An Approach for Speech Enhancement Using Deep Convolutional Neural Network

Arul.V.H
Department of Electronics Communication
Thejus Engineering College, Thrissur, Kerala, India
arul.vh01@gmail.com

V.G.Sivakumar
Department of Electronics Communication
Sathyabama University, Chennai, Tamil Nadu, India

Ramalatha Marimuthu
Department of Electronics Communication
Kumaraguru College of Technology, Tamil Nadu , India

Basabi Chakraborty
Department of Radio Physics andElectronics
Iwate Perfectural University, Japan

Abstract: Speech is a primary and universal medium to communicate with each other. The additive or background noise present in the channel humiliates the signal quality. In order to minimize undesirable background noises, speech enhancement techniques have been introduced. Accordingly, this paper proposes a speech enhancement approach using Deep Convolutional Neural Network (DCNN). At first, the noise signal is appended with the hygienic speech signal and the noisy speech signal is generated. Then, the next step is the framing, in which the Fractional Delta-Amplitude Modulation Spectrogram (FD-AMS) features are extracted from the frames. Finally, the extracted features are provided as the input to the DCNN, which generates the optimized estimation of the speech signal. The proposed method is analyzed using NOIZEUS database based on the metrics, Perceptual Evaluation of Speech Quality (PESQ) and Root Mean Square Error (RMSE). Also, the comparative analysis is performed with the existing speech enhancement techniques. From the results, it is shown that the proposed method obtains maximum PESQ and minimum RMSE than the existing techniques, which shows the superiority of the proposed speech enhancement.

Keywords: Speech enhancement, framing, feature extraction, Fractional Delta-Amplitude Modulation Spectrogram, Deep Convolutional Neural Network.

Nomenclature

Abbreviations/Acronyms	Description
DCNN	Deep Convolutional Neural Network
FD-AMS	Fractional Delta-Amplitude Modulation Spectrogram
PESQ	Perceptual Evaluation of Speech Quality
RMSE	Root Mean Square Error
NMF	Non-negative Matrix Factorization
ASR	Automatic Speech Recognition
MNMF	Multichannel Nonnegative Matrix Factorization
MVDR	Minimum Variance Distortionless Response
DNN	deep neural network
FFT	Fourier transform
POOL	pooling
conv	convolutional
FC	Fully Connected
TSNR	Two-Step Noise Reduction
HRNR	Harmonic Regeneration Noise Reduction
BPF	Band Pass Filter

1. Introduction

The communication among humans can be done in several ways. Among them, written text, images, and speech are the most significant sources of information. One of the fundamental functions of human beings is to communicate with each other through speech. The speech acts as the most suitable and

proficient source for communication. Speech expresses both the linguistic contents and other helpful information, like the mood of the speaker. The communication among humans is precise and easy if the listener and the speaker are close to each other in a gentle environment [27]. The listener suffers to understand the speech if the listener and the speaker are in distance or in a noisy background. The background noise in a speech signal degrades the lucidity and the worth of the speech. A noisy environment decreases the ability of the speakers and listeners to communicate with each other. To reduce the effects of this problem, speech enhancement techniques have been developed. Speech enhancement is a technique, which processes the noisy speech signals and aims to improve their perception. The capability of the speech enhancement methods depends on the tradeoff among the noise diminution and the speech alteration. Depending on the type of applications, the aim of the speech enhancement may vary, like to boost the voice communication device's performance, to improve intelligibility, to increase the overall speech quality, to decrease listener weakness, and so on. As a result, speech enrichment is required to evade the speech quality degradation and to address the drawbacks of human auditory systems [28].

Depends on the number of available microphones, the speech enhancement methods have been classified into two types, namely (1) single channel speech enhancement, (2) dual or multi-channel speech enhancement [7]. The first one utilizes a single microphone, on the other hand, the second one utilizes more than one microphone. When compared to the single channel speech enhancement techniques, multichannel speech enhancement techniques offers high performance due to its easy computations and convenient implementations [7] [8]. There are several types of noise occurred in the speech signal, such as impulsive noise, interfering speech, wideband noise, periodic noise, and so on. The literature presents the various types of speech enhancement methods, such as wavelet-based methods, model-based methods, filtering techniques, and so on [9] [10] [5]. The most popular speech enhancement techniques are statistical model-based methods [11], Gaussian or super-Gaussian, subspace-based approaches [12], the spectral subtraction method [13], and Wiener_ filtering method [14].

In general, every speech enhancement technique has a similarity in the estimation methods, such as minima controlled recursive averaging [15] and minimum statistics [16]. These estimation methods find out the noisy speech signals but they do not relevant for the non-stationary environment, where these methods incorrectly assume the statistical properties of the noise. One of the prominent noise enhancement techniques is the spectral subtraction technique, which determines the noise spectrum and subtracts it from the speech segments. However, this technique is limited to the quasi-stationarity of the detected noise [17]. The other important speech enhancement technique is the NMF, which is the core of the majority of the noise departure techniques [18].

This paper introduces the speech enhancement method through the utilization of FD-AMS features and DCNN. At first, the noisy speech signal is generated and converted into frames. Then, the features are extracted and provided as the input to the DCNN, which generates the optimized estimation of the speech signal. At last, the extracted features are presented as the input to the DCNN that creates the optimized estimation of the speech signal. The proposed method is analyzed using NOIZEUS database based on the metrics, such as PESQ and RMSE.

The structure of this paper: Section 2 surveys the existing noise enhancement techniques and presents the challenges in the existing noise enhancement techniques. The proposed noise enhancement technique is described in Section 3. The results are provided in Section 4. Finally, Section 5 concludes the paper.

2. Motivation

This section presents the literature review on speech enhancement methods and provides the challenges in the existing speech recognition techniques.

2.1 Literature Review

Here, four existing related works on speech enhancement are discussed. Arul Valiyavalappil Haridas *et al*. [1] utilized an FD-AMS and the D-matrix feature extraction for enhancing the speech signal. This method estimates the high-quality speech signal. Qi He *et al*. [2] determined the predictive parameters of the speech and noise by developing the codebook-driven Wiener filtering method. Here, the spectral shape codebook was utilized for modeling the predictive parameters. This technique was more practical, vigorous, and enhances the quality of the signal. However, this technique had the complexity issue in noise reduction. In [3], a single-channel speech enhancement method was developed by Johannes Stahl *et al*. This technique had three steps. Initially, a pitch synchronous segmentation was utilized to optimize the signal analysis. Then, the frequency and class dependent prior distributions solved the harmonic

retrieval problem. At last, a joint detection-estimation framework was developed to choose the proper signal model. This technique affected speech intelligibility and speech quality. Multichannel speech enhancement for enhancing ASR is developed by Kazuki Shimada *et al.* in [4]. Here, the initialization and updating of the MNMF parameters are done by online MVDR beamforming. This technique had a good performance than the DNN-based beamforming technique.

2.2 Challenges

This sub-section deliberates the challenges in conventional speech enhancement approaches. The decisive challenge in speech enhancement is concerning the quality of the enhanced speech signal [1]. The demerit of the spectral enhancement based approach is that it introduced the musical noise [5]. The performance of the speech enhancement system is influenced by surplus input and low output. Another major problem in speech enhancement is the user responsiveness, which occurs when the user begins to speak in the absence of the resources [6].

By considering the challenges in the existing noise enhancement techniques as motivation, this paper introduces a technique for noise enhancement, which is explained in the following section.

3. Proposed Work

This section explains a proposed speech enhancement technique, which utilizes the FD-AMS features and the DCNN for enhancing the speech signal. Initially, the noise is affixed to the dirt-free speech signal, such that it becomes the noisy speech signal. Afterward, the noisy speech signal is applied to the framing, which uses the Hanning window to extracts the frames from the noisy speech signal. After that, the extracted frames are applied to the feature extraction, where the FD-AMS features are extracted. Finally, the DCNN extracts the enhanced speech signal from the extracted features. Fig. 1 demonstrates the schematic representation of the proposed speech enhancement technique.
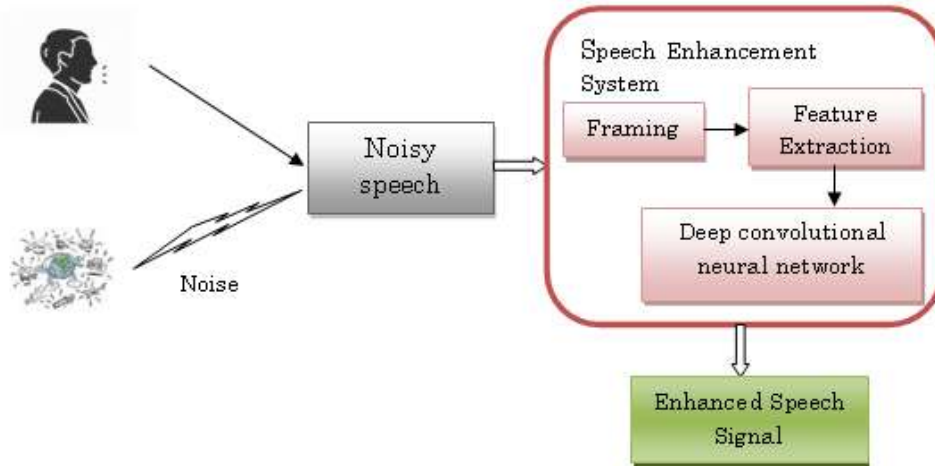
Fig. 1. *Schematic representation of the proposed system*

3.1 Pre-Processing

In pre-processing, the noisy input speech signal has been sampled and quantized to assist for the processes related to feature extraction. In general, the speech signal contains several types of noises, such as train noise, airport noise, etc.

3.2 Framing

It is the practice of converting the countless speech signal brook into blocks of samples, called frames. Here, the noisy speech signal is decomposed into 85 time-frequency segments by the BPF, each one contributing a channel. Then, the next step is the rectification, in which, the envelope for every channel is determined. At a time of the framing, the edges have the propensity to establish harmonics so; a tuning window is used for tuning the overlapped frames. Here, the Hanning window [26] is used for windowing, which eliminates unnecessary signals. The Hanning window size is $[1 \times 255]$ and the overlap rate is set as 0.5. Then, the frames are zero-padded and sent to the feature extraction procedure.

3.3 Feature Extraction

In this step, the signals are applied to FFT [26], which identifies the modulation spectrum of every frame. Then, the result of FFT and the triangular shape windows are multiplied, and the product of every frame is added to create the spectrum amplitudes as AMS features [19].

The AMS feature vector is denoted as $v(m,n)$. Here, the small alterations in the frequency and the time domains are taken into consideration, which includes the delta features in the feature vector. Equation (1) represents the feature vector obtained by the delta-AMS feature.

$$V(m,n) = [v(m,n), \Delta v_M(m,n), \Delta b_r(m,n)] \tag{1}$$

$$\Delta v_M(1,n) = v(2,n) - v(1,n), \text{ when } m = 1 \tag{2}$$

$$\Delta v_M(m,n) = v(m,n) - v(m-1,n) \tag{3}$$

where, the number of frames is represented as n and m be the length of the window and m varies from $2,...,M$. $\Delta v_M(m,n)$ is the delta feature calculated among the frequency and time. Then, the fractional calculus [20] is included in the delta-AMS feature and denoted as FD-AMS features. Hence, the equation (3) has been re-written as,

$$v(m,n) - v(m-1,n) \cong B^\sigma[\Delta v(m,n)] \tag{4}$$

where, $B^\sigma[\Delta v(m,n)]$ be the fractional calculus. The results of the FD-AMS is denoted as,

$$\Delta v_M(m,n) = B^\sigma[\Delta v(m,n)] \tag{5}$$

$$\Delta v_M(m,n) = \Delta v(m,n) - \frac{1}{2}\Delta v[(m-1,n)] - \frac{1}{6}(1-\sigma)[\Delta v(m-2,n)] - \frac{1}{24}\sigma(1-\sigma)(2-\sigma)\Delta v(m-3,n) \tag{6}$$

3.4 Noise Level Identification

In this step, the FD-AMS features are utilized for identifying the noise. In this step, the matrix $R(t)$ is formed by multiplying the columns and rows of the corresponding frames.

$$R(t) = \Delta^v AMS(t) \times [\Delta^v AMS(t)] \tag{7}$$

$$A(t) = R(t) \times \mu[\Delta^v AMS(t)] \tag{8}$$

where, the mean value of FD-AMS feature in the ith column is denoted as $\mu[\Delta^v AMS(t)]$. Then, the matrix $A^*(t)$ is calculated as given in equation (9),

$$A^*(t) = \frac{RealA(t)}{\mu(A(t))} \tag{9}$$

Afterward, the threshold value is generated by multiplying the matrix $A^*(t)$ and the constant 0.05, which is denoted in equation (10).

$$Threshold = 0.05 \times \mu(A) \tag{10}$$

Then, depending on the threshold condition, the original and noisy frames are differentiated.

$$N(t) = \begin{cases} Noisy; & \text{if } A^*(t) \leq Threshold \\ Original; & \text{if } A^*(t) \geq Threshold \end{cases} \tag{11}$$

If the matrix $A^*(t)$ is greater than or equal to the threshold then, the frame is original, otherwise, the frame has noise. Depending on the noisy frames, the optimization parameter is chosen for extracting the speech signal.

3.5 Noise Estimation

In this step, the mean value of the noise frames is identified and multiplied with the original speech signal.

$$\mu[N(t)] = \sum_{t \in N(t)}^{t=1} {}^x N(t) \tag{12}$$

$$N(t) = \begin{cases} t; & \text{if noisy} \\ 0; & \text{otherwise} \end{cases} \tag{13}$$

$$Y = \mu[N(t)] \times N(t) \tag{14}$$

where, x is the index value of the noisy frames, the number of frames in the noisy signal is denoted as t, and $\mu[N(t)]$ be the mean of the index value of the noisy frames.

3.6 FD-AMS Feature Subtraction and Noise Removal

In this step, the noisy frames are deducted from the noisy signal and the spotless signal is determined.

$$O(t) = N(t) \ Y \ (t) \tag{15}$$

where, $N(t)$ and Y (t) are the noisy speech signal and the estimated noise signal, respectively.

$$O(t) = \begin{cases} 0; & \text{if } O < 0 \\ O; & \text{Otherwise} \end{cases} ; 1 \leq t \leq 85 \tag{16}$$

Then, the DCNN identifies the optimized δ value.

$$w = (1 \quad \delta) \times O + \delta \quad \Delta^v AMS \tag{17}$$

where, the optimizing parameter is denoted as δ, the synthesized signal is represented as w and $\Delta^v AMS$ represents the FD-AMS feature.

3.7 Optimized Estimation of Speech Signal using DCNN

Here, DCNN is used for determining the optimized value of δ for extorting the optimized speech signal. The architecture of DCNN [21] has three layers, namely POOL layers, conv layers, and an FC layer. Each layer has its own functions.

Conv layers [31]: In these layers, a convolution operation is applied to the input, and the output obtained from these layers is provided to the next layer. The conv layers obtain the patterns in the input feature vector using the conv filters.

POOL layers [31]: It is a non-parametric layer and it does not have bias and weights. Pooling layer minimizes the data dimension for decreasing the number of parameters and computation.

FC layers [31]: FC layers connect each neuron in one layer to each neuron in another layer.

Finally, the output obtained from the DCNN is the optimized δ value.

4. Results Analysis

This section provides the results of the proposed method and the existing speech enhancement methods using the performance metrics.

4.1 Experimental Setup

The proposed speech enhancement approach is analyzed using NOIZEUS [22]. The evaluation metrics considered for analyzing the performance of the proposed technique are RMSE [29] and PESQ [30]. Also, the proposed method is compared with the high-tech speech enhancement systems, such as DF-AMS [1], TSNR [23], HRNR [23], Spectral Subtraction [24], and NMF [25].
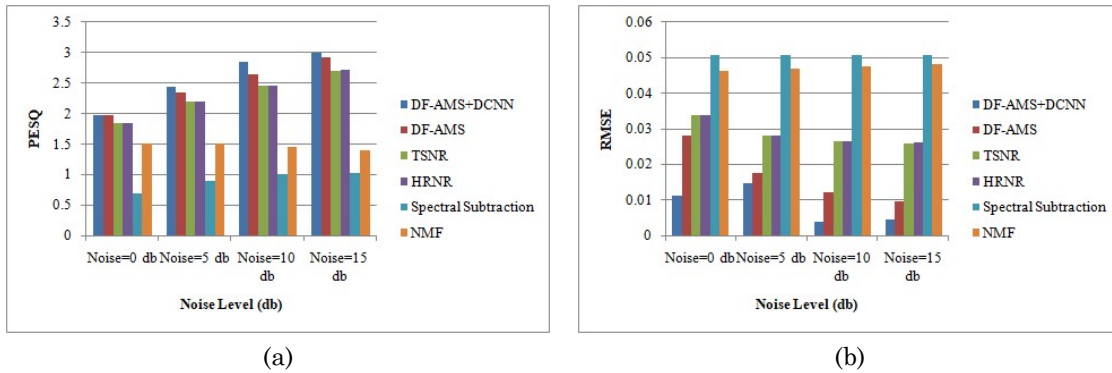
4.2 Comparative Analysis

For the experimentation, three types of noises, like Babble, airport, and car are included in the input speech signal and the performance is analyzed for various noise levels. Fig. 2 illustrates the comparative analysis using airport noise contaminated speech signal. Fig. 2 (a) depicts the PESQ analysis of the comparative methods. For 0 dB noise, the PESQ of DF-AMS+DCNN is 2.1267, DF-AMS is 1.9653, TSNR is 1.8153, HRNR is 1.8137, Spectral Subtraction is 0.7264, and NMF is 1.5591. The proposed DF-AMS+DCNN has the PESQ of 3.2898, for the noise level 15 dB. For the same noise level, the existing Spectral Subtraction, which shows the worst performance among the comparative methods, has the minimum PESQ of 1.0059. From the analysis, it can be exposed that the proposed method has the maximum PESQ, which shows the effectiveness of the proposed classifier. Fig. 2 (b) shows the RMSE analysis of the comparative methods. The proposed DF-AMS+DCNN has the RMSE of 0.0156 at noise level 0 dB, which is minimum than the RMSE of the existing techniques. When the noise level is 15 dB, the comparative methods have the RMSE of 0.0035, 0.0095, 0.0259, 0.0261, 0.0507, and 0.0479. Here also, the proposed method has a minimum RMSE than the existing methods.

(a)   (b)
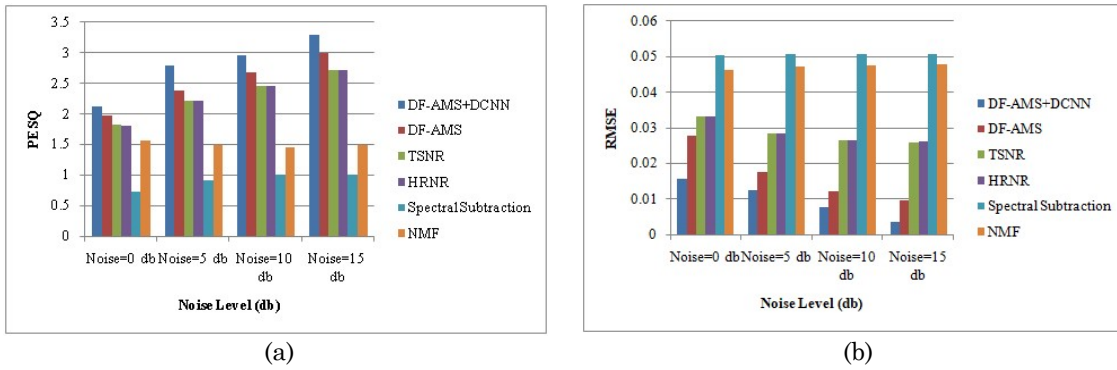
Fig. 2. *Comparative analysis using input signal with airport noise*

Fig. 3 illustrates the comparative analysis using the input speech signal with Babble noise. Fig. 3 (a) depicts the comparative analysis based on PESQ. At noise level 0 dB, the PESQ of the comparative methods, such as DF-AMS+DCNN, DF-AMS, TSNR, HRNR, Spectral Subtraction, and NMF is 1.9822, 1.9657, 1.8486, 1.8403, 0.6893, and 1.5056. The proposed DF-AMS+DCNN has the PESQ of 2.9944, for the noise level 15 dB. On the other hand, the existing approaches have the minimum PESQ than the proposed approach. Fig. 3 (b) shows the analysis of the comparative methods based on RMSE. The proposed DF-AMS+DCNN has the best performance with the RMSE of 0.0111. The existing DF-AMS has the best performance among the other existing methods with the RMSE of 0.0281, at the noise level 0 dB. At the noise level 15 dB, the comparative methods have the RMSE of 0.0044, 0.0095, 0.0259, 0.0261, 0.0507, and 0.0482.

(a)   (b)

Fig. 3. *Comparative analysis using input signal with Babble noise*

Fig. 4 illustrates the comparative analysis using the input speech signal with car noise. Fig. 4 (a) depicts the analysis of the comparative methods using PESQ. For the noise level 0 dB, the proposed DF-AMS+DCNN has the best performance with the PESQ of 1.9975. The existing Spectral Subtraction has the worst performance with the PESQ of 1.0549. Similarly, for the noise level of 15 dB, the proposed DF-AMS+DCNN has the PESQ of 2.9941. On the other hand, the existing methods have a minimum PESQ than the proposed approach. Fig. 4 (b) depicts the analysis of the comparative methods using RMSE. From the analysis, it is revealed that the proposed method has a better performance than the existing methods.

(a)   (b)
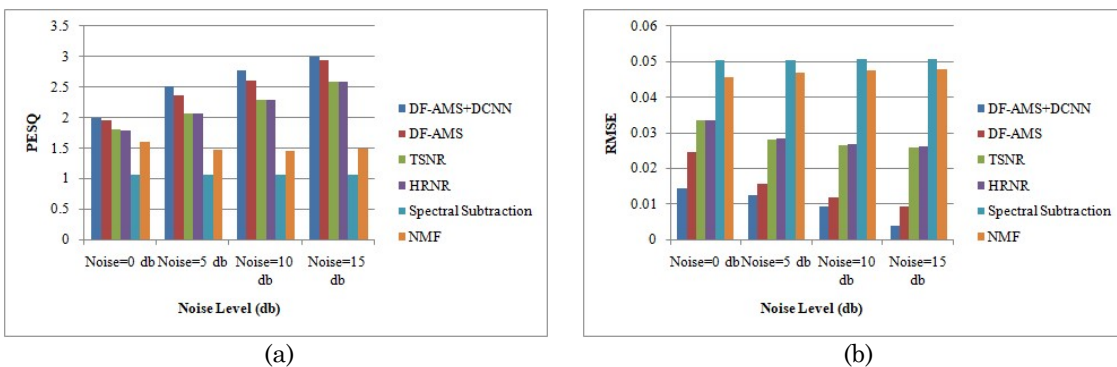
Fig. 4. *Comparative analysis using input signal with car noise*

5. Conclusion

This paper presents a speech enhancement technique with the utilization of FD-AMS features and the DCNN. Initially, the noise signal is incorporated into the hygienic speech signal and the noisy speech signal is generated. Then, the next step is the framing, in which the FD-AMS features are extorted from the frames. Finally, the extracted features are provided as the input to the DCNN, which generates the optimized estimation of the speech signal. The efficiency of the proposed speech enhancement approach is analyzed using NOIZEUS database for the performance metrics, PESQ and RMSE. Also, the comparative analysis is performed with the existing speech enhancement techniques. From the results, it can be exposed that the proposed classifier obtains maximum PESQ and minimum RMSE than the existing methods, which shows the effectiveness of the proposed method.

Compliance with Ethical Standards

Conflicts of interest: Authors declared that they have no conflict of interest.

Human participants: The conducted research follows the ethical standards and the authors ensured that they have not conducted any studies with human participants or animals.

References

[1] Arul Valiyavalappil Haridas, Ramalatha Marimuthu, and Basabi Chakraborty, "A Novel Approach to Improve the Speech Intelligibility Using Fractional Delta-amplitude Modulation Spectrogram," Cybernetics and Systems, vol. 49, no. 7, pp. 1-31, 2018.

[2] Qi He, Feng Bao, and Changchun Bao, "Multiplicative update of auto-regressive gains for codebook- based speech enhancement," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 3, pp. 457–68. 2017.

[3] Johannes Stahl and Pejman Mowlaee, "A Pitch-Synchronous Simultaneous Detection-Estimation Framework for Speech Enhancement," IEEE/ACM Transactions on Audio, Speech, and Language Processing (Volume: 26 , Issue: 2 , Feb. 2018) Page(s): 436 - 450

[4] Kazuki Shimada, Yoshiaki Bando, Masato Mimura, Katsutoshi Itoyama, Kazuyoshi Yoshii, and Tatsuya Kawahara, "Unsupervised Speech Enhancement Based on Multichannel NMF-Informed Beamforming for Noise-Robust Automatic Speech Recognition," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 27, no. 5, pp. 960 – 971, May 2019.

[5] Amol Chaudhari and S. B. Dhonde, "A review on speech enhancement techniques," In Proceedings of the International Conference on Pervasive Computing (ICPC), Pune, India, 2015.

[6] Neha Chadha, R.C. Gangwar, and Rajeev Bedi, "Current Challenges and Application of Speech Recognition Process using Natural Language Processing: A Survey," International Journal of Computer Applications, vol.131, no.11, pp. 28-31, December2015.

[7] Navneet Upadhyay, Abhijit Karmakar, "An Improved Multi-Band Spectral Subtraction Algorithm for Enhancing Speech in Various Noise Environments", In Proceedings of the International Conference on Design and Manufacturing, IConDM 2013, Elsevier, vol. 64, pp. 312-321, 13 November 2013.

[8] Yi Zhang, Yunxin Zhao, "Real and imaginary modulation spectral subtraction for speech enhancement", Journal on Speech Communication, Elsevier, vol. 55, pp. 509-522, 6 November 2012.

[9] Dr. Shaila D. Apte, "Speech Processing Applications", Speech and Audio Processing, Wiley India Edition.

[10] Sonia Sunny, David Peter S, K Poulose Jacob, "A New Algorithm for Adaptive Smoothing of Signals in Speech Enhancement", In Proceedings of the International Conference on Electronic Engineering and Computer Science, IERI Procedia, Elsevier, vol. 4, pp. 337–343, 12 December 2013. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[11] Deng, F., C. C. Bao, and F. Bao., "A speech enhancement method by coupling speech detection and spectral amplitude estimation," INTERSPEECH, Lyon, France, August 25–29, 3234–3238, 2013.

[12] Lev-Ari, H., and Y. Ephraim, "Extension of the signal subspace speech enhancement approach to colored noise," IEEE Signal Processing Letters, vol. 10, no. 4, pp. 104–106, 2003.

[13] Li, C., and W.-J. Liu., "A novel multi-band spectral subtraction method based on phase modification and magnitude compensation," In Proceedings of the IEEE International Conference on Acoustic, Speech, Signal Processing (ICASSP), Prague Congress Center, Prague, Czech Republic, May 22–27, 4760–4763, 2011.

[14] Loizou, P. C., "Speech enhancement: Theory and practice," Boca Raton, FL, USA: CRC Press, 2007.

[15] Cohen, I., "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," IEEE Transactions on Speech and Audio Processing, vol. 11, no. 5, pp. 466–475, 2003.

[16] Martin, R., "Noise power spectral density estimation based on optimal smoothing and minimum statistics," IEEE Transactions on Speech and Audio Processing, vol. 9, no. 5, pp. 504–512, 2001.

[17] Tobias, G., Y. Xin, J. M. M. Jessica, and B. Stefan, "Speech enhancement for hearingimpaired listeners using deep neural networks with auditory-model based features," In Proceedings of the 24th European Signal Processing Conference (EUSIPCO), Budapest, Hungary, August 16, 2300–2304, 2016.

[18] Wood, S. U. N., J. Rouat, S. Dupont, and G. Pironkov, "Blind speech separation and enhancement with GCC-NMF," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 4, pp. 745–55, 2017.

[19] Healy, E. W., S. E. Yoho, Y. Wang, and D. Wang, "An algorithm to improve speech recognition in noise for hearing-impaired listeners," Journal of Acoustical Society of America, vol. 134, no. 4, pp. 3029–3038, 2013.

[20] Bhaladhare, P. R., and D. C. Jinwala, "A clustering approach for the l-diversity model in privacy preserving data mining using fractional calculus-bacterial foraging optimization algorithm," Advances in Computer Engineering, 2014.

[21] G. Sateesh Babu, P. Zhao, and X.-L. Li, "Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life," Lecture Notes Comp. Scie. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9642, pp. 214–228, 2016.

[22] Loizou, P, "NOIZEUS: A noisy speech corpus for evaluation of speech enhancement algorithms," http://ecs.utdallas.edu/loizou/speech/noizeus/ .

[23] Plapous, C., 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. 2098–108. 2006.

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

[25] Nikolay, L., and K. Mikhail, "Non-negative matrix factorization with linear constraints for single-channel speech enhancement," Journal of Computer Science, pp. 446–450, 2013.

[26] S. R. Chintakindi, O. V. S. R. Varaprasad and D. V. S. S. S. Sarma, "Improved Hanning window based interpolated FFT for power harmonic analysis," TENCON 2015 - 2015 IEEE Region 10 Conference, Macao, 2015, pp. 1-5.

[27] Q. Huang, C. Bao, X. Wang and Y. Xiang, "DNN-Based Speech Enhancement Using MBE Model," 2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC), Tokyo, 2018, pp. 196-200.

[28] P. Hannon, M. Krini and I. Schalk-Schupp, "Advanced speech enhancement with partial speech reconstruction," 21st European Signal Processing Conference (EUSIPCO 2013), Marrakech, 2013, pp. 1-5.

[29] B. Sridhar and M. Z. Ali Khan, "RMSE comparison of path loss models for UHF/VHF bands in India," 2014 IEEE Region Symposium, Kuala Lumpur, 2014, pp. 330-335.

[30] D. Sharma, L. Meredith, J. Lainez, D. Barreda and P. A. Naylor, "A non-intrusive PESQ measure," 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Atlanta, GA, 2014, pp. 975-978.

[31] E. Jung, P. Chikontwe, X. Zong, W. Lin, D. Shen and S. H. Park, "Enhancement of Perivascular Spaces Using Densely Connected Deep Convolutional Neural Network," IEEE Access, vol. 7, pp. 18382-18391, 2019.