www.ijemr.net ISSN (ONLINE): 2250-0758, ISSN (PRINT): 2394-6962

Vandana Publications
IJEMR

Volume-7, Issue-3, May-June 2017
International Journal of Engineering and Management Research
Page Number: 455-459

Dual-Channel Noise Power Spectral Density Estimation by using Complex Coherence Method

Ch.Anoosha[1], D.Nagamani[2]
[1,2]Department of ECE, Anil Neerukonda Institute of Technology & Sciences, INDIA

ABSTRACT

This paper presents the enhancement of noisy speech signals picked up by a dual microphone mobile phone in hand-free position. In previous approaches, spectral subtraction noise PSD estimation and a dedicated spectral subtraction weighting to reduction of wind noise by means of spectral weighting often leads to severe degradations to the speech signal. Since the novel technique is presented which estimates the coherence of the speech and noise signals, which is usually not known in practice. In this paper proposed coherence method express a better performance compared to related approaches.

***Keywords*—** Complex coherence analysis, noise reduction, dual microphone noise PSD estimation,Spectral subtractuin,speech enhancement.

I. INTRODUCTION

Making a phone call in a noise environment often leads to significant degradations of the speech quality and speech intelligibility. This is even more severe during a phone call in hands-free position where the signal-to-background-noise ratio (SNR) is often much lower, because of the higher levels of background noise sources. This becomes even more severe in the case of wind noise [1]. Wind noise is generated by turbulences in the boundary layer around the used device and thus might be inaudible for the near-end speaker. Hence, it is necessary to reduce the distortions in the transmitted signals by means of noise reduction techniques. One part of noise reduction systems is the estimation of the background noise power spectral density (PSD), in the presence of speech. Given one microphone only, several well-established methods exist. The performance of single microphone algorithms in some way limited especially in the case of fast changing background noise. Hence, the latest generation of mobile phones there are frequently two microphones or in this contribution we propose a dual microphone setup for the detection and the estimation of wind noise in a speech signal [2]. While the first microphone is usually placed at the bottom of the phone, the second microphone is placed at the top or back of the device. Here, the different complex coherence properties of speech and noise are exploited which leads to a sufficient detection rate and good estimation results of the wind noise PSD [3]. More detailed information about complex coherence method l is given in Sec.2. In Sec.3 the proposed noise reduction system is described and Spectrograms are explained in Sec.4.and evaluation results are shown inSec.5 and a conclusion (Sec.6).

II. METHOD OF COMPLEX COHERENCE

In this section, the theoretical description of the coherence function and show how this function can be used as a criterion for noise reduction. After that, the proposed coherence-based method is described in detail.

2.1 COHERENCE FUNCTION

The substance of the coherence function is a collapsed power spectral density matrix. To fully appreciate the Intricacies of its definition, it is first necessary to review some basic concepts. They include the auto and cross correlation, and power spectral density. The coherence takes values between zero and one and is an indicator of how well two signals are correlation between each other at a particular frequency. Let us assume two microphones placed in a noisy environment in which the noise and the objective speech signals are spatially separated. In this case, the noisy speech signals, after delay compensation, can be defined as

$$y_i(m) = x_i(m) + n_i(m) \qquad (1)$$

where i denotes the microphone index, m is the sample-index and $x_i(m)$ and $n_i(m)$ represent the (clean) speech and noise components are in each microphone, respectively. After applying a short-time discrete Fourier transform on both sides of Eqn (1),it can be expressed in the frequency domain as

$$Y_i(w_l, k) = X_i(w_l, k) + N_i(w_l, k) \qquad (2)$$

where k is the frame index, $\omega_l = 2\pi l/L$ and $l = 0, 1, 2, ..., L - 1$, where L is the frame length in samples. In the following equations we exclude the subscript l for better clarity and call ω the angular frequency. In this paper, we consider the angular frequency range of $[-\pi, \pi]$

rather than [0, 2π]. The complex coherence function for two non-stationary processes is a normalized complex cross power spectral density function and the coherence properties of the speech and noise signals are exploited for the proposed noise reduction scheme.As the coherence function of two signals $Y_1(k)$and $Y_2(k)$is defined as

$$\Gamma_y = \frac{\Phi_{y1y2}}{\sqrt{\Phi_{y1y1}\Phi_{y2y2}}} \qquad (3)$$

From the Eqn (3) where $\Phi_{y1y2}$and Φ_{y1y1}, $\Phi_{y2y2}$are the cross- and auto PSDs of$y_1(k)$and $y_2(k)$.For an ideal diffuse noise field this function can be derived [5] as

$$\Gamma_{n,dif} = sinc\left(\frac{2\pi f d_{mic}}{c}\right) \qquad (4)$$

with the distance d_{mic} between two omnidirectional microphones at frequency f and the range velocity c. The speech signal is often assumed to be coherent ($\Gamma_{s,coh}$=1). However, these conditions are not contented in many real environments. Before the coherence-based method is described, we should point out that a coherence noise field is generated from a single well-defined directional sound source and in our case the equivalent directional microphones outputs are perfectly coherent except for a time delay.

III. NOISE REDUCTION FRAME WORK

A block diagram of the considered noise reduction system for dual-microphone mobile phones is depicted in Figure 1. Since state-of-the-art mobile communication is constrained to single channel transmission [7], only a single-channel output has to be computed from the dual-channel input signals [9]. We assume the input signal $x_i(m)$to be the sum of the desired speech $s_i(m)$and the noise $n_i(m)$wherem is the sample and i the microphone indexes$\left(x_i(m) = s_i(m) + n_i(m)\right)$.

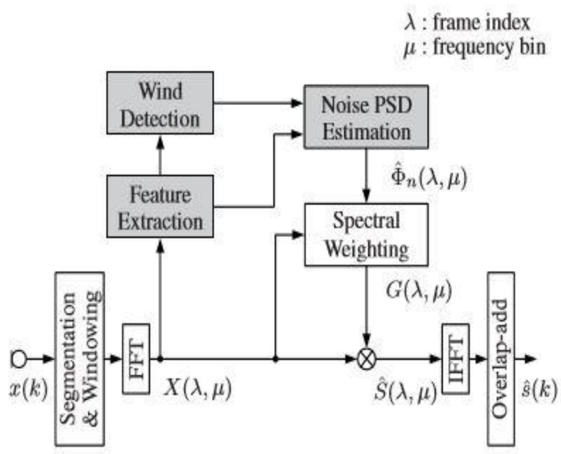

Figure 1: Wind noise reduction system

The signals are segmented, windowed and then transformed to the frequency domain. We obtain the signals $x_i(\lambda,\mu)$in the frequency domain, where λ is the frame and μ is the frequency bin index.

PARAMETER	SETTINGS
Sampling frequency	f_s=16KHZ
Frame length	L=320(20ms)
FFT length	M=512(Including zero-padding)
Frame overlap	50%(Hann window)

$\hat{\Phi}_{nn}(\lambda,\mu)$is the estimated noise PSD and G λ, μ) contains the output spectral weighting gains.$\hat{X}(\lambda,\mu)$ and $\hat{X}(m)$are the enhanced output signals in the frequency and time domain, respectively. The used simulation parameters are given in Table. In this paper we focus on the noise estimation module [5]. Thus, only the output of the noise PSD estimation $\hat{\Phi}_{nn}(\lambda,\mu)$ will be examined. Furthermore, a delay compensation of the useful speech between the microphones is presumed to have already been performed. The noise reduction system is mainly based on the two types,

(a) Spectral subtraction

Spectral subtraction is the most popular noise reduction method. This method operates in the frequency domain and assumes that the spectrum of the input noisy signal can be expressed as the sum of the speech spectrum and the noise spectrum. Figure 2 shows the block diagram for the spectral subtraction method. The noise spectrum is first estimated and then subtracted from the noisy speech spectrum to get the clean speech spectrum.

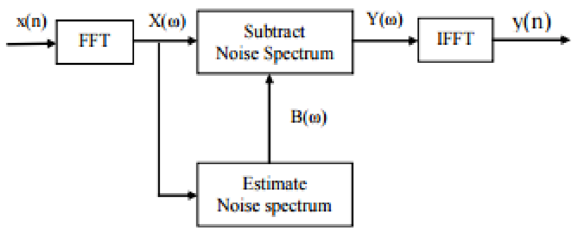

Figure 2: Block diagram of Spectral subtraction

then we can represent it as:

$$y(n) = s(n) + d(n)$$

(5)

Where s(n) is the desired signal and d(n) is the background noise.

(b) Hanning Window

The window function for causal Hanning Window can be represented as:

$$\text{WHann(n)=}\begin{cases} 0.5 - 0.5cos\left(\frac{2\pi n}{N-1}\right), 0 \le n \le N-1 \\ 0 \quad , \ otherwise \end{cases}$$

Whereas non-causal Hanning Window function is given by

$$W_{Hnn}(n) = \begin{cases} 0.5 + 0.5cos\left(\frac{2\pi n}{N-1}\right), 0 \le |n| \le \frac{N-1}{2} \\ 0\,, otherwise \end{cases}$$

(c) Fast Fourier Transform

The next step of this digital signal processing is fast Fourier transform that provides conversion of framed signal from time domain into N samples leading toward frequency domain. FFT is useful tool in digital signal analysis because it helps to reduce computational complexity; the signal recovered after Fast Fourier transform is often called as periodogram which can be further utilized in calculation of power contained in signal under processing.

(d)Wiener filtering

An alternative way to spectral subtraction for recovering an object sequence s(n) from a sequence x(n) = s(n) + b(n) is to find a linear filter h(n) such that the sequence minimizes the expected value of [7]. Under the condition that the signals s(n) and b(n) are uncorrelated and stationary, the frequency-domain solution to this stochastic optimization problem is given by the suppression filter.

$$H(w) = \frac{P_s(w)}{P_s(w)+P_W(w)} \quad (6)$$

which is referred to as the Wiener filter. When the signals s(n) and b(n) are uncorrelated and stationary, the Wiener filter provides noise suppression without considerable distortion in the estimated object.

3.1 Coherence Based Noise Estimation

In this section, we provide the implementation details of the proposed coherence-based method. The signals picked up by the two microphones are first processed in 20 ms frames with 10ms overlap using a Hann window and transformed into the frequency domain with a FFT size of 512.After computing the short-time Fourier transform of the two signals, the PSDs computed based on the following two first order recursive equations. For the auto- and cross-PSDs which are needed in the following the short-term estimates are calculated by recursive smoothing of the input signals as

$$\Phi_{x_i x_j}(\lambda,\mu) = \alpha_s \Phi_{x_i x_j}(\lambda-1,\mu) + (1-\alpha_s)x_i(\lambda,\mu)\cdot x_j(\lambda,\mu)^* \quad (7)$$

The $\{\}^*$ operation denotes the complex conjugate of the signal and the smoothing factor αs is chosen to 0.5. As initially mentioned [4], the coherence properties of the input signals $x_{1/2}(\lambda,\mu)$ can be used for the noise estimation in the higher frequency range. For the sake of brevity the frame and frequency indices (λ and μ) are omitted in the following equations.

3.2 Wind noise detection:

In the proposed method it is necessary to reliably detect signal segments containing wind noise. Based on the phase variance of the complex coherence function a wind noise detection mechanism is presented. Fig.3 explains the complex coherence of a noisy dual channel signal. Here the smoothing constant was chosen to α_s = 0.5. The phase variance and the magnitude of the complex cross power spectral density (PSD) showed in below Figures is express the signal segment in noisy speech signal.

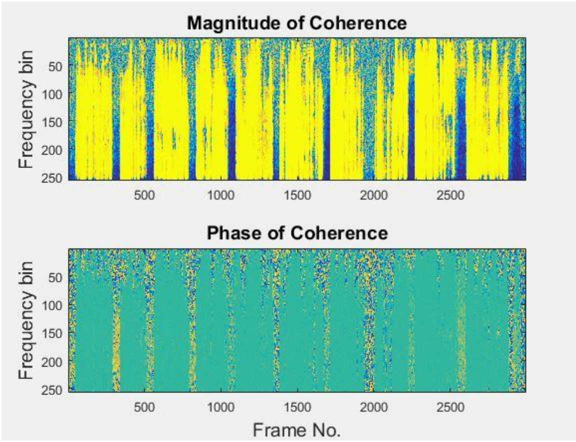

Figure 3: Noisy speech signal and complex coherence

Where the microphone signals are combined in order to generate a single directional signal, and where a first correlation signal is generated composed of the auto correlation function values from one of said microphone signals and where a second correlation signal is generated composed of auto correlation function values from the directional signal, and where the value of the first correlation signal is compared to the value of the second correlation signal, and that a wind noise indicator is activated whenever the value of the second correlation signal is higher than the value of the first correlation signal. From the figures 2 we are understanding that phase is uneven so there is no useful information but the magnitude is shows randomly.

therefore, the distribution of the phase within a time period (e.g., one frame) is investigated in the following.

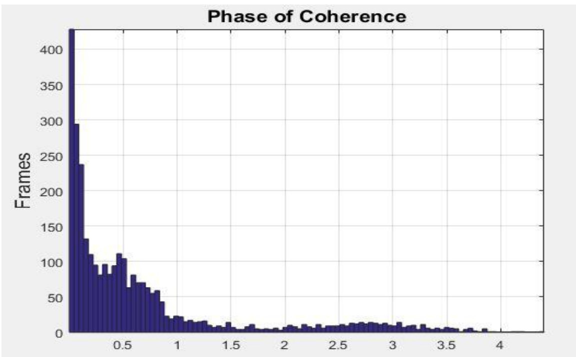

Figure 4: Noisy speech signal

The distributions of the phase measured over a segment of 2 seconds are depicted in Fig.4, derived from clean speech and wind noise signals. It is obvious that the phase of speech results in peak at $\Phi(\lambda,\mu) = 0^0$ whereas the phase of wind signal is nearly uniformly distributed in the interval − π ... π. A measurable quantity for the distribution of a signal is the variance. The phase of coherence wind signal is corrupted by noise in speech signal. To measure the phase of distribution is express above figure. The phase of wind signal is non-uniformly distributed. For a uniform distribution between ± π, thusis normalized by this factor leading to values between 0 and 1

3.3 Wind noise PSD estimation:

In general, the estimation of the PSD $\widehat{\Phi}^2{}_n(\lambda, \mu)$ of a time varying signal is often realized via recursive smoothing of the noise component N (λ, μ) in consecutive signal frames as

$$\Phi_{x_i x_j}(\lambda, \mu) = \lambda\Phi_{x_i x_i}(\lambda - 1, \mu) + (1 - \lambda)|x_i(\lambda, \mu)|^2 \qquad (8)$$

Where the smoothing factor α(λ) can take values between 0 and 1 and can be chosen fixed or adaptive. $|N(\lambda, \mu)|^2$ is called a noise period gram and is not directly accessible since the input signal contains both speech and wind noise. The performance of the noise estimation is heavily dependent on the smoothing factor α (λ) in Eq.8; On the one hand, a small smoothing factor allows fast tracking of the wind noise. This has the drawback that speech segments which are wrongly detected as wind noise have a great influence on the estimated noise PSD [6]. On the other hand, a large smoothing factor reduces the effect of wrong detection during speech activity. However, this leads to slow adaptation of the noise estimate. Thus, an adaptive computation of α(λ) is favorable where low values are chosen during wind activity in speech pauses and high values during speech activity.

IV. SPECTROGRAMS

Speech spectrograms are a useful tool for observing the structure of the residual noise and speech distortion in the outputs of speech enhancement algorithms. Example spectrograms of clean and noisy speech and also those of the outputs coherence-based methods are presented for speech embedded in speech-weighted noise and competing-talkers respectively. The Fig., 5 and 6 show that we estimate the noise, This was done without introducing much distortion in the speech signal. The superiority of the proposed method over the beamformer is more apparent by comparing the spectrograms at low frequencies, where our method manages to recover the target speech signal components more accurately. These evaluations suggest that speech enhanced [6] with our method will be more pleasant to human listeners than speech processed by the beam former. This outcome is in agreement with the improvement in speech quality shown below Fig 5 and the plot shows that, speech signal plus noise signal, is taken as an input signal (green color), to make it strong; we are enhancing the speech signal, the enhanced output of the speech signal is shown in black color. The signal is in time-domain signal.

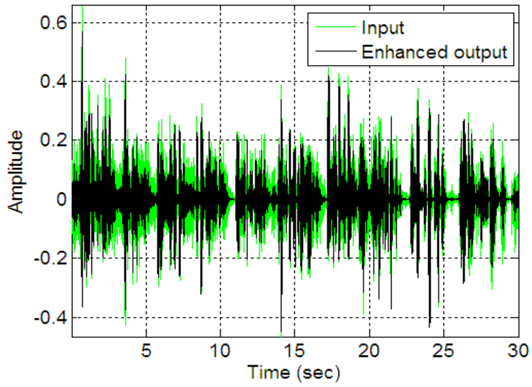

Figure 5: Enhanced noisy speech signal

The above Fig, 5 is noisy speech signal. The figure plots the power spectral density of input noisy speech signal. In the noisy speech signal ,the figure we are done in time domain and after the power plots in noise case the power is reduced after that enhanced.

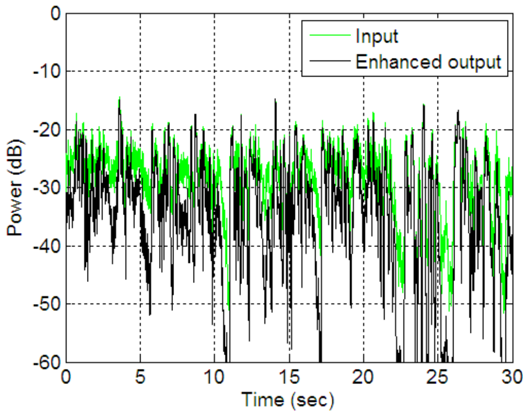

Figure 6: Power calculation of noisy speech signal

The spectrogram of the speech signal plus noise is shown in Fig., 5. From above simulation, we can say that, in top plot, the noise is more in signal, so we cannot extract the signal properly, but after enhancement, the bottom of the Fig., 7 shows that, even though noise is present (red color), we can also detect the signal.

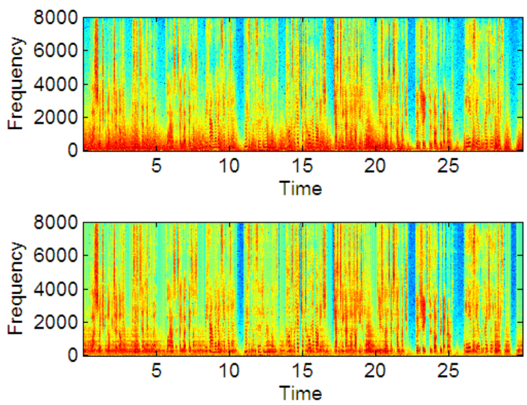

Figure 7: Spectrograms of input and output signal

V. EXPERIMENTS AND RESULTS

For this evaluation, wind noise was recorded with a mock-up mobile phone mounted in hand-held position on an artificial head (HEAD acoustics HMS II.3 & HHP III) on a windy day with wind speeds up to 20m/s. In order to have a reference for the evaluation the noisy speech was generated noise recordings at different SNR values. The thresholds f1 and f2 for the classification were set to 200 and 600 Hz, respectively. The limits were chosen somewhat higher than shown in Fig.6, and ensure less misclassification of speech as wind noise and thus lead to lower speech distortion. The range of the adaptive smoothing factor was set to α_{min}= 0.1 and α_{max}= 0.5. The required PSD [4] of the input signal $\Phi_x(\lambda,\mu)$was calculated by recursive smoothing as defined in Eq., with a fixed smoothing factor of 0.5.

VI. CONCLUSIONS

The proposed dual-microphone algorithm utilizes the coherence function between the input signals and yields a filter, whose coefficients are computed, based on the real and imaginary parts of the coherence function. The proposed algorithm makes no assumptions about the placement of the noise sources and addresses the problem in its general form .In this contribution a system was proposed which exploits the short-term spectral energy distributions to detect and reduce wind noise in noisy speech signal. Based on the complex coherence to estimate the wind noise PSD in a dual microphone signal is presented. An evaluation with wind noise recordings shows that the proposed method outperforms the state-of-the-art complex coherence a approach for background noise estimation and leads to similar results as other approaches especially designed for wind noise reduction with a significantly lower computational complexity.

REFERENCES

[1] C. Nelke, C. Beaugeant, and P. Vary, "Dual microphone noise psd estimation for mobile phones in hands-free position exploiting the coherence and speech presence probability," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., May 2013, pp. 7279-7283.
[2] N. Chatlani, C. Beaugeant, C. Nelke, and P. Vary, “Single microphone wind noise PSD estimation using signal centroids,” in Proc. of IEEE Intern. Conf. on Acoustics, Speech, and Signal Process. (ICASSP), (Florence, Italy), May 2014.
[3] P. Vary and C. Nelke “Measurement, analysis and simulation of wind noise signals for mobile communication devices,” in Proc. of Intern. Workshop on Acoustic Echo analysis.
[4] J. Taghia, J. Taghia, N. Mohammadiha, J. Sang, V. Bouse, and vR. Martin, “An evaluation of noise power spectral density estimation algorithms in adverse acoustic environments,” in Proc.of IEEE Intern. Conf. on Acoustics, Speech, and Signal Process. (ICASSP), Prague, Czech Republic, 2011.
[5] B.F. Cron and C. H. Sherman, “Spatial-correlation functions for various noise models,” Journal of the Acoustical Society of America, vol. 34, no. 11, pp. 1732–1736, 1962.
[6] M. Jeub, C.M. Nelke, H. Kr¨uger, C. Beaugeant, and P. Vary,“Robust dual-channel noise power spectral density estimation, ”in Proc. of European Signal Processing Conf. (EUSIPCO),Barcelona, Spain, 2011.
[7] S. Franz and J. Bitzer, “Multi-channel algorithms for wind noise reduction and signal compensation in binaural hearing aids,” in Proc. of Intern. Workshop on Acoustic Echo and Noise Control (IWAENC), (Tel Aviv, Israel), 2010.
[8] P. Aarabi and G. Shi, “Phase-based dual-microphone robust speech enhancement,” IEEE Trans. Syst., Man, Cybern. B, vol. 34, no. 4, pp. 1763 –1773, 2004.
[9] J.S. Erkelens, R.C. Hendriks, and R. Heusdens, “On the estimation of complex speech DFT coefficients without assuming independent real and imaginary parts,” Signal Processing Letters, IEEE, vol. 15, pp. 213 –216, 2008.