Applied Acoustics 83 (2014) 40–46

Contents lists available at ScienceDirect

Applied Acoustics

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

# A statistical analysis of power-level-difference-based dual-channel post-filter estimator

Shiwei Wang, Chengshi Zheng *, Renhua Peng, Xiaodong Li

*Communication Acoustics Laboratory, Institute of Acoustics, Chinese Academy of Sciences, 100190 Beijing, China*
*Acoustics and Information Technology Laboratory, Shanghai Advanced Research Institute, Chinese Academy of Sciences, 201210 Shanghai, China*

ARTICLE INFO

*Article history:*
Received 12 September 2013
Received in revised form 12 February 2014
Accepted 24 February 2014
Available online 12 April 2014

*Keywords:*
Power level difference
Probability density function
Post-filter estimator
Non-stationary noise

ABSTRACT

By deriving the explicit expression of the probability density function (p.d.f.), this paper presents a statistical analysis of the power-level-difference-based dual-channel post-filter (PLD-DCPF) estimator. The derivation is based on the joint p.d.f. of the auto-spectra of a two-dimensional stationary Gaussian process with mean zero, where the theoretical expression is verified by numerical simulations. Using this theoretical p.d.f. expression, this paper studies the impacts of the correlative parameters on the amount of noise reduction and speech distortion. According to both the theoretical analysis results and the simulation results, four schemes are proposed to improve the performance of the traditional PLD-DCPF estimator.



## 1. Introduction

Multi-channel speech enhancement (MC-SE) often has much better performance than single-channel speech-enhancement (SC-SE), especially in dealing with non-stationary noise and improving speech intelligibility, since more information such as spatial information and directivity can be obtained and exploited [1–6]. Some of the MC-SE algorithms combine a spatial filtering with a post-filter estimator since the amount of noise reduction by spatial filtering decreases significantly as the increasing of the reverberation time [3, Ch.4], where a post-filter estimator is often needed to further reduce the residual noise components. There are mainly three kinds of post-filter estimators, which include the coherence-based [5–9], the phase-based [10,11] and the power-based estimators [12–14]. Generally, all of these post-filter estimators perform better than the existing SC-SE algorithms in non-stationary environments.

Among these post-filter estimators, the power-level-difference-based dual-channel post-filter (PLD-DCPF) estimator is a simple and effective way for non-stationary noise reduction [13]. It assumes that the powers of the desired speech signals at the two microphones are quit different, while those of the noise signals are nearly the same. Based on these assumptions, the PLD-DCPF estimator uses the power level difference (PLD) of the observed signals at the two microphones as a criterion for noise reduction, where this post-filter estimator has a outstanding performance on noise reduction when the speech and noise signals satisfy its fundamental assumptions and all the parameters can be estimated accurately. However, the two assumptions are difficult to meet in real applications. First, the powers of the noise signals at the two microphones may be different due to microphone mismatch. Second, considering that both the speech and the noise signals are stochastic, only an approximation of the parameters can be obtained due to the limited number of available data. The power mismatch of the noise in the two channels and the estimation errors of the parameters may decrease the performance of the PLD-DCPF estimator.

It becomes an attractive research topic to study the mechanisms of the traditional post-filter estimators. Lots of researchers have already made great efforts to study statistical properties of several well-known post-filter estimators, such as the well-known Zelinski post-filter estimator and the noise-field-coherence (NFC)-based post-filter estimator, and so on. Statistical properties of these post-filter estimators can provide some insight into their performance under the stochastic model and can give us some guidelines to improve their performance [15–17]. Zheng et al. have studied statistical properties of the traditional coherence-based post-filter estimators in isotropic noise field [17].

Although some improved versions have already been proposed for the PLD-DCPF estimator [14], its statistical properties have not been well studied until now. To the best of our knowledge, there

* Corresponding author at: Communication Acoustics Laboratory, Institute of Acoustics, Chinese Academy of Sciences, 100190 Beijing, China. Tel.: +86 10 82547945; fax: +86 10 62553898.

*E-mail address:* cszheng@mail.ioa.ac.cn (C. Zheng).

http://dx.doi.org/10.1016/j.apacoust.2014.02.009
0003-682X/

are not any theoretical guidelines to improve the PLD-DCPF estimator for real applications until now. Based on this fact, this paper studies its statistical properties and reveals the influence of the noise power mismatch and the estimation errors of the parameters on its performance in theory. These analysis results can give us a clear path to improve the performance of this post-filter estimator. We derive the explicit expression of the probability density function (p.d.f.) of this post-filter estimator, where a key function for this derivation is the joint p.d.f. of the auto-spectra of a two-dimensional stationary Gaussian process [18]. Using the theoretical p.d.f. expression, we study the impacts of the noise power mismatch and the estimation errors of the parameters on both noise reduction (NR) and speech distortion (SD). Finally, four schemes are proposed to improve the performance of the PLD-DCPF estimator based on these analysis results.

The rest of this paper is organized as follows. Section 2 gives a brief introduction of the PLD-DCPF estimator, and a detailed statistical analysis of this post-filter estimator is presented in Section 3. In Section 4, the theoretical results are verified by the simulation results. Some potential applications are presented in Section 5. Discussions are given in Section 6.

## 2. Background

The PLD-DCPF estimator uses the PLD of the observed signals at the two microphones as a criterion for noise reduction. There are two assumptions in the traditional PLD-DCPF estimator. First,the powers of the desired speech at the two microphones are quite different; Second, the powers of the noise at the two microphones are nearly the same.

Let the received signals at the two microphones, respectively, be:

$$y_1(n) = s(n) + d_1(n), \tag{1}$$

and

$$y_2(n) = h_{12}(n) * s(n) + d_2(n), \tag{2}$$

where $s(n)$ denotes the desired speech at the first microphone. $h_{12}(n)$ is the acoustic transfer function (ATF) of the desired speech between the two microphones. $d_1(n)$ and $d_2(n)$ are the noise received by the first and the second microphones, respectively. The frequency-domain of (1) and (2) can be, respectively, given by:

$$Y_1(k,l) = S(k,l) + D_1(k,l), \tag{3}$$

and

$$Y_2(k,l) = H_{12}(k,l)S(k,l) + D_2(k,l), \tag{4}$$

where $Y_i(k,l)$ and $D_i(k,l)$, with $i = 1,2$, are the short-time Fourier transforms (STFTs) of $y_i(n)$ and $d_i(n)$, respectively. $S(k,l)$ and $H_{12}(k,l)$ are the STFTs of $s(n)$ and $h_{12}(n)$, respectively. $k$ is the frequency index and $l$ is the frame index.

It is well-known that there are two assumptions in the original PLD-DCPF estimator [13]. First, $|H_{12}(k,l)|$ should be much smaller than 1 for all $k$ and $l$. Second, $E\{|D_1(k,l)|^2\} \approx E\{|D_2(k,l)|^2\}$ should hold true for all $k$ and $l$. Under these two assumptions, the PLD-DCPF estimator can be derived from the Wiener filter, which is given by [13, (21)]:

$$G(k,l) = \frac{|\Delta P_Y(k,l)|}{|\Delta P_Y(k,l)| + \theta\left(1 - |H_{12}(k,l)|^2\right)P_{D_s}(k,l)}, \tag{5}$$

where $\theta$ is a constant factor to adjust the noise reduction level of the Wiener filter. When $\theta = 1$, the PLD-DCPF estimator (5) equals to the Wiener filter. $P_{D_s}(k,l)$ is the estimated stationary noise power spectral density (SNPSD) at the first microphone, and $|\Delta P_Y(k,l)|$ is the PLD of the received signals at the two microphones, which can be given by:

$$\Delta P_Y(k,l) = P_{Y_1Y_1}(k,l) - P_{Y_2Y_2}(k,l), \tag{6}$$

where $P_{Y_iY_i}(k,l)$, with $i = 1,2$, are the auto-spectra power spectral density (PSD) estimation of $Y_i(k,l)$, which can be estimated by a recursive scheme:

$$P_{Y_iY_i}(k,l) = \lambda_y P_{Y_iY_i}(k,l-1) + (1-\lambda_y)I_{Y_iY_i}(k,l), \quad i = 1,2, \tag{7}$$

where $I_{Y_iY_i}(k,l) = |Y_i(k,l)|^2$ and $\lambda_y$ is a constant smoothing factor.

## 3. Statistical analysis of the PLD-DCPF estimator

In order to study statistical properties of the PLD-DCPF estimator, this section derives the p.d.f. of $G(k,l)$ in (5) mathematically. Notice should be given that the frequency index $k$ is discarded in the following sections when no confusion can arise.

### 3.1. The joint p.d.f. of the auto-spectra of a two-dimensional stationary Gaussian process

Without loss of generality, we can assume that $(y_1(n),\ y_2(n))$ is a two-dimensional stationary Gaussian process, then the real and the imaginary part of $Y_i(l)$, with $i = 1,2$, are statistically independent Gaussian random variables. Hence, the periodograms of the auto-spectra of the noisy speech $I_{Y_iY_i}(l)$ follow the $\chi^2$ distributions with 2 degrees of freedom, which can be given by [19, (5)]:

$$f_{I_{Y_iY_i}(l)}(x) = \frac{U(x)}{\sigma^2_{Y_i}(l)} \exp\left(-\frac{x}{\sigma^2_{Y_i}(l)}\right), \quad i = 1,2, \tag{8}$$

where $\sigma^2_{Y_i}(l) = E\{I_{Y_iY_i}(l)\}$ and $U(x)$ denotes the unit step function. As can be seen from (7), $P_{Y_iY_i}(l)$ are estimated by smoothing $I_{Y_iY_i}(l)$ over time. Martin introduced the concept of equivalent degrees of freedom and verified that $P_{Y_iY_i}(l)$ follow the $\chi^2$ distributions with $2L$ degrees of freedom, which can be given by [19]:

$$f_{P_{Y_iY_i}(l)}(x) = \frac{LU(x)}{\sigma^2_{Y_i}(l)\Gamma(L)}\left(\frac{xL}{\sigma^2_{Y_i}(l)}\right)^{L-1} \exp\left(-\frac{xL}{\sigma^2_{Y_i}(l)}\right), \quad i = 1,2, \tag{9}$$

where $\Gamma(\bullet)$ is the complete Gamma function. $L$ should be a positive number that represents half of the equivalent degrees of freedom. The relationship between $L$ and $\lambda_y$ can be derived by using [19, (25) and (28)]:

$$L = \frac{1+\lambda_y}{1-\lambda_y}\frac{1}{1+2\sum_{m=1}^{\infty}\lambda_y^m \rho(m)}, \tag{10}$$

where $\rho(m)$ is given by [19, (20)]:

$$\rho(m) = \frac{\left(\sum_{n=0}^{N-1} h(n)h(n+mM)\right)^2}{\left(\sum n = 0^{N-1} h(n)\right)^2}, \tag{11}$$

where $h(n)$, with $n = 0,\ldots,N-1$, is the window function, $N$ is the FFT length and $M$ is the frame shift. If the FFT length, the frame shift and the window function are chosen beforehand, $\rho(m)$ is only determined by (11) that should be a positive number. Thus $L$ is only dependent on $\lambda_y$, which can be found from (10).

Define

$$a = \left(L \cdot P_{Y_1Y_1}(l)\right)/\left(\delta \cdot \sigma^2_{Y_1}(l)\right), \tag{12}$$

and

$$b = \left(L \cdot P_{Y_2Y_2}(l)\right)/\left(\delta \cdot \sigma^2_{Y_2}(l)\right), \tag{13}$$