Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 5s (2025)
https://www.jneonatalsurg.com

Advancements in Noise Removal: A Review of Traditional and AI-Driven Audio Signal Processing Techniques

Nilav Darsan Mukhopadhya[1], Suparna Panchanan[2], Ranjan Banerjee[3], Subhadip Kowar[4], Debmalya Mukherjee[5], Most Mahabuba Islam[6]

[1]Computer Science and Engineering - CS & DS, Brainware University
Email ID: contactnilav@gmail.com
[2]Computer Science and Engineering - CS & DS, Brainware University
Email ID: suparna_mou2k@yahoo.co.in
[3]Computer Science and Engineering, Brainware University
Email ID: rnb.cse@brainwareuniversity.ac.in
[4]Computer Science and Engineering - CS & DS, Brainware University
Email ID: subhadipkowar4@gmail.com
[5]Computational Sciences Department, Brainware University
Email ID: dbm.cs@brainwareuniversity.ac.in
[6]Computer Science and Engineering(AI & ML), Brainware University
Email ID: bwubta21092@brainwareuniversity.ac.in

.Cite this paper as: Nilav Darsan Mukhopadhya, Suparna Panchanan, Ranjan Banerjee, Subhadip Kowar, Debmalya Mukherjee, Most Mahabuba Islam, (2025) Advancements in Noise Removal: A Review of Traditional and AI-Driven Audio Signal Processing Techniques. *Journal of Neonatal Surgery*, 14 (5s), 585-589.

ABSTRACT

Noise removal is one of the more prominent research areas of signal processing. Its diversified application scope from diverse falls into the list under speech enhancements, restorations of music, telecommunications aids, and a hearing aid, which reduced unwanted noises and developed through minimal deterioration of its original methods over time. There is a review of the traditional techniques as well as some of the modern techniques that include spectral subtraction, Wiener filtering, and adaptive filtering. Deep learning-based methods in recent years among such approaches are convolutional neural networks and recurrent neural networks. These discussions include capabilities and limitations related to computation efficiency, real-time processing, and effectiveness in the presence of noisy environments. In conclusion, there is a presentation of future directions for research through the use of hybrid models as well as the artificial intelligence-driven approach to eliminate noise for better and adaptive applications.

1. INTRODUCTION

Noise reduction from audio signals is required to enhance the speech clarity, improve the quality of music, and ensure clear communication in a number of applications such as telecommunication and hearing aids. Since several years back, researchers have designed many techniques for noise reduction techniques, from the classical methods of signal processing to modern machine learning-based techniques.There exist traditional methods like spectral subtraction, Wiener filtering, and adaptive filtering that has widely been used but is plagued with non-stationary noise. Deep learning ongoing studies have dramatically improved noise removal in recent years with special promise with the DNN, CNN, RNN, and GAN architectures far outperforming the traditional approaches significantly. Hybrid approaches with the best of both worlds will solve current problems with even more efficiency and promisingly. This review covers evolution of the technique based on analysis of its effectiveness and limitations, in addition to guidelines towards effective real-time noise cancellation.

2. TRADITIONAL NOISE REMOVAL TECHNIQUES

2.1 Spectral Subtraction

Spectral subtraction is the most widely used noise reduction technique applied on audio signals. Boll first introduced it in 1979. It works based on the estimation of the noise spectrum and subtracted from the noisy speech signal in the frequency domain. It is particularly effective for stationary noises, such as background hums or white noise, and fan noise, but can Introduce artifacts like "musical noise.

This spectral subtraction technique assumes that a noisy speech signal is composed of clean speech and noise by transforming the signal into the frequency domain using the Short-Time Fourier Transform (STFT), we have

$$Y(t)=X(t)+N(t) \quad (1)$$

$$Y(f)=X(f)+N(f) \quad (2)$$

Where Y(f), X(f), and N(f) represent the frequency spectra of the noisy speech, clean speech, and noise, respectively. The phase is retained, and the inverse STFT reconstructs the denoised audio signal.

2.2 Wiener Filtering

Scalart & Filho, (1996), is a statistical method, applying the Wiener filtering method whose aim is minimizing the mean square error between clean and noisy signals. The filter relies on previous knowledge of noise and hence its optimality would be under a low-noise environment. In the case of the Wiener filter, which is in frequency domain, this tries to make the best estimates of clean speech signal, it removes noise through prior a-priori SNR estimates.

$$H(f) = Sx(f)/ Sx(f)+Sn(f)$$

Here, Sx(f) and Sn(f) denote the power spectral densities of the clean speech and noise respectively. This formulation ensures that frequencies with a higher SNR are preserved, while those dominated by noise are attenuated.

2.3 Adaptive Filtering

The statistical characteristics of signal and noise control the adaptive filters. The use of adaptive filter is seen for real-time reduction of noise characteristics that vary by time. It is popular by the following adaptations:

- Least Mean Squares : The algorithm consists of minimizing difference between desired and actual signals and a gradient flow.
- Recursive Least Squares: The convergence of RLS is faster than that of LMS since it tries to minimize the mean square error recursively.
- Kalman Filtering: Estimates the state of a system dynamically makes it more appropriate for tracking time-varying noise.

Traditional methods such as spectral subtraction and Wiener filtering are used considerably because they are simple and low-computational. Spectral subtraction makes noise estimations and removes them from the Signal, however introduces musical noise artifacts most of the times. Wiener filtering, minimizing mean squared error, does better at keeping speech intelligibility but fails at non-stationary noise.

3. MACHINE LEARNING-BASED NOISE REMOVAL TECHNIQUES

3.1 Deep Neural Networks (DNNs)

Deep learning has revolutionized noise removal completely using big datasets for training models in order to Identify noise from clean signals. Techniques based on DNN have been proven effective (Xu et al., 2015).Learning patterns in noisy audio data allow the performance over traditional methods.They are extremely computationally intensive and require huge amounts of training data.

3.2 Convolutional Neural Networks (CNNs)

The most famous CNN-based approach towards denoising is Denoising CNN. Following residual learning, they have used audio signals by interpreting the audio waveform or spectrogram as a 2D signal. Audio signals are converted into a spectrogram; it is the 2D representation of frequency vs. time. The CNNs are applied to learn spatial patterns in the spectrogram. It removes noise by learning the mapping from noisy spectrograms to clean ones. CNNs work very well for Gaussian noise and any other form of degradation. Because of the ability that it possesses to capture features in a spatially effective way, CNNs are used immensely to apply image restoration, which includes denoising. It was implemented with the use of a spectrogram transform to remove noise from audio signals by casting them into 2D representations (Park et al., 2017). CNNs are very good at capturing spatial features of noise and speech components and permit high-quality denoising. However, they fail with non-stationary changes in noise.

Autoencoders are neural networks that learn to encode an image into a lower-dimensional representation and then reconstruct it. Denoising Autoencoders (DAE) is one type of autoencoder that was trained to remove noise by reconstructing clean images from noisy inputs. These generally consist of encoder-decoder architecture. They could be very helpful in learning the data-specific noise reduction patterns.

3.3 Recurrent Neural Networks (RNNs)

RNNs (Hershey et al., 2016) are much more computationally efficient in modeling the temporal dependencies found in the audio signal, and it is consequently used in speech enhancement noise reduction. They have been known to yield better performance than other traditional methods, although a lot of computations have to be carried out. RNNs form a class of neural networks which is applied on tasks involving prediction in sequences that follow sequential nature in input and have dependence with the previous time steps. Audio signals,Especially speech and music, have temporal dependencies, such that the value of the signal at a particular time step depends on previous values. This property makes RNNs an ideal architecture for modeling audio data because they can capture such dependencies through their recurrent connections.

In audio denoising, an RNN scans the noisy audio signal step-by-step and learns to filter. Coupled architecture, although it overcame the noise patterns over time, was allowed to take the essential features of the signal. It updated its hidden state using both the present input and the previous hidden state of time steps; however, RNNs suffer from some issues such as the vanishing and exploding gradient while learning long sequences.

3.4 Long Short-Term Memory (LSTM)

Long Short-Term Memory networks is a family of RNNs who was specially designed in an attempt to remedy the lack in normal RNNs that concentrate intensely on vanishing gradients problem. An important property that gives LSTMs the possibility of learning long dependencies is being more powerful in relation to usual RNNs with applications as those of audio denoising that may appear far apart across various time steps.

The flow of the information in such networks is mainly regulated by means of a stack of gates the input gate, forget gate, and output gate. This lets the LSTMs keep preserving the memory along time while appropriately updating it since some features ought to be given higher importance during training. Otherwise, those meaningless and noisy may not be helpful during processing but waste the network. Especially, noise removal has very high applicability for an audio denoise.

An LSTM-based audio denoising model works by passing noisy audio features such as spectrograms or Mel-frequency cepstral coefficients to the input of a network. It then processes over time the feature using the LSTM to learn the transformation of the noisy input to clean output. Using LSTM to model short as well as long-term dependencies enables the network to remove noisy audio content, and whatever really matters - for instance, clarity of speech or quality of music - is retained in place.

3.5 Generative Adversarial Networks (GANs)

Recently, Pascual et al. (2017) introduced a promising technique with GAN for audio denoising, such that the quality of the cleaned audio signals should be evaluated with a discriminator against a generator network trying to clean the audio signal. The primary challenge in their state-of-art performance is its training stability and convergence. The Generative Adversarial Network (GAN) is one of the most recent inventions in the field of deep learning, where Goodfellow et al., in 2014, used this concept with the name.Class of machine learning models designed to generate new, synthetic instances of data that resemble a given training dataset. The architecture of GANs consists of two neural networks — the generator and the discriminator — that compete in a game-theoretic setting, hence the term "adversarial." The generator learns to create realistic data samples, while the discriminator learns to distinguish between real and fake data. This adversarial process leads to the generation of highly realistic outputs make GANs super powerful for various generative activities, such as creating images, videos, audio.

The generator is a neural network that accepts, as input, a random noise vector-known as the latent vector-and transforms it into data sample, as an image or an audio, a video-the copy of the characteristics of the real data of the training set. It should produce synthetic data the discriminator couldn't distinguish from the real data.

Discriminator: It is a neural network that learns whether the input received is actual or another created through the generator. In other words, this discriminator gives feedback to the generator to increase the quality of its output.

Deep learning-based techniques have been more efficient by using neural networks like CNNs, RNNs, and transformer models. These models learn noise patterns from data.

Generalizing to different kinds of noise can be improved upon. Autoencoders, u-net, wave net, and GANs have worked superbly for improvement in speech clarity. However, deep learning-based techniques are significantly data-intensive as well as computation-intensive and thereby present a challenge for real-time applications.

4. HYBRID APPROACHES

Up-to-date studies indicate hybrid approaches to blend shallow and deep learning-based methods for enhanced robustness. For instance, the blending of Wiener filtering with deep learning is more appropriate in noise reduction and remains computationally viable (Kim and Stern, 2018). Analogously, the blending of CNNs with adaptive filtering achieves enhanced generalizability to various noise (Zhao et al., 2020).

Frequency-domain denoising utilizes spectral subtraction as the first step. Following

first reduction, one employs a deep neural network (DNN), i.e., Convolutional Neural Network (CNN), to further process the result denoised from learning residual noise. The DNN will be able to eliminate artifacts and enhance signal quality, particularly when it is in non-stationary states of noise.

Wiener filtering supplies the preliminary approximation of noise elimination. Later on, RNNs or LSTMs are used to additional process the signal considering temporal patterns of audio data. RNNs suit appropriately to fit in sequential data and thus suit particularly well for de-noising the audio signals since the nature of noise is varying with time.

Hybrid approaches combine deep learning with conventional methods for an attempt to find a trade-off between computation cost and noise removal. Augmentation of Wiener filtering with deep learning networks enhances noise estimation, and addition of spectral subtraction as part of preprocessing enhances overall performance. Non-supervised and self-supervised learning approaches also offer consistent solutions for real-world deployment in case of inadequate labeled training set.

5. RESULTS AND DISCUSSION

The execution of noise reduction algorithms shows erratic performance based on the nature of the noise,computational complexity, and real-time adaptability. Spectral subtraction as a low computational algorithm is bedeviled by musical noise artifacts in non-stationary noise environments. Wiener filtering as an ideal filter under ideal conditions is bedeviled by noisy estimation of noise, particularly under severe dynamic conditions. Adaptive filtering, particularly LMS and RLS algorithms, are ideal for real-time implementations with enhanced performance under non-stationary noise conditions.Though with increased computations and cautious parameter tuning. Spectral subtraction-based hybrid approaches combining Wiener filtering and adaptive filtering provide improved noise reduction performance, balancing computational complexity and noise reduction.Deep learning approaches such as convolutional neural networks (CNNs) and recurrent neural networks(RNNs) have provided best-performing noise reduction performance, especially in speech enhancement.But they require insatiable training data and extremely computational resources.Generally, choice of the noise reduction method is application-specific.computational complexity, and the characteristics of noise. There is a need for more work to further enhance robustness, decrease computational cost, and achieve optimal performance in real-world applications.Performance metrics like SNR, PESQ, STOI, and MOS were employed for estimating the effectiveness of such technologies. Experimental results indicate that deep learning methods would prevail over traditional methods but must be adequately trained and supported in terms of hardware while being deployed in actual situations. Traditional methods remain superior under computational economies-oriented scenarios.Special care must be taken in adapting to multi variety noise environments. Classical methods work optimally with highly controlled environments but are not effective with dynamic noise. Deep learning models, especially those trained on multivariate datasets, exhibit greater adaptability. They are thus optimal to apply in voice assistants, teleconferencing, and hearing aids.Even with the progress, it is still challenging to balance noise removal and speech distortion.Over-denoising results in the creation of artifacts, thus compromising speech quality. Future research should be aimed at designing algorithms for preserving speech integrity and filtering noise. In addition, reduced dependency on large labeled datasets with the use of unsupervised and semi-supervised learning makes it more viable. Reinforcement learning, meta-learning, and multimodal techniques based on vision and contextual data for improved denoising are the directions for future research. Lip-reading technology combined with audio processing is able to enhance speech intelligibility in noisy environments. Hardware technological developments like TPUs and GPUs also allow deep learning-based real-time noise reduction.

6. FUTURE DIRECTIONS

Existing future studies on noise reduction from audio signals must focus on the development of stronger and adaptive approaches. The fusion of deep learning with traditional techniques for noise reduction has provided good results, but much more work needs to be done so that real-time performance is further enhanced. Application of hybrid approach based on spectral subtraction, Wiener filtering, and adaptive filtering could yield better output leveraging the strengths of these approaches. Furthermore, advancements in the architecture of neural networks such as transformer-based and generative adversarial GANs (networks) have the potential to improve noise suppression a bit with distortion.

Another very important area is real-time processing efficiency because most of the applications to eliminate noise such as teleconferencing and hearing aids require fast and low-latency resolution. Optimization of deployment efficiency for deep learning in edge devices and for mobile devices will be very important in helping cost savings for noise reduction. Apart from this, context-aware noise reduction must be explored where systems are able to identify between good background noise and poor background noise. This would offer better user experience in a variety of applications, such as augmented reality (AR) and virtual reality (VR) audio environments.

Finally, the development of benchmarking data and metrics will be essential in order to offer a level playing field for the comparison of different noise removal methods. Having standardized data will allow researchers to compare performance directly under different conditions of noise and push research efforts in the field.

7. CONCLUSION

Audio denoising research has compared different methods, such as conventional signal processing, deep learning algorithms, and compound methods. Each of the methods has strength and weakness.because efficiency would be a function of variables such as the noise type, capacity to process, and deployment situation.In short, the paper mentions that while traditional approaches enjoy the advantage of simplicity and online implementation, deep learning approaches attain better denoising with the penalty of higher

computational demands. The difference between such paradigms is filled by hybrid approaches and self-supervised learning. More research and technological innovations will keep on challenging audio denoising, telecommunication, assistive hearing aids, etc.

REFERENCES

[1] . Haykin, S. (2002). Adaptive Filter Theory. Pearson Education.

[2] Hershey, J. R., Chen, Z., Le Roux, J., & Watanabe, S. (2016). Deep clustering and conventional networks for music separation. IEEE Transactions on Audio, Speech, and Language Processing, 24(3), 693-703.

[3] Kim, C., & Stern, R. M. (2018). Power-normalized cepstral coefficients (PNCC) for robust speech recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing,24(7), 1315-1329.

[4] Loizou, P. C. (2013). Speech Enhancement: Theory and Practice. CRC Press.

[5] Park, S., Lee, T., & Nam, J. (2017). Music denoising using convolutional neural networks. IEEE Transactions on Multimedia, 19(7), 1803-1816.

[6] Pascual, S., Bonafonte, A., & Serrà, J. (2017). SEGAN: Speech Enhancement Generative Adversarial Network. Interspeech 2017, 3642-3646.

[7] Zhao, X., Wang, Y., & Tan, K. (2020). A hybrid approach for noise reduction using convolutional neural networks and adaptive filtering. IEEE Signal Processing Letters,27, 1255-1259.