication systems

Joyraj Chakraborty[1,2], Martin Reed[1], Nikolaos Thomos[1]
[1]School of Computer Science and Electronic Engineering
University of Essex, UK
Email: {joyraj.chakraborty},{mjreed},{nthomos}@essex.ac.uk

Geoff Pratt[2], Nigel Wilson[2]
[2]CML Microcircuits (UK) Ltd.
Maldon, UK
Email: {GPratt},{NWilson}@cmlmicro.com

***Abstract*—Narrow-band digital personal radio systems are used for speech communication in challenging environments where background noise, such as machinery or emergency sirens, can pose significant problems for speech intelligibility. This paper proposes a machine learning based noise suppression approach that utilises a neuro-fuzzy logic-based neural network for noise estimation and reduction. The technique is shown to give significant improvements in noise suppression compared to a non-adaptive noise suppression approach. The choice of a neuro-fuzzy logic neural network is motivated by the need for a low-power implementation suitable for mobile, power constrained, terminals. To validate this, the algorithm has been tested in a real-time system showing that it can be implemented in constrained devices unlike more complex machine learning techniques that are unsuitable for low-power digital personal radio systems.**

I. Introduction

Narrow-band speech communication systems are an essential tool within emergency services and industries such as construction. These systems typically use wireless terminals with constrained power requirements due to their personal mobile use, generally termed digital personal mobile radio (Digital-PMR) [1]. Under many Digital-PMR use cases, background noise can be a significant problem, for example, sirens in emergency service use or machinery noise in industrial applications. Without sufficient suppression of the background noise speech intelligibility can be highly impaired; this is exacerbated by low bit-rate voice codecs that are mandated by standards that support the limited channel capacity of such systems [1]. Consequently, background noise suppression is an important feature and is the subject of this paper.

Classical noise suppression techniques have been used for decades in speech systems such as spectral subtraction schemes [2], and adaptive filtering techniques [3], [4]. Moreover, improved approaches can found in literature including automatic gain control [5] and Kalman filtering [6]. However, recently machine learning (ML) approaches have been proposed in the literature such as deep neural networks (DNN) [7], recurrent neural networks (RNN) [8], deep denoising AutoEncoders (DDAE) [9], fuzzy based deep learning [10], convolutional neural network (CNN) [11] or generative adversarial networks (GAN) [12]. However, these models require significant amounts of storage and computational resources for training and operation, which greatly exceeds the resources available in mobile personal radio systems. Thus, this paper proposes a noise suppression mechanism that is suitable for constrained devices such as those used for Digital-PMR. The contributions of this paper are: i) charaterization of background noise using a low-complexity ML algorithm, *Adaptive Neuro-Fuzzy Inference System* (ANFIS); ii) use of a database of pre-captured, typical, background noises to improve the performance of the system; iii) demonstration of the system using real voice and background noise samples showing significant improvements in both PESQ [13] and signal-to-noise (SNR) scores compared to an existing, non-adaptive, noise suppression approach; and iv) confirmation that the system can operate for low-power systems through implementation in a 32-bit embedded CPU utilising modest power consumption.

II. Related Work

This section presents an overview of existing methodologies used for speech enhancement in wide-band and narrow-band communication systems. While speech denoising is a well-established technique in audio processing, so far, more focus has been given on wide-band speech signals. Denoising narrow-band speech signals present unique challenges, including limited frequency range and the presence of narrow-band noise sources such as electrical interference or other types of environmental noise. To address these challenges, researchers have found that employing a Wiener filter can be an optimal solution and this is why it has been widely utilized in communication systems. Several studies have been conducted on the Wiener filter and coupled with spectral subtraction [14], [15], [16]. It has been found that the combined method works better compared to using a single Wiener filter. The work in [17] has demonstrated that using combined spectral subtraction and Wiener filter methods in the wavelet domain is superior to using a single Wiener filter. In another study [18], voice activity detection (VAD) was adopted to overcome the limitation of assumption Gaussian signal that does not hold when processing speech signals in real-time communication systems with a small frame size. In VAD, it is assumed that noise is present throughout the entire period, and the noise is estimated using VAD during the silent periods. In [19], the authors proposed generative dictionary learning for speech enhancement. However, these methods fail when noise is non-stationary or has low SNR. For a more comprehensive review of traditional approaches to speech enhancement systems, we recommend referring to the survey conducted in [20]. Recently, automatic noise class detection using artificial neural networks has shown promising results [21]. In [22], Nonnegative Matrix Factorization and Robust Principal Component Analysis were shown to be efficient for speech enhancement. In [23], the Adaptive Neuro-Fuzzy Inference System (ANFIS) was used as a simple ML filter for speech enhancement. However,

it has only been applied with white noise, which requires subtraction for each frame and may result in artifact creation. A recent review on supervised learning algorithms for single-channel speech enhancement can be found in [24]. From the above, it appears that ML techniques provide more effective objective measures for speech enhancement. However, utilizing ML techniques for this purpose can be complex, and requires significant amounts of training data to efficiently reduce noises [24]. As ANFIS has been shown to effectively overcome non-stationary and low SNR noisy scenarios, this paper explores a hybrid method based on this method for accurate noise estimation, leveraging small frames of noisy speech samples to detect noise types from the noise database history. The estimated noise is then used for denoising the speech frames.

III. System and Problem Overview

Half-duplex communication systems consist of two nodes, a transmitter and a receiver, that alternately transmit and receive information. The transmitted speech is produced in the presence of background noise which distorts the intended speech signal. The transmitted signal is further degraded by noise during communication. The received signal is then decoded through a speech decoder which attempts to retrieve the original speech signal. The output of the front end is then passed to a speaker to reproduce the speech.

Let us denote by $x(i)$ the speech signal and by $n(i)$ the background noise signal. Thus, we have

$$w(i) = x(i) + n(i) \tag{1}$$

The encoder compresses the input signal $w(i)$ and generates the encoded signal $z(i)$. We can represent the encoder as a function $f[]$, such that

$$z(i) = f[w(i)] \tag{2}$$

Now, let us consider the output signal $y(i)$ after passing through the decoder that is represented by $g[]$. The output signal $y(i)$ can be written as:

$$y(i) = g[f[w(i)] \approx x(i) + n(i) + e(i) \tag{3}$$

where $e(i)$ is the remaining error signal after passing $z(i)$ through the decoder, which is due to the introduced codec and transmission noise.

The noise component $n(i)$ is assumed to be independent of the speech signal, so we can write:

$$E[R(x(i), n(i)] = 0 \tag{4}$$

In summary, the output signal after passing through the encoder can be represented as the encoded speech signal plus remaining noise after decoding.

IV. Proposed Methodology

The proposed noise suppression consists of two stages: noise estimation and noise suppression. Both of these stages are carried out online, where the noise estimation aims to characterise the noise in the signal $y(i)$ by comparing this speech+noise signal with stored noise history. The denoising front end includes a noise estimation module that accurately detects the type of noise in the signal based on the history noise samples in the database and a small frame of noisy speech samples. This information is used to train on multiple noise chunks for noise estimation by using an efficient ML algorithm that is described below. In the second stage, the estimated noise is then suppressed in real-time using a combination of Wiener filtering and spectral subtraction to mitigate environmental disturbances.

Let us assume, the first noisy frame at time instant k is denoted $\mathbf{n}(k)$ and $\mathbf{N}$ is a matrix containing a collection of historical noise samples in each column. The goal is to find the column vector $\mathbf{n}_i$ in $\mathbf{N}$ that best matches the current noisy speech frame $\mathbf{n}(k)$. One way to achieve this is by computing the Euclidean distance between $\mathbf{n}(k)$ and each column vector in $\mathbf{N}$, $\mathbf{N}_i$,and selecting the column vector with the smallest distance as the best match:

$$i^* = \arg\min_i \|\mathbf{n}(k) - \mathbf{N}_i\|_2 , \tag{5}$$

where i^* is the index of the best-matching column vector in $\mathbf{N}$, and $\|\cdot\|_2$ denotes the Euclidean norm. Once the best-matching noise sample is identified, then an ML algorithm can be used for training. In our work, ANFIS [25] is selected as it is known to have good performance with low-complexity when the number of inputs is modest, as in this case [26]. ANFIS is a hybrid model that combines the strengths of neural networks and fuzzy logic to create a powerful tool for nonlinear system modeling [27]. The ANFIS model can be trained using a set of input-output data, and the parameters can be adjusted using the backpropagation algorithm. Let us assume that we have taken a few chunks from the best-matching historical column vector of noisy signals, denoted by $v(n)$ at time instant n. We can use the ANFIS model to estimate the noise signal based on the historical chunks of the noisy speech signals.

An example of ANFIS model for two inputs and two membership functions is shown in Fig. 1. In the context of noise estimation in speech signals, ANFIS can be an effective approach for modeling the statistical properties of the noise. The ANFIS model consists of multiple layers, including the input layer, a membership function layer, a rule layer, a normalization layer, and an output layer. During the training process, the model learns the optimal parameters for each layer, based on the input data and the desired output. The input layer takes the noisy speech frame and two frames from the noise database as input. The input layer applies appropriate scaling and normalization to the input data to prepare it for processing by the subsequent layers. The second layer is the membership function layer using a clustering method (i.e., grid partitioning), which maps the input data to a set of *linguistic variables*, such as "high noise" or "low noise". This layer helps to represent the input data in a way that is more meaningful for the subsequent layers to process. There are many different types of membership functions that can be used, but two commonly used functions in ANFIS are the Gaussian and the triangular membership functions. Here, we use the Gaussian membership function as it can effectively capture the statistical properties. The third layer is the rule layer, which applies fuzzy logic rules to the output of the membership function layer to generate a set of inference rules. These rules capture the relationships between the linguistic variables and the desired output (i.e., the estimated noise in the signal). The fourth layer is the normalization layer, which simply scales the outputs of

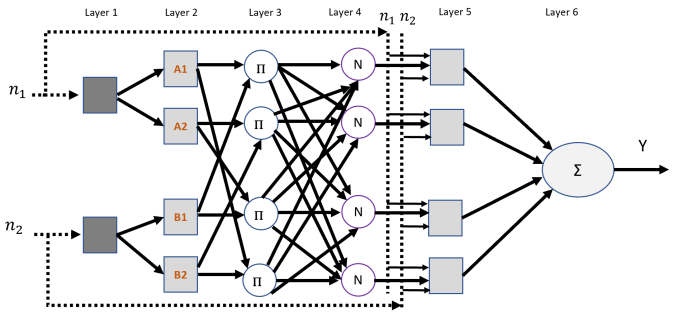


Fig. 1. Example of two inputs ANFIS structure

the rule layer to ensure that they sum to one. The final layer is the output layer, which generates the estimated noise based on the input data and the learned parameters of the model.

The input variables to the ANFIS model are the first noisy speech frame and two chunks of the best-matching historical signals, the output variable is the estimate of the statistical noisy frame at time instant k. Then the ANFIS model can be represented as a set of fuzzy if-then rules, such that: If x_1 is A_1 and x_2 is A_2 and $\ldots$ and x_N is A_N, then $y = f(x_1, x_2, \ldots, x_N)$ where $x_1, x_2, ..., x_N$ are the first noisy speech frame and a few chunks from the best-matching column vector, $A_1, A_2, ..., A_N$ are the fuzzy sets defined over the input variables, and $f(x_1, x_2, ..., x_N)$ is a nonlinear function that maps the input variables to the output variable $\mathbf{n}$.

The parameters of the ANFIS model can be adjusted using the backpropagation algorithm together using the gradient descent algorithm, which iteratively updates the parameters of the membership functions and the adaptive parameters. The goal is to minimize the mean squared error between the estimated noise signal and the true noise signal. The PSD of the output (ANFIS) noise signal is the estimate of the noise power in the noisy speech signal, which is then used to calculate the Wiener filter coefficients. The combination of the Wiener filter and spectral subtraction estimates the clean speech signal by using a transfer function that minimizes the mean square error between the noisy signal and the estimated clean signal. This is expressed as

$$\hat{\mathbf{x}}(k,\omega) = \alpha(\omega)\mathbf{W}(\omega)\mathbf{y}(k,\omega) + (1-\alpha(\omega))\mathbf{S}(\omega)\mathbf{X}(k,\omega) \quad (6)$$

where $\hat{\mathbf{x}}(k,\omega)$ is the estimated clean speech signal at time instant k and frequency ω, $\mathbf{W}(\omega)$ is the Wiener filter, $\mathbf{S}(\omega)$ is the spectral subtraction filter, $\mathbf{y}(k,\omega)$ is the noisy speech signal at time instant k and frequency ω, and $\alpha(\omega)$ is a weighting function that controls the contribution of each filter at different frequencies. We have determined though experimentation that the optimal range for the value of $\alpha(\omega)$ is 0.8 to 0.9. The Wiener filter and spectral subtraction filter are coupled together by the weighting function to achieve better noise suppression performance.

$$\mathbf{W}(\omega) = \frac{\mathbf{P}yy(\omega)}{\mathbf{P}yy(\omega) + \mathbf{P}_{nn}(\omega)} \quad (7)$$

where $\mathbf{P}yy(\omega)$ is the power spectral density (PSD) of the noisy speech signal and $\mathbf{P}nn(\omega)$ is the PSD of the noise signal.

$$S(\omega) = \frac{\mathbf{P}_{nn}(\omega)}{\mathbf{P}yy(\omega) + \mathbf{P}_{nn}(\omega)} \quad (8)$$

where $\mathbf{P}nn(\omega)$ represents the PSD of the noise signal and $\mathbf{P}yy(\omega)$ is the frequency-domain representation of the noisy speech signal.

V. Experimental set up and results

A. Experimental set up

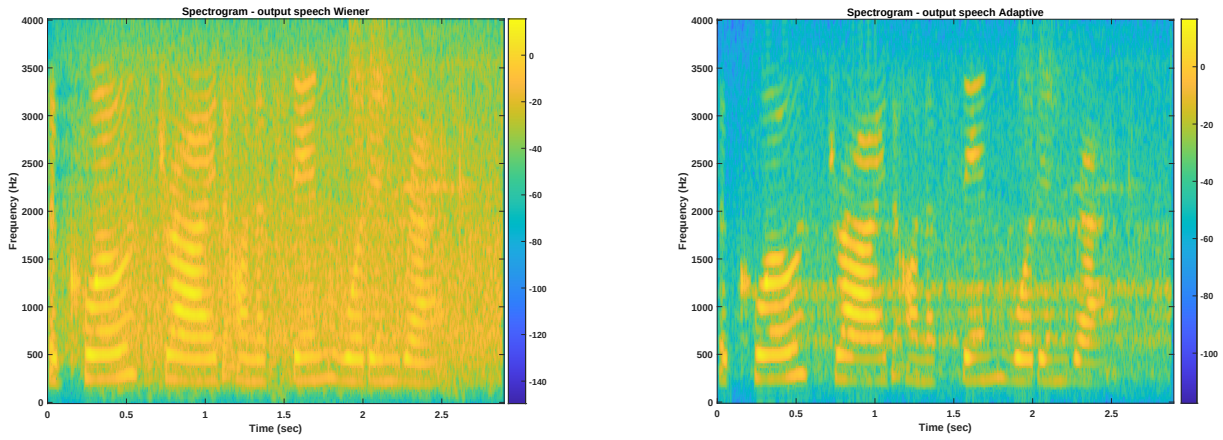
To evaluate the performance of our proposal we used a half-duplex hardware AMBE2+ codec [28][1] as this is widely used in Digtal-PMR. This device is designed to process audio signals in real-time. It can also introduce signal distortions and channel impairments, such as packet loss or delay, to mimic real-world communication scenarios.

In our study, we conducted experiments under three distinct types of noise conditions: *busy street*, *siren*, and *exhibition hall*. For each noise condition, we mix the clean speech signals with various levels of noise that correspond to different signal-to-noise ratios (SNRs). By introducing these different types of noise, we aimed to evaluate the effectiveness of our proposed denoising methodology in improving the quality of speech across different noise types and SNR levels. In our experiments, each signal was sampled at a rate of 8 kHz and encoded/decoded by the device as a 20 ms chunk as is commonly used in Digital-PMR [1]. For the clean speech signals, we used the widely used TIMIT Acoustic-Phonetic Continuous Speech Corpus [29], which includes twenty clean speech signals from ten male and ten female speakers. In our experiment, we played the twenty predefined noisy speech signals into this device, and stored the received signal after the receiver decoder. We evaluated the signal quality using standard metrics such as SNR and perceptual evaluation of speech quality (PESQ) [13].

B. Results

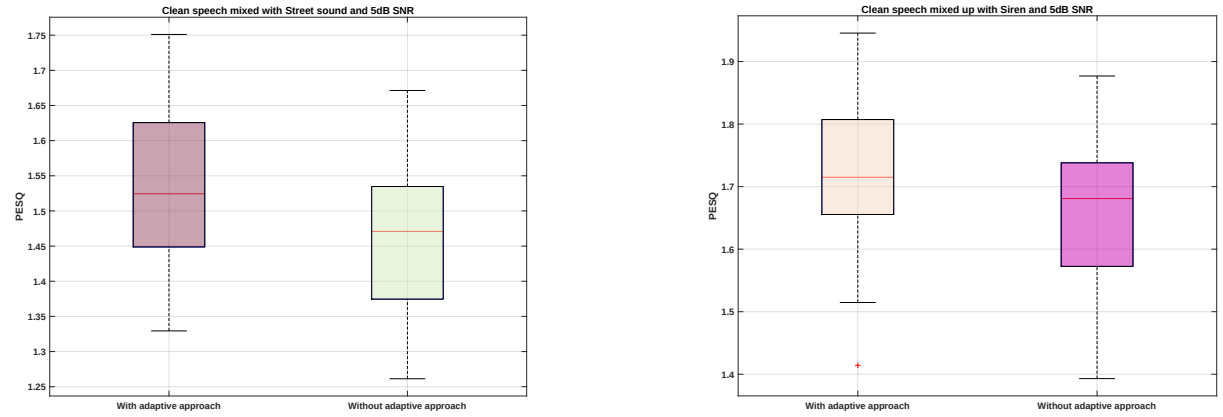
To test speech enhancement techniques, we mixed clean speech with non-stationary street sound. We ensured that the silent frames were very short, resulting in less noise. In general, considering the first frame as a representative noise sample works well for most cases. However, in our study, we take into account the worst-case scenario where this is not the case. In doing this, we are improving upon the general approach. Our proposed method selects the first 20 ms frame as the noisy frame and searches for the best matching frame in the historical noise samples column vector, then extracts a few frames from that noise sample and combines it with the initial noise sample to obtain the best noise statistics through training. It should be noted that the effectiveness of the ANFIS model for noise estimation may vary based on the accuracy of the noise samples and the noisy speech frame being analyzed. We used a spectrogram to show the effectiveness of our adaptive method. A spectrogram is particularly useful in the context of speech denoising as it allows for the analysis of changes in the frequency components of a signal over time. Thus, we evaluated our example noisy speech by analyzing its spectrogram (Fig. 2). Our observation reveals that our adaptive ANFIS with noise database approach proves to be better for the non-stationary noise statistics, as it significantly reduces the background noise and reveals the dominant frequency components of the speech.Therefore, this ANFIS based adaptive

[1]supplied by DVSI https://www.dvsinc.com/soft_products/ambe_p2.shtml



(a) (b)

Fig. 2. Spectrograms showing the signal after noise suppression for: (a) a Weiner filter (b) the proposed ANFIS following the noise database approach.



(a) (b)

Fig. 3. PESQ score comparison between the proposed ANFIS with noise database approach and native suppression of the codec (without adaptive approach) for: (a) street sounds and (b) siren background noise.

approach was chosen to suppress the noise after decoder and before digital to analogue module.

TABLE I. TABLE SHOWING THE COMPARISON BETWEEN NORMAL (NATIVE SUPPRESSION OF THE CODEC) AND ADAPTIVE METHODS FOR THREE DIFFERENT NOISY SCENARIOS IN TERMS OF PESQ AND SNR VALUES.

Noisy Scenario	Method	PESQ	SNR
Street Sound	Normal	1.46	2.5
	Adaptive	1.52	3.3
Siren Sound	Normal	1.67	2.9
	Adaptive	1.73	3.6
Exhibition Sound	Normal	1.47	2.5
	Adaptive	1.53	3.1

The objective evaluation used SNR while (emulated)-subjective evaluation was performed using the speech quality metric, PESQ [13]. Our proposed method achieved a significant improvement in speech quality after the decoder, as evidenced by the evaluation results shown in Fig. 3. The PESQ scores increased by 0.07 - 0.08 points on average compare to native suppression of the codec. It is widely acknowledged that PESQ does not give an accurate impression at the low MOS scores found in low-bit rate systems such as Digital-PMR [30]. However, we would like to comment that it gives a significant subjective improvement. A wider set of results across the three noise samples are presented in Table I.

We have deployed our algorithm on an open-source 32-bit microprocessor architecture, although there were optional custom instructions added. Our experiments suggest that the noise estimation and suppression stages could be performed with acceptable power consumption. This, therefore, confirms that the proposal is potentially suitable for practical deployment in a Digital-PMR system.

VI. CONCLUSION

In this work, a low-complexity, neuro-fuzzy based machine learning method utilising a database of historic background noises was proposed. This system was used for noise estimation and suppression using a coupled Wiener and spectral filter. The system adapts to varying background noise within the speech signal, unlike traditional non-adaptive approaches. Our proposed method was shown to effectively suppress background noise in a narrow-band Digital-PMR system in terms of both objective and (emulated) subjective speech quality metrics. Implementation in a real-time low-power 32-bit CPU showed that the proposed method is both effective and suitable for use in power-constrained terminals.

ACKNOWLEDGMENT

This work was supported by Innovate UK under Knowledge Transfer Partnership number 012507.

REFERENCES

[1] *Electromagnetic compatibility and Radio spectrum Matters (ERM); Digital Mobile Radio (DMR) Systems; Part 1: DMR Air Interface (AI) protocol*, ETSI, TS 102 361-1.

[2] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 2, pp. 113–120, apr 1979.

[3] Y. Ephraim and D. Malah, "Speech enhancement using optimal nonlinear spectral amplitude estimation," in *ICASSP '83. IEEE International Conference on Acoustics, Speech, and Signal Processing*. Institute of Electrical and Electronics Engineers.

[4] K. Paliwal and A. Basu, "A speech enhancement method based on kalman filtering," in *ICASSP '87. IEEE International Conference on Acoustics, Speech, and Signal Processing*. Institute of Electrical and Electronics Engineers.

[5] Y. Nagata, T. Fujioka, and M. Abe, "Speech enhancement based on auto gain control," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 1, pp. 177–190, jan 2006.

[6] P. Meyer, S. Elshamy, and T. Fingscheidt, "A multichannel kalman-based wiener filter approach for speaker interference reduction in meetings," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, may 2020.

[7] Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, "An experimental study on speech enhancement based on deep neural networks," *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 65–68, jan 2014.

[8] F. Weninger, F. Eyben, and B. Schuller, "Single-channel speech separation with memory-enhanced recurrent neural networks," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, may 2014.

[9] R. E. Zezario, J.-W. Huang, X. Lu, Y. Tsao, H.-T. Hwang, and H.-M. Wang, "Deep denoising autoencoder based post filtering for speech enhancement," in *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, nov 2018.

[10] C. L. P. Chen, C.-Y. Zhang, L. Chen, and M. Gan, "Fuzzy restricted boltzmann machine for the enhancement of deep learning," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 6, pp. 2163–2173, dec 2015.

[11] T. Kounovsky and J. Malek, "Single channel speech enhancement using convolutional neural network," in *2017 IEEE International Workshop of Electronics, Control, Measurement, Signals and their Application to Mechatronics (ECMSM)*. IEEE, may 2017.

[12] H. Phan, I. V. McLoughlin, L. Pham, O. Y. Chen, P. Koch, M. D. Vos, and A. Mertins, "Improving GANs for speech enhancement," *IEEE Signal Processing Letters*, vol. 27, pp. 1700–1704, 2020.

[13] "Itu-t rec. p. 862," 2000, objective quality measurement of speech codecs. [Online]. Available: https://www.itu.int/rec/T-REC-P.862/en

[14] W. Guang-yan, Z. Xiao-qun, and W. Xia, "Musical noise reduction based on spectral subtraction combined with wiener filtering for speech communication," in *IET International Communication Conference on Wireless Mobile & Computing (CCWMC 2009)*. IET, 2009.

[15] L. Nahma, P. C. Yong, H. H. Dam, and S. Nordholm, "Improved a priori snr estimation in speech enhancement," in *2017 23rd Asia-Pacific Conference on Communications (APCC)*. IEEE, dec 2017.

[16] H. Pardede, K. Ramli, Y. Suryanto, N. Hayati, and A. Presekal, "Speech enhancement for secure communication using coupled spectral subtraction and wiener filter," *Electronics*, vol. 8, no. 8, p. 897, aug 2019.

[17] F. Ykhlef, A. Guessoum, and D. Berkani, "Combined spectral subtraction and wiener filter methods in wavelet domain for noise reduction," in *Proceedings of Second International Symposium on Communications, Control and Signal Processing*, 2006, pp. 13–15.

[18] M. Hoffman, Z. Li, and D. Khataniar, "GSC-based spatial voice activity detection for enhanced speech coding in the presence of competing speech," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 2, pp. 175–178, 2001.

[19] C. D. Sigg, T. Dikk, and J. M. Buhmann, "Speech enhancement using generative dictionary learning," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 6, pp. 1698–1712, aug 2012.

[20] A. Chaudhari and S. B. Dhonde, "A review on speech enhancement techniques," in *2015 International Conference on Pervasive Computing (ICPC)*. IEEE, jan 2015.

[21] R. Jaiswal, "Automatic noise class detection for improving speech quality using artificial neural networks," in *2022 7th International Conference on Communication and Electronics Systems (ICCES)*. IEEE, jun 2022.

[22] Y. HU, X. ZHANG, X. ZOU, M. SUN, Y. ZHENG, and G. MIN, "Semi-supervised speech enhancement combining nonnegative matrix factorization and robust principal component analysis," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E100.A, no. 8, pp. 1714–1719, 2017.

[23] J. Thevaril and H. Kwan, "Speech enhancement using adaptive neuro-fuzzy filtering," in *2005 International Symposium on Intelligent Signal Processing and Communication Systems*. IEEE, 2005.

[24] N. Saleem and M. I. Khattak, "A review of supervised learning algorithms for single channel speech enhancement," *International Journal of Speech Technology*, vol. 22, no. 4, pp. 1051–1075, oct 2019.

[25] J.-S. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 665–685, 1993.

[26] G. Özkan and M. İnal, "Comparison of neural network application for fuzzy and ANFIS approaches for multi-criteria decision making problems," *Applied Soft Computing*, vol. 24, pp. 232–238, nov 2014.

[27] M. O. Okwu and O. Adetunji, "A comparative study of artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS) models in distribution system with nondeterministic inputs," *International Journal of Engineering Business Management*, vol. 10, p. 1847979018768421, 2018. [Online]. Available: https://doi.org/10.1177/1847979018768421

[28] Digital Voice Systems, Inc. (DVSI), "Ambe+2 vocoder: Next generation voice compression," Whitepaper, January 2019. [Online]. Available: https://www.dvsinc.com/wp-content/uploads/2019/01/AMBE2-Whitepaper.pdf

[29] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus [speech database]," National Institute of Standards and Technology, 1993, available at: https://catalog.ldc.upenn.edu/LDC93S1.

[30] A. Rix, *Comparison between subjective listening quality and P.862 PESQ score*, white Paper by Psytechnics available online at https://www.sageinst.com/assets/download_doc/1639138826-wp_sub_v_pesq.pdf.