International Journal of Speech Technology (2022) 25:745–758
https://doi.org/10.1007/s10772-022-09987-4

Single-channel speech enhancement using implicit Wiener filter for high-quality speech communication

Rahul Kumar Jaiswal[1] · Sreenivasa Reddy Yeduri[1] · Linga Reddy Cenkeramaddi[1]

Received: 24 August 2021 / Accepted: 17 June 2022 / Published online: 1 August 2022

Abstract
Speech enables easy human-to-human communication as well as human-to-machine interaction. However, the quality of speech degrades due to background noise in the environment, such as drone noise embedded in speech during search and rescue operations. Similarly, helicopter noise, airplane noise, and station noise reduce the quality of speech. Speech enhancement algorithms reduce background noise, resulting in a crystal clear and noise-free conversation. For many applications, it is also necessary to process these noisy speech signals at the edge node level. Thus, we propose implicit Wiener filter-based algorithm for speech enhancement using edge computing system. In the proposed algorithm, a first order recursive equation is used to estimate the noise. The performance of the proposed algorithm is evaluated for two speech utterances, one uttered by a male speaker and the other by a female speaker. Both utterances are degraded by different types of non-stationary noises such as exhibition, station, drone, helicopter, airplane, and white Gaussian stationary noise with different signal-to-noise ratios. Further, we compare the performance of the proposed speech enhancement algorithm with the conventional spectral subtraction algorithm. Performance evaluations using objective speech quality measures demonstrate that the proposed speech enhancement algorithm outperforms the spectral subtraction algorithm in estimating the clean speech from the noisy speech. Finally, we implement the proposed speech enhancement algorithm, in addition to the spectral subtraction algorithm, on the Raspberry Pi 4 Model B, which is a low power edge computing device.

Keywords Edge computing · Non-stationary noise · Raspberry Pi · Spectral subtraction · Speech analysis · Stationary noise · Wiener filtering

1 Introduction

Speech is the vocalized form of human communication. It is also one of the preferred methods of communication between humans and machines in a variety of speech processing applications, including speech recognition (Schultz et al., 2021), speech coding (Kleijn et al., 2018), and speech quality estimation of disordered tracheoesophageal speech of patients (Ali et al., 2020). When the input speech signal is in its original (clean) form, these applications perform admirably. However, due to the presence of background noise in the surroundings, the input speech signal suffers due to the presence of noises such as stationary noise (fan, white noise) and non-stationary noise (exhibition, station, drone, helicopter, airplane), among others. As a result, in order to achieve better performance, it is necessary to improve the speech before injecting it into any specific speech processing applications.

Speech enhancement algorithms improve the speech quality that is degraded by additive noise such as exhibition noise, drone noise, etc., as shown in Fig. 1. Speech identification (Sheft et al., 2008), speech recognition system (Moore et al., 2017), unmanned aerial vehicle (UAV)-based search and rescue system (Deleforge et al., 2019) are among the speech enhancement applications. Speech enhancement improves the quality and intelligibility of noisy speech signals (Das et al., 2020). It can be used as a pre-processing block as well as a filter to remove background noise, as in the cellular phone (Ogunfunmi et al., 2015). Further, the

✉ Linga Reddy Cenkeramaddi
linga.cenkeramaddi@uia.no

Rahul Kumar Jaiswal
rahul.jaiswal@uia.no

Sreenivasa Reddy Yeduri
sreenivasa.r.yeduri@uia.no

[1] Department of Information and Communication Technology, University of Agder, Grimstad, Norway

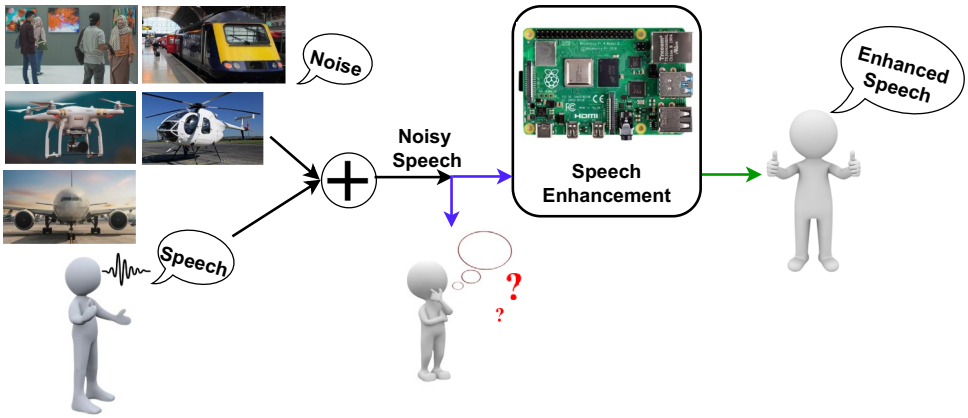

Fig. 1 Single channel speech enhancement system

spectral characteristics of the noise, as well as the number of available microphones, have an impact on the speech enhancement algorithms. A single microphone or channel, in general, is considered for mobile applications due to its cost and size (Loizou, 2013). Moreover, noise in the real-world can be stationary or non-stationary, with varying spectral properties. A speech enhancement algorithm is said to be effective when the noise in a noisy speech signal is estimated accurately without introducing perceptible distortion. An inaccurate noise estimation can distort the speech signal and musical noise (Loizou, 2013).

Spectral subtraction (SS) is a traditional and widely used speech enhancement algorithm for improving the speech signal which is degraded by the additive background noise (Boll, 1979; Loizou, 2013). The algorithm employs a forward and inverse Fourier transform of the input noisy speech signal. However, the algorithm suffers with the problem of "musical noise" (Loizou, 2013) present in the enhanced speech signals. A multi-band spectral subtraction algorithm for improving the fan-noise-degraded speech has been proposed in Saldanha and Shruthi (2016). However, the presence of non-stationary noise degrades the performance of this algorithm. In Yan et al. (2020), an iterative graph spectral subtraction algorithm to suppress the noise using graph spectral subtraction (GSS) has been proposed. The GSS leverages the differences in the spectrum of speech graphs and noise graph signals. However, this algorithm is explored only with the stationary noise, such as, white, and pink noise, and the non-stationary noise, such as, babble noise. They have not considered the aerial environmental noise to investigate the suitability of the algorithm.

Wiener filtering (WF) follows a conceptually similar approach for speech enhancement. A Wiener filter-based speech enhancement algorithm for additive white Gaussian noise (AWGN) and colored noise has been presented in Abd El-Fattah et al. (2014). The filter transfer function in Abd El-Fattah et al. (2014) is based on speech signal statistics; namely, the local mean and local variance. This algorithm, on the other hand, is implemented in the time domain. In the frequency domain, a speech enhancement algorithm that combines voiced speech probability with wavelet decomposition has been proposed in Bhowmick and Chandra (2017). This algorithm, however, is incapable of optimizing the multi-taper spectrum, resulting in poor speech signal denoising. Similarly, a combination of Wiener filter and Karhunen–Loeve transform (KLT) for enhancing the degraded speech has been introduced in Srinivasarao and Ghanekar (2020). A time-frequency mask-based parametric Wiener filter in Chiea et al. (2019) considers the trade-off between speech distortion and noise reduction for designing and minimizing the cost function in order to obtain optimal solution for noise suppression. However, these methods do not provide the degree of flexibility that allows the engineer to control the estimate of noise power spectral density (PSD) to suppress the musical noise components.

To make the spectral estimation smooth in order to reduce the musical noise during speech enhancement, Hu and Loizou (2004) employed low-variance spectral estimators based on wavelet thresholding the multi-taper spectrum of speech sample. Speech-shaped noise and car noise are investigated for the suitability of the algorithm. However, in Hu and Loizou (2004), the effect of aerial environmental noises which are having different spectral characteristics have not been analysed. In Charoenruengkit and Erdöl (2010), the effect of spectral estimate variance on the quality of speech enhancement system has been analysed. Reducing the variance of the spectral estimator improves the speech enhancement system. In Charoenruengkit and Erdöl (2010), babble, car, jet airplane, and white noise are considered. However, the aerial environment noises have not been considered. Spectral subtraction algorithm and SNR-dependent noise estimation approach have been considered in Islam et al. (2018) for speech enhancement. The magnitude and phase of the noisy speech spectrum are modified in order to enhance the speech. Moreover, babble and street noise are considered, however, the aerial environment noises have not been considered. The effect of musical noise on speech enhancement is also missing in Islam et al. (2018). In Kanehara

et al. (2012), three different estimators to mathematically investigate the amount of musical noise generated during speech enhancement are investigated. A practical example is shown for the white and babble noise, but not for the aerial environment noises. Moreover, none of the above algorithms provide degree of flexibility to estimate the noise which is the major parameter in speech processing.

To this end, a single-channel speech enhancement algorithm based on implicit Wiener filter (IWF) with recursive noise estimation technique is proposed (Jaiswal & Romero, 2021), which provides degree of flexibility to the engineer to estimate the noise for speech enhancement. The algorithm is implemented in frequency domain, only for the speech degraded by non-stationary noise. However, the speech degraded by stationary noise is not considered in Jaiswal and Romero (2021). It also does not take into account the noisy environments created by the aerial environments such as UAV noise, helicopter noise, and airplane noise. In addition, it lacks embedded hardware implementation using low-cost edge computing node, Raspberry Pi. Motivated by this, in this work, we propose and extend the implicit Wiener filter-based speech enhancement algorithm that provides the degree of flexibility to suppress the noise to enhance the speech. The key contributions of this paper are as follows:

- The proposed implicit Wiener filter-based speech enhancement algorithm is analysed in the presence of stationary noise such as the AWGN noise and real-world non-stationary noisy environments such as exhibition and station.
- In addition, we consider novel aerial environmental noises such as drone, helicopter, and airplane to evaluate the performance of proposed speech enhancement algorithm.
- The proposed speech enhancement algorithm provides degree of flexibility using α, which is the noise smoothing parameter, and γ, which is the noise adjustable parameter, to estimate the noise in order to enhance the speech accurately.
- Through extensive results, we compare the performance of the proposed speech enhancement algorithm with the spectral subtraction algorithm and achieve relatively better performance. The enhanced speech obtained with proposed speech enhancement algorithm shows the smoother spectrum.
- Finally, we implement the proposed speech enhancement algorithm on the Raspberry Pi 4 Model B hardware. This, in turn, shows that the proposed speech enhancement algorithm can be implemented on the low computational embedded platforms.

The rest of the paper is organized as follows. The literature survey is described in Sect. 2. A review on the spectral subtraction algorithm is presented in Sects. 3 and 4 presents a review on the implicit Wiener filter in frequency domain. Section 5 describes the noise estimation technique, and Sect. 6 describes the experimental dataset. The evaluation methodology for the speech enhancement algorithm is outlined in Sect. 7. The implementation of the speech enhancement algorithms using edge computing system is described in Sect. 8. The results are presented and discussed in Sect. 9. Finally, Sect. 10 provides the concluding remarks and the scope for future work.

2 Related work

A deep neural network (DNN)-augmented colored-noise Kalman filter-based speech enhancement system has been proposed in Yu et al. (2020) that models both clean speech and noise as auto-regressive process. The parameters of auto-regressive processes comprising of linear prediction coefficients and driving noise variances which are obtained by training multi-objective DNNs (Yu et al., 2020). It is important to note here that the existing DNN-based speech enhancement models extract only local features from the noisy speech in a non-causal way. Thus, in Yuan (2020), a time-frequency smoothing neural network has been proposed. The proposed network in Yuan (2020) works on time-frequency correlation in the improved minima controlled recursive averaging (MCRA)-based feature calculation using long short-term memory (LSTM) and convolutional neural network (CNN) (Shrestha and Mahmood, 2019). A generative adversarial networks (GANs) (Creswell et al., 2018) based speech enhancement algorithm to estimate the clean signal from the corrupted signal has been proposed in Pascual et al. (2019). In You and Ma (2017), a modified scheme has been proposed that includes a smoothing adaptation to the frame signal-to-noise ratio (SNR) and a re-estimation of previous SNR in order to reduce the artifacts for speech enhancement. However, these DNN-based models require a huge amount of speech data for training. Therefore, we propose a speech enhancement algorithm which achieves a good performance with limited amount of data.

A multi-band algorithm to the spectral subtraction is proposed in Kamath et al. (2002), where the speech spectrum is divided into multiple non-overlapping bands. Afterwards, spectral subtraction has been performed independently in each band. Finally, the modified frequency bands are recombined to obtain the enhanced speech. This algorithm has been investigated on the colored noise alone. Further, the multi-band algorithm works on the fact that the noise does not affect the speech signal uniformly. In Asano et al. (2000), a speech enhancement algorithm based on the subspace approach has been proposed. The proposed algorithm in Asano et al. (2000) reduces the ambient noise by eliminating

the noise-dominant eigenvalues. The basic principle of the subspace approach is that the clean signal might be confined to a subspace of the noisy Euclidean space. Therefore, the vector space of the noisy signal is decomposed into "signal" and "noise" subspaces using the singular value decomposition (SVD) or eigenvector–eigenvalue factorization (Loizou, 2013). However, above algorithms do not consider the non-stationary noise and the aerial environmental noise, which have different spectral characteristics, for speech enhancement.

The use of Raspberry Pi for the experimental evaluation of speech signals has gained attention, due to its low cost for edge computing applications. A UAV-based voice recognition system is considered for recognizing the humans buried under the rubble after a massive disaster (Yamazaki et al., 2019). Here, the Raspberry Pi is mounted on UAV to process the voice signals (Yamazaki et al., 2019). A real-time implementation of advanced binaural noise reduction algorithm is demonstrated using Raspberry Pi in Azarpour et al. (2017). In Drakopoulos et al. (2019), a DNN-based noise suppression scheme for audio signals is demonstrated using Raspberry Pi. Motivated by Yamazaki et al. (2019), Azarpour et al. (2017), Drakopoulos et al. (2019), in this work, the experimental evaluations of both algorithms, that is, proposed speech enhancement algorithm and spectral subtraction algorithm are performed using Raspberry Pi 4 Model B board.

3 Basic spectral subtraction algorithm

The spectral subtraction is one of the widely used algorithms for single-channel speech enhancement (Loizou, 2013). It efficiently estimates the clean speech spectrum from the noisy speech spectrum by subtracting the estimate of the noise spectrum. The implementation of spectral subtraction algorithm is performed with the following assumptions. (i) speech signals are assumed to be stationary; (ii) speech and noise are uncorrelated; and (iii) phase of the noisy speech is unchanged. Let $y[n]$ represent the noisy speech which is defined as

$$y[n] = s[n] + d[n], \tag{1}$$

where $s[n]$ denotes the clean speech signal and $d[n]$ is the noise signal. Since our focus is on speech enhancement in frequency domain, the expression of discrete short-time Fourier transform (STFT) of $y[n]$, $s[n]$, and $d[n]$ are $Y[\omega, k]$, $S[\omega, k]$, and $D[\omega, k]$, respectively. Here, $Y[\omega, k]$, $S[\omega, k]$, and $D[\omega, k]$ are noisy, clean, and noise signal in frequency domain, respectively. $Y[\omega, k]$ is obtained as

$$Y[\omega, k] = S[\omega, k] + D[\omega, k], \tag{2}$$

where $k \in \{1, 2, ..., N\}$ denotes the frame index with N being the number of frames and ω denotes the discrete angular frequency of a frame. With $P_{yy}[\omega, k]$ and $\hat{P}_{dd}[\omega, k]$ being the noisy speech power spectrum and estimated noise power spectrum, respectively, the estimate of the clean speech power spectrum, $\hat{P}_{ss}[\omega, k]$, is obtained as Jaiswal and Romero (2021)

$$\hat{P}_{ss}[\omega, k] = P_{yy}[\omega, k] - \hat{P}_{dd}[\omega, k] \tag{3}$$

Finally, the enhanced speech signal is obtained by computing the inverse short-time Fourier transform (Inverse STFT) of the square root of $\hat{P}_{ss}[\omega, k]$, using the phase of the noisy speech signal. A detailed derivation of the spectral subtraction algorithm is presented in Jaiswal and Romero (2021).

Spectral subtraction algorithm has a trade off between speech information and interference. To avoid speech distortion, it must be done carefully. There is a possibility that we may loose some clean speech information while subtracting the noise from the noisy signal. If we subtract too much, some speech information may be lost; if too little is subtracted, much of the interfering noise (musical noise) may be present. Musical noise is the noise with increasing variance that remains present in the estimated speech signal and may cause listening fatigue (Vaseghi, 2008).

4 Implicit Wiener filter in frequency domain

The estimate of the clean speech obtained from spectral subtraction algorithm is influenced by the musical noise due to the negative subtraction value. Thus, Wiener filter, which is a linear estimator, minimizes the mean square error between the original (clean) and the estimated speech signal and takes care of the direct subtraction (Haykin, 1996; Loizou, 2013).

Given the clean speech power spectrum, $P_{ss}[\omega]$, and noise power spectrum, $P_{dd}[\omega]$, the condition for optimal transfer function of the Wiener filter $H_{WF}[\omega]$[1] for speech enhancement in frequency domain is given as Loizou (2013), Jaiswal and Romero (2021)

$$\left[H_{WF}[\omega] = \frac{P_{ss}[\omega]}{P_{ss}[\omega] + P_{dd}[\omega]}\right] \tag{4}$$

For estimating the clean speech, $P_{ss}[\omega]$ and $P_{dd}[\omega]$ must be estimated accurately. Therefore, an additional flexibility is provided to these estimates by introducing two adjustable parameters β and γ. These parameters are achieved by so-called modified or parametric Wiener filter, resulting in the following implicit estimator (Lim & Oppenheim, 1979):

[1] The detailed derivation of the transfer function of the Wiener filter $H_{WF}[\omega]$ is presented in our work (Jaiswal & Romero, 2021).

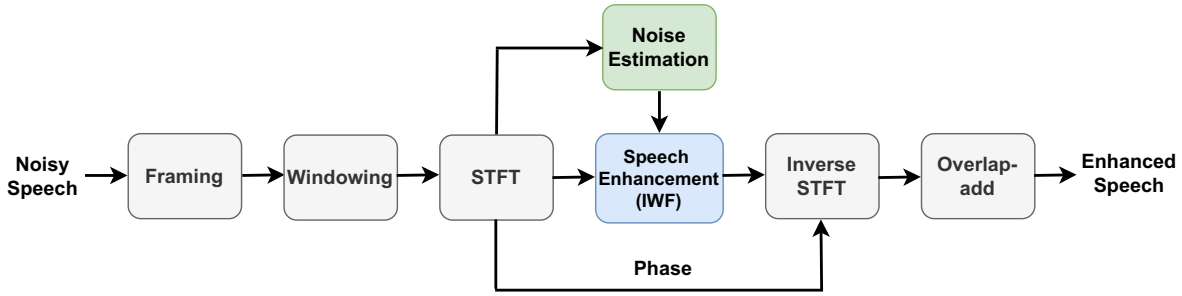

Fig. 2 Block diagram of the proposed Implicit Wiener filer-based speech enhancement algorithm

$$H[\omega] = \left[\frac{P_{ss}[\omega]}{P_{ss}[\omega] + \gamma P_{dd}[\omega]}\right]^{\beta} \tag{5}$$

Here, β represents the noise suppression factor and γ is the noise adjustable parameter that controls the amount of perceived noise. The value of γ is calculated using segmental SNR (SNRseg) (see Eq. 6) of each frame of the noisy speech signal as, $\gamma = 4 - 0.15\,\text{SNRseg}$. For the frame having high segmental SNR, γ should be small which is the case when speech is present. However, for the frame having low segmental SNR, γ should be large which is the case when low-energy frames or pauses are present. The segmental SNR is defined as Loizou (2013)

$$\text{SNRseg} = \frac{10}{M}\sum_{m=0}^{M-1}\log\left(1 + \frac{\sum_{n=Nm}^{Nm+N-1} s^2[n]}{\sum_{n=Nm}^{Nm+N-1}(s[n] - \hat{s}[n])^2}\right), \tag{6}$$

where $s[n]$ and $\hat{s}[n]$ are the clean and the enhanced speech signal, respectively. N is the frame length and M is the number of frames in the speech signal.

Furthermore, to accommodate the non-stationarity of the speech signal, it is convenient to introduce the following approximation (Lim & Oppenheim, 1979)

$$P_{ss}[\omega] \approx |\hat{S}[\omega]|^2. \tag{7}$$

Here, we approximate the true power spectral density of $s[n]$ by its spectral energy. After substituting (7) in (5) with $\beta = 1/2$, $\hat{S}[\omega]$[2] is obtained as Jaiswal and Romero (2021)

$$|\hat{S}[\omega]| = \left[|Y[\omega]|^2 - \gamma P_{dd}[\omega]\right]^{1/2}. \tag{8}$$

Finally, the enhanced speech signal is estimated by taking the inverse short-time Fourier transform of $|\hat{S}[\omega]|$ on a frame-by-frame basis. We use overlap-add method (Daher et al., 2010) to recombine spectra of individual frame with the phase of noisy speech signal. The block diagram of proposed Implicit Wiener filer-based speech enhancement algorithm is shown in Fig. 2.

5 Noise estimation

Different types of noises are encountered in daily life. Each type of noise has a distinct behavior and spectral characteristics. For example, AWGN noise is stationary whereas, exhibition, station, drone, helicopter, and airplane noise are non-stationary due to the constantly changing spectral characteristics. Accurate noise estimation from the noisy speech results in an improved speech quality. To estimate the noise spectrum, the speech enhancement algorithm employs a first order recursive equation (Loizou, 2013), which averages the previous noise estimates and the current noisy speech spectrum as

$$\hat{P}_{dd}[\omega, k] = \alpha\hat{P}_{dd}[\omega, k-1] + (1-\alpha)P_{yy}[\omega, k] \tag{9}$$

where α $(0 \le \alpha \le 1)$ denotes the smoothing parameter. $P_{yy}[\omega, k]$, $\hat{P}_{dd}[\omega, k]$, and $\hat{P}_{dd}[\omega, k-1]$ denote the the short-time power spectrum of the noisy speech, estimate of the noise power spectrum in ω^{th} frequency bin of current frame, and estimate of the past noise power spectrum, respectively. Here, the value of α is selected analytically. For different values of α, segmental SNRs (Loizou, 2013) are calculated for each frame of the noisy speech sample. Then, the transition in segmental SNRs is observed to select the value of α. Usually, segmental SNR decreases as the α increases.

6 Experimental dataset

In the presence of both stationary and non-stationary noises, the performance of the algorithms for speech enhancement is evaluated. We generate AWGN noise at different SNRs of 0 dB, 2.5 dB, and 5 dB for the stationary scenario. We use exhibition, station, drone, helicopter, and airplane noise at SNRs of 0 dB, 2.5 dB, and 5 dB for non-stationary

[2] The detailed derivation of $\hat{S}[\omega]$ is presented in our work (Jaiswal & Romero, 2021).

scenarios. Noisy samples for the exhibition and the station are taken from the noisy speech corpus NOIZEUS (Hu & Loizou, 2006), in which the noise samples are used from the Aurora dataset (Hirsch & Pearce, 2000). NOIZEUS corpus was created when the phonetically balanced IEEE English sentences were uttered by three male and three female speakers, respectively. The drone noise is taken from the drone audio dataset (Al-Emadi et al., 2019). The helicopter and airplane noise are taken from the environmental sound classification (ESC) dataset (Piczak, 2015).

We also consider two clean speech utterances, one is pronounced by a male speaker and another by a female speaker. The male utterance is "A good book informs of what we ought to know" and the female utterance is "Let us all join as we sing the last chorus". For the stationary scenario, we generate AWGN noise with SNRs of 0 dB, 2.5 dB, and 5 dB and combine it with each clean speech utterance to obtain noisy speech sample. Similarly, for the non-stationary scenario, exhibition and station noise, obtained from the Aurora dataset, are added with each clean speech utterance at 0 dB, 2.5 dB, and 5 dB SNR in order to obtain noisy speech sample. In addition, drone, helicopter, and airplane noise are also added to each clean speech utterance at 0 dB, 2.5 dB, and 5 dB SNR to obtain the corresponding noisy speech sample. The speech samples are narrow-band, with frequency of 8 kHz and an average duration of 3 seconds. The samples are saved in .WAV (16 bit PCM, mono) format.

7 Evaluation methodology of the speech enhancement algorithm

After generating noisy speech samples as discussed in Sect. 6, each speech sample is divided into multiple frames having a frame duration of 25 ms with 50% overlap. Each frame has a Hamming window with a duration of 25 ms. The windowed speech frames are analyzed using a 256-sampled short-time Fourier transform (STFT). The noise is estimated using Eq. (9). Since stationary and non-stationary noises have different time-frequency distributions and spectral characteristics, they reflect different impacts on the speech signals. Consequently, we calculate the frame-wise segmental SNR (see equation (6)) of each noisy speech sample (uttered by a male and a female speaker) to obtain the best value of smoothing parameter, α, for estimating the noise of each noise type. In the implicit Wiener filtering, we also consider initial 5 frames of the noisy speech sample as noise/silence to estimate the noise power spectral density (PSD) using equation (9). A simple voice activity detector is used to update the noise PSD.

The performance of speech enhancement algorithm is evaluated with four objective speech quality measures such as perceptual evaluation of speech quality (PESQ), log-likelihood ratio (LLR), cepstral distance (CD), and weighted spectral slope distance (WSS). For testing, we also have the original (clean) speech sample. PESQ (Hu & Loizou, 2006) compares the clean speech sample and the enhanced speech sample and generates a quality score that ranges from -0.5 to 4.5. A higher value of PESQ indicates better speech quality.

LLR is a spectral distance measure used to measure the mismatch between formants of the clean and the enhanced speech sample (Loizou, 2013). It usually ranges between 0 and 2, and is defined as Loizou (2013), Jaiswal and Romero (2021)

$$d_{LLR}(a_s, \bar{a}_{\hat{s}}) = \log_{10}\left(\frac{\bar{a}_{\hat{s}}^T R_s \bar{a}_{\hat{s}}}{a_s^T R_s a_s}\right), \tag{10}$$

where a_s^T and $\bar{a}_{\hat{s}}^T$ denote the linear prediction coefficients (LPC) of the clean and the enhanced speech sample, respectively. T denotes transpose. R_s represents the auto-correlation matrix of the clean speech sample.

CD (Loizou, 2013; Hu & Loizou, 2007) provides an estimate of the log spectral distance between two spectra. It has the range of [0, 10] and it is calculated as Loizou (2013), Jaiswal and Romero (2021)

$$d_{CD}(c_s, \bar{c}_{\hat{s}}) = \frac{10}{\log_e 10}\sqrt{2\sum_{k=1}^{p}(c_s[k] - c_{\hat{s}}[k])^2}, \tag{11}$$

where $c_s[k]$ and $\bar{c}_{\hat{s}}[k]$ represent the cepstrum coefficients (obtained from LPC) of the clean and the enhanced speech sample, respectively. p denotes the maximum order of the LPC coefficients.

WSS (Loizou, 2013) is based on the weighted difference between the spectral slopes in each band. It has the range of [0, 150]. With $S_s[k]$, and $\bar{S}_{\hat{s}}[k]$ being the spectral slope of the clean and the enhanced speech sample, respectively, and $W[k]$ being the weight of the band k, WSS is obtained as Loizou (2013), Jaiswal & Romero (2021)

$$d_{WSS}(C_s, \bar{C}_{\hat{s}}) = \sum_{k=1}^{36} W[k](S_s[k] - \bar{S}_{\hat{s}}[k])^2. \tag{12}$$

In addition to the objective performance measures, the informal listening test of the enhanced speech signal is also performed for both speech enhancement algorithms.

8 Implementation using the edge computing system

This section presents the implementation of speech enhancement algorithms on an edge computing system, that is, the Raspberry Pi (Azarpour et al., 2017).

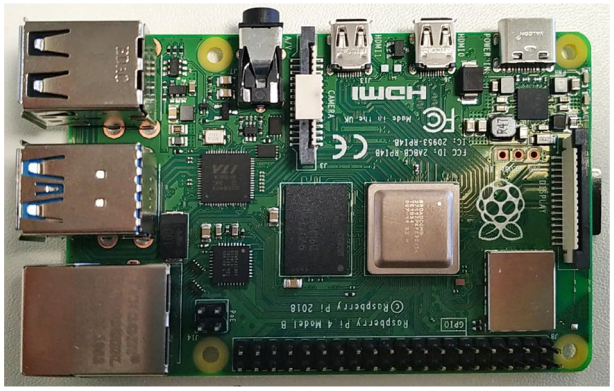

Fig. 3 Experimental system demonstrating Raspberry Pi 4 Model B board

For the experimental evaluation of the considered speech enhancement algorithms, we employ the Raspberry Pi 4 Model B, which acts as a mini computer, as shown in the Fig. 3. It is a faster, highly powerful, re-engineered, and completely upgraded model. It is equipped with ARM7 quad core processor having 1.5 GHz speed and supports upto 8 GB SDRAM. It has two USB 3.0 and two USB 2.0 ports. It supports Ethernet, WiFi, and bluetooth at 2.4 GHz and 5 GHz operating frequencies. It also has a camera, audio, and composite video ports.

Since Raspberry Pi (RPi) works with C/C++, we install the MATLAB support package for RPi hardware. This package is used for the configuration and the deployment of MATLAB code into the RPi board for the experimental evaluation. The MATLAB support package for RPi hardware is used to make a network connection between RPi and computer. Here, we connect the RPi to the computer using Ethernet. Further, this package compiles the MATLAB functions into C functions and deploys the C code into the RPi. The Raspberry Pi model B is compatible with MATLAB release R2020a and higher. Finally, the Raspberry Pi Resource Monitor App is used to track the resource consumption of the speech enhancement algorithms.

For the simulation analysis in MATLAB, we use *audioread* function to extract the sample values and frequency from a given speech sample. However, this function is not supported for deploying into RPi. Thus, for the experimental implementation of both speech enhancement algorithms, we read the sample values and frequency extracted in MATLAB. Table 1 lists the resources consumed in the RPi hardware for the experimental evaluation of both spectral subtraction and proposed speech enhancement algorithms.

Table 1 Resource consumption of Raspberry Pi 4 Model B for the experimental evaluation of both SS and IWF-based speech enhancement algorithms

Name	Availability	Utilization	
		SS	IWF
ARMv7 processor rev 3	4 cores @ 1500 MHz	27%	28%
RAM	4 GB	3%	3%
SD Card	256 GB	1%	1%
Code execution time	–	13 s	13 s

9 Results and discussions

The speech enhancement algorithms are implemented in MATLAB R2021b on Windows 10 laptop having Intel Core i5 8^{th} generation processor, Intel UHD Graphics 620, and 16 GB of memory.

Tables 2, 3, 4, 5, 6, and 7 present the segmental SNR of the spectral subtraction with recursive noise estimation algorithm for AWGN, exhibition, station, drone, helicopter, and airplane noise at SNRs of 0 dB, 2.5 dB, and 5 dB with different values of smoothing parameter "α", for the speech uttered by both a male and a female speaker. It is observed from Tables 2, 3, 4, 5, 6, and 7 that for the increasing value of α, the SNRseg decreases for each input SNR scenario, showing deviation in the SNRseg. However, the SNRseg decreases at the extreme point $\alpha = 1$. Thus, from Tables 2, 3, 4, 5, 6, and 7, we observe that $\alpha = 0.9$ is the most appropriate value.

Tables 8 and 9 present the results of PESQ, LLR, CD, and WSS for the speech uttered by both male and female speakers, respectively. These tables are for both the spectral subtraction and the implicit Wiener filter-based speech enhancement algorithms. Here, the speech signal is degraded by stationary noise (AWGN) and non-stationary noise (exhibition, station, drone, helicopter, and airplane) at SNRs 0 dB, 2.5 dB, and 5 dB, respectively. We also verified these performance metrics using the RPi experimental setup as shown in Fig. 3.

It can be noticed from Table 8 that when a male speaker pronounces speech utterance then the implicit Wiener filter-based speech enhancement algorithm is performing better than the spectral subtraction algorithm. The same is observed for both stationary and non-stationary noise at each input SNR except the station noise at 0 dB and helicopter noise at 2.5 dB, where the spectral subtraction

Table 2 Segmental SNR of spectral subtraction with recursive noise estimation algorithm for AWGN noise at SNRs 0 dB, 2.5 dB, and 5 dB

SNR	α									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Male										
0 dB	6.152	6.024	5.942	5.891	5.864	5.838	5.834	5.855	5.867	5.652
2.5 dB	6.276	6.165	6.071	6.044	6.017	6.008	6.004	6.027	6.074	5.922
5 dB	6.390	6.329	6.243	6.181	6.182	6.181	6.229	6.242	6.255	6.224
Female										
0 dB	6.486	6.341	6.269	6.226	6.181	6.155	6.187	6.151	6.194	5.907
2.5 dB	6.588	6.486	6.427	6.359	6.350	6.322	6.377	6.425	6.442	6.234
5 dB	6.684	6.618	6.569	6.546	6.536	6.565	6.577	6.644	6.690	6.572

Table 3 Segmental SNR of spectral subtraction with recursive noise estimation algorithm for exhibition noise at SNRs 0 dB, 2.5 dB, and 5 dB

SNR	α									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Male										
0 dB	6.174	6.022	5.936	5.882	5.843	5.819	5.801	5.796	5.805	5.624
2.5 dB	6.384	6.273	6.215	6.182	6.168	6.169	6.183	6.211	6.261	6.021
5 dB	6.388	6.264	6.205	6.173	6.159	6.161	6.179	6.218	6.289	6.245
Female										
0 dB	6.512	6.369	6.278	6.217	6.178	6.154	6.150	6.175	6.239	5.856
2.5 dB	6.617	6.536	6.490	6.462	6.449	6.446	6.458	6.491	6.576	6.405
5 dB	6.718	6.625	6.570	6.537	6.525	6.525	6.542	6.590	6.657	6.608

Table 4 Segmental SNR of spectral subtraction with recursive noise estimation algorithm for station noise at SNRs 0 dB, 2.5 dB, and 5 dB

SNR	α									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Male										
0 dB	6.237	6.100	6.025	5.977	5.946	5.936	5.935	5.956	6.006	5.640
2.5 dB	6.451	6.321	6.251	6.212	6.193	6.189	6.193	6.206	6.247	6.092
5 dB	6.440	6.341	6.296	6.275	6.278	6.295	6.333	6.402	6.506	6.271
Female										
0 dB	6.474	6.334	6.250	6.192	6.150	6.126	6.117	6.136	6.185	5.854
2.5 dB	6.612	6.457	6.362	6.293	6.250	6.221	6.207	6.208	6.219	6.122
5 dB	6.698	6.623	6.584	6.562	6.556	6.566	6.590	6.648	6.773	6.555

Table 5 Segmental SNR of spectral subtraction with recursive noise estimation algorithm for drone noise at SNRs 0 dB, 2.5 dB, and 5 dB

SNR	α									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Male										
0 dB	6.450	6.292	6.195	6.128	6.083	6.048	6.019	6.005	5.996	5.797
2.5 dB	6.564	6.431	6.353	6.298	6.265	6.240	6.222	6.220	6.223	6.131
5 dB	6.685	6.581	6.521	6.480	6.458	6.443	6.436	6.446	6.463	6.481
Female										
0 dB	6.626	6.511	6.445	6.401	6.375	6.360	6.360	6.377	6.378	6.136
2.5 dB	6.725	6.635	6.582	6.549	6.532	6.526	6.536	6.563	6.577	6.463
5 dB	6.825	6.764	6.725	6.705	6.696	6.695	6.713	6.744	6.765	6.805

Table 6 Segmental SNR of spectral subtraction with recursive noise estimation algorithm for helicopter noise at SNRs 0 dB, 2.5 dB, and 5 dB

SNR	α									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Male										
0 dB	6.286	6.144	6.067	6.024	5.997	5.992	5.989	5.996	6.013	5.769
2.5 dB	6.402	6.280	6.215	6.183	6.166	6.168	6.171	8.189	6.209	6.080
5 dB	6.528	6.427	6.375	6.353	6.345	6.349	6.361	6.385	6.404	6.403
Female										
0 dB	6.597	6.499	6.436	6.395	6.362	6.341	6.332	6.339	6.353	6.181
2.5 dB	6.683	6.611	6.565	6.539	6.516	6.503	6.503	6.522	6.543	6.519
5 dB	6.768	6.723	6.697	6.683	6.672	6.669	6.679	6.705	6.731	6.872

Table 7 Segmental SNR of spectral subtraction with recursive noise estimation algorithm for airplane noise at SNRs 0 dB, 2.5 dB, and 5 dB

SNR	α									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Male										
0 dB	6.338	6.208	6.144	6.108	6.090	6.079	6.081	6.093	6.095	5.665
2.5 dB	6.645	6.339	6.283	6.254	6.245	6.245	6.256	6.270	6.281	5.984
5 dB	6.583	6.484	6.435	6.413	6.409	6.413	6.430	6.454	6.464	6.315
Female										
0 dB	6.593	6.490	6.429	6.391	6.367	6.350	6.344	6.355	6.365	5.950
2.5 dB	6.685	6.607	6.564	6.537	6.524	6.515	6.522	6.548	6.571	6.273
5 dB	6.778	6.724	6.698	6.687	6.682	6.684	6.701	6.737	6.768	6.618

Table 8 PESQ, LLR, CD and WSS of enhanced speech using Spectral subtraction (SS) and Implicit Wiener filter (IWF)-based speech enhancement algorithms for the speech pronounced by a Male speaker

Type of noise	Input SNR	PESQ		LLR		CD		WSS	
	(dB)	SS	IWF	SS	IWF	SS	IWF	SS	IWF
AWGN	0	1.563	1.657	1.824	1.657	9.579	9.110	106.541	93.099
	2.5	1.908	1.932	1.740	1.566	9.254	8.711	61.879	86.722
	5	1.656	2.008	1.736	1.485	9.394	8.520	99.245	82.164
Exhibition	0	1.496	1.473	1.566	1.483	7.856	7.555	95.302	89.479
	2.5	1.906	1.658	1.430	1.387	7.538	7.579	54.626	93.347
	5	1.803	1.980	1.323	1.143	7.055	6.303	99.414	79.732
Station	0	1.683	1.496	1.070	1.167	5.834	6.045	88.213	73.604
	2.5	2.047	1.944	1.388	1.287	7.646	7.391	49.926	84.377
	5	1.910	2.275	0.883	0.830	5.101	5.030	82.288	67.854
Drone	0	1.663	1.942	1.753	1.492	9.361	8.460	119.340	116.619
	2.5	2.088	2.238	1.619	1.404	8.808	8.158	85.830	111.763
	5	1.754	2.419	1.629	1.296	8.980	7.745	106.827	102.352
Helicopter	0	1.825	1.807	1.201	1.049	7.002	6.118	102.432	108.025
	2.5	2.312	2.102	0.821	0.917	5.355	5.602	68.859	97.942
	5	2.065	2.338	1.106	0.810	6.670	5.184	93.885	88.767
Airplane	0	2.027	2.366	1.074	0.848	6.653	5.596	92.426	61.481
	2.5	2.427	2.591	0.711	0.775	5.141	5.284	46.443	56.642
	5	2.222	2.687	1.006	0.744	6.430	5.149	85.152	53.985

algorithm has better PESQ, LLR and CD. It is also observed from Table 8 that the WSS of the spectral subtraction algorithm for each noise type at 2.5 dB is better. However, at other SNRs, WSS of the implicit Wiener filter algorithm is better. This indicates a poor noise reduction by the spectral subtraction algorithm, resulting in severe perceptual dissimilarity. Implicit Wiener filter algorithm performs better for airplane noise at 5 dB, giving highest

Table 9 PESQ, LLR, CD and WSS of enhanced speech using Spectral subtraction (SS) and Implicit Wiener filter (IWF)-based speech enhancement algorithms for the speech pronounced by a Female speaker

Type of noise	Input SNR	PESQ		LLR		CD		WSS	
	(dB)	SS	IWF	SS	IWF	SS	IWF	SS	IWF
AWGN	0	1.594	1.839	1.530	1.397	9.068	8.378	119.713	103.075
	2.5	1.815	1.899	1.298	1.323	8.508	8.195	63.850	93.645
	5	1.803	2.227	1.504	1.228	8.984	7.868	109.863	81.030
Exhibition	0	1.563	1.617	1.015	0.999	5.881	5.901	106.839	92.325
	2.5	1.849	1.696	0.953	1.086	6.195	6.713	59.202	91.358
	5	1.816	2.019	0.925	0.886	5.587	5.397	105.653	77.759
Station	0	1.610	1.778	0.816	0.907	5.217	5.568	114.247	96.60
	2.5	1.905	1.801	0.948	1.124	6.254	6.649	53.765	82.504
	5	1.860	2.267	0.800	0.718	5.022	4.833	95.878	69.258
Drone	0	1.698	1.981	1.563	1.258	8.912	7.950	131.652	116.563
	2.5	2.181	2.379	1.190	1.076	8.035	7.342	84.035	98.933
	5	1.873	2.545	1.456	0.999	8.469	6.997	117.652	90.250
Helicopter	0	1.778	2.054	1.171	0.897	7.014	6.005	109.900	99.249
	2.5	2.364	2.333	0.571	0.784	4.980	5.380	64.786	86.267
	5	1.955	2.621	1.161	0.677	6.775	4.769	101.228	73.173
Airplane	0	1.760	2.178	1.209	0.880	7.209	5.877	102.774	69.538
	2.5	2.221	2.264	0.663	0.838	5.495	5.644	55.942	66.255
	5	1.898	2.424	1.124	0.765	6.743	5.281	93.567	59.873

PESQ, lowest LLR and WSS among other noise types. It also performs better for station noise at 5 dB, giving lowest CD among other noise types. This reflects that the speech degraded by the airplane noise at 5 dB is highly improved by the implicit Wiener filter algorithm. Finally, it is noticed that the Implicit Wiener filter algorithm exhibits better performance in reducing the non-stationary noise than the stationary noise.

It can be noticed from Table 9 that when a female speaker pronounces the speech utterance then the implicit Wiener filter-based speech enhancement algorithm outperforms the spectral subtraction algorithm. This can be observed for both the stationary and the non-stationary noise at each input SNR, expect for the case of exhibition and station noise at 2.5 dB, where the spectral subtraction algorithm has better PESQ. The LLR, CD, and WSS of the spectral subtraction algorithm for the AWGN, exhibition, station, helicopter and airplane noise at 2.5 dB are better. However, at other SNRs, LLR, CD, and WSS of the implicit Wiener filter algorithm are better. This indicates a poor noise reduction by the spectral subtraction algorithm, resulting in severe perceptual dissimilarity. Implicit Wiener filter algorithm performs outstanding for helicopter noise at 5 dB, giving highest PESQ and lowest LLR and CD among other noise types. It also performs outstanding for airplane noise at 5 dB, giving lowest WSS among other noise types. This reflects that the speech degraded by the helicopter noise at 5 dB is improved more accurately by the implicit Wiener filter algorithm. Finally, it is noticed that the implicit Wiener filter algorithm exhibits better performance in reducing the non-stationary noise than the stationary noise.

Figure 4 shows the comparison of enhanced speech using the implicit Wiener filter-based speech enhancement algorithm. This comparison is for the speech uttered by both male and female speakers, and degraded by each type of noise at 5 dB. It is remarkable in Fig. 4a that the speech quality (PESQ) and LLR of male uttered speech degraded with airplane noise at 5 dB are better than the other noise types. However, the CD of male uttered speech degraded with station noise at 5 dB is better. This indicates that the speech degraded due to the airplane and the station noise at 5 dB, are enhanced more accurately with different measures employed. Similarly, it is also remarkable in Fig. 4b that the speech quality (PESQ), LLR, and CD of female uttered speech degraded with helicopter noise at 5 dB are better than the other noise types. This indicates that the speech degraded due to the helicopter noise at 5 dB is enhanced more accurately. In addition, the speech quality (PESQ) and WSS of the male uttered speech degraded with airplane noise at 5 dB are better than the female uttered speech using implicit Wiener filter-based speech enhancement algorithm. However, the LLR and CD of the female uttered speech degraded with the helicopter noise at 5 dB are better than the male uttered speech using implicit Wiener filter-based speech enhancement algorithm. This indicates that the male and the female uttered speeches are enhanced better with different quality measures in different noisy environments. It leads to design a noise-specific or context-specific speech

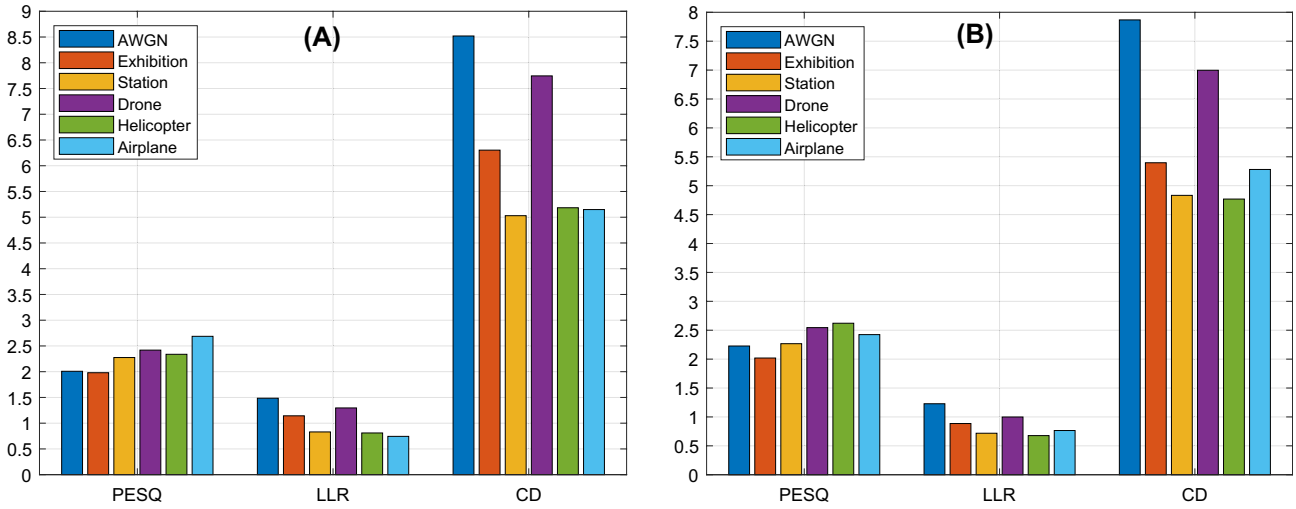

Fig. 4 PESQ, LLR and CD of the enhanced speech using implict Wiener filter-based speech enhancement algorithm, for **a** the speech uttered by a male speaker and degraded by each type of noise at 5 dB, and **b** the speech uttered by a female speaker and degraded by each type of noise at 5 dB

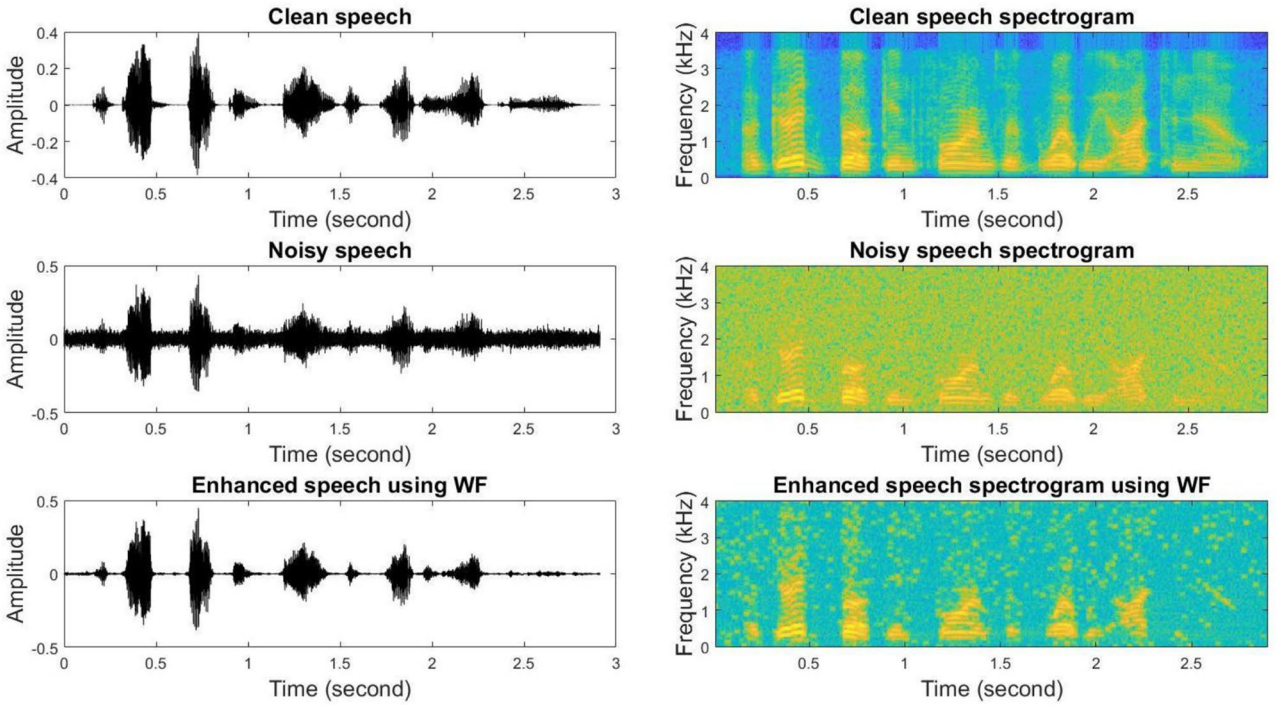

Fig. 5 Time domain and spectrogram representations of the clean, noisy, and enhanced speech using the implicit Wiener filter-based speech enhancement algorithm, for the speech pronounced by a male speaker and degraded by AWGN noise at 5 dB

enhancement algorithm. Overall, it again directs to deploy the implicit Wiener filter-based speech enhancement algorithm to reduce the non-stationary noise and perform speech enhancement tasks.

Figures 5 and 6 show the time domain and spectrum representation of the clean, noisy, and enhanced speech signals using the implicit Wiener filter-based speech enhancement algorithm for the speech uttered by a male

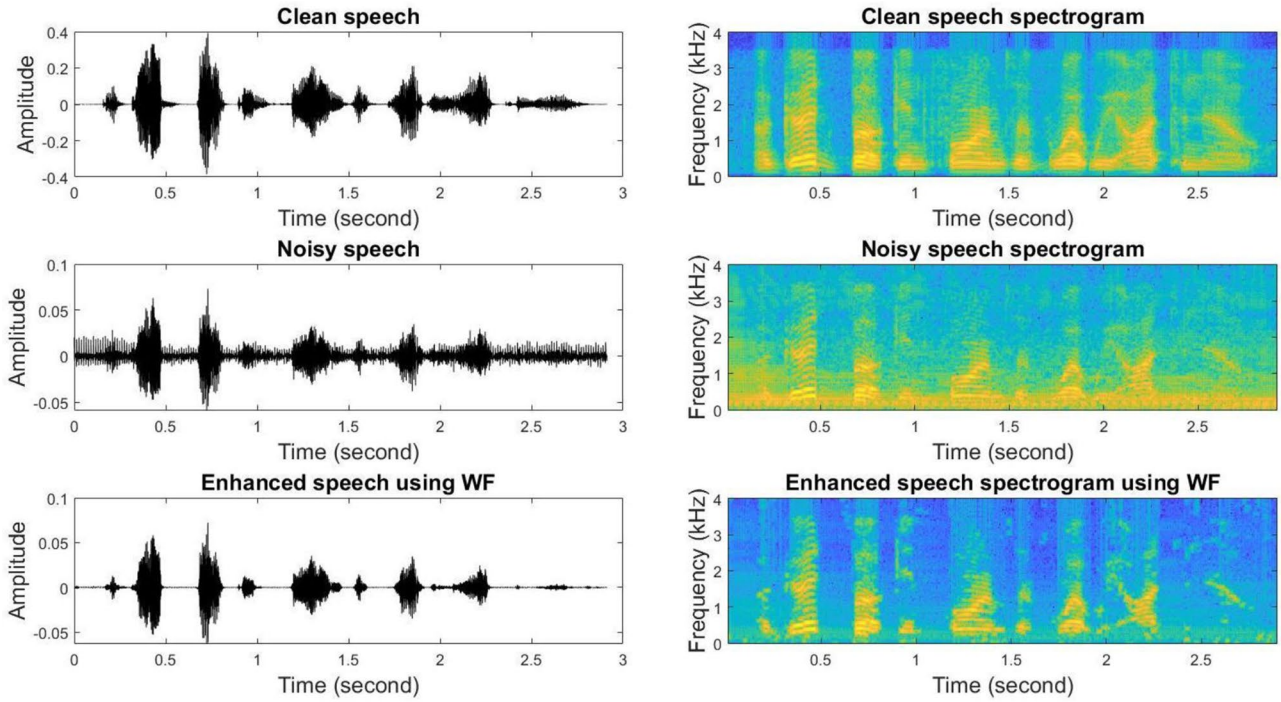

Fig. 6 Time domain and spectrogram representations of the clean, noisy, and enhanced speech using the implicit Wiener filter-based speech enhancement algorithm, for the speech pronounced by a male speaker and degraded by airplane noise at 5 dB

speaker and degraded by the AWGN and the airplane noise at 5 dB, respectively. It can be observed from Figs. 5 and 6 that the speech enhanced by the implicit Wiener filter-based speech enhancement algorithm has improved signal estimation, showing superior performance for the non-stationary noise than the stationary noise. Moreover, it is worth noting that the spectral components that are being filtered out using the proposed speech enhancement algorithm constructs an insignificant amount of musical noise in some of the speech samples, which is not very susceptible to the hearing, and is acceptable.

10 Conclusions and future work

In this paper, we have proposed and extended the implicit Wiener filter-based algorithm for speech enhancement, in the presence of the non-stationary noise (exhibition, station, drone, helicopter, and airplane) and the stationary noise (additive white Gaussian noise). The proposed speech enhancement algorithm employs a first order recursive noise estimation equation in order to estimate noise from the degraded speech. The equation employs a smoothing parameter for continuously updating noise in each frame. To obtain the most appropriate value of the smoothing parameter, segmental SNR is calculated in each frame of the noisy speech spectrum. The implementation results show that for various types of noise degradations tested, the proposed speech enhancement algorithm outperforms the spectral subtraction algorithm. Moreover, the envelop of the estimated/enhanced speech signal is close to the envelop of the clean speech spectrum. The enhanced speech signal is shown to have similar perceptual quality as the clean speech signal. The spectrogram of the enhanced speech is also similar to the clean speech spectrogram. In addition, the informal listening test of the enhanced speech signal using the proposed speech enhancement algorithm demonstrates a clear sound as compared to the spectral subtraction algorithm. The construction of the musical noise in the enhanced speech signal is too small which is acceptable. Furthermore, it is also shown that the proposed speech enhancement algorithm supports the low power edge computing device, such as, the Raspberry Pi.

As a result, the proposed speech enhancement algorithm may be a promising future candidate to enhance the speech signal of various speech processing applications, such as, speech quality prediction metric, speech recognition system, speech identification system, speech coding system, etc., where a crystal clear speech is required for the efficient communication. One can easily integrate these speech enhancement algorithms as a pre-processing block in order to improve the system performance, when implementing using the edge computing system. In addition,

we also intend to examine different noise estimation techniques for designing and developing the speech enhancement algorithms.

Acknowledgements This work was supported by the INCAPS project: 287918 of INTPART program from the Research Council of Norway. The authors would like to thank the anonymous reviewers for their feedback, which are incorporated in the manuscript to significantly improve the quality of this work. They also thank the editor for coordinating the review process.

Funding Open access funding provided by University of Agder.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

References

Abd El-Fattah, M. A., Dessouky, M. I., Abbas, A. M., Diab, S. M., El-Rabaie, E. S. M., Al-Nuaimy, W., Alshebeili, S. A., & Abd El-Samie, F. E. (2014). Speech enhancement with an adaptive Wiener filter. *International Journal of Speech Technology, 17*(1), 53–64.

Al-Emadi, S., Al-Ali, A., Mohammad, A., & Al-Ali, A. (2019). Audio based drone detection and identification using deep learning. In *Proceedings of the international wireless communications & mobile computing conference* (pp. 459–464).

Ali, Y. S. E., Parsa, V., Doyle, P., & Berkane, S. (2020). Low-complexity disordered speech quality estimation. *International Journal of Speech Technology, 23*(3), 585-594.

Asano, F., Hayamizu, S., Yamada, T., & Nakamura, S. (2000). Speech enhancement based on the subspace method. *IEEE Transactions on Speech and Audio Processing, 8*(5), 497–507.

Azarpour, M., Siska, J., & Enzner, G. (2017). Real-time binaural speech enhancement demo on raspberry pi. In *Proceedings of the IEEE international conference on acoustics, speech and signal processing* (ICASSP) (pp. 6572–6573).

Bhowmick, A., & Chandra, M. (2017). Speech enhancement using voiced speech probability based wavelet decomposition. *Computers & Electrical Engineering, 62*, 706–718.

Boll, S. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 27*(2), 113–120.

Charoenruengkit, W., & Erdöl, N. (2010). The effect of spectral estimation on speech enhancement performance. *IEEE Transactions on Audio, Speech, and Language Processing, 19*(5), 1170–1179.

Chiea, R. A., Costa, M. H., & Barrault, G. (2019). New insights on the optimality of parameterized wiener filters for speech enhancement applications. *Speech Communication, 109*, 46–54.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE Signal Processing Magazine, 35*(1), 53–65.

Daher, A., Baghious, E. H., Burel, G., & Radoi, E. (2010). Overlap-save and overlap-add filters: Optimal design and comparison. *IEEE Transactions on Signal Processing, 58*(6), 3066–3075.

Das, N., Chakraborty, S., Chaki, J., Padhy, N., & Dey, N. (2020). Fundamentals, present and future perspectives of speech enhancement. *International Journal of Speech Technology, 24*(4), 883–901.

Deleforge, A., Di Carlo, D., Strauss, M., Serizel, R., & Marcenaro, L. (2019). Audio-based search and rescue with a drone: highlights From the IEEE Signal Processing Cup 2019 Student Competition [SP Competitions]. *IEEE Signal Processing Magazine, 36*(5), 138–144. https://doi.org/10.1109/MSP.2019.2924687.

Drakopoulos, F., Baby, D., & Verhulst, S. (2019). Real-time audio processing on a Raspberry Pi using deep neural networks. In *Proceedings of the international congress on acoustics*.

Haykin, S. (1996). *Adaptive filter theory* (5th ed.). Prentice-Hall.

Hirsch, H. G., & Pearce, D. (2000). The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions. In *Proceedings of the automatic speech recognition: Challenges for the new millenium, ISCA Tutorial and Research Workshop* (*ITRW*).

Hu, Y., & Loizou, P. C. (2004). Speech enhancement based on wavelet thresholding the multitaper spectrum. *IEEE Transactions on Speech and Audio Processing, 12*(1), 59–67.

Hu, Y., & Loizou, P. C. (2006). Subjective comparison of speech enhancement algorithms. In *Proceedings of the EEE international conference on acoustics speech and signal processing proceedings* (Vol. 1, pp. 153–156).

Hu, Y., & Loizou, P. C. (2007). Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on Audio, Speech, and Language Processing, 16*(1), 229–238.

Islam, M. T., Shahnaz, C., Zhu, W. P., Ahmad, M. O., et al. (2018). Speech enhancement in adverse environments based on non-stationary noise-driven spectral subtraction and snr-dependent phase compensation. arXiv preprint arXiv:1803.00396.

Jaiswal, R., & Romero, D. (2021). Implicit Wiener filtering for speech enhancement in non-stationary noise. In *11th international conference on information science and technology* (*ICIST*), IEEE (pp. 39–47).

Kamath, S., Loizou, P., (2002). A multi-band spectral subtraction method for enhancing speech corrupted by colored noise. In *ICASSP*. IEEE.

Kanehara, S., Saruwatari, H., Miyazaki, R., Shikano, K., & Kondo, K. (2012). Comparative study on various noise reduction methods with decision-directed a priori snr estimator via higher-order statistics. In *Proceedings of The Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, IEEE (pp. 1–6).

Kleijn, W. B., Lim, F. S., Luebs, A., Skoglund, J., Stimberg, F., Wang, Q., & Walters, T. C. (2018). Wavenet based low rate speech coding. In *Proceedings of the IEEE international conference on acoustics, speech and signal processing* (*ICASSP*) (pp. 676–680).

Lim, J. S., & Oppenheim, A. V. (1979). Enhancement and bandwidth compression of noisy speech. *Proceedings of the IEEE, 67*(12), 1586–1604.

Loizou, P. C. (2013). *Speech enhancement: Theory and practice* (2nd ed.). CRC Press.

Moore, A. H., Parada, P. P., & Naylor, P. A. (2017). Speech enhancement for robust automatic speech recognition: Evaluation using a

baseline system and instrumental measures. *Computer Speech & Language, 46*, 574–584.

Ogunfunmi, T., Togneri, R., & Narasimha, M. (2015). *Speech and audio processing for coding, enhancement and recognition*. Springer.

Pascual, S., Serrà, J., & Bonafonte, A. (2019). Time-domain speech enhancement using generative adversarial networks. *Speech Communication, 114*, 10–21. https://doi.org/10.1016/j.specom.2019.09.001

Piczak, K. J. (2015). ESC: Dataset for environmental sound classification. In *Proceedings of the ACM international conference on multimedia* (pp. 1015–1018).

Saldanha, J. C., & Shruthi, O. R. (2016). Reduction of noise for speech signal enhancement using spectral subtraction method. In *Proceedings of the IEEE international conference on information science* (*ICIS*) (pp. 44–47).

Schultz, B. G., Tarigoppula, V. S. A., Noffs, G., Rojas, S., van der Walt, A., Grayden, D. B., & Vogel, A. P. (2021). Automatic speech recognition in neurodegenerative disease. *International Journal of Speech Technology 24*(3) , 771–779.

Sheft, S., Ardoint, M., & Lorenzi, C. (2008). Speech identification based on temporal fine structure cues. *The Journal of the Acoustical Society of America, 124*(1), 562–575.

Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. *IEEE Access, 7*, 53040–53065.

Srinivasarao, V., & Ghanekar, U. (2020). Speech intelligibility enhancement: A hybrid Wiener approach. *International Journal of Speech Technology, 23*(3), 517–525.

Vaseghi, S. V. (2008). *Advanced digital signal processing and noise reduction* (4th ed.). Wiley.

Yamazaki, Y., Tamaki, M., Premachandra, C., Perera, C. J., Sumathipala, S., & Sudantha, B. H. (2019). Victim detection using UAV with on-board voice recognition system. In *Proceedings of the IEEE international conference on robotic computing* (*IRC*) (pp. 555–559). https://doi.org/10.1109/IRC.2019.00114

Yan, X., Yang, Z., Wang, T., & Guo, H. (2020). An iterative graph spectral subtraction method for speech enhancement. *Speech Communication, 123*, 35–42. https://doi.org/10.1016/j.specom.2020.06.005

You, C. H., & Ma, B. (2017). Spectral-domain speech enhancement for speech recognition. *Speech Communication, 94*, 30–41. https://doi.org/10.1016/j.specom.2017.08.007

Yu, H., Zhu, W. P., & Champagne, B. (2020). Speech enhancement using a DNN-augmented colored-noise Kalman filter. *Speech Communication, 125,* 142–151. https://doi.org/10.1016/j.specom.2020.10.007.

Yuan, W. (2020). A time-frequency smoothing neural network for speech enhancement. *Speech Communication, 124*, 75–84. https://doi.org/10.1016/j.specom.2020.09.002