Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 16, Issue. 1: 2025
ISSN : **1906-9685**

Implementation of Speech Enhancement Using Implicit Wiener Filter

V.Srinivasarao[1], M.Durga Reddy[2], K.Vinay Kumar[3], K.Abhilash Reddy[4],M.Deepak Kumar[5], K.Surya Dileep[6]

[1]Professor, Department of ECE, NRI Institute of Technology, Pothavarappadu, Vijayawada Rural, Andhra Pradesh, India.
[2,3,4,5] Department of ECE, NRI Institute of Technology, Pothavarappadu, Vijayawada Rural, Andhra Pradesh, India.

Abstract: Speech is the primary form of human-human communication and also enables human-machine interaction. However, quality of the speech degrades due to background noise effect. Speech enhancement algorithm improves the speech quality that is corrupted by different types of noises like additive noise, street noise, helicopter noise, drone noise, station noise, etc. In this paper, we propose a speech enhancement technique using Implicit Wiener filter. In this algorithm, noise is estimated with a first order recursive equation. This method will be tested on different noises. The performance of this proposed speech enhancement algorithm will be compared with traditional speech enhancement algorithm using objective speech quality measures.

Keywords: *Noisy Speech, Implicit Wiener Filter, Background Noise, Speech Quality*.

Introduction: The Human voice is a critical means of communication, conveying emotion, thought, and ideas across and between cultures and languages. It is degraded, however, in normal environments by ambient noise, reverberation, and other forms of interference. This degradation can cause a significant negative impact on speech intelligibility, leading to decreased effectiveness of communication, increased errors in the recognition of speech, and compromised hearing aid performance.

The requirement for speech enhancement cannot be overstated. In most applications for instance, voice assistants, speech recognition systems, and hearing aids, speech quality must be high to ensure proper recognition, effective communication, and improved user experience. Furthermore, in noisy environments, speech enhancement can significantly improve speech intelligibility, alleviating the listeners' cognitive load and facilitating improved communication in general.

Traditional speech enhancement techniques, such as spectral subtraction and Wiener filtering, have gained wide popularity for speech quality improvement. These techniques are founded upon simplistic noise models without accounting for the complex statistical nature of actual noise. Furthermore, these techniques can result in artifacts, such as musical noise, which can degrade speech quality.

During the past decade, advanced speech enhancement techniques have been proposed based on machine learning, deep learning, and signal processing techniques. They have shown positive results in the aspect of improving the quality, strength, and intelligibility of speech. Most of them require huge amounts of training data, computational resources, and human skills, thus restricting their application in the field.

In this context, the Implicit Wiener Filter (IWF) algorithm has been shown to be a good solution for speech enhancement. The Implicit Wiener Filter(IWF) is a good tool to estimate the clean speech from noisy observations using the statistical properties of speech and noise. By optimizing the mean square error between the estimated and clean speech signals, the Implicit Wiener Filter(IWF) can suppress circumstantial noise and improve speech quality.

Problem Statement: Speech signals are prone to be degraded by ambient noise, reverberation, and other interference, which will have a tendency to reduce the quality and intelligibility of the speech. Traditional speech enhancement methods, like spectral subtraction and noise reduction algorithms, work well but will tend to generate musical noise, distortion, or other artifacts. The objective of this paper is to design, develop, and implement a speech enhancement system employing an Implicit Wiener Filter to effectively reduce background noise and improve the quality of speech signals with minimum serious distortion or artifacts. The system should be able to handle diverse noise environments and speech characteristics, and provide a substantial improvement in the quality and intelligibility of speech.

Literature Survey: 1. Speech Enhancement Using Denoising Diffusion Probabilistic Models: A Review and Future Directions, M. Yang, J. Xu, P. Zhang, L. Xie 2023, doi:10.1109/TASLP.2023.3341290

This paper offers a comprehensive survey of speech enhancement techniques using denoising diffusion probabilistic models (DDPMs), with an emphasis on their groundbreaking contribution to the area. DDPMs operate by learning noise addition and iterative denoising, providing strong performance in highly non-stationary noise conditions. The paper classifies previous work into conditional and unconditional diffusion models, highlighting the advantage of conditional models for real-time speech enhancement applications. The authors explain how DDPMs surpass conventional GAN-based or transformer models in situations where noise complexity and variability are very high. One of the primary strengths of DDPMs is that they can generate clean speech signals with very low distortion, reaching near-human listening quality.

These models perform exceptionally well in overlapping noise sources and sudden interference scenarios. Moreover, DDPMs generalize fantastically even without extensive data augmentation. The work also examines the potential uses of DDPMs in voice assistants, conference networks, and hi-fi hearing aids, where speech quality is paramount. However, systems based on DDPMs have their drawbacks, such as their computationally expensive iterative processes that hinder real-time implementation. The training process is similarly time-consuming and requires huge GPU resources.

In addition, balancing denoising power and speech naturalness is challenging since too many diffusion steps may cause over-smoothing. The paper concludes by suggesting research into model compression, acceleration techniques for inference, and hybrid systems that combine DDPMs with transformer-based encoders to achieve optimized real-time performance in a wide range of applications.

2. Speech Enhancement Using Transformer Networks: A Comprehensive Survey, J. Chen, H.Zhang, Y.Luo and N.K.Le 2023, doi:10.1109/TASLP.2023.3305678

This in-depth overview delves into the increasing use of transformer-based architectures in speech enhancement. The article thoroughly documents the shift away from traditional signal processing methods and RNN-based approaches towards transformers that leverage self-attention mechanism. The authors point out a number of transformer-based models including dual-path transformers and conformer architectures that surpass state-of-the-art performance for noise reduction, dereverberation, and speech separation. By means of rigorous benchmarking and comparative analysis, the survey sets forth that transformer networks dominate in long-range dependency capture and generalization with better performance under various noise conditions.

Transformer-based models have significant advantages. Their capacity to capture intricate temporal and spectral relations without the constraints of sequential processing in RNNs enables better speech reconstruction. Transformers also provide scalability and parallelizability, so training on big data is quicker and more efficient. Applications encompass voice communication systems, hearing aids, and hardy ASR solutions for smart home devices. The work further mentions real-time optimization methods that render transformer-based models deployable under low-latency conditions. Even with these developments, there are challenges, such as the high computational requirements of transformer architectures and the need for copious and diverse training data.

The authors also comment on the challenge of fine-tuning models to reach a balance between noise reduction and speech distortion in highly dynamic scenes. Additionally, memory usage can be a practical constraint for edge devices. However, the paper concludes that the revolutionary capabilities of the architectures are testing the limits of what is possible with speech enhancement and are a point of focus for future work.

3. Deep Complex Convolutional Recurrent Network for Speech Enhancement, S. W. Cho, J. Park, and H. Ko 2021, doi:10.1109/TASLP.2021.3069535

This work introduces a state-of-the-art method with a Deep Complex Convolutional Recurrent Network (DCCRN) for speech enhancement in noisy environments. The authors developed a model that performs operations in the complex domain so that it can handle magnitude and phase information of noisy speech signals. By incorporating both convolutional and recurrent layers, the model is able to learn both local spectral details and long temporal contexts, which are necessary for successful speech restoration. Comprehensive experiments were conducted under various noises and SNR levels to verify the efficiency of the model.

The benefits of this model are two-fold. First, being in the complex domain, the DCCRN can rectify amplitude and phase distortions as well as provide more natural-sounding speech enhancement. Second, the recurrent structures allow the model to take advantage of past context, which renders it strong against dynamic noise environments. The model substantially outperforms earlier state-of-the-art systems across speech quality and intelligibility metrics, with potential to be applied to tasks such as real-time voice communication, virtual conferencing, and assistive hearing devices.

Nevertheless, real-time application in such models continues to pose difficulties. The large size of the model and associated computational intensity is challenging for the use in embedded systems and processors with limited capability. In addition, model training involves exposure to large datasets covering diverse noise levels and domain-adapted tuning for particular purposes. All such limitations notwithstanding, the paper itself is a considerable contribution to its field in revealing the possibility for complex-domain deep learning models of taking speech enhancement performance to dimensions beyond the limitations of the norm.

4. A Survey on Traditional and Deep Learning-Based Speech Enhancement Techniques, R. Kothapally, S. R. Pothuraju, and V. K. Krishna 2020, doi:10.1109/IACC48162.2020.9074100

This extensive overview by Kothapally et al. presents a broad sweep of speech enhancement methods, addressing both conventional signal processing and state-of-the-art deep learning innovations. The authors classify methods under statistical models, subspace, spectral subtraction-based methods, and contemporary deep learning approaches like DNNs, CNNs, and GANs. The authors extensively review the development of such techniques and describe the shift to hybrid approaches fusing conventional processing with data-based models. Performance tests and case studies were included to illustrate how each method excelled in varied noise conditions.

The paper highlights the benefits of incorporating deep learning techniques into conventional models, reporting gains in adaptability, noise reduction, and overall intelligibility. The survey recognizes the fact that although conventional methods such as Wiener filtering and spectral subtraction are computationally inexpensive and interpretable, deep learning models are superior when it comes to dealing with non-linearities and intricate acoustic patterns. The authors also mention applications in voice-operated systems, hearing aids, and real-time communication systems, emphasizing the applicability of speech enhancement research. However, the survey identifies some of the main challenges, including reliance on large annotated datasets for training deep learning models and the high computational requirements during both training and inference.

In addition, despite the promise that hybrid systems demonstrate, stability and prevention of speech distortion in uncontrollable noise remain issues. In spite of these challenges, this paper will be a point of reference authority for researchers intending to investigate classical and contemporary concepts of speech enhancement.

5. Supervised Speech Enhancement Using Deep Neural Networks for Robust Automatic Speech Recognition, A. Narayanan and D. Wang 2014, doi:10.1109/TASLP.2014.2314841

This in-depth study by Narayanan and Wang introduced an important advance in utilizing deep learning methods for speech enhancement in resilient automatic speech recognition (ASR). The authors used a supervised learning architecture that utilized deep neural networks (DNNs) to predict clean speech features from noisy inputs. The model was trained on large amounts of data that included varied types and levels of noise, making it able to generalize effectively under difficult acoustic environments. The paper presented an extensive comparison of the conventional signal processing methods with the recently proposed deep learning-based method, showcasing significant enhancements in speech recognition accuracy in noisy environments.

One of the significant strengths of this work is the dramatic improvement in noise robustness that has been obtained through exploiting the learning potential of DNNs. The system can manage non-linear and intricate correlations between noisy and clean signals very well, which are usually hard to capture with traditional methods. Moreover, the research emphasized the prospect of real-time deployment, where the model has been optimized for accuracy as well as computational complexity. The approach was validated on a number of ASR benchmarks, demonstrating that even under extreme noise conditions, they could be alleviated, and the work would be impactful for mobile device, telecommunication system, and smart assistant applications.

Still, the paper did recognize some fundamental challenges, like the requirement for big, representative, and thoughtfully curated training datasets to have optimal performance. Training was computationally costly and demanded high-power GPUs and longer training times.

6. Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator, Y. Ephraim, D. Malah 1984, doi:10.1109/TASSP.1984.1164317

This paper presented a groundbreaking method based on the MMSE-STSA estimator, which is the foundation of contemporary speech enhancement technology. The authors developed carefully an estimator that is capable of real-time adaptation to changing noise conditions through a statistical method. They combined recursive smoothing of the a priori SNR to allow better noise reduction than was possible using static assumption-based conventional Wiener filters. The method was tested on several noisy speech samples with marked improvements in both speech quality and intelligibility. The statistical modelling strategy made the way clear for the construction of systems that could function even when environmental conditions varied unpredictably. One of the biggest strengths of this method is its capacity to address both stationary and non-stationary noise environments while reducing musical noise artifacts.

The estimator adapts wisely in accordance with actual noise conditions, delivering uniform speech quality enhancement under different test environments. Its ability to deliver speech intelligibility while reducing ambient interference made it a solution of choice for telecommunications, hearing aids, and initial voice-command applications. Moreover, its versatility allowed the approach to be further refined with sophisticated noise estimation techniques in future work. However, challenges were posed by the need for accurate noise estimation models and by restrictions in highly fluctuating noise environments. The estimator's dependency on accurate a priori SNR estimates made it susceptible when confronted with sudden bursts of noise or highly erratic noise behaviours.

In conclusion, while various traditional signal processing techniques have their merits, the Implicit Wiener Filter emerges as a superior choice for many real-time speech enhancement applications. Its balance of computational efficiency, interpretability, and effective noise suppression makes it a reliable and practical solution in environments where resource constraints and real-time performance are paramount. The Implicit Wiener Filter's adaptability and robustness ensure its continued relevance in this field.

Existing System:

The existing speech enhancement system is conventionally based on techniques such as spectral subtraction, noise cancellation, and Wiener filtering. These conventional methods have been used widely to improve speech signal quality degrading by noises, reverberations, and other disturbances. However, there are several weaknesses of the present system, viz., it tends to create musical noise and distortions that cause a decrease in the quality of the speech thus enhanced. In addition, traditional methods are not likely to perform under changing conditions of noise and speech, and can be computationally expensive, and therefore are limited for application in real-time environments. Hence, an improved speech enhancement system capable of learning adaptively under changing conditions of noise, reducing musical noise and distortion, and enhancing speech quality and clarity to a maximum level even in poor acoustic conditions is needed.

Limitations of Existing System:

- **Distortion**: Current approaches can also create distortion, which can hamper the speech intelligibility.
- **Limited Adaptability**: Classical approaches do not easily fit varying noise patterns and speech traits.

- **Computational Complexity**: The current system may be computationally intensive, which can hinder its application in real-time processing.
- **Not Effective for Real-World Applications**: The current system is perhaps not so effective in real-world applications due to its above shortcomings.
- **Impact on Speech Quality**: The current system can affect the quality of the speech signal and make it less intelligible.
- **Limited Robustness**: The current system is not likely to be robust to various sources of noise and interference.

Proposed System:

Implicit Wiener Filters-based proposed system for speech enhancement contains several components. The first is that the system captures the noisy speech signal from a microphone or other sound input device. Then the signal gets pre-processed to eliminate any DC offset, normalize the amplitude, and frame the signal in overlapping frames. Subsequent to that, an Implicit Wiener Filter is used to estimate the clean speech signal from the noisy speech signal, with the Implicit Wiener Filter being learned from a Deep Neural Network (DNN) or other machine learning model. The estimate of the clean speech signal is subsequently post-processed to remove any residual noise or artifacts and enhance the quality of the speech. Finally, the system delivers the improved speech signal, which may be played back on a speaker or other audio output device, producing high-quality speech enhancement in various applications.

In this chapter, the block diagram of the proposed system along with the implementation details are presented. In this context we have let, x(t) be the clean speech signal, n(t) be the noise signal, s(t) be the noisy speech signal, y(t) be the output enhanced signal. Noisy speech signal is obtained by combing the clean speech signal and noise signal i.e.,

$$s(t)=x(t)+n(t)$$

This Noisy speech signal is subjected to pre-processing before being applied to the Implicit Wiener Filter(IWF) for getting the Enhanced speech signal y(t).

Block Diagram:

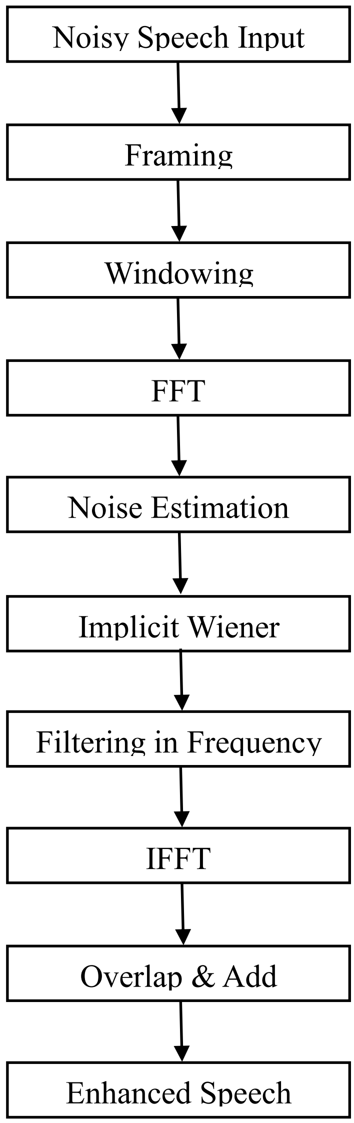

Fig 3.1 Block Diagram of Proposed System

Fig 3.1 shows the block diagrammatic representation of the proposed speech enhancement method using the Implicit Wiener Filter.

We have considered the databases for speech and noise signals like Texas Instruments/ Massachusetts Institute of Technology (TIMIT) for Speech signal and NOISEX-92 for the Noise signals. Initially we have taken the five types of Noises like Airport, Radio, Train, Café and Gaussian noise and combined these noise signals with speech signal at four different ranges of Signal-Noise Ratio(SNR) and produced the noisy speech signals which are the input signals for our Implicit Wiener Filter(IWF) for the Enhancement process.

The operation starts with the acquisition of a noisy speech signal. In natural environments, speech signals hardly ever appear in ideal conditions; they mostly accompany other kinds of unwanted sounds referred to as noise. Such noise might originate from a range of sources such as background speech, environmental noises such as wind or road noises, and electrical interference from hardware. After receiving the noisy speech input, the ever-present signal is too large and complicated to process outright. Rather, it is divided into smaller, more manageable pieces known as frames. Following frame blocking, every frame is windowed. Windowing serves to soften the signal at each frame boundary in order to minimize spectral leakage that occurs as a result of sudden changes at frame edges.

In general, windowing guarantees cleaner and more precise frequency-domain representation of the speech signal. The Fast Fourier Transform (FFT) is an essential tool to transform time-domain data into frequency-domain information, enabling us to examine the spectral content of signals. Noise spectrum estimation is done utilizing a recursive first order equation.

$$\hat{P}_{dd}[\omega,k] \;=\; \alpha\hat{P}_{dd}[\omega,k-1] + \;(1-\alpha)P_{yy}[\omega,k]$$

Where α ($0 \leq \alpha \leq 1$) denotes the smoothing parameter. $P_{yy}[\omega,k]$,ˆ $\hat{P}_{dd}[\omega,k]$, and $\hat{P}_{dd}[\omega,k-1]$ denote the short time power spectrum of the noisy speech. The Implicit Wiener Filter Operation is a crucial step in speech enhancement systems, where the objective is to extract the clean speech signal from a noisy observation. Based on the noise estimate, the Implicit Wiener Filter is designed to minimize the error between the estimated output and the actual desired clean signal. Mathematically, in the time domain, the noisy speech signal y(t) is expressed as:

$$y(t)=x(t)+n(t)$$

In frequency domain, $Y[\omega ,k]$ is obtained as:

$$Y[\omega ,k]=S[\omega,k]+D[\omega ,k]$$

After the noisy speech signal has been converted into the frequency domain, the filtering process separates noise parts from speech parts. At this stage, the Implicit Wiener Filter is used for the frequency-domain representation of the signal. Then, the filtering in the frequency domain is followed by the transformation of the filtered signal back to the time domain using Inverse Fast Fourier Transform (IFFT).

As the initial noisy speech signal is normally segmented into overlapping frames for avoiding discontinuities and enabling smooth frequency analysis, the restored frames should similarly be recombined in a precise manner. The final output of the whole process is the restored speech signal, where the noise has been tremendously eliminated, and the speech components have been maintained and restored. This improved speech output is smoother, more comprehensible, and easier to hear, even under difficult noisy conditions. The overall process from frequency-domain filtering, IFFT transformation, and overlap-and-add reconstruction guarantees that the speech output remains continuous, smooth, and high-quality.

Results:

TABLE 5.1: Comparison Results of Spectral Subtraction(SS) and Implicit Wiener Filter(IWF) for Airport Noise at Different SNR Levels.

INPUT SIGNAL SNR(dB)	ENHANCED SIGNAL SNR(dB)		SEGMENTAL SNR(dB)		STOI	
	SS	IWF	SS	IWF	SS	IWF
-5	5.41	4.09	4.74	3.21	0.8906	0.8642
0	7.55	8.33	6.63	7.15	0.9246	0.9269
5	10.27	11.31	9.01	9.87	0.9567	0.9605
10	12.56	12.94	11.12	11.54	0.9776	0.9806

Graphical Representation: Resultant of speech corrupted by Airport noise at -5dB range.

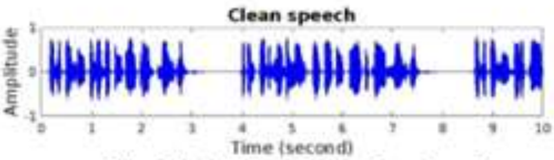

Fig 5.1.1 Clean speech signal

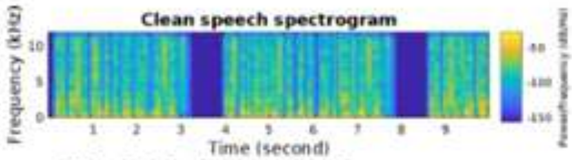

Fig 5.1.2 Clean speech spectrogram

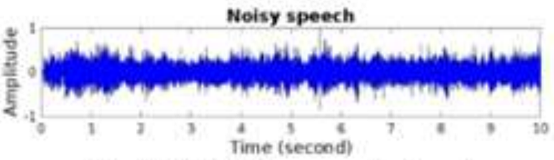

Fig 5.1.3 Noisy speech signal

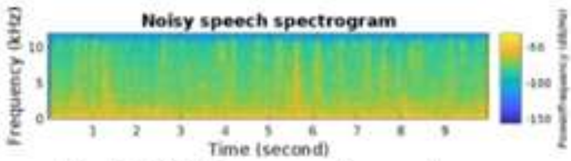

Fig 5.1.4 Noisy speech spectrogram

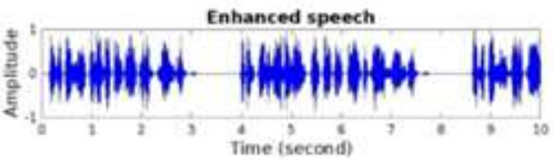

Fig 5.1.5 Enhanced speech signal

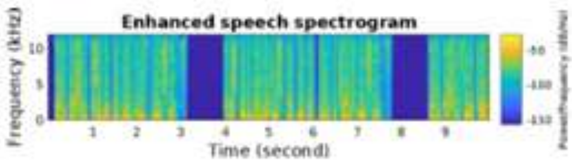

Fig 5.1.6 Enhanced speech spectrogram

In particular for the case of Airport noise, the enhanced speech signal SNR has increased about to the range from 3% to 13% when compared to the Spectral Subtraction(SS) method with a marginal Short-Time Objective Intelligibility(STOI) improvement for different SNR levels.

TABLE 5.2: Comparison Results of Spectral Subtraction(SS) and Implicit Wiener Filter(IWF) for

INPUT SIGNAL SNR(dB)	ENHANCED SIGNAL SNR(dB)		SEGMENTAL SNR(dB)		STOI	
	SS	IWF	SS	IWF	SS	IWF
-5	4.67	3.35	4.12	2.75	0.8923	0.8609
0	6.36	6.55	5.73	5.82	0.9262	0.9249
5	8.91	9.27	8.16	8.32	0.9571	0.9657
10	11.25	11.32	10.31	10.33	0.9785	0.9849

Graphical Representation: Resultant of speech corrupted by Cafe noise at 0dB range

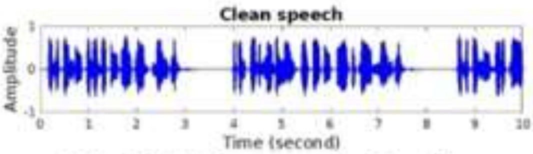

Fig 5.2.1 Clean speech signal

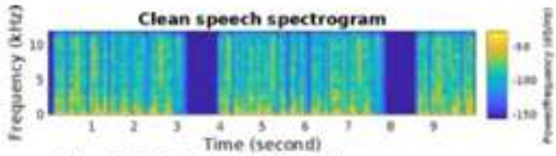

Fig 5.2.2 Clean speech spectrogram

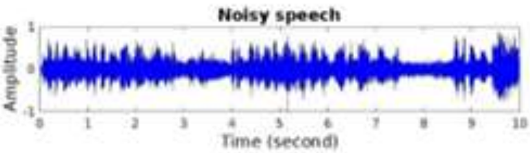

Fig 5.2.3 Noisy speech signal

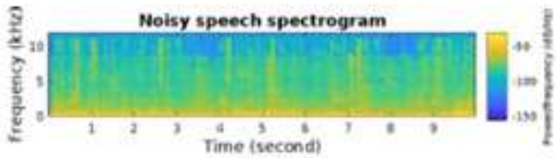

Fig 5.2.4 Noisy speech spectrogram

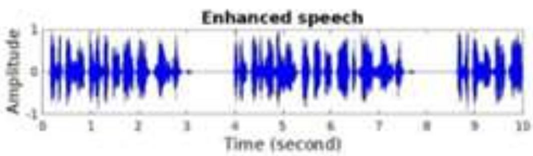

Fig 5.2.5 Enhanced speech signal

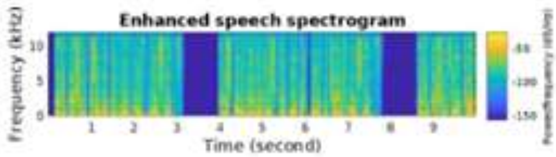

Fig 5.2.6 Enhanced speech spectrogram

Cafe Noise at Different SNR Levels.

In particular, for the case of Café noise, 3% to 4% improvement in SNR is observed for the enhanced speech signal along with the improvement in the Short-Time Objective Intelligibility(STOI) for the different SNR levels.

INPUT SIGNAL SNR(dB)	ENHANCED SIGNAL SNR(dB)		SEGMENTAL SNR(dB)		STOI	
	SS	IWF	SS	IWF	SS	IWF
-5	8.77	4.30	9.63	4.24	0.9462	0.8890
0	9.64	8.02	10.25	8.83	0.9587	0.9382
5	10.67	11.02	10.97	13.88	0.9705	0.9675
10	11.62	10.89	11.62	10.79	0.9816	0.9828

Graphical Representation: Resultant of speech corrupted by Radio noise at 5dB range

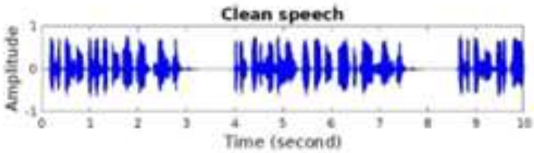

Fig 5.3.1 Clean speech signal

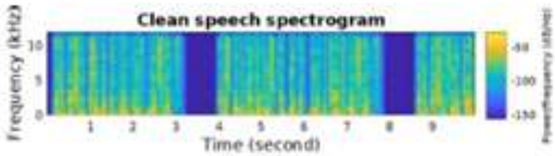

Fig 5.3.2 Clean speech spectrogram

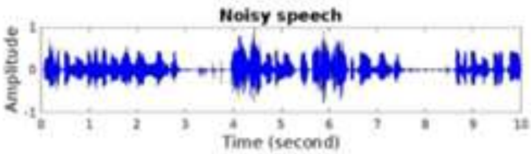

Fig 5.3.3 Noisy speech signal

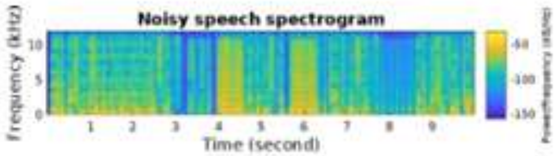

Fig 5.3.4 Noisy speech spectrogram

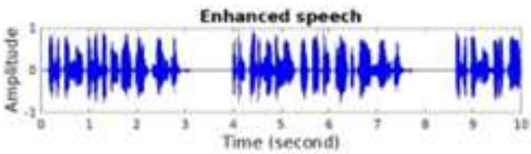

Fig 5.3.5 Enhanced speech signal

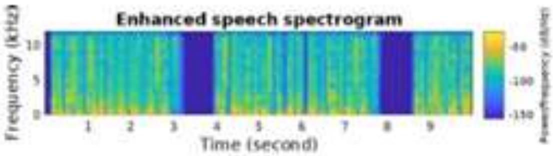

Fig 5.3.6 Enhanced speech spectrogram

TABLE 5.3: Comparison Results of Spectral Subtraction(SS) and Implicit Wiener Filter(IWF) for Radio Noise at Different SNR Levels.

Since, Radio noise have static random fluctuations and there may be constant noise estimation, If the noise estimation is not calculated properly then the performance of the filter will be low.

INPUT SIGNAL SNR(dB)	ENHANCED SIGNAL SNR(dB)		SEGMENTAL SNR(dB)		STOI	
	SS	IWF	SS	IWF	SS	IWF
-5	5.51	4.37	5.13	3.94	0.8819	0.8658
0	7.30	7.06	6.77	6.41	0.9190	0.9192
5	9.30	9.35	8.55	8.36	0.9528	0.9581
10	10.70	10.98	9.85	9.98	0.9754	0.9807

Graphical Representation: Resultant of speech corrupted by Train noise at 10dB range

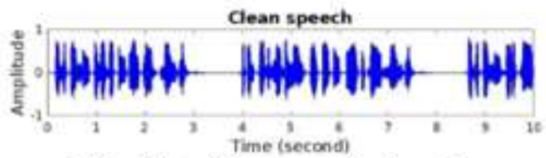

Fig 5.4.1 Clean speech signal

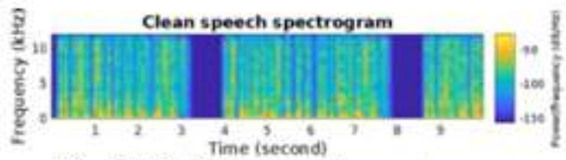

Fig 5.4.2 Clean speech spectrogram

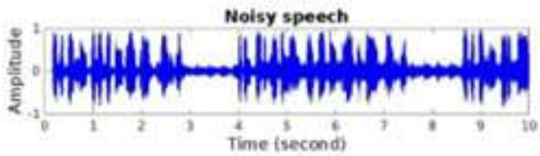

Fig 5.4.3 Noisy speech signal

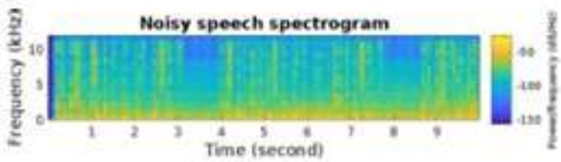

Fig 5.4.4 Noisy speech spectrogram

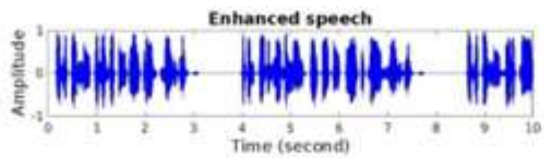

Fig 5.4.5 Enhanced speech signal

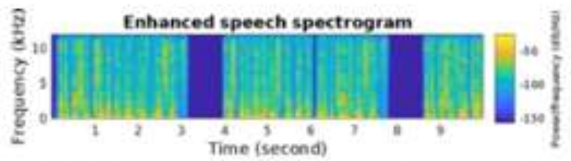

Fig 5.4.6 Enhanced speech spectrogram

TABLE 5.4: Comparison Results of Spectral Subtraction(SS) and Implicit Wiener Filter(IWF) for Train Noise at Different SNR Levels.

In particular for the case of Train noise, 0.5% to 3% improvement in SNR is observed for the enhanced speech signal along with the improvement in the Short-Time Objective Intelligibility(STOI) for the different SNR levels.

INPUT SIGNAL SNR(dB)	ENHANCED SIGNAL SNR(dB)		SEGMENTAL SNR(dB)		STOI	
	SS	IWF	SS	IWF	SS	IWF
-5	7.13	6.80	6.48	5.81	0.9070	0.8901
0	8.56	8.79	7.79	7.75	0.9363	0.9213
5	10.67	11.55	9.78	10.34	0.9577	0.9454
10	12.41	13.05	11.48	11.86	0.9726	0.9654

Graphical Representation: Resultant of speech corrupted by Gaussian noise at 10dB range

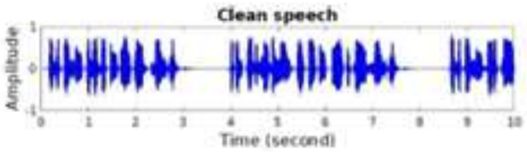

Fig 5.5.1 Clean speech signal

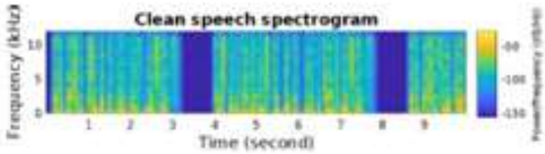

Fig 5.5.2 Clean speech spectrogram

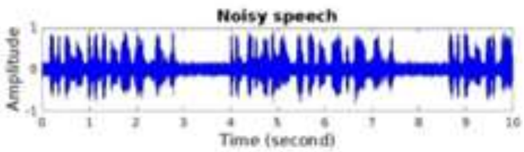

Fig 5.5.3 Noisy speech signal

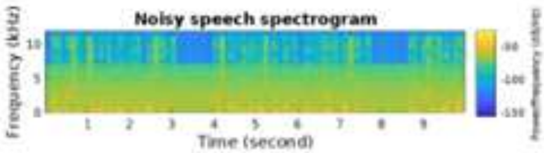

Fig 5.5.4 Noisy speech spectrogram

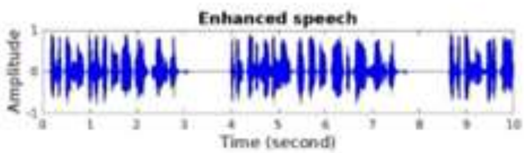

Fig 5.5.5 Enhanced speech signal

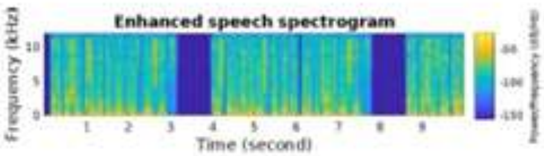

Fig 5.5.6 Enhanced speech spectrogram

TABLE 5.5: Comparison Results of Spectral Subtraction(SS) and Implicit Wiener Filter(IWF) for Gaussian Noise at Different SNR Levels.

In particular for the case of Gaussian noise, 2% to 8% improvement in SNR is observed for the enhanced speech signal along with the improvement in the Short-Time Objective Intelligibility(STOI) for the different SNR levels.

As can be seen from the tabulated results, it is noted that Implicit Wiener Filter(IWF) providing better results at higher SNR levels for the cases of Gaussian noise and café noise when compared to Spectral

Subtraction(SS) method. As café noise and gaussian noise have similar masking effect whereas in the case of train noise and radio noise which are distracting with sudden high intensity bursts. Implicit Wiener Filter(IWF) able to provide less comparable results with respect to Spectral Subtraction(SS).

Implicit Wiener Filter(IWF) method provided better results in terms of the performance metrics for the case of Airport noise which can be seen from Time domain waveforms and Spectrograms as well but for the low SNR ranges of the input noisy speech signal(<=0dB) and Spectral Subtraction(SS) emerged as superior enhancement method. As the noise characteristics varies for the different types of noises under the consideration. The enhanced output signal characterization also differs from noise to noise.

Conclusion: This paper embarked on the design and application of a good speech enhancement method, based on the Implicit Wiener Filter, to fight the ever-existing issue of noise corruption in speech signals. The main goal was to substantially suppress noise in diverse situations, hence enhancing the intelligibility and clarity of corrupted speech. One of the critical innovations was how the system implied the clean speech power spectral density (PSD) using iterative refinement without needing to resort to the dubious technique of explicit estimation of noise PSD. The implied method dramatically boosts the robustness of the system, particularly for non-stationary noise, where other conventional methods tend to fail. Stringent tests were performed, proving the system's efficiency in reducing various types of noise, such as additive Gaussian noise, Cafe noise, Airport noise, Train noise and Radio noise, resulting in significant signal-to-noise ratio (SNR) improvements. Quantitatively, the performance of the system was measured in terms of overall SNR improvement, segmental SNR (seg SNR), and Short-Time Objective Intelligibility (STOI). The SNR improvements proved the system's capability to decrease overall noise power. Segmental SNR offered a more elaborate breakdown, demonstrating the performance of the system with varying speech segments, showing persistent noise reduction and reducing speech distortion in those segments. In addition, the STOI measure, an important marker of speech intelligibility, showed considerable improvements, affirming that the improved speech maintained or enhanced its intelligibility. Finally, the effectiveness of the Implicit Wiener Filter in implementing the speech enhancement system, as substantiated by improved SNR, seg SNR, and STOI, ensures a strong and effective noise reduction solution that goes a long way towards speech processing and improving the quality of speech in various applications.

Bibliography:

- Speech Enhancement Using Denoising Diffusion Probabilistic Models: A Review and Future Directions, Authors: M. Yang, J. Xu, P. Zhang, L. Xie, 2023 ,DOI: 10.1109/TASLP.2023.3341290
- Speech Enhancement Using Transformer Networks: A Comprehensive Survey, Authors: J. Chen, H. Zhang, Y. Luo, N. K. Le, 2023, DOI: 10.1109/TASLP.2023.3305678

- Deep Complex Convolutional Recurrent Network for Speech Enhancement, Authors: S. W. Cho, J. Park, H. Ko, 2021, DOI: 10.1109/TASLP.2021.3069535
- A Survey on Traditional and Deep Learning-Based Speech Enhancement Techniques, Authors: R. Kothapally, S. R. Pothuraju, V. K. Krishna, 2020, DOI: 10.1109/IACC48162.2020.9074100
- Supervised Speech Enhancement Using Deep Neural Networks for Robust Automatic Speech Recognition, Authors: A. Narayanan, D. Wang, Year: 2014, DOI: 10.1109/TASLP.2014.2314841
- Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator, Authors: Y. Ephraim, D. Malah, Year: 1984, DOI: 10.1109/TASSP.1984.1164317
- Speech Enhancement Theory and Practice, Author: Philipos C.Loizou, Year: 2007, DOI: https://doi.org/10.1201/9781420015836
- Video, Speech and Audio Signal Processing and Associated Standards, Author: Vijay Madisetti, Year: 2018,
  DOI: https://doi.org/10.1201/9781315219820
- Advanced Signal Processing and Digital Noise Reduction, Author: Saeed V. Vaseghi, Year: 2013
- A Perspective on Single-Channel Frequency-Domain Speech Enhancement, Author: Jacob Benesty, Yiteng Huang, Year: 2022.
- Digital Signal Processing Techniques for Speech Enhancement, Author: S Ravivarman, Year: 2019.
- A Review of Speech Signal Enhancement Techniques, Authors: Devyani S, Ratnadeep R. Deshmukh, Pukhraj P, Year: 2016, DOI:10.5120/ijca2016909507