2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10340687

# Denoising Cough Sound Recordings Using Neural Networks*

Laya Jose[1], Shlomo Berkovsky[1], Hao Xiong[1], Cecilia Mascolo[2], and Roneel V. Sharan[1]

*Abstract*— **Objective cough sound evaluation is useful in the diagnosis and management of respiratory diseases. However, the performance of cough sound analysis models can degrade in the presence of background noises common in everyday environments. This brings forward the need for cough sound denoising. This work utilizes a method for denoising cough sound recordings using signal processing and machine learning techniques, inspired by research in the field of speech enhancement. It uses supervised learning to find a mapping between the noisy and clean spectra of cough sound signals using a fully connected feed-forward neural network. The method is validated on a dataset of 300 manually annotated cough sound recordings corrupted with babble noise. The effect of various signal processing and neural network parameters on denoising performance is investigated. The method is shown to improve cough sound quality and intelligibility and outperform conventional denoising methods.**

## I. INTRODUCTION

Cough is one of the most common and familiar symptoms of various respiratory diseases, such as pertussis, pneumonia, and COVID-19. In clinical practice, cough assessment can be performed using various methods, such as visual analogue scales, verbal descriptive scores, and quality-of-life questionnaires [1]. However, these are subjective assessments, depending on the patient's attention to the symptoms and the understanding and interpretation of the cough by a physician.

Objective cough sound analysis methods, such as cough frequency measurement using cough monitors [2] and smartphone-based artificial intelligence algorithms for cough sound assessment [3], are promising tools that can aid medical practitioners in the diagnosis of respiratory diseases. There is also the possibility of using cough sound analysis as a screening tool for COVID-19 [4]. Since the outbreak of COVID-19, there has been a substantial increase in telehealth consultations [5], where such objective cough sound assessment technologies can be invaluable.

However, the presence of everyday background noise can degrade the quality of the cough sound signals and, therefore, aggravate objective cough sound assessment. Cough signal denoising is, therefore, important before objective cough sound analysis can be performed. Cough signal denoising has, however, received limited attention, although denoising audio signals in other applications, speech analysis, in particular, has been ongoing for decades. As speech and cough share similarities in the generation process [6], our work is inspired by prior speech denoising and enhancement studies.

Conventional speech enhancement methods, such as Wiener filtering and spectral subtraction, have several limitations, like dependence on the nature of the background noise and statistical properties of the target signal [7]. In recent years, data-driven methods, such as deep neural networks [8], [9], have become increasingly popular for speech enhancement. This is due to their effectiveness in capturing the nonlinear relationship between the clean and noisy speech, including in non-stationary noisy environments, where the performance of conventional denoising methods degrades. Our work is inspired by these studies which focused on mapping between noisy and clean speech spectra for speech enhancement using neural networks.

In this work, we use a deep neural network technique to denoise the cough signals for enhancing the quality and intelligibility of the cough signals obtained from a crowdsourced dataset. In particular, we study the use of a fully connected feed-forward neural network (FNN) for cough denoising. The method is evaluated using a dataset of manually annotated cough recordings. We study the effect of various signal processing and neural network parameters on the denoising performance. The performance of the neural network-based cough denoising method is also compared against a conventional denoising method. The results show enhancement in cough sound quality and intelligibility, and, therefore, the denoising method has the potential to improve objective cough sound analysis in the presence of background noise.

## II. MATERIALS AND METHODS

### A. Dataset

This work exploits a publicly available COUGHVID dataset [10]. The COUGHVID dataset was crowdsourced and contains 27,550 audio recordings of cough sounds. The data has been collected through a web application for a study of COVID-19 cough sound. The cough recordings have been submitted by subjects, who also self-reported additional data, such as age, gender, COVID-19 status, and symptoms.

For denoising the cough recordings, this work considered a subset of 200 cough recordings from the COUGHVID dataset, which are used in [11]. These recordings were manually screened to eliminate those containing non-cough sounds or background noises. This resulted in 114 usable recordings. In addition, we screened recordings with a high probability of cough, as identified by [11], for additional 186 recordings, resulting in a total of 300 cough recordings exploited in this work. These recordings were downsampled to 16 kHz.

*This work was supported by the Google Research Scholar Program. *(Corresponding author: Roneel V. Sharan.)*

[1]Australian Institute of Health Innovation, Macquarie University, Sydney, NSW 2109, Australia (e-mail: {laya.jose, shlomo.berkovsky, hao.xiong, roneel.sharan} @mq.edu.au).

[2]Department of Computer Science and Technology, University of Cambridge, Cambridge, United Kingdom (e-mail: cm542@cam.ac.uk).

979-8-3503-2447-1/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: Macquarie University. Downloaded on January 17,2024 at 19:34:10 UTC from IEEE Xplore. Restrictions apply.

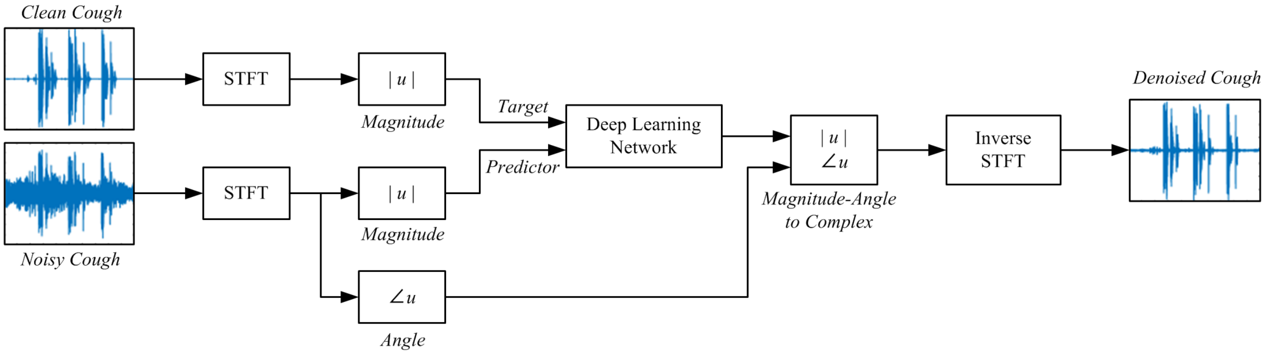

Fig. 1. An overview of the supervised deep learning method used for denoising cough sound audio recordings. The magnitude spectra of the noisy and clean cough recordings are the predictor and target inputs of the deep learning network, respectively, and the denoised spectra is the output.

TABLE I
OVERVIEW OF THE DATASET USED IN THIS WORK

| Description | Value |
|---|---|
| Number of recordings | 300 |
| Average duration (seconds) | 9.35±1.36 |
| Number of frames (256 points, 75% overlap) | 700,594 |
| Gender (male / female / unknown) | 169/90/41 |
| Average age (years) | 39.24±14.62 |

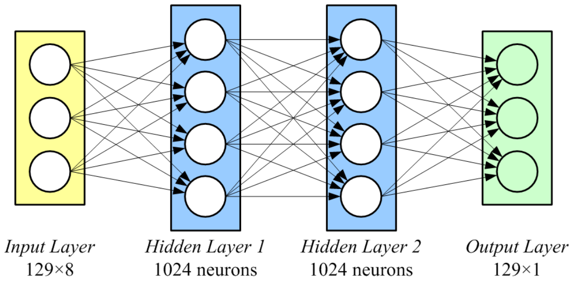

Fig. 2. An overview of the FNN architecture. The FNN has an input layer, two fully connected layers, and an output layer.

A descriptive characterization of the 300 recordings used in this work is provided in Table I. The average duration of the recordings is 9.35±1.36 seconds. The recordings provide a total of 700,594 frames of length 256 points with a 75% overlap between adjacent frames. Of the 300 subjects, 169 are male, 90 are female, and 41 did not report their gender. The average age of the participants is 39.24±14.62 years.

### B. Cough Denoising Method

Fig. 1 provides an overview of the cough denoising method used in this work. The noisy cough recordings are obtained by adding speech babble noise, available from the NOISEX-92 database, to the clean cough recordings. In similar to [8] and [9], noise is added at a signal-to-noise ratio (SNR) of 0 dB. The magnitude spectra of the noisy and clean cough recordings form the predictor and target of the FNN, respectively. The output of the FNN is the magnitude spectrum of the denoised signal. The FNN in this case is a regression network that uses the predictor input to minimize the mean square error between the output of the network and the input (target) [8], [9].

The cough recordings are transformed to the frequency domain using a short-time Fourier transform (STFT), with a window length of 256 points, an overlap of 75% between adjacent frames, and a Hamming window. The size of the Fourier transform is the same as the window length. The size of the spectral vector is reduced to 129 by dropping the symmetric half. The denoised cough signal is obtained by converting the denoised spectra to the time domain using the phase of noisy signal and inverse STFT [8].

An overview of the architecture of the FNN used in this work is shown in Fig. 2. The size of the predictor input is 129×8, as the input is reduced to 129 by removing the symmetric half and it consists of 8 consecutive noisy STFT vectors, such that the prediction of each STFT output (129×1) is based on the current noisy STFT vector and the previous 7 STFT vectors. The predictor matrices and target vectors are normalized using their mean and standard deviation values. The input layer is followed by 2 fully connected layers, each with 1024 neurons. Each fully connected layer is followed by a batch normalization layer and a ReLU layer. The output layers include a fully connected layer of size 129, same as the target vector, and a regression layer. The network is trained using the adaptive moment estimation algorithm with an initial learning rate of $1 \times 10^{-3}$, mini batch size of 128, and maximum number of epochs of 3. In addition, we use a learn rate drop factor and learn rate drop period of 0.9 and 1, respectively. The network training stops after the maximum number of epochs is reached.

### C. Setup and Metrics

The denoising experiments use all 300 cough recordings. The performance of the denoising method is evaluated in 5-fold cross-validation, whereby frames from 60 recordings are used for testing in each fold. Of the remaining 240 recordings, 90% are used for training and the remaining 10% – for validation. We investigated the effect of various signal processing and network parameters on the denoising performance. The FNN was implemented in MATLAB R2022b and trained on NVIDIA Quadro P6000 GPU.

Authorized licensed use limited to: Macquarie University. Downloaded on January 17,2024 at 19:34:10 UTC from IEEE Xplore. Restrictions apply.

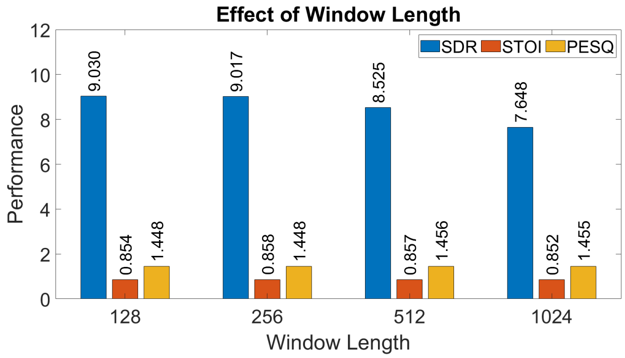

Fig. 3. Denoising performance with different window lengths.

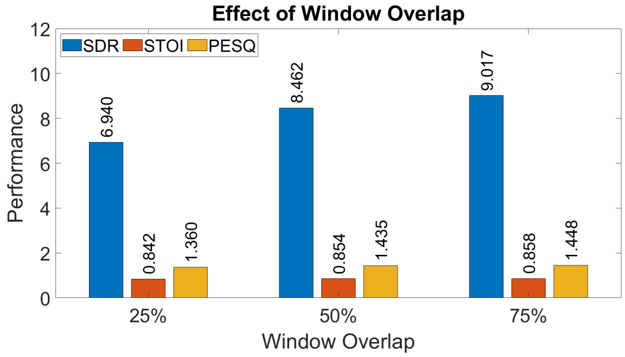

Fig. 4. Denoising performance with different window overlaps.

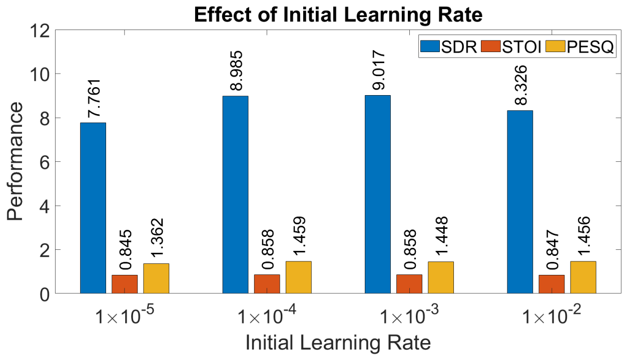

Fig. 5. Denoising performance with different learning rates.

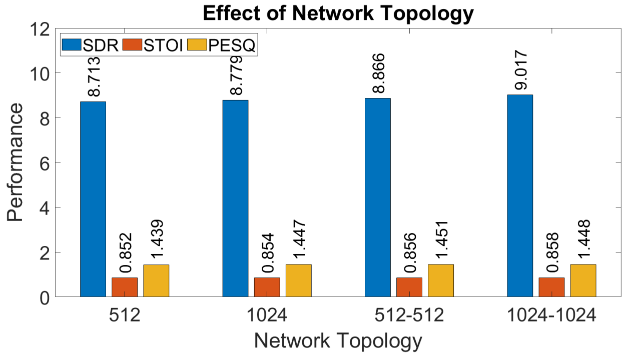

Fig. 6. Denoising performance with different network topologies.

The signal-to-distortion ratio (SDR) [12] is used as the main measure of the denoising performance. SDR measures the error between the clean and denoised cough signals in dB. It is a widely-used performance metric in audio denoising tasks. In addition, since cough and speech share similarities in the generation process, we use two common speech intelligibility measures. These are short-time objective intelligibility (STOI) [13] and perceptual evaluation of speech quality (PESQ) [14]. STOI is based on a correlation coefficient between the temporal envelopes of the clean and denoised signals in short-time overlapping segments, ranging between 0 and 1. PESQ is an integration of two perceptual analysis measurement systems with values in the range of 1 to 4.5, indicating the quality of the denoised signals. We report the average value for these metrics, where higher values indicate a better denoising performance.

## III. RESULTS

We first compare the denoising performance with various signal processing parameters. In particular, we compare different window lengths and overlaps between adjacent windows. Denoising performance with the window length of 128, 256, 512, and 1024 points, and overlap of 75% are compared in Fig. 3. The highest SDR is achieved at the window length of 128, only slightly better than at 256 points. The STOI value is highest at the window length of 256 while the PESQ values at the window lengths of 128 and 256 are identical. As such, the denoising performance with the window length of 256 yields comparable performance to the window length of 128, with an advantage of using fewer frames. We now fix the window length to 256 points and measure the denoising performance with the window overlaps of 25%, 50%, and 75%. The denoising performance in Fig. 4 shows that the window overlap of 75% outperforms the 25% and 50% overlaps for all the performance metrics.

Next, we investigate the effect of initial learning rate and network topology on the denoising performance of FNN. We experiment with four initial learning rates: $1\times10^{-5}$, $1\times10^{-4}$, $1 \times 10^{-3}$, and $1 \times 10^{-2}$. The denoising performance in Fig. 5 shows that the highest SDR and STOI are achieved with the initial learning rate of $1 \times 10^{-3}$, albeit with a slightly lower PESQ than with $1\times10^{-4}$ and $1\times10^{-2}$. We now fix the initial learning rate to $1\times10^{-3}$ and experiment with different network topologies. We experiment with one and two fully connected layers in the network, with 512 and 1024 neurons in each layer. As shown in Fig. 6, the best overall denoising performance is achieved with two fully connected layers, each having 1024 neurons. The SDR and STOI values at this network configuration are the highest, while the PESQ value is only slightly lower than when the number of neurons in each of the two layers is 512.

In Fig. 7(a), we provide an illustration of a clean cough recording and, in Fig. 7(b), a noisy version of this recording with babble noise at 0 dB SNR. The denoised cough recording using the FNN algorithm, with window length of 256, window overlap of 75%, initial learning rate of $1 \times 10^{-3}$, and 1024 neurons in each of the two layers, is illustrated in

Authorized licensed use limited to: Macquarie University. Downloaded on January 17,2024 at 19:34:10 UTC from IEEE Xplore. Restrictions apply.

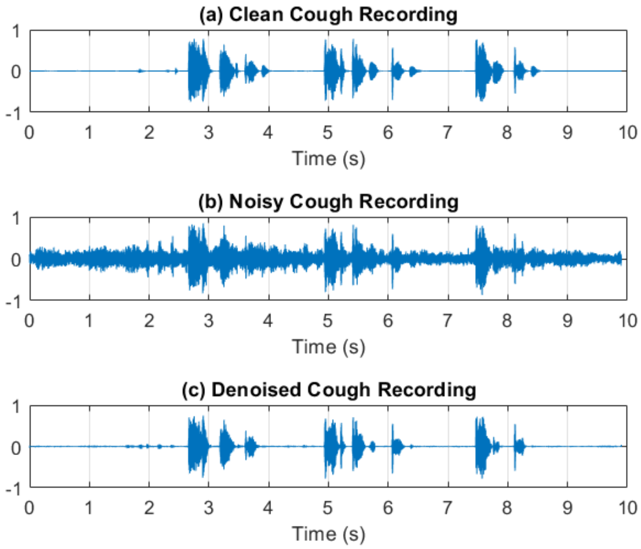

Fig. 7. Waveform of a (a) clean cough recording, (b) noisy cough recording, and (c) denoised using FNN cough recording.

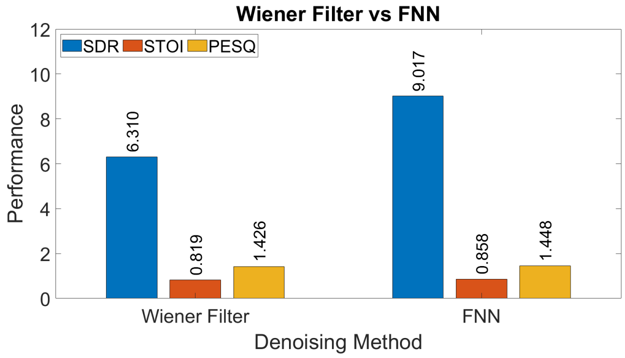

Fig. 8. Denoising performance of Wiener filter and FNN.

Fig. 7(c). Visual analysis of the denoised signal of Fig. 7(c) shows that the denoising method is effective in removing the background noise shown in Fig. 7(b) while preserving the cough sound signal characteristics shown in Fig. 7(a).

Finally, in Fig. 8, we compare the performance of the FNN-based cough sound recording denoising method used in this work against Wiener filtering [15], a commonly used method for audio denoising. The FNN method outperforms the Wiener filtering denoising with respect to all three metrics, offering a relative improvement of 42.90% in SDR, 4.76% in STOI, and 1.54% in PESQ.

## IV. Conclusion

This paper presented a method for denoising cough sound recordings using a fully connected FNN. Through various experiments, we determined the most appropriate parameterization of the FNN in terms of the window length, window overlap, initial learning rate, and network configuration. These parameter settings produced an SDR of 9.017 dB, STOI of 0.858, and PESQ of 1.448, outperforming the conventional Wiener filtering denoising method.

Earlier works on objective cough sound analysis exclude noisy cough recordings [4]. However, everyday environments, such as homes and hospitals, where an objective cough sound assessment technology would be used carry background noises. Cough sound denoising, such as the one employed in this work, can make objective cough sound assessment possible in such noise-prone environments.

Our work, however, has limitations. The dataset is limited to 300 cough recordings. A larger dataset of cough recordings could provide a stronger generalizability of the network. In addition, in this work we have studied only one type of noise, babble noise, and only at one SNR, 0 dB. However, other noise types and levels can be present in real-life. In the future, we plan to extend our study to overcome these limitations and also explore other deep learning based denoising methods and network architectures. Although our experiments are not exhaustive, they shed light on the application of neural networks for cough sound denoising purposes and, therefore, demonstrate their potential in improving objective cough sound evaluation in the presence of background noises.

## References

[1] S. Leconte *et al.*, "Validated methods of cough assessment: a systematic review of the literature," *Respiration*, vol. 81, no. 2, 2011.

[2] J. I. Hall *et al.*, "The present and future of cough counting tools," *J. Thorac. Dis.*, vol. 12, no. 9, pp. 5207–5223, 2020.

[3] P. Porter *et al.*, "A prospective multicentre study testing the diagnostic accuracy of an automated cough sound centred analytic system for the identification of common respiratory disorders in children," *Respir. Res.*, vol. 20, no. 1, Art. no. 81, 2019.

[4] C. Brown *et al.*, "Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2020, pp. 3474–3484.

[5] S. Omboni *et al.*, "The worldwide impact of telemedicine during COVID-19: current evidence and recommendations for the future," *Conn. Health*, vol. 1, no. 1, pp. 7–35, 2022.

[6] R. V. Sharan *et al.*, "Cough sound analysis for diagnosing croup in pediatric patients using biologically inspired features," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2017, pp. 4578–4581.

[7] Y. Wang *et al.*, "Speech enhancement from fused features based on deep neural network and gated recurrent unit network," *EURASIP J. Adv. Signal Process.*, vol. 2021, no. 1, Art. no. 104, 2021.

[8] D. Liu, P. Smaragdis, and M. Kim, "Experiments on deep learning for speech denoising," in *Proc. Annu. Conf. Int. Speech Commun. Assoc. (INTERSPEECH)*, 2014, pp. 2685–2689.

[9] S. R. Park and J. W. Lee, "A fully convolutional neural network for speech enhancement," in *Proc. Annu. Conf. Int. Speech Commun. Assoc. (INTERSPEECH)*, 2017, pp. 1993–1997.

[10] L. Orlandic, T. Teijeiro, and D. Atienza, "The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms," *Sci. Data*, vol. 8, no. 1, Art. no. 156, 2021.

[11] R. V. Sharan, H. Xiong, and S. Berkovsky, "Detecting cough recordings in crowdsourced data using CNN-RNN," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, 2022, pp. 1–4.

[12] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, 2006.

[13] C. H. Taal *et al.*, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE Trans. Audio Speech Lang. Process.*, vol. 19, no. 7, pp. 2125–2136, 2011.

[14] A. W. Rix *et al.*, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2001, pp. 749–752.

[15] C. Plapous, C. Marro, and P. Scalart, "Improved signal-to-noise ratio estimation for speech enhancement," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 6, pp. 2098–2108, 2006.

Authorized licensed use limited to: Macquarie University. Downloaded on January 17,2024 at 19:34:10 UTC from IEEE Xplore. Restrictions apply.