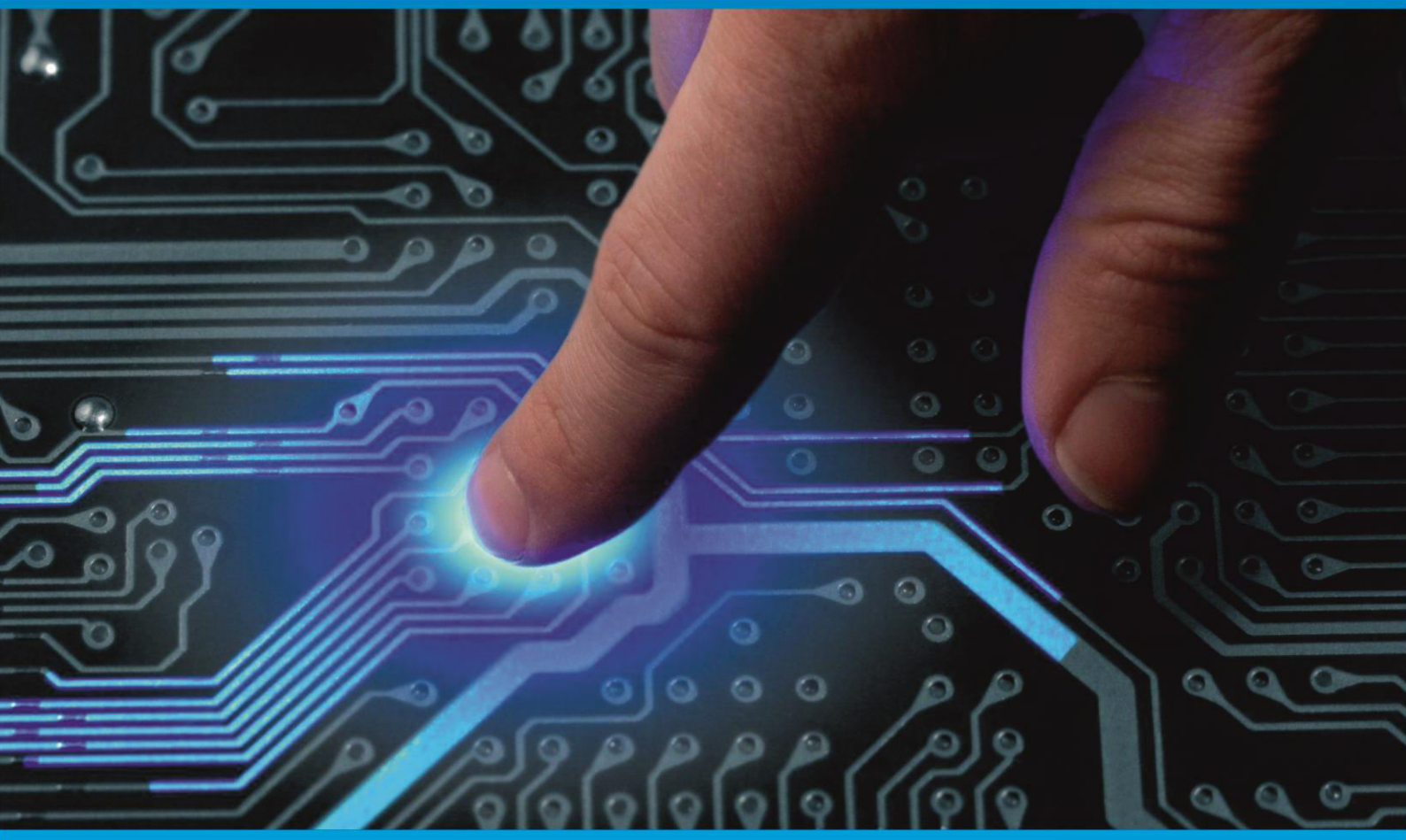
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 4, April 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904045 |

Background Noise Suppression Algorithms for Speech Communications

Vijayakumar Pattanashetti[1], Paramjeet Singh[1], Manasa H[1], Vishal Mishra[1], Gincy Varghese C[2]

UG Student, Dept. of ECE, SOET, CMR University, Bengaluru, India[1]

Assistant Professor, Dept. of ECE, SOET, CMR University, Bengaluru, India[2]

ABSTRACT: Background noise during online meetings, video conferencing, audio and calls impact mobile communications adversely. Randomness of the noise, which is often described in terms of its first and second order statistics [1], makes it difficult to remove without employing dedicated, sophisticated background noise suppression algorithms. Noise suppression seen attention and developments in recent years, including for active and background noises, for numerous applications such as digitalization of historical audio, consumer applications, military, mission critical communications and hearing aids. Accuracy and latency play very significant role in the real-time implementation of such background noise suppression algorithms. The paper attempts to review few such noise suppression algorithms, including raw signal processing techniques to deep learning models, along with the uses case scenarios and comparison among them. Such noise suppression systems may find useful applications in voice and speech recognition in AI based voice assistants such as Alexa and Siri as well.

KEYWORDS: noise suppression, deep learning, speech enhancement, speech processing, recurrent networks, spectral subtraction

I. INTRODUCTION

Background noise is any sound other than the primary sound such as traffic noise, alarms, extraneous speech, animals' noises, and noises from refrigerators, air conditioners, power supplies and motors. Reduction or suppression of such noises contribute significantly towards effective communication, and noise control which is a significant consideration, particularly in sound reproduction and ultrasound based medical imaging. Various speech enhancement, speech recognition and deep learning techniques and algorithms are employed to eliminate such background noises in real-time.

Speech enhancement involves processing of speech to improve the quality and intelligibility. It depends on noise or noise spectrum estimation and reduction, which is a very challenging problem as noise characteristics may vary randomly in time. It is therefore required to develop a versatile algorithm that works in diversified environments. The most significant efforts that can be made by speech enhancement techniques is in reducing the noise degradation of the signal [2]. The degradation could be due to room echoes, additive random noise, multiplicative or convolutional noise, parallel speakers, etc... So, the choice of algorithm or technique depends on the context or scenario of the problem. The paper reviews such context-based background noise suppression algorithms.
Sections 2, 3 and 4 include review on the raw speech processing techniques, sophisticated speech enhancement algorithms and deep learning algorithms respectively. Further in section 5, speech enhancement/noise suppression considerations and metrics are discussed.

Nodes in MANET have limited battery power and these batteries cannot be replaced or recharged in complex scenarios. To prolong or maximize the network lifetime these batteries should be used efficiently. The energy consumption of each node varies according to its communication state: transmitting, receiving, listening or sleeping modes. Researchers and industries both are working on the mechanism to prolong the lifetime of the node's battery. But routing algorithms plays an important role in energy efficiency because routing algorithm will decide which node has to be selected for communication.

The main purpose of energy efficient algorithm is to maximize the network lifetime. These algorithms are not just related to maximize the total energy consumption of the route but also to maximize the life time of each node in the network to increase the network lifetime. Energy efficient algorithms can be based on the two metrics: i) Minimizing total transmission energy ii) maximizing network lifetime. The first metric focuses on the total transmission energy

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904045 |

used to send the packets from source to destination by selecting the large number of hops criteria. Second metric focuses on the residual batter energy level of entire network or individual battery energy of a node [1].

II. SPEECHENHANCEMENT TECHNIQUES AND ALGORITHMS

A. *Tranform Domain Approach*

In [3], generalized transform domain approach for noise reduction is presented. There exists many different transforms, however Fourier, cosine, Karhunen-Loeve andHadamardtranforms caught research attention for noise reduction. The typical flow of transform domain approach is as follows: i) Reformulation of the noise reduction problem into a more generic transform domain problem, where any unitary matrix can be used to serve as a transform, and ii) Design different optimal and suboptimal filters in thegeneralized transform domain. The important points to be considered in signal denoising applications include eliminating noise from signal to improve the SNR, and preserving the shape and characteristics of the original signal. The approach is used to eliminate the additive noise from noisy speech signal to make speech immune to additive noise. Also, it may be used to extract voice of interest among multiple simultaneous voices. That is, when the pitch values of speakers are given, Discrete Fourier Transform (DFT) of the mixed signal is determined and the harmonics of the fundamental frequencies of the speakers are tracked. Then, DFT outputs are isolated, and Inverse DFT (IDFT) is applied to obtain the voice of interest.

B. *Spectral subtraction*

Spectral subtraction involves subtraction of estimated noise spectrum from the spectrum of the noisy speech. The noise power is estimated during the absence of voice, and for a very accurate estimate, the signal recovery is accurate. The magnitude of the signal is combined with the phase obtained for noisy signal. Then, the signal is obtained by applying Inverse Discrete Fourier Transform (IDFT). The noise correlated with signal can be removed using adaptive cancellation [4]. Few variants of spectral subtraction method include Magnitude Spectral Subtraction (MSS), Nonlinear Spectral Subtraction (NSS), and Minimum Mean Square Error Estimation (MMSE), Multi-Band Spectral Subtraction (MBSS) and Wide-Band Spectral Subtraction (WBSS).

In [5], the author(s) proposed slight modification to the spectral subtraction that uses spectral smoothing, formant intensification and comb filtering for better performance and quality perception. However, it is suggested to consider the inconsistency between objective and subjective measures.

In [6], the author(s) presented a spectral subtraction method in short-time modulation domain as a modification to overcome the musical noise due to mismatch between the estimated noise and the true noise in spectral subtraction. It is suggested to use of other advanced techniques such as MMSE, Kalman filtering etc. in modulation domain to improved performance.

In [7], spectral subtraction is performed on the real and imaginary spectra separately in the modulation frequency domain (MRISS) which resulted in better performance than other variants of spectral subtraction.

C. *Stationary and adaptive filters*

Stationary filters use a bank of notch filters such as comb filters for removal of periodic noise. Applications include audio signal processing, including delay, flanging. Adaptive filters are the forward prediction error filters used as the inverse filters to filter out periodic noises. The frequency response of the filter will have minima at the frequencies of the periodic components. The adaptive noise cancellation assumes the use of two microphones. A primary microphone is used for the noisy input signal, while asecondary microphone is used for noise that is not correlated to the speech signal, but is correlated to the noise at the primary microphone. Adaptive algorithms are used widely because they converge rapidly. There exist two popular adaptive filtering algorithms viz., recursive least squares (RLS) algorithm & normalized least mean squares (NLMS) algorithm. But these algorithms have high computational complexity, and stability & poor adaptive rate problems respectively [8]. Therefore, tradeoff between computational complexity and fast convergence specific to the application decides the adaptive algorithm to be used.

In [9], algorithm based on adaptive filtering with averaging (AFA) is proposed for noise cancellation. It concluded that AFA algorithm has low computational complexity, high convergence rate comparable to that of the RLS algorithm and possible robustness in fixed-point implementations.

The algorithm, proposed in [10], adaptively estimates the instantaneous noise spectrum from an autoregressive signal model as opposed to the widely-used constant noise spectrum fingerprint approach. The adaptive algorithm is able to work without user interaction and significantly reduces stationary & non-stationary noises in real-time.

D. *Comb filters*

A comb filter is implemented by adding a delayed version of the signal to itself, leading to constructive and destructive interference. The frequency response of a comb filter is a series of regularly spaced notches which appears

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904045 |

as a comb, hence the name. Comb filters are used in delay, flanging and digital waveguide synthesis. These are of two types viz., feedforward and feedback, based on the direction in which signals are delayed before they are added to the input. These may be implemented in discrete time or continuous time, however discrete comb filters are widely employed in speech applications. A comb filter is used in pitch separation to extract desired speaker's harmonics when the pitch values are known [11]. Thus, comb filters find applications for noise suppression in the environments where noise is due to multiple speakers.

E. *Blind signal separation*

Proposed blind signal separation algorithm, in [12], uses design of the separation matrix and a post-filtering noise canceller, implemented in frequency domain for better computational efficiency. The algorithm promises real-time functionality when an FPGA-like hardware accelerator is used for computing critical functions of the algorithm. The algorithm implementation on embedded system finds applications in modern multimedia systems.

F. *HE-LPC based speech enhancement algorithm*

In [2], background noise reduction algorithm based on Harmonic Excitation Linear Predictive Coder (HE-LPC) speech coder is proposed. It discusses speech coding system that is capable of reducing the background noise during the HE-LPC speech analysis and synthesis process without theuse of any additional noise suppression or speech enhancement algorithms. The HE-LPC speech coder is modified to isolate the background noise from the speech signal as efficiently as possible. As HE-LPC speech coder is a parametric representation of the speech signal in the frequency domain, information such as pitch, spectral envelope and residual harmonic spectral amplitudes may be used to perform background noise reduction. These harmonic spectral amplitudes are modified such that they account for the presence of background noise. This increases the signal to noise ratio for each harmonic. It is concluded that the proposed algorithm improved the speech quality and the intelligibility significantly under ambient background noise conditions.

G. *Mel filter*

In [13], application of mel filter for noise reduction in the speech signal is presented. This takes the advantage of accurate estimation of mel filter, thus improving SNR and speech quality. Use of mel filter requires reconstruction of the speech signal by applying Discrete Short Time Fourier Transform (DSTFT). Post removal of noise using the mel filter, it is used to convert time domain signal to frequency domain signal for reconstruction purpose. Major advantages of use of mel filter include mel-frequency (perceptual scale) scaling, and better resolution at lower frequencies & low resolution at higher frequencies.

III. DEEP LEARNING ALGORITHMS FOR SPEECH ENHANCEMENT

A. *Separating background noise with deep learning*

In [14], application of deep learning for noise suppression is introduced, and a regression method that learns to produce a ratio mask for every audio frequency which deletes extraneous noise leaving back the human voice is explained. The typical flow include: i) Data acquisition: Synthesis of noisy speech dataset by mixing clean speech with noise, ii) Training the DNN, and iii) Inference: Produce a mask (binary, ratio or complex). In the following years, the deep learning architectures evolved in multiple dimensions for better performance.

In [15], it is concluded that the DNN architecture presented performs comparatively better under diversified noise environments at 2Hz. It is stated that average Mean Opinion Score (MOS) was found to be 1.4 points more on noisy speech which is against to the typical case.

B. *Speech enhancement using recurrent networks*

[16] proposed a deep recurrent auto encoder neural network model for denoising in robust automatic speech recognition systems. It was featured that the model out performs compared to the basic DNN models and is flexible enough to learn any type of noise, but limited by the training dataset. That is, the model's performance is not limited by any assumptions such as how noise affects the signal or the type of noise environments. The paper alsostudies the effect of secondary parameters of the architecture on the performance, and concludes that both the temporally recurrent connections and multiple hidden layers are essential for both training and generalization to unseen noise types.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904045 |

C. *Speech enhancement using bidirectional LSTM networks*

In [17], the context sensitivity of the bidirectional long short-term memory (BLSTM) networks is exploited, by combining BLSTM based probabilistic feature generation with the bottleneck leading to lower error rates in spontaneous speech recognition. The bottleneck principle facilitates generation of tandem feature vectors of arbitrary size employing the activations of a narrow hidden layer as features. The activations of the output layer are ignored, and only the activations of the forward and backward bottleneck layer are processed during the feature extraction. Then, forward and backward bottleneck layer activations are concatenated to form one feature vector which is further decorrelated and dimensionality reduced employing Principal Component Analysis (PCA).

D. *Hybrid approaches to speech enhancement*

The paper, [18] presents DNN to estimate the ideal ratio masks at individual time-frequency bins. These are then used to design potential speech enhancement systems for drones. The proposed DNN-TF algorithm deduces the noise dominance probabilities at individual time-frequency bins from the DNN estimated time-frequency ratio masks, and incorporates them into a time-frequency spatial filtering to reduce ego noise. The performance of the proposed algorithm is better than single and multi-channel DNN based speech enhancement techniques. The algorithm makes the best use of the information on the direction of arrival of the target, and the time-frequency sparsity of the ego noise and speech signals to effectively reduce the ego noise in very low-SNR cases with robust performance for very close direction of arrival of the target to the ego noise sources. The algorithm finds best applications in embedded speech systems, aerial vehicles and remote non-aerial vehicles.

In [19], complex linear coding (CLC) based DNN model to suppress noise in monophonic speech is presented, known as CLCNet. The algorithm involves three steps: i) Define Linear Predictive Coding (LPC) inspired Complex Linear Coding (CLC) which is applied in the complex frequency domain, ii) Design framework that incorporates complex spectogram input and coefficient output, and iii) Define low latency parametric normalization for complex valued spectrograms. It was demonstrated to reduce noise within individual frequency bands while preserving the speech harmonics. The proposed algorithm proved that it works best for low latency, real-time, low SNR scenarios, and/or low-resolution spectrograms.

E. *GRU based speech enhancement method:*

In [20], a novel teacher-student learning algorithm for preprocessing, online noise tracking of improved minima controlled recursive averaging (IMCRA) and deep learning of nonlinear interactions between speech and noise is proposed. Initially, a teacher DNN learns the ideal ratio masks (IRMs) using simulated training pairs of clean and noisy speech dataset. Then, a student DNN learns the improved speech presence probability using the estimated IRMs from the teacher DNN in the IMCRA method. It was concluded that the bidirectional gated recurrent units (BGRUs) based student DNN achieves 18.85% less relative word error rate (WER) than the unprocessed, untrained system for a real test dataset. Whereas, zero-latency student DNN in causal mode results in 7.94% reduction in relative

WER, and the computing cycles required were found to be 670 times less than BGRU based student DNN. If the zerolatency student DNN is compactly designed in a causal processing mode under a complex and non-causal teacher DNN, noisy speech data can be directly used to adapt the regression based enhancement algorithm to further improve speech recognition performance for noisy speech.

IV. CONSIDERATIONS AND METRICS

A. *Latency*

Latency is defined as the time required to process the request and respond accordingly, or the time elapsed between the request and the response. Low latency is very important in mobile communications as humans can tolerate up to 200ms of end-to-end latency when conversing, otherwise while on calls [15]. Compute and network are two important considerations that impact the latency. Compute latency is the latency contributed by the computations required by the algorithms, thus makes DNNs challenging.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904045 |

B. *DNN Architecture & Audio Sampling Rate*

The number of layers and hyper-parameters in DNN accounts for the latency. For better performance, DNN needs to have significantly more layers and hyper-parameters, thus contributing to latency. Also, the performance of DNN depends on the audio sampling rate as well. The higher the sampling rate, the more hyper parameters needed for DNN. However, narrowband, where most of our mobile communications happen, requires less data per frequency making DNN acceptable for real-time speech enhancement [21].

C. *Performance Measure*

Testing the performance of speech enhancement algorithms is challenging as it's subjective, i.e. varies from person to person depending upon their hearing capabilities due to age, training or other factors. Due to lack of open and consistent benchmarks for noise suppression, most researchers use Mean Opinion Score (MOS), Perceptual Evaluation of Speech Quality (PESQ), and Short-TimeObjective Intelligibility (STOI) measure for comparing results, where the simple metric score is obtained for the given set of clean and noisy speech [22].

V. CONCLUSION

In this paper, we reviewed background noise suppression or speech enhancement techniques and algorithms that are employed to eliminate degradation of quality of the speech due to noise. Transform domain approaches are quite basic ones with minimal performance, whereas spectral subtraction is widely used for specific applications [7]. However, musical noise due to mismatch between the estimated noise and the true noise limits its performance [23]. Hence spectral subtraction seen many modifications to improve its performance for a specific application by eliminating its limitations. Stationary or adaptive filters are used in the scenarios where periodic noise is expected, such as environments where noise is due to periodic rotation of machinery. Since its required to have versatile algorithms for better performance in diversified environments, the attempt of deep learning and sophisticated speech enhancement algorithms is also presented.

Noise suppression is seeing recent advancements and research attention, while offering many challenges. Suppression of both inbound and outbound noise in real-time is one such challenge, that is elimination of noise being received at both mic and speakers. Due to acoustic and voice variances not typical for these noise suppression algorithms, inbound noise suppression becomes highly challenging [15]. Inbound noise suppression is scopeful for military, industrial and consumer applications.

ACKNOWLEDGMENT

We extend our sincere gratitude and thanks to Prof.Gincy V., Asst. Professor, Dept. of ECE, SOET, CMR University, Bengaluru for her guidance and support. We also thank Prof.Muralishankar R. for his guidance and encouragement.

REFERENCES

1. J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," in *Proceedings of the IEEE*, vol. 67, no. 12, pp. 1586-1604, Dec. 1979.
2. S. Yeldener and J. H. Rieser, "A background noise reduction technique based on sinusoidal speech coding systems," *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100)*, Istanbul, Turkey, 2000, pp. 1391-1394 vol.3.
3. J. Benesty, J. Chen and Y. A. Huang, "Noise Reduction Algorithms in a Generalized Transform Domain," in *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 11091123, Aug. 2009.
4. M. Bahoura, "Pipelined Architecture of Multi-Band Spectral Subtraction Algorithm for Speech Enhancement," *Electronics*, vol. 6, no. 4, p. 73, Sep. 2017.
5. H. T. Hu, F. J. Kuo and H. J. Wang, "Supplementary Schemes to spectral subtraction for speech enhancement," *Journal on Speech Communication, Elsevier*, vol. 36, pp. 205-218, 7 January 2002.
6. K. Paliwal, K. Wo´jcican and B. Schwerin, "Single-channel Speech enhancement using spectral subtraction in the short-time Modulation domain," *Journal on Speech Communication, Elsevier*, vol. 52, pp. 450-475, 19 February 2010.
7. Y. Zhang and Y. Zhao, "Real and imaginary modulation spectral Subtraction for speech enhancement," *Journal on Speech Communication, Elsevier*, vol. 55, pp. 509-522, 6 November 2012.
8. S. Hadei and M. lotfizad, "A Family of Adaptive Filter Algorithms in Noise Cancellation for Speech Enhancement," *International Journal of Computer and Electrical Engineering*, pp. 307-315, 2010.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904045 |

9. G. Iliev and N. Kasabov, "Adaptive filtering with averaging in noise cancellation for voice and speech recognition," in *Proc. ICONIP/ANZIIS/ANNES'99 Workshop*, pp. 22-23, 2001.
10. C. Wiesener, T. Flohrer, A. Lerch and S. Weinzierl, "Adaptive Noise Reduction for Real-time Applications," in *Proceedings of the 128th Audio Engineering Society Convention*, vol. 3, 2012.
11. Nehorai and B. Porat, "Adaptive comb filtering for harmonic signal enhancement," in *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 5, pp. 1124-1138, October 1986.
12. K. Yiu and S. Low, "On a Real-Time Blind Signal Separation Noise Reduction System," *International Journal of Reconfigurable Computing*, vol. 2018, pp. 1-9, 2018.
13. Y. Manasa and R. Palaparthi, "Minimization of Noise in Speech Signal Using Mel- Filter," in *International Journal of Engineering Development and Research*, vol. 5, no. 2, pp. 2321-9939, 2020.
14. Y. Xu, J. Du, L. Dai and C. Lee, "A Regression Approach to Speech Enhancement Based on Deep Neural Networks," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 1, pp. 7-19, Jan. 2015.
15. Davit Baghdasaryan. (2018, October, 31). *Real-Time Noise Suppression Using Deep Learning* [Online]. Available: https://developer.nvidia.com/blog/nvidia-real-time-noisesuppression-deep-learning/
16. Maas, Q. Le, T. Neil, O. Vinyals, P. Nguyen and A. Ng, "Recurrent Neural Networks for Noise Reduction in Robust ASR," in *13th Annual Conference of the International Speech Communication Association 2012, INTERSPEECH 2012*, vol. 1, 2012.
17. M. Wöllmer, Z. Zhang, F. Weninger, B. Schuller and G. Rigoll, "Feature enhancement by bidirectional LSTM networks for conversational speech recognition in highly non-stationary noise," *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, 2013, pp. 6822-6826.
18. L. Wang and A. Cavallaro, "Deep Learning Assisted Time-Frequency Processing for Speech Enhancement on Drones," in *IEEE Transactions on Emerging Topics in Computational Intelligence*, pp. 1-11, 2020.
19. H. Schröter, T. Rosenkranz, A. N. Escalante-B, M. Aubreville and A. Maier, "CLCNET: Deep Learning-Based Noise Reduction for Hearing aids using Complex Linear Coding," *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020, pp. 6949-6953.
20. Y. Tu, J. Du and C. Lee, "Speech Enhancement Based on Teacher– Student Deep Learning Using Improved Speech Presence Probability for Noise-Robust Speech Recognition," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 12, pp. 20802091, Dec. 2019.
21. S. Nossier, J. Wall, M. Moniri, C. Glackin and N. Cannings, "An Experimental Analysis of Deep Learning Architectures for Supervised Speech Enhancement," *Electronics*, vol. 10, no. 1, p. 17, 2021.
22. R. Streijl, S. Winkler and D. Hands, "Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives," *Multimedia Systems*, vol. 22, no. 2, pp. 213-227, 2014.
23. Ekaterina and B. Simak, "Noise Reduction Based on Modified Spectral Subtraction Method," in *IAENG International Journal of Computer Science*, vol. 38, 2011.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details