Applied Acoustics 109 (2016) 1–4

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Technical note

Speech enhancement in dual-microphone mobile phones using Kalman filter

Wahbi Nabi [a,*], Noureddine Aloui [b], Adnane Cherif [a]

[a] *Innov'Com Laboratory, Sciences Faculty of Tunis, University of Tunis El-Manar, Tunis, Tunisia*
[b] *Centre for Research on Microelectronics & Nanotechnology, Sousse Technology Park, Tunisia*

ARTICLE INFO

Article history:
Received 17 August 2015
Received in revised form 25 February 2016
Accepted 29 February 2016
Available online 5 March 2016

Keywords:
Speech enhancement
Dual-microphone
Coherence function
Kalman filter

ABSTRACT

In this paper, a dual-microphone speech enhancement algorithm for the mobile phones is proposed. The adopted method exploits the coherence function algorithm and the Kalman filter. This algorithm has a simple implementation that does not need a prediction of interfering signals statistics. In addition, this algorithm can be used in small devices with so closely distance between the two microphones. Moreover, the use of such algorithm allows reducing multiple noise sources at many azimuths positions. Finally, the new algorithm proves its performances referring to the perceptual evaluation of speech quality and the time domain waveforms.

1. Introduction

Billions of people around the world using mobile phones have problems in telephone conversation when the speech is degraded by the noise. Hence, different speech enhancement techniques were involved to reduce the interfering signals in mobile phones. Particularly, single microphone noise reduction methods prove its effectiveness during fifty years in many fields as they are easy to apply [1,2]. However, the performance of single channel speech enhancement algorithms is limited on the non stationary noise case due to the problems related to the estimation of this noise using such algorithms [2]. Thus, speech quality can be improved using multi-channel speech enhancement algorithms [3]. For that, more the number of microphones is increased more the quality of speech will be improved. In contrast, a great number of microphones is difficult to implement in cellular phones and requires more computational complexity. The use of dual-microphone speech enhancement algorithms can solve these two problems cited previously. Moreover, these algorithms are specified by its good performances in term of speech quality and intelligibility.

The literature is enriched by many works which treat several dual-channel speech enhancement methods. Among them, we can cite the work given by Youssefian et al. [4] which depicted a dual-channel speech enhancement algorithm using power level difference for near field. This algorithm exploits the difference of the power signals in the two microphones as a criterion for noise reduction. In [5], Youssefian and Loizou presented a dual-microphone speech enhancement algorithm based on the coherence function. The proposed strategy treats the coherence between the target and noise signals as a criterion for noise reduction and can be used for narrowly spaced microphones. Koldovsky [6] proposed a noise reduction dual-microphone in mobile phones using a bank of pre-measured target-cancelation filters. This method is based on a target cancelation filters exploited to estimate the noise, which is then subtracted from the noisy speech using Wiener filter or power level difference algorithm. A dual-microphone speech enhancement method in mobile phones is proposed in [7]. This method is based on the inter-microphone Posteriori SNR Difference (PSNRD) for Speech Presence Probability (SPP) estimation and a MVDR Filter for noise reduction. Finally, the given work by Prajna et al. [8] presented a new algorithm based on gravitational search algorithm (GSA). This approach used heuristic algorithm for noise reduction. Compared to this works, we present in this paper a dual-channel speech enhancement algorithm dedicated to mobile phone applications using the coherence function and the Kalman filter. This method can be applied with closely spaced microphones in the mobile phone (see Fig. 1).

The rest of the paper is organized as follows. Section 2 depicts the coherence function. The proposed dual-microphone speech enhancement algorithm using Kalman filter is presented in Section 3. In Section 4, we present the result of the proposed algorithm in comparison with other methods using an objective criterion (PESQ) and the time domain waveforms. Section 5 concludes this work.

* Corresponding author.
E-mail address: wahbi.nabi@gmail.com (W. Nabi).

http://dx.doi.org/10.1016/j.apacoust.2016.02.009
0003-682X/

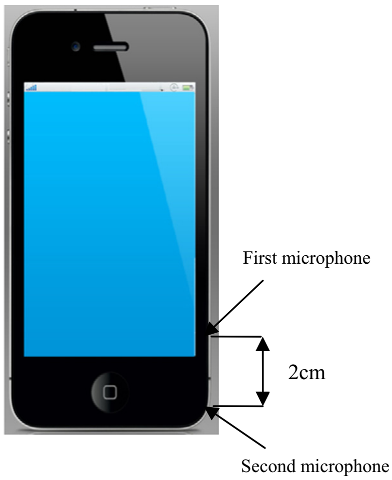

Fig. 1. Illustration of mobile phone with closely microphones.

2. Coherence function

In this section, we are interested in presenting theoretically the coherence function of the noisy signals. Then, we are interested in depicting the prediction coherence function used in our work.

Let us assume that two microphones are positioned in a noisy environment. The two signals received by the two microphones after delay compensation are described as follows:

$$x_1(m) = s_1(m) + n_1(m) \quad (1)$$

$$x_2(m) = s_2(m) + n_2(m) \quad (2)$$

where $x_1(m)$ and $x_2(m)$ are respectively the noisy signal of the first microphone and the second microphone, m is the simple index, $s_1(m)$ and $s_2(m)$ are clean signals obtained at each microphone and finally $n_1(m)$ and $n_2(m)$ are respectively the noise signal of the first microphone and the second microphone.

The Fourier transform of the two noisy signals can be described as follows:

$$X_1(n,k) = S_1(n,k) + N_1(n,k) \quad (3)$$

$$X_2(n,k) = S_2(n,k) + N_2(n,k) \quad (4)$$

where n represents the frame index and k is the frequency bin.

The coherence function algorithm is a basic dual microphone speech enhancement method proposed in [9,10]. This technique is based on a correlation between speech signals in the two microphones instead of the noise signals are uncorrelated. The coherence function between the signals x_1 and x_2 received by the two microphones is presented as follows:

$$F_{coh}(n,k) = \frac{P_{x_1x_2}(n,k)}{\sqrt{P_{x_1}(n,k)P_{x_2}(n,k)}} \quad (5)$$

where $P_{x_1x_2}(n,k)$ presents the cross power spectral density (CPSD) of the two noisy signals x_1 and x_2, $P_{x_1}(n,k)$ and $P_{x_2}(n,k)$ present respectively the power spectral density (PSD) of x_1 and x_2.

The objective of the coherence technique is to determine if the speech signal at a specific frequency bin is present or absent whereas it is proportional to the magnitude of the coherence function. The speech signal is ruling when the magnitude is close to one and the noise signal is ruling when the magnitude is close to zero. The last hypothesis is true when the noise signals at the two microphones are not too much correlated.

Usually, the correlation of the noise signals at the two microphones increases when the distance between them decreases [14]. Thus, the two microphones are too much coherent in our work because they are so closely spaced. The design of the two microphones with target sound sources and the approximate coherence function adopted by this work are detailed in [5]. The two microphones are placed on a dummy head with a distance of 2-cm between them. The clean speech source is placed at 0° azimuth and the noise source is at $\beta°$ azimuth. The distance between the two sources and the microphones are 1.2 m. Based on the last configuration, the approximate coherence function is computed as:

$$\begin{aligned}\widetilde{F}_{x_1x_2} \approx {} & [\cos(\omega\tau) + j\sin(\omega\tau)]\frac{\widetilde{SNR}}{1+\widetilde{SNR}} \\ & + [\cos(\omega\tau\cos\beta) + j\sin(\omega\tau\cos\beta)]\frac{1}{1+\widetilde{SNR}}\end{aligned} \quad (6)$$

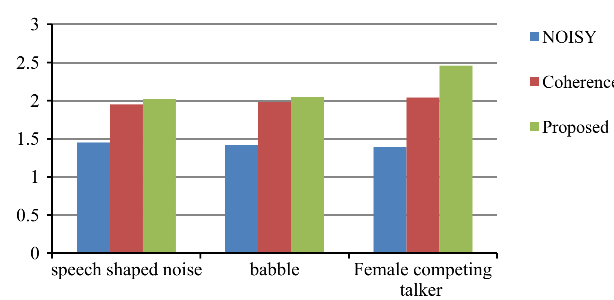

Fig. 3. PESQ scores given when signal is degraded by one noise source positioned at 135° (SNR = −3dB).

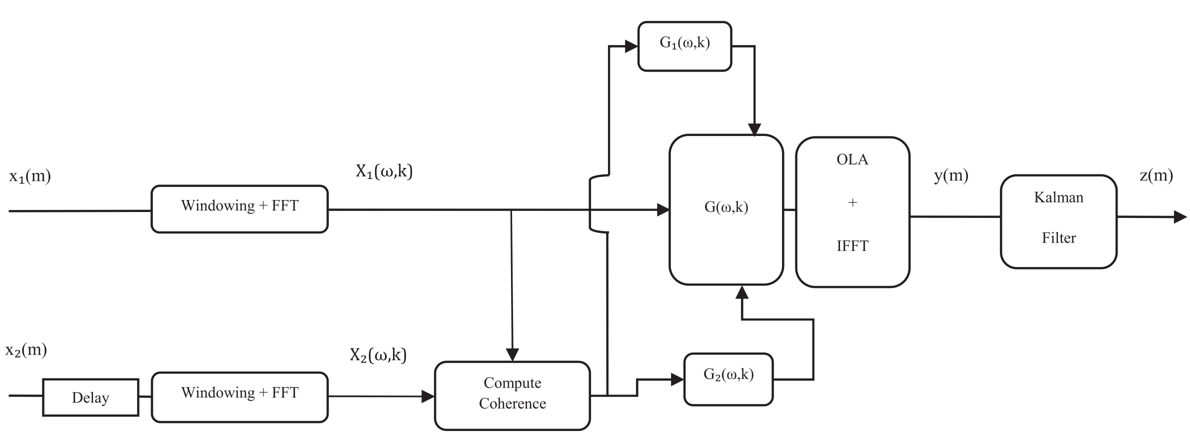

Fig. 2. Block diagram of the proposed algorithm.

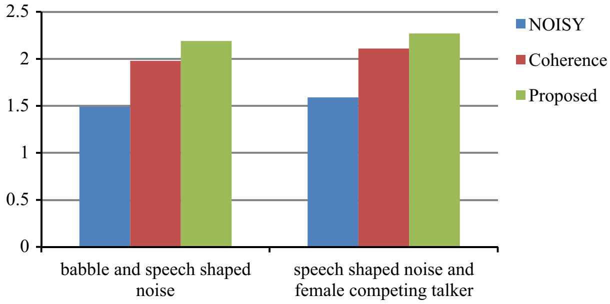

Fig. 4. PESQ scores given when signal is degraded by two noise sources positioned respectively at 90° and 180° (SNR = −3 dB).

β is the angle of incidence and $\omega = 2\pi d/D$ where D is the frame length in samples and ω define the angular frequency

3. The proposed speech enhancement algorithm

The proposed dual-channel speech enhancement algorithm consists of two steps. The first step is the dual-microphone speech enhancement algorithm based on the coherence function proposed in [5]. The second step is the Kalman filter which presents a good solution to the linear MMSE difficulties for the stochastic system by obtaining an optimal prediction of the clean speech [11].

Fig. 2 presents the block diagram of the proposed speech enhancement algorithm. It shows the time delay compensation applied by many dual-microphone speech enhancement methods. Moreover, the figure describes the approximate coherence function defined by (6).

4. Test and results

In this section, we evaluate the performance of the proposed algorithm. For that, we used the Perceptual Evaluation of the Speech Quality (PESQ) [12] and the time domain waveforms as criteria of evaluation. The PESQ is cohesion between two other objective measures: PAMS and PSQM99. PESQ estimation maps mean opinion score (MOS) predicts to a range between −0.5 (bad) and 4.5 (distortion-less).

In order to test the proposed algorithm, we used a MATLAB environment. For the listening and objective test, we extracted the sound sentences (about 7–12 words) from [13]. The proposed

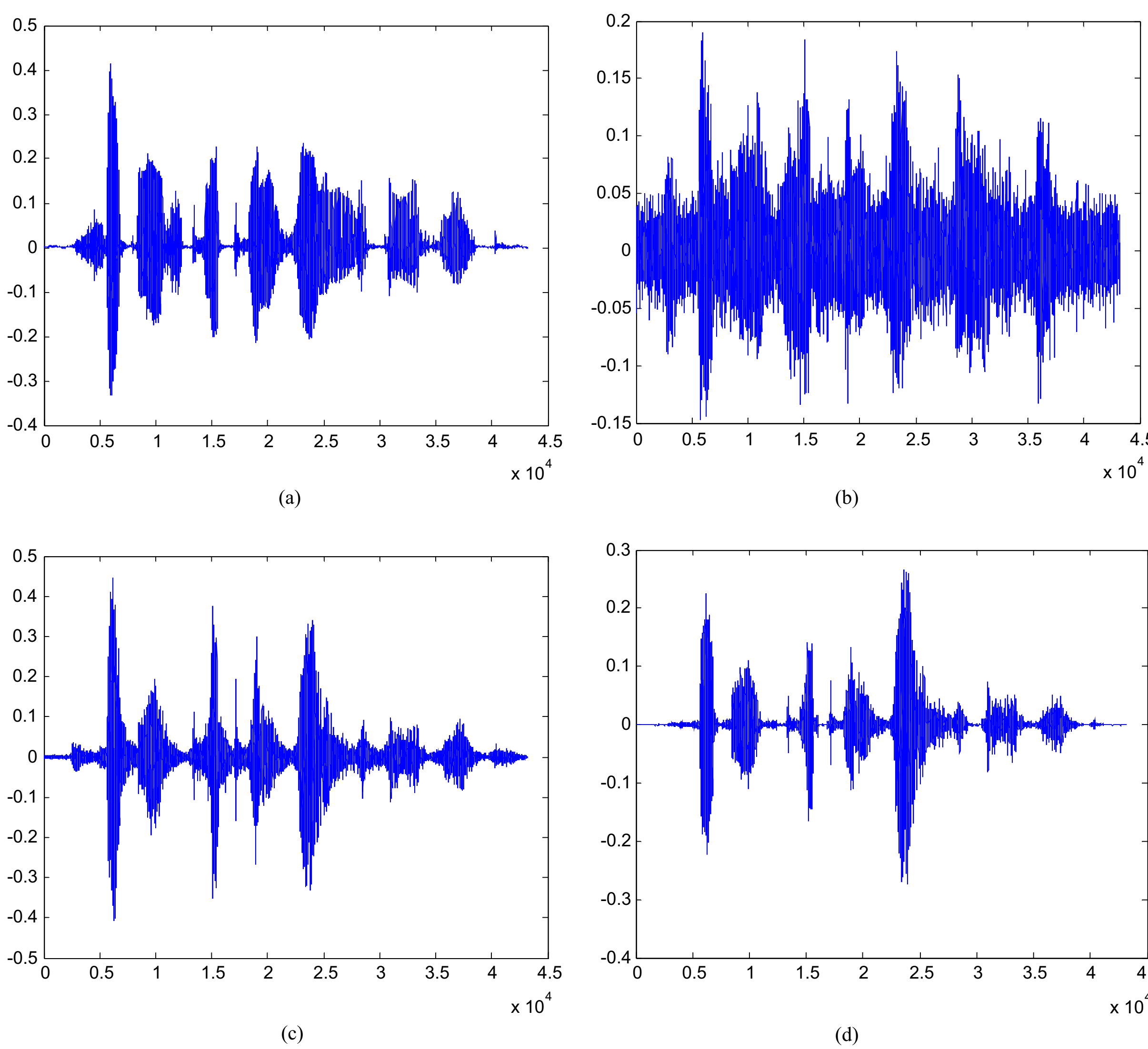

Fig. 5. Enhancement results of speech degraded respectively by speech-shaped noise and female competing talker at 90° and 180° while SNR is −3 dB. (a) Clean signal. (b) Noisy signal while SNR is −3 dB. (c) Enhanced signal by the algorithm based on the coherence function. (d) Enhanced signal by the proposed algorithm.

algorithm was exploited in five scenarios at input-SNR = −3 dB. These five scenarios are presented as follows:

– Three scenarios when speech is degraded by one noise source.
– Two scenarios when speech is degraded by two noise sources.

Finally, we compared the proposed method with the speech enhancement algorithm based on the coherence function described in [5] using the PESQ scores and the time domain waveforms.

Fig. 3 shows the PESQ scores for the noisy signal and the enhanced mentioned signals when the speech is degraded by one noise source located at 135°. Three types of interfering signals were used for the test: speech-shaped noise, babble noise and female competing talker. As clearly shown in the figure below, PESQ scores of the proposed algorithm are better than PESQ scores given by the algorithm based on the coherence function.

Fig. 4 presents the PESQ scores for the noisy signal and the enhanced mentioned signals when the speech is degraded by two noise sources located respectively at 90° and 180°. Two scenarios were used for this test. In the first scenario, babble and speech-shaped noises were placed respectively at 90° and 180°. In the second scenario, a speech-shaped noise and a female competing talker were placed respectively at 90° and 180°. Clearly, the PESQ scores of the proposed algorithm outperform the PESQ scores of the method based on the coherence function.

Fig. 5 depicts respectively the time domain waveforms of the clean signal, the noisy signal and the enhanced mentioned signals when the speech is degraded by a speech-shaped noise and a female competing talker. The two noise sources are placed respectively at 90° and 180°. It's clear from the figure that the proposed algorithm eliminates more interfering signals than the algorithm based on the coherence function.

5. Conclusion

This work presented an improved dual-microphone speech enhancement algorithm for mobile communications. The proposed algorithm can be applied with so closely distance between the two microphones. So, it can be used in small mobile phones. First, we depicted the approximate coherence function used in this algorithm. Then, we explained the determined method exploiting the Kalman filter in details. Next, we proved the effectiveness of the proposed algorithm in term of speech quality using the PESQ scores and the time domain waveforms. Finally, the ability of implementation and intelligibility advantage make this method a good way for future commercial mobile phones.

References

[1] Benesty J, Makino S, Chen J. Speech enhancement. Berlin: Springer-Verlag; 2005.
[2] Loizou PC. Speech enhancement: theory and practice. New York: CRC Press; 2007.
[3] Brandstein M, Ward D. Microphone arrays: signal processing techniques and applications. Berlin: Springer-Verlag; 2001.
[4] Youssefian N, Akbari A, Rahmani M. Using power level difference for near field dual-microphone speech enhancement. Appl Acoust 2009;70(11):1412–21.
[5] Youssefian N, Loizou PC. A dual-microphone speech enhancement algorithm based on the coherence function. IEEE Trans Audio Speech Lang Process 2012;20(2):599.
[6] Koldovsky Z, Tichavsky P, Botka D. Noise reduction in dual-microphone mobile phones using a bank of pre-measured target-cancelation filters. In: Proceedings of the IEEE international conference on acoustics, speech and signal processing, ICASSP-2013, Canada; May 2013. p. 679–83.
[7] Fu ZH, Fan F, Huang JD. Dual-microphone noise reduction for mobile phones. In: Proceedings of the IEEE international conference on acoustics, speech and signal processing, ICASSP-2013, Canada; May 2013. p. 7239–43.
[8] Prajna K, Rao GSB, Reddy KVVS, Maheswari RU. A new approach to dual channel speech enhancement based on gravitational search algorithm (GSA). Int J Speech Technol 2014;17(4):341–51.
[9] Allen JB, Berkley DA, Blauert J. Multi microphone signal processing technique to remove room reverberation from speech signals. J Acoust Soc Am 1977;62(4):912–5.
[10] Le Bouquin R, Faucon G. Using the coherence function for noise reduction. Proc IEE Commun Speech Vision 1992;139(2):276–80.
[11] Goh Z, Tan KC, Tan BTG. Kalman-filtering speech enhancement method based on a voiced-unvoiced speech model. IEEE Trans Speech Audio Process 1999;7(5):510–24.
[12] Hu Y, Loizou PC. Perceptual Evaluation of Speech Quality (PESQ), and objective method for end-to-end of speech quality assessment of narrowband telephone network and speech codecs. ITU-T Rec; 2000. p. 862.
[13] http://ecs.utdallas.edu/loizou/speech/COH_Audio_Demo [Feb 25, 2016].
[14] Davis GM. Noise reduction in speech applications. New York: CRC Press; 2002.