SN Computer Science (2023) 4:71
https://doi.org/10.1007/s42979-022-01484-4

ORIGINAL RESEARCH

Single Channel Speech Enhancement Using Masking Based on Sinusoidal Modeling

Ashishkumar Gudmalwar[1] · **Ch V. Rama Rao**[2]

Received: 24 April 2022 / Accepted: 31 October 2022 / Published online: 29 November 2022

Abstract

This paper concentrated on enhancement of noisy speech signal which is observed through single microphone under back ground noise environment. The presence of background noise degrades the quality and intelligibility observed speech signal. The conventional noise reduction methods such as power spectral subtraction, Wiener filtering, and masking etc., have been developed based on spectral magnitudes. These methods are developed by representing the noisy speech signal in time–frequency domain. In this work, sinusoidal modelling approach is used to analyse the observed speech signal and spectral weighting filter gain is estimated for masking of sinusoidal components which are not related to the speech signal. Here, filter gain is calculated in each frame using the sinusoidal components. The performance of the proposed system is analyzed in terms of Perceptual Evaluation of Speech Quality (PESQ) and Segmental Signal to Noise Ratio (SegSNR) values. It is evident from experiments that the mask which is estimated based on Wiener filter including cross correlation terms providing higher values to PESQ and segmental SNR for considered noise environment compared to existing approaches.

Keywords Speech enhancement · Sinusoidal modeling and masking

Introduction

Speech Enhancement (SE) is vitally important in speech communication while speech is observed through a single microphone. SE is a challenging work in many applications such as recognition of speech utterances [1], hearing aids and hands-free mobile communications. The objective of SE is to improve the intelligibility and quality of noise corrupted signal. Since the presence of background noise will degrade the performance of speech communication systems. In most of the applications, speech signals are observed through a single microphone. Due to this reason, single channel SE has attracted a lot of research attention from researchers. The existing statistical methods for single channel SE such as power spectral subtraction [2], minimum mean squared error (MMSE) estimation [3, 4] and optimally modified log-spectral amplitude (OM-LSA) speech estimator [5, 6] and Wiener filtering [7] are using in the past several decades. The time-frequency representation of speech signal and estimation of the mask for removing the background noise is given in [8–10]. The mask or spectral weighting filter gain is estimated based on the magnitude of time frequency coefficients. In this process, the estimation of noise signal power and the apriori signal to noise ratio (SNR) [3, 11–14] is required. Sometimes the value of a posteriori SNR [15] is also required. As per the literature it is a regular practice to obtain the enhanced speech signal magnitudes using the enhancement method. The enhanced speech signal is reconstructed by including the noisy signal phase information along with the obtained enhanced amplitudes.

Recently, deep learning-based algorithms have been developed to use in speech processing applications. In [16], ideal ratio mask is estimated using a Deep Neural Network (DNN) for obtaining the enhanced speech signal magnitudes. The enhanced speech signal is reconstructed

Ch V. Rama Rao contributed equally to this work.

This article is part of the topical collection "Advances in Applied Image Processing and Pattern Recognition" guest edited by K C Santosh.

✉ Ashishkumar Gudmalwar
g.ashishkumar@nitm.ac.in

Ch V. Rama Rao
chvramarao@nitw.ac.in

[1] National Institute of Technology, Meghalaya 793003, India

[2] National Institute of Technology, Warangal 506004, India

by combining the phase of the noisy speech signal with the enhanced amplitudes. The complex ideal ratio mask [17] is estimated using DNN for obtaining the real and imaginary components of the enhanced speech signal. In [18], the Ideal Binary Mask (IBM) mask is developed for mapping the noisy features from separating the target speech from the observations. DNN based systems shown good performance in enhancing the noisy speech. However, DNN based systems require high processing time. Another factor which limits the use of deep learning-based systems for SE in real-world applications is the computational complexity.

This work concentrated on development of single channel SE system based on analysis of noisy speech signal using sinusoidal modelling. In [19] discussed on analysis and synthesis of speech signal using sinusoidal representation. In this paper, investigations are carried out for enhancing the noisy speech signal by estimating the time-frequency mask using sinusoidal components in each frame. The significant sinusoidal components are retained in the process of analysis of noisy speech signal.

Signal Model and Notation

In real time observation of speech signal through a single microphone, observed signal is a noisy signal. Since the environmental noise signals which are appearing in the background are added with the clean speech signal. The observed noisy speech signal $y(n)$ the time domain is given by

$$y(n) = s(n) + d(n) \tag{1}$$

here $s(n)$ and $d(n)$ represents the clean speech and noise signals respectively. In addition, the observed signal contains the slow and fast varying components. The slowly varying components are related to noise and other part is associated with speech signal. The speech signal contains the time dependent spectral components. In general SE methods are developed by assuming the speech signal is stationary. In order to achieve stationarity, the noisy speech signal is processed through framing and analysis. In this process, the noisy speech signal is segmented using the windowing. Fourier transform is applied for each segment of the noisy speech signal which is known as Short-Time Fourier Transform (STFT) to get behaviour of the signal in the spectral domain. The STFT of the observed speech signal at lth frame and kth frequency bin is given in the Eq. (2)

$$Y(l,k) = S(l,k) + D(l,k) \tag{2}$$

here $Y(l, k)$, $S(l, k)$ and $D(l, k)$ are spectral coefficients of observed, clean speech and noise signals respectively. In order to enhance the noisy speech signal, researchers assume that clean speech and noise signals are uncorrelated. As per the spectral weighting concept, the enhanced spectral component of the speech signal is estimated by multiplying the observed signal with filter gain which is shown in (3)

$$\tilde{S}(l,k) = G(l,k).Y(l,k) \tag{3}$$

where $G(l, k)$ and $\tilde{S}(l, k)$ denotes weighting filter gain function and enhanced speech signal respectively in the spectral domain. The filter gain is used as a mask which is having smaller values near to zero for noise spectral components in the noisy speech spectrum. This gain is obtained by applying the concept of Minimum-Mean Square Error (MMSE) estimation. As per the MMSE criteria, the cost function is defined as shown in Eq. (4) by considering the clean speech and estimated speech signal components.

$$\epsilon = E\{(\tilde{S}(l,k) - S(l,k))^2\} \tag{4}$$

By substituting the Eq. (3) in Eq. (4), the cost will be

$$= E\{(G(l,k)(S(l,k) + D(l,k)) - S(l,k))^2\} \tag{5}$$

The Eq. (5) is reduced to Eq. (6) by applying the assumption that the noise and speech signal are uncorrelated

$$\epsilon = (1 - G(l,k))^2 E\{S^2(l,k)\} + G^2(l,k)E\{D^2(l,k)\} \tag{6}$$

The most favourable values for the filter gain $G(l, k)$ is calculated by differentiating the Eq. (6) with respect to $G(l, k)$ and setting the result to zero.

$$G_0(l,k) = \frac{\epsilon(l,k)}{\epsilon(l,k) + 1} \tag{7}$$

where $\epsilon(l, k)$ denotes the apriori signal to noise ratio which is obtained by estimating the $\frac{E\{S^2(l,k)\}}{E\{D^2(l,k)\}}$ [4]. In this work, the filter gain is estimated using the amplitudes of sinusoidal components in each frame.

Sinusoidal Modeling

The observed noisy speech in each frame is represented by $\{s(n)\}_{n=0}^{N-1}$. Here n and N denotes the sample index and number of samples in each frame respectively. The noisy speech signal in each frame in terms of sinusoidal components can be represented as

$$s(n) = \sum_{p=1}^{L} A_p \cos(,\omega_p + \psi_p) + d(n), \qquad 0 \le n \le (N-1) \tag{8}$$

where p is the index of the sinusoidal component and L denotes the number of sinusoidal components. Each sinusoidal component is characterized by the magnitude A_p, frequency ω_p and phase ω_p respectively. We represent these parameters in a vector form for all sinusoidal components as

$[\boldsymbol{\alpha}, \boldsymbol{\omega}, \boldsymbol{\psi}]$ of dimension with $L \times 3 \boldsymbol{\alpha} = \left\{A_p\right\}_{p=1}^{L}, \boldsymbol{\omega} = \left\{\omega_p\right\}_{p=1}^{L}$ and $\boldsymbol{\psi} = \left\{\psi_p\right\}_{p=1}^{L}$. The noisy speech signal spectrum is obtained by applying STFT. The spectral components are transformed into the Mel-scale for resembling the human auditory system. The spectral peaks with the highest magnitude are selected [20]. By doing like this, most perceptually relevant sinusoidal components per band has been selected. The N-point discrete Fourier transform (DFT) vector for each frame is given by

$$\boldsymbol{v}_p = \left[1, e^{j\omega_1}, e^{j\omega_2(2)}, \ldots . e^{j\omega_p(N-1)}\right] \tag{9}$$

We define

$$\boldsymbol{V} = \left[\boldsymbol{v}_1 \;\; \boldsymbol{v}_1^* \;\; \boldsymbol{v}_2 \;\; \boldsymbol{v}_2^* \ldots . \boldsymbol{v}_L^*\right]^t \tag{10}$$

here $(.)^*$ represents the complex conjugate operation and $\boldsymbol{V}$ is a $2L \times N$ Vandermonde matrix whose rows are filled with $\boldsymbol{V}_i$ given by Eq. (5). The main goal of the sinusoidal modelling is to retain the components which are required for reconstructing the signal. The observed noisy speech signal is represented in terms of sinusoids using Eq. (7)

$$S = \boldsymbol{V}^t \boldsymbol{\rho} \tag{11}$$

where $\boldsymbol{\rho} = [A_1 e^{\psi_1} A_1 e^{-\psi_1} A_2 e^{\psi_2} A_2 e^{-\psi_2} \ldots . A_L e^{-\psi_L}]$ in each frame and L denotes the number sinusoidal components retained. The retained sinusoidal components in each frame represent the peaks in the noisy speech spectrum. The peaks are obtained based on the following constraint.

$$\omega_p = \underset{\omega \in \Omega_p}{\operatorname{argmax}} \; |\, S(l,k) \,| \quad \text{and} \quad A_p e^{j\psi_p} = S(l,k) \tag{12}$$

where ω_p is a set of all continuous frequencies within the pth band.

Estimation of Mask

In conventional approaches, the mask is estimated based on time–frequency representation of speech signals for reducing the background noise. In the process of estimation of spectral components of the clean speech signal, the spectrum of noisy speech signal is multiplied with the estimated filter gain function. For estimating the filter gain function all spectral components of the noisy signal is used. In general, the observed signal at destination is noise corrupted signal due to the background noise and channel effects. To recover the original signal at destination, existing approaches using the complete set of spectral components of the noisy signal. This will lead to requirement of more storage. This problem can be overcome if we reconstruct the clean signal with the reduced set of spectral components of the observed noisy

Table 1 Comparison of obtained PESQ values with NOIZEUS database

Noise type and SNR (dB)		IBM	CIRM	WM	CCWM
Babble	0	1.43	1.53	1.31	1.5
	5	1.81	1.58	1.69	1.79
	10	2.04	1.76	2.19	2.24
	15	2.12	1.85	2.11	2.20
Car	0	1.26	1.04	1.3	1.43
	5	1.64	1.4	1.78	1.78
	10	1.89	1.67	1.85	1.64
	15	2.13	1.82	2.11	2.31
Train	0	1.35	1.37	1.22	1.33
	5	1.6	1.44	1.86	1.85
	10	1.8	1.6	1.94	2.10
	15	2.03	1.74	2.10	2.23
Restaurant	0	1.42	1.24	1.29	1.36
	5	1.83	1.6	1.89	1.89
	10	2.02	1.8	2.09	2.11
	15	2.16	1.88	2.23	2.33
Airport	0	1.51	1.48	1.45	1.49
	5	1.8	1.56	1.79	1.83
	10	1.97	1.75	1.98	1.99
	15	2.07	1.83	2.01	2.17

Table 2 Comparison of obtained SegSNR values with NOIZEUS database

Noise type and SNR(dB)		IBM	CIRM	WM	CCWM
Babble	0	−4.31	−6.24	−1.5	−1.4
	5	−2.9	−5.16	−0.8	−1.25
	10	−1.13	−4.1	1.06	0.55
	15	0.39	−3.08	1.39	1.13
Car	0	−4.58	−6.23	−1.59	−0.02
	5	−2.98	−5.36	0.06	0.03
	10	−1.61	−4.10	0.91	0.36
	15	−0.24	−3.12	1.47	1.5
Train	0	−4.41	−6.12	−0.09	−0.06
	5	−2.61	−5.11	0.03	0.09
	10	−2.11	−4.32	0.05	0.69
	15	−0.22	−3.03	0.04	1.57
Restaurant	0	−4.9	−5.95	−0.09	−0.09
	5	−3.16	−5.27	0.03	0.05
	10	−1.41	−4.05	0.04	0.09
	15	−0.22	−3.28	0.06	1.28
Airport	0	−4.6	−0.84	−0.5	−0.49
	5	−3.08	−0.56	−0.1	−0.09
	10	−1.28	−0.22	−0.04	0.03
	15	−0.5	−0.06	0.06	1.3

signal. In this work, observed noisy signal is analysed using sinusoidal modelling. In this analysis, some set of sinusoidal components are retained in each frame. Each sinusoidal component is represented with its amplitude, frequency and phase respectively.

Here, the mask is estimated using the sinusoidal components, which are obtained by analyzing the speech signals in sinusoidal modeling. The sinusoidal based mask is estimated in the following ways:

Ideal Binary Mask (IBM): which is estimated using the Eq. (9) [18]

$$G_{IBM} = \begin{cases} 1 \text{ if SNR(l, k)} > 5\,\text{dB} \\ 0 \text{ otherwise} \end{cases} \tag{13}$$

Here, in each frame, we are defining the signal to noise ratio (SNR) by considering the selected sinusoidal related to speech signal and noise components.

Complex Ideal Ratio Mask (CIRM) [17, 21]: this mask is estimated using the Eq. (10).

$$G_{CIR}(l,k) = \frac{|A_d|}{|A_y|} e^{j(\psi_d - \psi_y)} \tag{14}$$

Here, A_d and A_y represents the magnitudes of sinusoidal components of noise and noisy speech signal. Similarly, ψ_d and ψ_y denotes phase of the noise and noisy speech signals respectively. Wiener Mask (WM): which is estimated using the Eq. (11)

$$G_{WM}(l,k) = \frac{\xi(l,k)}{\xi(l,k)+1} \tag{15}$$

where $\xi(l,k)$ is known as apriori SNR, which estimated depending using the sinusoidal components of speech and noise signals. The Cross-correlation Compensated Wiener Mask (CCWM): this mask is estimated using Eq. (12) as per the discussion in [22]

$$G_{CWM}(l,k) = \frac{\xi(l,k) + \delta \frac{E[A_y A_d]}{E[A_d^2]} - 1}{\gamma(l,k)} \tag{16}$$

Here, $\gamma(l,k)$ denotes the posterior SNR and which is obtained using the sinusoidal component magnitudes of noisy speech and noise signals.

Results and Discussion

This section deals with the performance analysis of the developed masking filters based on sinusoidal components. The following databases are used in the performance analysis.

Table 3 Comparison of obtained PESQ values with Librispeech database

Noise type and SNR (dB)		IBM	CIRM	WM	CCWM
Babble	0	0.78	1.36	1.42	1.49
	5	0.99	1.47	1.52	1.54
	10	1.20	1.54	1.50	1.6
	15	1.31	1.60	1.56	1.84
Machine gun	0	0.78	1.35	1.39	1.47
	5	1.01	1.4	1.45	1.49
	10	1.06	1.37	1.45	1.52
	15	1.24	1.39	1.49	1.54
Factory 1	0	0.79	1.32	1.5	1.52
	5	0.92	1.36	1.55	1.56
	10	1.17	1.57	1.56	1.6
	15	1.34	1.46	1.57	1.64
Factory 2	0	1.06	1.4	1.5	1.49
	5	1.09	1.42	1.51	1.55
	10	1.16	1.45	1.53	1.57
	15	1.22	1.5	1.57	1.65
F16	0	0.74	1.29	1.5	1.56
	5	0.95	1.4	1.54	1.57
	10	1.17	1.45	1.57	1.62
	15	1.28	1.47	1.64	1.67

Table 4 Comparison of obtained SegSNR values with Librispeech database

Noise type and SNR (dB)		IBM	CIRM	WM	CCWM
Babble	0	−0.21	−0.91	−0.24	−0.2
	5	−0.26	−0.55	−0.27	−0.19
	10	−0.28	−0.32	−0.28	−0.22
	15	−0.28	−0.22	−0.26	−0.24
Machine gun	0	−0.42	−0.74	−0.28	−0.27
	5	−0.5	−0.69	−0.27	−0.25
	10	−0.55	−0.51	−0.41	−0.26
	15	−0.45	−0.36	−0.29	−0.25
Factory 1	0	−0.33	−0.07	−0.20	−0.21
	5	−0.32	−0.68	−0.28	−0.22
	10	−0.31	−0.40	−0.27	−0.25
	15	−0.28	−0.27	−0.26	−0.25
Factory 2	0	−0.45	−0.73	−0.24	−0.16
	5	−0.39	−0.45	−0.20	−0.2
	10	−0.33	−0.29	−0.22	−0.19
	15	−0.31	−0.22	−0.22	−0.19
F16	0	−0.21	−0.91	−0.28	−0.24
	5	−0.26	−0.55	−0.26	−0.22
	10	−0.28	−0.32	−0.24	−0.20
	15	−0.28	−0.27	−0.22	−0.19

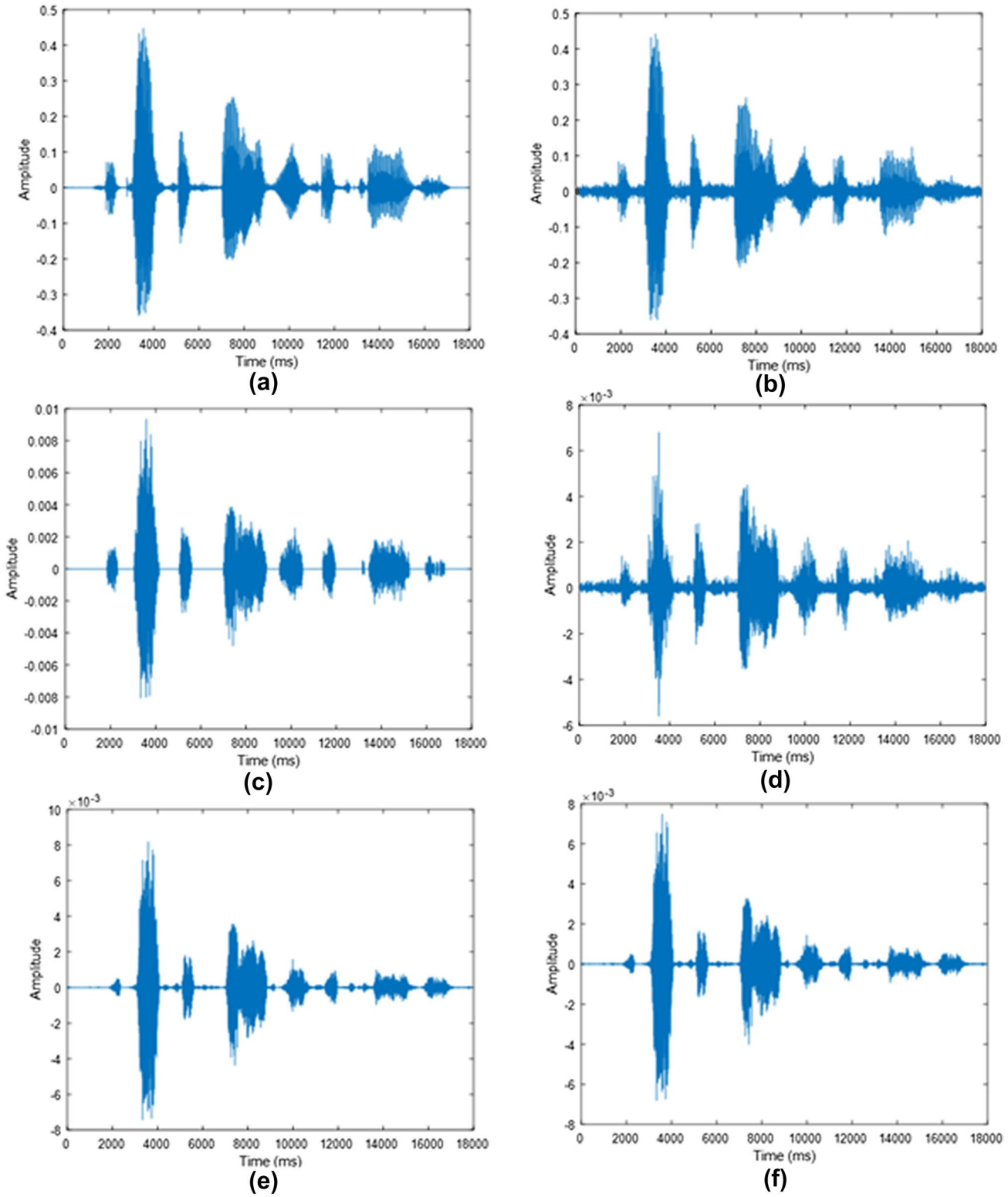

Fig. 1 Time domain waveforms of (**a**) clean signal (**b**) signal with airport noise at 0 dB SNR (**c**) enhanced signal using IBM (**d**) enhanced signal using CIRM (**e**) enhanced signal using WM and (**f**) enhanced signal using CCWM

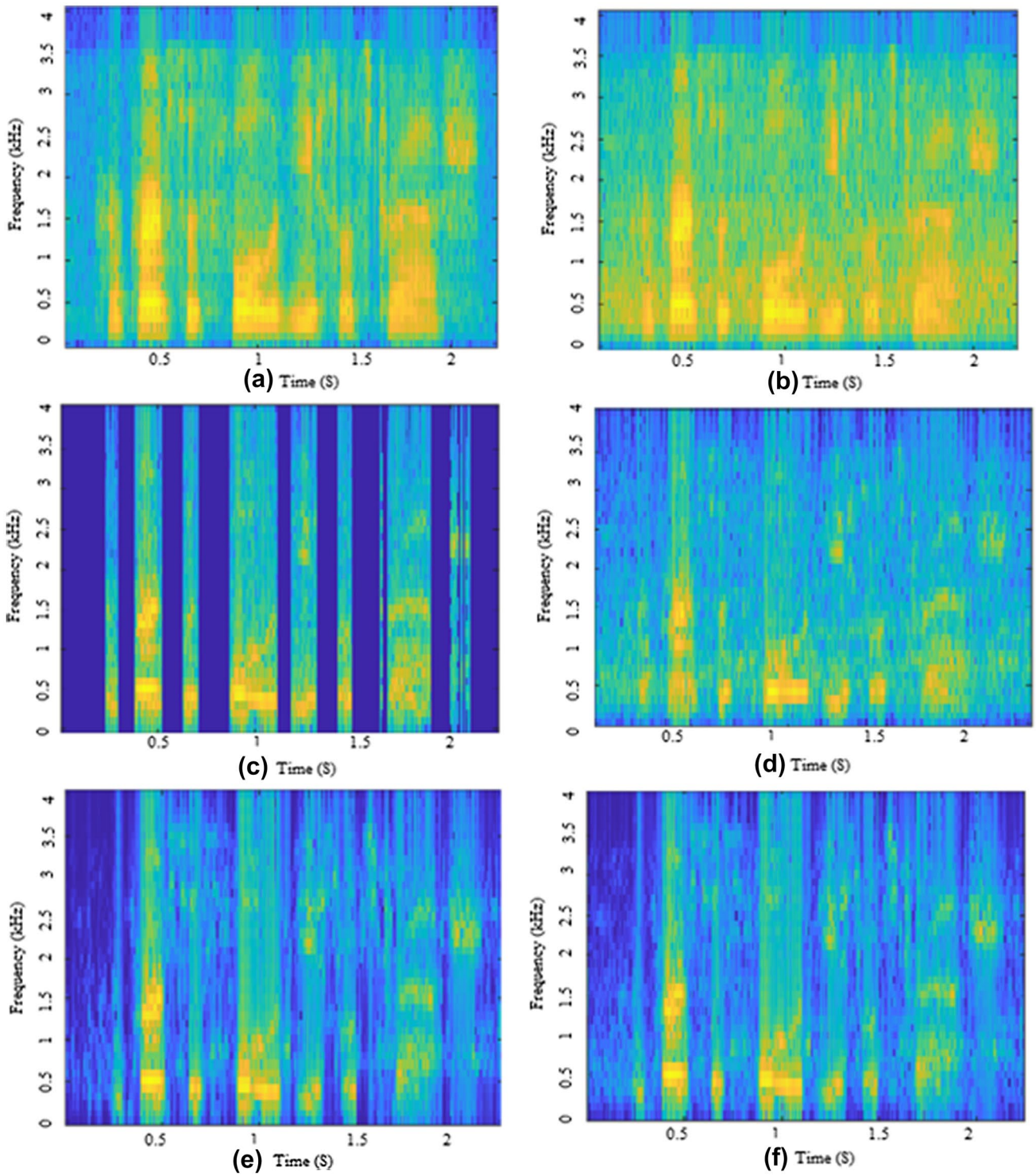

Fig. 2 Spectrograms of (**a**) clean signal (**b**) signal with airport noise at 0 dB SNR (**c**) enhanced signal using IBM (**d**) enhanced signal using CIRM (**e**) enhanced signal using WM and (**f**) enhanced signal using CCWM

- NOIZEUS database [23]: This database contains 30 IEEE sentences corrupted by eight different real environment noises at various SNRs. In this work, considered speech utterance which is corrupted by babble, car, restaurant, train and airport noises at 0, 5, 10 and 15 dB values.
- Librispeech Database [24]: LibriSpeech training dataset consist of about 1000 h of read audio books. The

dev and test sets were split into simple ("clean") and harder ("other") subsets. corpora.

Here the objective quality measures like Perceptual Evaluation of Speech Quality (PESQ) [25] and Segmental Signal to Noise Ratio (SegSNR)) [26] values are considered in the evaluation process. The PESQ measure is one of the objective metrics used in the evaluation of SE algorithms which is related to human-perception. The SegSNR determines the real level noise in the enhanced speech more precisely and it correlates with the perception of the noisy speech by humans.

Table 1 shows the comparison of obtained PESQ scores with speech sample from NOIZEUS database. From table it is observed the cross-correlation compensated Wiener mask is providing the better PESQ values as compared to other approaches. Table 2 shows the obtained SegSNR values with speech samples from NOIZEUS database. Similarly, speech samples are considered from Librispeech database which are corrupted with babble, machine gun, factory 1, factory 2, f16 noises at 0, 5, 10 and 15 dB values respectively. Table 3 shows the obtained PESQ values with speech sample from Librispeech database. Table 4 illustrates the comparison of obtained SegSNR values with speech sample from Librispeech database. The improvement in the segmental SNR values with the cross-correlation compensated Wiener filter-based mask is observed from Table 2 Since the correlation exists between speech and noise signals [22]. The conventional approaches assumed that speech and noise signals are uncorrelated but it is not valid in real noise environments.

Here comparison is always presented by representing the clean, noisy speech and enhanced signal with various approaches in time and frequency domains. Figure 1 illustrates the time domain waveforms of clean, noisy signal which is corrupted with airport noise at 0 dB SNR and enhanced signals using various approaches.

Similarly, spectrograms are presented in Fig. 2 To have better quality in the enhanced speech signal, we need to retain the transitions between the speech sounds. Form the Fig. 2 it is observed that transitions between sounds in the speech utterances are retained while enhancing the noisy signal with WM and CCWM approaches.

Conclusion

This paper proposed sinusoidal modeling of noisy speech signal and development of filter gains for masking of background noise. Various masks like ideal binary mask, Complex Ideal Ratio Mask, Wiener Mask and Cross-correlation compensated Wiener mask are developed using magnitude, frequency and phase of sine waves in each frame. The performance is evaluated in terms of PESQ and SegSNR values. From the experimental results, it is clear that cross-correlation compensated Wiener filter based mask is providing better results compared other approaches.

Funding This work is supported by the Science and Engineering Research Board (SERB), Department of Science and Technology (DST), Government of India and, file no. is EEQ/2018/001338, dated 27th February 2019.

Declarations

Conflict of interest Authors have received the Science and Engineering Research Board (SERB), Department of Science and Technology (DST), Government of India and, file no. is EEQ/2018/001338, dated 27th February 2019.

References

1. Makino S. Speech enhancement. Berlin: Springer; 2005.
2. Boll S. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans Acoust Speech Signal Process. 1979;27(2):113–20.
3. Ephraim Y, Malah D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Trans Acoust Speech Signal Process. 1984;32(6):1109–21.
4. Ephraim Y, Malah D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Trans Acoust Speech Signal Process. 1985;33(2):443–5.
5. Cohen I, Berdugo B. Noise estimation by minima controlled recursive averaging for robust speech enhancement. IEEE Signal Process Lett. 2002;9(1):12–5.
6. Cohen I. Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging. IEEE Trans Speech Audio Process. 2003;11(5):466–75.
7. Lim J, Oppenheim A. All-pole modeling of degraded speech. IEEE Trans Acoust Speech Signal Process. 1978;26(3):197–210.
8. Scalart P. et al. Speech enhancement based on a priori signal to noise estimation. In: 1996 IEEE international conference on acoustics, speech, and signal processing conference proceedings, vol 2. IEEE; 1996. p. 629–32.
9. Lotter T, Vary P. Speech enhancement by map spectral amplitude estimation using a super-gaussian speech model. EURASIP J Adv Signal Process. 2005;2005(7):1–17.
10. Fodor B, Fingscheidt T. Speech enhancement using a joint map estimator with gaussian mixture model for (non-) stationary noise. In: IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE. 2011. p. 4768–71.
11. Cohen I. Speech enhancement using super-gaussian speech models and noncausal a priori snr estimation. Speech Commun. 2005;47(3):336–50.
12. Suhadi S, Last C, Fingscheidt T. A data-driven approach to a priori snr estimation. IEEE Trans Audio Speech Lang Process. 2010;19(1):186–95.
13. Elshamy S, Madhu N, Tirry W, Fingscheidt T. An iterative speech model-based a priori snr estimator. In: Sixteenth annual conference of the international speech communication association; 2015.

14. Elshamy S, Madhu N, Tirry W, Fingscheidt T. Instantaneous a priori snr estimation by cepstral excitation manipulation. IEEE/ACM Trans Audio Speech Lang Process. 2017;25(8):1592–605.
15. Gerkmann T, Breithaupt C, Martin R. Improved a posteriori speech presence probability estimation based on a likelihood ratio with fixed priors. IEEE Trans Audio Speech Lang Process. 2008;16(5):910–9.
16. Wang Y, Narayanan A, Wang D. On training targets for supervised speech separation. IEEE/ACM Trans Audio Speech Lang Process. 2014;22(12):1849–58.
17. Williamson DS, Wang Y, Wang D. Complex ratio masking for monaural speech separation. IEEE/ACM Trans Audio Speech Lang Process. 2015;24(3):483–92.
18. Wang D. On ideal binary mask as the computational goal of auditory scene analysis. In: Speech separation by humans and machines. Berlin: Springer; 2005. p. 181–97.
19. McAulay R, Quatieri T. Speech analysis/synthesis based on a sinusoidal representation. IEEE Trans Acoust Speech Signal Process. 1986;34(4):744–54.
20. Mowlaee P, Sayadiuan A, Sheikhzadeh H. Fdmsm robust signal representation for speech mixtures and noise corrupted audio signals. IEICE Electron Expr. 2009;6(15):1077–83.
21. Tan K, Wang D. Learning complex spectral mapping with gated convolutional recurrent networks for monaural speech enhancement. IEEE/ACM Trans Audio Speech Lang Process. 2019;28:380–90.
22. Rao CVR, Murthy MR, Rao KS. Speech enhancement using sub-band cross-correlation compensated wiener filter combined with harmonic regeneration. AEU Int J Electron Commun. 2012;66(6):459–64.
23. https://ecs.utdallas.edu/loizou/speech/noizeus/
24. Panayotov V, Chen G, Povey D, Khudanpur S. Librispeech: an ASR corpus based on public domain audio books. In: 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE; 2015. p. 5206–5210
25. Rix AW, Beerends JG, Hollier MP, Hekstra AP. Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs. In: 2001 IEEE international conference on acoustics, speech, and signal processing. Proceedings (Cat. No. 01CH37221), vol 2. IEEE; 2001. p. 749–752.
26. Hu Y, Loizou PC. Evaluation of objective quality measures for speech enhancement. IEEE Trans Audio Speech Lang Process. 2007;16(1):229–38.