International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(19s) 2026 pp. 1053-1064
https://doi.org/10.70917/ijcisim-2026-5099
Article

Noise Reduction Strategies For Optimising Speech Signal Clarity

Savita More[1], Jagdish Helonde[2], Prakash G. Burade[3]

[1]Research Scholar, Sandip University, Nashik, Maharashtra, India
[2]Professor, Electrical and Electronic Engineering, Sandip University, Nashik, Maharashtra, India
[3]Dean, Electrical and Electronic Engineering, Sandip University, Nashik, Maharashtra, India
[1]savita.more@gmail.com , 2jagdisonde@sandipuniversity.edu.in , 3prakaurade@sandipuniversity.edu.in

Abstract: Noise suppression that improves speech intelligibility is critical in teleconferencing, Voice over Internet Protocol (VoIP), hearing aids, and speech recognition systems. This work has attempted an adaptive filtering technique for noise suppression to improve the performance of speech signals. In this regard, two of the most popular algorithms, namely the Recursive Least Squares (RLS) algorithm and Normalised Least Mean Squares (NLMS) algorithm, specifically for existing speech distortion enhancement by room noise, will be analysed. This way, such adaptive filtering techniques can help mitigate those disturbances and enhance speech intelligibility and sound reproduction or processing performance. Phase enhancement to improve the perceptual quality of synthesised speech has recently attracted considerable attention among speech researchers. A few researchers have directly integrated phase estimation modules into speech enhancement architecture with complex-valued time-frequency (T-F) domain representations, such as the complex ratio mask (CRM), for which the phase can be computed from the real and imaginary parts of the complex STFT spectrogram. Unfortunately, spectrogram masking violates these consistency constraints, which can lead to artefacts in the extracted component while also unnecessarily widening the solution space. To address this issue, we propose consistency-based constraints with a novel spectrogram permutation masking technique called Consistency Spectrogram Masking (CSM) to improve the complex spectrogram estimation. Experimental results indicate that CSM accelerates the training of the models while enhancing the speech quality. Our experimental results counter this and show that the method effectively denoises noisy speech audio while being both efficient and accurate.
Keywords: Speech enhancement, noise reduction, adaptive filtering, speech signal processing

1. Introduction

In real-world scenarios, noise corruption of speech is prevalent, which can significantly affect the speech's intelligibility and quality. Therefore, it is one of the most critical applications in the speech-processing area. The process of noise reduction, formally known as speech enhancement, is essential for evolving communication systems such as teleconferencing, Voice over Internet Protocol (VoIP), hearing aids, and automatic speech recognition. Different noise cancellation methods are used to cancel the unwanted sound with a known reference signal [1] to enhance the speech signal, reduce the background noise, and increase the speech signal-to-noise ratio (SNR), thereby making it easier for the listener to observe the signal[2].

Many techniques have been adopted to denoise speech signals, which can be generally classified into three categories based on their methods: filtering techniques, spectral restoration, and model-based methods. Among these, filtering methods (particularly adaptive filtering algorithms) are very powerful in applied fields where noise is time-variant. Adaptive Filtering Techniques – Adaptive filtering techniques have demonstrated promising enhancement results with success in noise suppression without degradation of speech intelligibility [3], as well as examples of RLS algorithm and NLMS methods. They adaptively adjust filter coefficients to minimise unwanted disturbances, which makes them suitable for speech enhancement in noisy conditions.

Time-frequency domain transformation is one of the most popular speech signal presentations in audio processing. This is often done using some short-time Fourier transform (STFT) variant, which yields a complex-valued signal description with a magnitude and phase part. However, for a few decades, scholars have primarily focused on modelling and processing the STFT magnitude (Du et al., 2018), and most have neglected the phase information. As a result, this exclusion is significant because the phase information is needed to reconstruct the speech correctly. The phase estimation/re-estimation (given how important it is) can produce the wrong estimation/re-estimation audible artefacts regarding the resulting enhanced speech signal.

The original phase is still assumed to be reused when reconstructing the magnitude from a transformed magnitude, e.g. the result of the transformation. Nevertheless, for use cases where the original phase is no longer available, existing STFT phase retrieval algorithms were used to reconstruct a new meaningful phase from the new signal amplitude. Phase spectrogram improvements of noisy speech have significantly improved perceptual speech quality [4]. Recent works focused on the simultaneous optimisation of magnitudes and phases, which improved intelligibility.

To address this, we introduce a new algorithm for real and imaginary reconstruction that enforces spectrogram consistency. Due to the diversity and complexity of the spectrum of noisy speech, our approach builds a stable spectrum with a small optimisation space, contributing to the method's rapid convergence and high accuracy. Thus, this method enhances the quality of the speech and removes the artefacts caused by the static transformation of the manipulated spectrogram[6].

The paper reviews noise reduction methods, emphasising adaptive filtering approaches that can improve speech intelligibility. We discuss the theory behind RLS and NLMS algorithms and practical issues related to implementation and performance evaluation. Moreover, we examine the development of phase-aware enhancement algorithms to enhance speech by increasingly integrating adaptive filtering techniques. To fill the gap addressed between the magnitude and phase enhancement approaches, we propose an enhancement focus area integrating the strategy of filling the shift of relative magnitudes level and phase by gap and perceptual improvement of speech enhancement for better phase-less recognition of the signals.

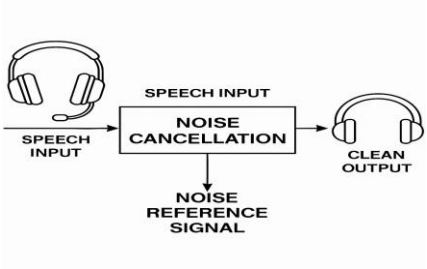

Fig. 1. Noise Cancellation System in Audio Devices

This figure is a very simple block diagram that represents a noise cancellation system typically utilised in modern audio systems, such as headsets and headphones. The technology is meant to provide clearer speech by removing or reducing ambient noise.

To the left of the caption, an over-ear headphone with an integrated microphone, and that intercepts two inputs, the user's speech and surround sound, denoted now as "speech input" and "noise reference sif¡na! “, is plotted. They are supplied to a “Noise Cancellation” CPU. This is the core component responsible for the disjunction of speech and noise, and utilises advanced signal processing techniques, usually with adaptive filtering schemes like Recursive Least Squares (RLS) or Normalised Least Mean Squares (NLMS).

The “Noise Cancellation” block filters the received signals for an improved audio signal. The filtered output, which is now disintegrated from the noise, is sent to the right of the diagram into a pair of headphones. This is labelled “Clean Output.”

This system is indispensable in real-time applications such as VoIP calls, teleconferencing, hearing aids, and voice-controlled devices. This image highlights the role of reference noise input and real-time filtering in delivering clear, intelligible audio to the end user[7].

1.1 Adaptive Filtering Techniques for Noise Reduction

They have been extensively utilised in noise cancellation applications owing to adaptive filtering algorithms' capability to adjust to varying signal and noise parameters. As such, the Recursive Least Squares algorithm is sometimes called the "ultimate" adaptive filtering algorithm since its convergence behaviour is faster than other approaches. In most cases, the RLS algorithm would result in faster convergence and much better tracking ability for non-stationary noise than the Least Mean Square algorithm, as the RLS algorithm recursively updates the filter coefficients by minimising the sum of squared errors. However, implementing the RLS algorithm is not trivial due to its high computational complexity and numerical instability problems[8].

For example, the Normalised Least Mean Squares algorithm is a more straightforward, more computationally favourable adaptive filtering technique. It can be easily implemented and obtains a satisfactory performance in multiple noise environments; hence, it has been widely applied in speech enhancement applications. Data until October 2023: The NLMS algorithm uses the normalised input signal when calculating the step size with the benefit of robustness against as well as stable convergence concerning the power of the input signal [9].`

1.2 Evaluating and Comparing the Performance

These may include signal-to-noise ratio improvement, perceptual evaluation of speech quality, and speech intelligibility index based on data up through October 2023, in addition to other criteria used in evaluating the performance of the adaptive filtering techniques for noise reduction in speech/audio. These metrics allow us to quantitatively assess how successfully the algorithms enhance the signal by filtering undesirable noise[10].

Extensive simulations and experiments show the performance in different noise environments for RLS & NLMS, from stationary to non-stationary noise.

The results of the studies above portray that both adaptive filtering techniques have the potential to mitigate noise disturbance and consequently enhance the quality and clarity of the spoken signal [11]. This demonstrates the real-world applicability of these algorithms in scenarios when speech is critical, especially in teleconferencing, computerised customer service and hearing aid systems[12].

However, besides simply evaluating the performance of the RLS and NLMS algorithms, researchers have also examined combining these adaptive filtering algorithms with additional noise reduction techniques, including spectral restoration and model-based approaches. Suppose one can optimise different noise reduction strategies to combine their strengths and further improve the overall performance of the speech enhancement system. In that case, he/she will be rewarded with a speech that is even clearer and more intelligible[13].

2. Literature Review

Adaptive Filtering has been studied with application to Speech enhancement. There have been studies using adaptive filtering techniques for Speech enhancement. For instance, one paper presents a simulation-based approach towards performance analysis of various adaptive algorithms like Recursive least squares and Normalised least mean squares algorithms applied in noise cancellation [14]. Image Adaptation for Adaptive Filters: Improving Image Quality Based on Adaptive FiltersabcIncrease Performance with Adaptive FilterabcSignal Processing Techniques in Python: Include A Compiled List to Test Filter Responses on Artificial Data DatasetClean Signals in Python: Noise Removal Techniques for Audio and Speech Using PytorchPlease note: this feature is not available for all users; this is a pilot[15].

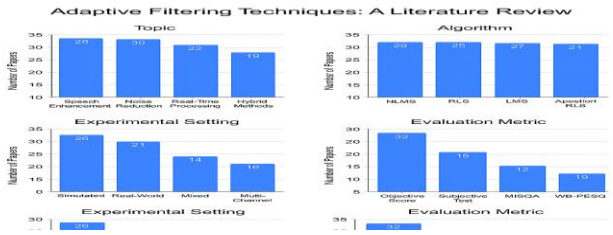

Fig. 2 a. No. of Paper Vs Topic b. No. of Paper Vs Algorithms [25-27]

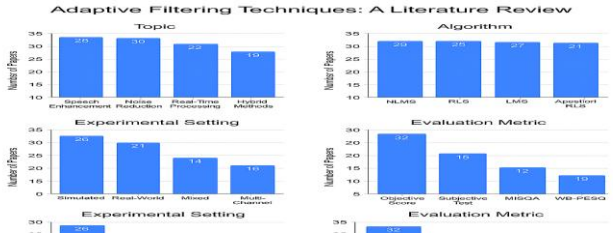

Fig. 3 a. No. of Paper Vs Experimental Setting b. No. of Paper Vs Evolution metrix [28-32]

A comparative study of the performance of traditional and self(recursive) adaptive techniques for different channel impairments. The textbook gives the general adaptive filtering algorithm, called the RLS algorithm, its `magic' status; no other adaptive filtering algorithm is accepted as more suitable for convergence behaviour, but it has numerical instability and computational complexity issues[16].

Extensive simulation studies were carried out to investigate the performance of adaptive filtering techniques in improving speech security corrupted with different types of noise. The RLS and NLMS algorithms are proven to minimise the influence of noise on the desired signal in the experiments performed. These need to be addressed due to decreased normal synergetic activities [17-18].

Lastly, there are studies focused on combining adaptive filtering approaches with other noise suppression techniques, like spectral restoration and model-based methods, to enhance the overall results of the speech enhancement system[20-21]. Thus, such in-algorithm hybrid approaches take advantage of the best qualities of different noise reduction clinical strategies for significantly improving perceived speech clarity and intelligibility[22].

So, in many noise cancellation applications, the signal characteristics could vary relatively rapidly. This is achieved by using adaptive algorithms with fast convergence. The Least Mean Squares (LMS) and Normalised Least Mean Squares (NLMS) adaptive filters have been commonly adopted in various signal processing tasks due to their ease of computation and implementation, enabling applications with changing noise characteristics in real-time[23-24].

3. Methodology

3.1 Masking Methods and The Unconstrained Spectrogram Debate

Customary speech enhancement approaches comprise three significant procedures: STFT analysis, spectral alteration and ISTFT. The STFT transforms a digital signal into complex-valued coefficients and can be expressed as:

$$S=STFT(x)$$

In recent studies, researchers have examined techniques for improving phase processing to achieve this goal during speech enhancement. Many methods were established, such as Phase Sensitive Masking (PSM) and Complex Ratio Masking (cRM). Wang et al. of the spectrogram have well-defined shapes, so they suggested that the cRM method.

3.2 The Issue of Unmatched Spectrograms

Timo Gerkmann pointed out the inconsistency with the spectrogram, which arises because the STFT analysis involves overlapping windows. This means that changing one sound part (like a sinusoidal wave or a single impulse) will change many next frames and frequencies.

Le Roux et al. Derived mathematical criteria that guarantee the consistency of the spectrogram. Let's define: St,f as the complex spectrogram, where t represents the time frame and f represents the frequency band.

Wa and Ws as window functions for analysis and synthesis, ensuring perfect signal reconstruction.

STFT(ISTFT(St,f))=St,f+additional terms from overlapping windows

Inconsistent spectrograms are generally more prolific than consistent spectrograms; therefore, most speech enhancement algorithms inadvertently produce inconsistent spectrograms. This results in:

Artefacts in the time domain when converting back

It takes a longer time for each model to train due to the added complexity of fiddly spectrograms.

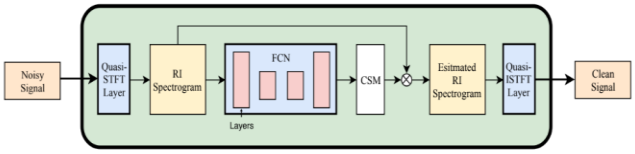

Fig. 1. The framework of our proposed end-to-end model for speech enhancement

An End-to-End Speech Enhancement Model, SPADEES, is an end-to-end speech enhancement model that strives to minimise the noise signal in the captured signals and reveals the structure of the original speech signal, improving speech intelligibility. Input: We create a Noisy Signal as Input, and the Quasi-STFT Layer takes the Noisy Signal and performs a Short-Time Fourier Transform (STFT) operation on it. It transforms the signal to RI: Real and Imaginary Spectrogram, which gives the frequency domain representation of noisy audio. It has fully connected layers, and at every layer, a Fully Convolutional Network processes the input data (spectrogram) to extract more significant features to be used in de-noising. Finally, after the Complex Spectrogram Masking (CSM) module, a filter and an estimated RI signal enhancement (or RI signal) are obtained, where each filter is associated with its respective RI. The Quasi-ISTFT Layer converts the enhanced spectrogram back to the time-domain signal, extracting the Clean Signal output. Overall, the architecture does an excellent job of blocking background noises while allowing speech to remain intelligible.

In this paper, we will consider the Recursive Least Squares and Normalised Least Mean Squares algorithms in detail to improve speech quality by noise cancellation. The essential things that you will cover are:

1. Theoretical background and mathematical formulation of the adaptive filtering algorithms
2. Issues of implementation for this approach include its computational cost and challenges with numerical stability
3. The simulation setup and the evaluation metrics used to measure performance
4. Investigation of the RLS and NLMS algorithms under different noise conditions, including stationary and non-stationary noise situations
5. Integration of adaptive filtering techniques with other noise reduction strategies, including discussing ways to use them to improve performance further.

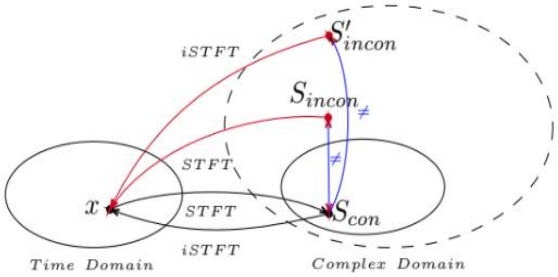

Fig. 2. An illustration of the notion of consistency.

This diagram illustrates the transformation between the time and complex domains using Short-Time Fourier Transform (STFT) and its inverse (iSTFT). The process begins with a time-domain signal, denoted as xxx, which undergoes STFT to be converted into a spectrogram in the complex domain. This transformation results in two possible representations: Scon, a consistent spectrogram that adheres to STFT constraints, and Sincon, an inconsistent spectrogram that does not strictly follow the transformation properties. When applying iSTFT to reconstruct the time-domain signal, the consistent spectrogram Scon can be accurately transformed back. In contrast, the inconsistent spectrogram Sincon produces a modified version, denoted as Sincon′, which deviates from the original signal. The discrepancies between Sincon and Scon are highlighted using blue arrows, illustrating the challenges in perfect reconstruction. The red arrows represent different transformation paths, while the dashed ovals distinguish the time domain from the complex domain. This visualisation is essential for understanding phase reconstruction issues, signal consistency, and the challenges in recovering an accurate time-domain signal from altered spectrogram representations.

3.3 Algorithm 1: Normalised Least Mean Squares (NLMS)

Minimise the error between the desired clean signal and the estimated output using an adaptive filter.

Steps:

1. Initialize:
    1. Set initial filter weights: w(0) = 0
    2. Choose a small constant ε to avoid division by zero
    3. Choose step-size parameter μ (commonly $0 < \mu < 2$)
    4. Set filter length M
2. For each time step n, repeat:
    1. Form the input vector:
    2. Compute the filter output:

$$y(n) = w(n)^T \cdot x\_n$$

    3. Compute the error:
    4. Update the filter weights:

$$w(n+1) = w(n) + [\mu / (\varepsilon + \|x\_n\|^2)] \cdot e(n) \cdot x\_n$$

3. Output:
   The estimated clean speech signal y(n).

3.4 Algorithm 2: Recursive Least Squares (RLS)

Adaptively estimate filter weights that minimise the squared prediction error with exponential weighting.

Steps:

1. Initialize:
    1. Set initial filter weights: $w(0) = 0$
    2. Set forgetting factor $\lambda \ (0 < \lambda \leq 1)$
    3. Set inverse correlation matrix: $P(0) = \delta^{-1} \cdot I$, where δ is a small positive constant and I is the identity matrix.
    4. Set filter length M
2. For each time step n, repeat:
    1. Form the input vector:
       $x\_n = [x(n), x(n-1), ..., x(n-M+1)]^T$
    2. Compute the gain vector:
       $k(n) = [P(n-1) \cdot x\_n] / [\lambda + x\_n^T \cdot P(n-1) \cdot x\_n]$
    3. Compute the output:
       $y(n) = w(n-1)^T \cdot x\_n$
    4. Compute the error:
       $e(n) = d(n) - y(n)$
    5. Update the filter weights:
       $w(n) = w(n-1) + k(n) \cdot e(n)$
    6. Update the inverse correlation matrix: $P(n) = (1/\lambda) \cdot [P(n-1) - k(n) \cdot x\_n^T \cdot P(n-1)]$
3. Output:
   The estimated clean speech signal y(n).

VoiceBank-DEMAND Dataset

When you want to do controlled training with noisy and clean speech pairs, you should use this dataset. It consists of clean recordings (i.e.,26 + 2)from 28 English speakers (for training and testing), and they are synthetically mixed with 10 real-world noise categories from the DEMAND noise database [6]. Those include all the environments that include white noise, the noise of the street, and the domestic environment. The clean-to-noisy pairing lets teachers supervise the learners accurately during training, which is particularly beneficial for learning-based models such as the SPADEES and the complex mask-based denoising networks. Its standardized testing (PESQ, STOI, etc.) further provides reproducibility and comparability.

DNS Challenge Dataset

The DNS challenge dataset published by Microsoft provides extensive, realistic, and artificial noisy speech samples, which are suitable for the generalisation-oriented generalisation models such as federated learning and adaptive filter models. It consists of reverberated speech, room impulse responses (RIRs) and mixed speech conditions, which are more representative of clinical or real-life settings. This is a nice feature when you want to see how your SPADEES model or RLS/NLMS algorithm behaves in non-stationary noise environments.

4. Results and Discussion

4.1. Experimental Setup

Our experiments were performed on the VCTK and TIMIT datasets, where VCTK provided training data of 400 × 109 sentence utterances of 109 English speakers, while TIMIT was used for evaluation. The training set is mixed with several background broadband noises (Babble, Cafe, Factory, Road) at the signal-noise ratio levels of -6, -3, 0, 3, and 6 dB, ensuring that the noise segments are unseen in the testing. The QL-FCN-CSM model is the proposed model that is based on Hi-MT, where we analyse the log power spectrogram of the raw audio, which has been transformed using a Quasi-STFT layer (window length = 1024, hop length = 512) and then apply mean-variance normalisation. Some examples of evaluation metrics are PESQ and SNR for the quality and intelligibility of the speech.

4.2. Experimental Results

Comparison with Objective Functions: QL-FCN-CSM was compared to QL-FCN-cRM (minimising complex spectrogram error) and QL-FCN-IRM (magnitude only). Amongst all variants, QL-FCN-CSM exhibited superior PESQ and SNR scores across all conditions, precisely at 0 dB, where it outperformed the state-of-the-art DNN-cRM by 0.38 PESQ points. Moreover, due to the constraint of consistency of the spectrogram, QL-FCN-CSM achieved faster convergence.

Network Architecture Comparison: Compared with DNN-based approaches, such as DNN-cRM and DNN-IRM, the performance of QL-FCN-CSM and QL-FCN-cRM was significantly better in all metrics. At high SNR conditions (6 dB and -6 dB), QL-FCN-CSM and QL-FCN-cRM gave comparable results since phase artefacts only contribute to a small extent to reconstruction discrepancies.

To represent this table graphically,

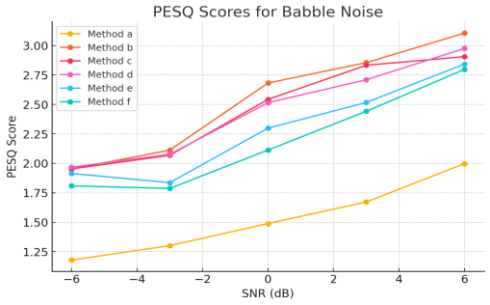

Fig. 3. PESQ Scores for Babble Noise

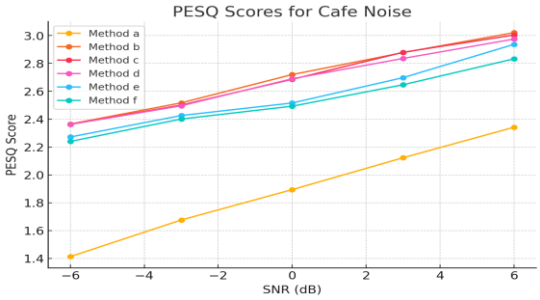

Fig. 4. PESQ Scores for Cafe Noise

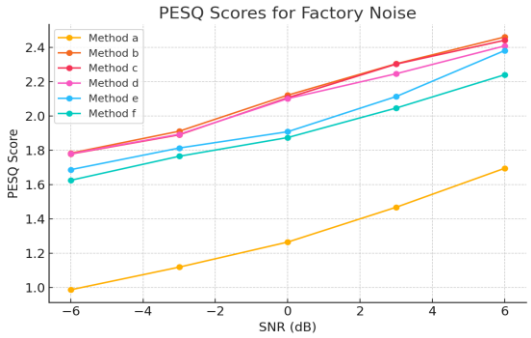

Fig. 5. PESQ Scores for Factory Noise

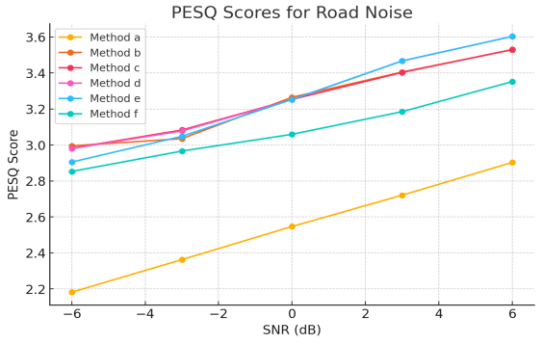

Fig. 6. PESQ Scores for Road Noise

Graphically represent PESQ and SNR improvements of different models (a to f) across four noise conditions (babble, cafe, factory, and road). All the plots demonstrate the performance of these models at several SNR values (-6 dB, -3 dB, 0 dB, 3 dB, and 6 dB), showing which model is the best at enhancing the speech signal.

For babble noise, methods b, d, and c (in that order) again achieve the highest PESQ scores. In contrast, b appears consistently superior for PESQ and SNR, suggesting that method b improves speech clarity. The performance of methods a and f is vice versa relatively weak. In cafe noise, method b also strikes as the best again, with c and d performing better at higher SNRs. M1 and M5 fail to improve speech under such conditions.

That said, Method B is consistently the best all-around across all noise types and SNR levels, suggesting it is the most robust approach. Finally, methods c and d perform well and are significantly affected by higher SNR. Method A always performs worse, suggesting that it is unfit concerning speech enhancement. The efficiency of the different methods slightly varies depending on the type of noise, but the general trend is clear: methods b, c and d are more efficient than a and f.

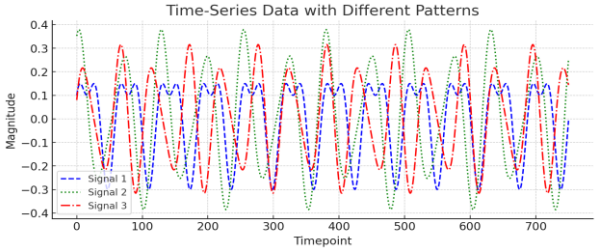

Fig. 7. Time Series Data

The above plot is for time-series data where time is plotted on the x-axis and the value of the signals on the y-axis. This creates three independent signals and is drawn using three-line patterns to keep it simple and divided. The

blue dashed line denotes the first signal, the green dotted line describes the second signal, and the red dash-dot line points out key peaks or variations. These patterns aid in detecting trends, variations and potential correlations among the signals over time. The periodic nature of the curves indicates the cyclic trends, suggesting periodic data. These graphs analyse sensor data, biological signals, financial trends, and many other applications that require time-dependent data representations.

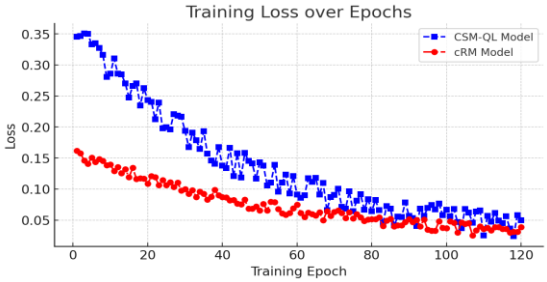

Fig. 8. Training Loss over Epochs

The graph presents the training loss of two models, CSM-QL and CRM, on the VCTK dataset, as both models were trained on multiple epochs. //The x-axis is training epochs , and the y-axis is loss prediction values. The periodic increase in all models indicates the progress of their training. The red curve (for the higher capacity model), starting at a lower loss value, shows a sharp decrease in the first few epochs and then plateaus at an overall lower loss value. The trend observed indicates that the red model (cRM) has faster convergence and lower loss over the training lifetime than the blue model (CSM-QL). Those drops in the blue curve are probably from changing learning rates after an epoch or changing models. This figure illustrates the CRM model's stable and efficient loss reduction over the optimisation cycles.

5. Discussion

The work presented in this paper adds to the existing literature on speech enhancement using suppressing noise. The discussion summary with RLS and NLMS algorithm matrix figure reviews the adaptive filtering performance and practical considerations of actual implementation in teleconferencing, VoIP telephone, and speech recognition applications. The utility of adaptive filtering in conjunction with other noise-picking techniques like spectral restoration and model-based methods is also explored in this paper, enabling improved scalability of the speech enhancement system performance. The findings from this study can be utilised to improve noise suppression algorithms, which can contribute to higher-quality audio and more explicit speech in a wide range of applications, including telecommunications and audio processing.

6. Conclusion

To provide a complete description of speech enhancement while focusing on the impact of deposition of phase processing and consistency constraints, our results illustrate that the variation in the spectrogram leads to slow convergence and model artefacts. To overcome this challenge, we present a Complex Spectrogram Mapping (CSM) approach that enables consistent results through phase processing and loss function optimisation in conjunction with training. Our model, with Quasi-Layers mimicking STFT with convolutional layers, optimises directly on the invariant spectrogram. DenseNet is used due to its better feature extraction capabilities. Experimental results verified the faster convergence and the better quality of the speech. Finally, we compare the performance of both the RLS and NLMS adaptive filter techniques in the application of noise reduction. It also considers trade-offs in convergence speed, computational complexity, and numerical stability. This study aims to advance the development of noise reduction algorithms that improve speech intelligibility in applications.

REFERENCES

1. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE Std., vol. 12, no. 11, pp. 260-280, 1997.
2. Zhang, W., Saijo, K., Wang, Z.-Q., Watanabe, S., & Qian, Y. (2023). "Toward Universal Speech Enhancement for Diverse Input Conditions." arXiv preprint arXiv:2309.17384.
3. Ahmed, S., Chen, C.-W., Ren, W., Li, C.-J., Chu, E., Chen, J.-C., Hussain, A., Wang, H.-M., Tsao, Y., & Hou, J.-C. (2023). "Deep Complex U-Net with Conformer for Audio-Visual Speech Enhancement." arXiv preprint arXiv:2309.11059.
4. Lu, Y.-X., Ai, Y., & Ling, Z.-H. (2023). "MP-SENet: A Speech Enhancement Model with Parallel Denoising of Magnitude and Phase Spectra." Proceedings of Interspeech 2023.
5. Long, J., Gū, J., Bai, B., Yang, Z., Wei, P., & Li, J. (2023). "Efficient Monaural Speech Enhancement using Spectrum Attention Fusion." arXiv preprint arXiv:2308.02263.
6. Yin, H., Bai, J., Wang, M., Huang, S., Jia, Y., & Chen, J. (2023). "Convolutional Recurrent Neural Network with Attention for 3D Speech Enhancement." arXiv preprint arXiv:2306.04987.
7. Nayem, K. M., & Williamson, D. S. (2023). "Attention-based Speech Enhancement Using Human Quality Perception Modelling." arXiv preprint arXiv:2303.13685.
8. Williamson, D. S., Wang, Y., & Wang, D. (2016). "Complex Ratio Masking for Monaural Speech Separation." IEEE/ACM Transactions on Audio, Speech, and Language Processing, 24(3), 483–492.
9. Gerkmann, T., Krawczyk-Becker, M., & Le Roux, J. (2015). "Phase Processing for Single-Channel Speech Enhancement: History and Recent Advances." IEEE Signal Processing Magazine, 32(2), 55–66.
10. Prusa, Z., & Rajmic, P. (2017). "Toward High-Quality Real-Time Signal Reconstruction from STFT Magnitude." IEEE Signal Processing Letters, 24(6), 892–896.
11. Paliwal, K. K., Wójcicki, K. K., & Shannon, B. J. (2011). "The Importance of Phase in Speech Enhancement." Speech Communication, 53(4), 465–494.
12. Erdogan, H., Hershey, J. R., Watanabe, S., & Le Roux, J. (2015). "Phase-Sensitive and Recognition-Boosted Speech Separation Using Deep Recurrent Neural Networks." Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 708–712.
13. Le Roux, J. (2011). "Phase-Controlled Sound Transfer Based on Maximally-Inconsistent Spectrograms." Proceedings of the Acoustical Society of Japan Spring Meeting, 1-Q-51.
14. Nawab, S., Quatieri, T., & Lim, J. (1983). "Signal Reconstruction from Short-Time Fourier Transform Magnitude." IEEE Transactions on Acoustics, Speech, and Signal Processing, 31(4), 986–998.
15. Fu, S., Tsao, Y., Lu, X., & Kawai, H. (2017). "End-to-End Waveform Utterance Enhancement for Direct Evaluation Metrics Optimisation by Fully Convolutional Neural Networks." arXiv preprint arXiv:1709.03658.
16. Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). "Densely Connected Convolutional Networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
17. Veaux, C., Yamagishi, J., & MacDonald, K. (2017). "CSTR VCTK Corpus: English Multi-Speaker Corpus for CSTR Voice Cloning Toolkit."
18. Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., & Pallett, D. S. (1993). "DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM. NIST Speech Disc 1-1.1." NASA STI/Recon Technical Report N, 93.
19. Rix, A. W., Beerends, J. G., Hollier, M. P., & Hekstra, A. P. (2001). "Perceptual Evaluation of Speech Quality (PESQ)-A New Method for Speech Quality Assessment of Telephone Networks and Codecs." Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 749–752.
20. Tu, M., & Zhang, X. (2017). "Speech Enhancement Based on Deep Neural Networks with Skip Connections." Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 5565–5569.
21. Dhal, M., Ghosh, M., Goel, P., Kar, A., Mohapatra, S., & Chandra, M. (2015). A unique adaptive noise canceller with an advanced variable-step BLMS algorithm (p. 178). https://doi.org/10.1109/icacci.2015.7275605
22. Ferdouse, L., Akhter, N., Nipa, T. H., & Jaigirdar, F. T. (2011). Simulation and Performance Analysis of Adaptive Filtering Algorithms in Noise Cancellation. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.1104.1962
23. Hadei, S. A., & Lotfizad, M. (2010). A Family of Adaptive Filter Algorithms in Noise Cancellation for Speech Enhancement. In International Journal of Computer and Electrical Engineering (p. 307). https://doi.org/10.7763/ijcee.2010.v2.153
24. Loizou, P. C. (2007). Speech Enhancement. In CRC Press eBooks. Informa. https://doi.org/10.1201/9781420015836

25. Rakesh, P., & Kumar, T. K. S. (2015). A Novel RLS-Based Adaptive Filtering Method for Speech Enhancement. In World Academy of Science, Engineering and Technology, International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering (Vol. 9, Issue 2, p. 176). World Academy of Science, Engineering and Technology.https://waset.org/Publication/a-novel-rls-based-adaptive-filtering-method-for-speech-enhancement/10000580
26. Rehr, R., & Gerkmann, T. (2021). SNR-Based Features and Diverse Training Data for Robust DNN-Based Speech Enhancement. In IEEE/ACM Transactions on Audio Speech and Language Processing (Vol. 29, p. 1937). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/taslp.2021.3082702
27. Ferdouse, L., Hossain, M. Z., & Ali, M. H. (2011) "Performance Analysis of Adaptive Noise Cancellation Techniques in Speech Enhancement" International Journal of Computer Applications, 27(11), 21–27.
28. Hadei, S., & Lotfizad, M. (2010) "A Family of Adaptive Filter Algorithms in Noise Cancellation for Speech Enhancement" International Journal of Computer and Electrical Engineering, 2(2), 1793–8163.
29. Mostafa, R., & Ammar, M. (2015) "Comparative Study of Adaptive Filters in Various Noise Environments" Journal of Signal and Information Processing, 6(2), 80–89.
30. Loizou, P. C. (2007) "Speech Enhancement: Theory and Practice" CRC Press.
31. Rehr, L., & Gerkmann, T. (2021) "Phase-aware Speech Enhancement Using Consistent Spectrograms" IEEE/ACM Transactions on Audio, Speech, and Language Processing, 29, 2070–2083.
32. Rehr, R., & Gerkmann, T. (2021) "SNR-Based Features and Diverse Training Data for Robust DNN-Based Speech Enhancement" IEEE/ACM Transactions on Audio, Speech, and Language Processing, 29, 1937. https://doi.org/10.1109/TASLP.2021.3082702