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

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).

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:

$$y_1(n) = s(n) + d_1(n), \quad (1)$$

and

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

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:

$$Y_1(k,l) = S(k,l) + D_1(k,l), \quad (3)$$

and

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

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)]:

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

where θ 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:

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

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:

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

where $I_{Y_iY_i}(k,l) = |Y_i(k,l)|^2$ and λ_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 χ^2 distributions with 2 degrees of freedom, which can be given by [19, (5)]:

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

where $\sigma_{Y_i}^2(l) = E\{I_{Y_iY_i}(l)\}$ 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 χ^2 distributions with $2L$ degrees of freedom, which can be given by [19]:

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

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 λ_y can be derived by using [19, (25) and (28)]:

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

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

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

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 λ_y, which can be found from (10).

Define

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

and

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

where $\delta = 1 - \gamma^2, \gamma^2 = \left|E\{I_{Y_1Y_2}(l)\}\right|^2 / \left(\sigma_{Y_1}^2(l) \cdot \sigma_{Y_1}^2(l)\right)$ is the magnitude-squared coherence (MSC) of $Y_1(l)$ and $Y_2(l)$, with γ the square root of the MSC, $I_{Y_1Y_2}(l) = Y_1(l)Y_2^*(l)$ is the cross-spectrum of $Y_1(l)$ and $Y_2(l)$. To be mentioned, a and b can be approximately regarded as a form of the auto-spectra of $y_1(n)$ and $y_2(n)$. Furthermore, the joint p.d.f. of a and b can be derived from [18, (4.110)], which is given by:

$$p(a,b) = \begin{cases} \frac{\delta^L e^{-(a+b)}(ab)^{0.5(L-1)}}{\gamma^{L-1}\Gamma(L)} I_{L-1}\left(2\gamma\sqrt{ab}\right), & \text{if } a \geqslant 0 \ \& \ b \geqslant 0 \\ 0, & \text{else} \end{cases}, \quad (14)$$

where $0 < \gamma < 1$. (14) could be regarded as the joint p.d.f. of the auto-spectra of the two-dimensional stationary Gaussian process $(y_1(n), y_2(n))$.

3.2. The relationship between a, b and $G(l)$

Once we can reveal the relationship between a, b and $G(l)$, that is to say, if we can use a and b to express $G(l)$, then the p.d.f. of $G(l)$ can be derived by using (14).

Let $\sigma_{Y_i}^2(l) = E\left\{|Y_i(l)|^2\right\}$, $\sigma_{D_i}^2(l) = E\left\{|D_i(l)|^2\right\}$ and $\sigma_S^2(l) = E\left\{|S(l)|^2\right\}$. We further define the following five parameters:

$$\begin{cases} v(l) = \sigma_{Y_2}^2(l)/\sigma_{Y_1}^2(l), \\ v_s(l) = \sigma_S^2(l)|H_{12}(l)|^2/\sigma_S^2(l) = |H_{12}(l)|^2, \\ v_d(l) = \sigma_{D_2}^2(l)/\sigma_{D_1}^2(l), \\ \xi_1(l) = \sigma_S^2(l)/\sigma_{D_1}^2(l), \\ R(l) = P_{D_s}(l)/\sigma_{D_1}^2(l), \end{cases} \quad (15)$$

where $v(l)$, $v_s(l)$ and $v_d(l)$ measure the power ratio of the noisy signals, that of the desired speech signals and that of the noise signals at the two microphones, respectively. $\xi_1(l)$ is the signal-to-noise-ratio (SNR) at the first microphone and $R(l)$ defines the ratio of the SNPSD to the NPSD at the first microphone. It should be pointed out that all parameters in (15) are defined for each frequency and we have already discarded the frequency index k before. In the remainder of this paper, we will also discard the frame index l of the five parameters defined in (15) when no confusion can arise.

Substituting (12), (13) and (15) into (5), we can rewrite (5) as follows:

$$G(l) = \frac{|a - vb|}{|a - vb| + \theta\left(1 - |H_{12}(l)|^2\right)\frac{RL}{(1-\gamma^2)(\xi_1+1)}}. \quad (16)$$

Define

$$C_1 = \theta\left(1 - |H_{12}(l)|^2\right)\frac{RL}{(1 - \gamma^2)(\xi_1 + 1)}, \quad (17)$$

then (16) can be rewritten as:

$$G(l) = \frac{|a - vb|}{|a - vb| + C_1}. \quad (18)$$

As can be seen from (18), $G(l)$ can be regarded as a function of a and b, thus we can use the joint p.d.f. of a and b in (14) to derive the p.d.f. of $G(l)$.

3.3. Derivation of the p.d.f. of $G(l)$

In order to derive the p.d.f. of $G(l)$, we need to derive the p.d.f. of $|a - vb|$ from (14) first, which can be seen in (18). For this purpose, we define:

$$A = a - vb, \ B = a + vb, \quad (19)$$

then the joint p.d.f. of A and B can be derived from (14) with the help of [20, (8-8)], which is given by:

$$p(A,B) = \frac{\delta^L e^{-\frac{(v+1)B+(v-1)A}{2v}}\sqrt{B^2 - A^2}^{L-1}}{2(2\gamma)^{L-1}\sqrt{v}^{L+1}\Gamma(L)}I_{L-1}\left(\gamma\sqrt{\frac{B^2 - A^2}{v}}\right), \quad (20)$$

where $0 < \gamma < 1$.

The p.d.f. of A is the marginal p.d.f. of A and B, which can be obtained by integrating B in (20), given by:

$$p(A) = \frac{(\delta|A|)^L e^{-\frac{(v-1)A}{2v}}}{2(2\gamma)^{L-1}\sqrt{v}^{L+1}\Gamma(L)}$$
$$\times \int_1^{+\infty} e^{-\frac{(v+1)B|A|}{2v}}\sqrt{B^2 - 1}^{L-1} I_{L-1}\left(\gamma|A|\sqrt{\frac{B^2 - 1}{v}}\right)dB. \quad (21)$$

Then the p.d.f. of $|A| = |a - vb|$ can de obtained as:

$$p(|A|) = p(A) + p(-A), A \geqslant 0$$
$$= \int_1^{+\infty} e^{-\frac{(v+1)BA}{2v}}\sqrt{B^2 - 1}^{L-1} I_{L-1}\left(\gamma A\sqrt{\frac{B^2 - 1}{v}}\right)dB$$
$$\times \frac{(\delta A)^L\left(e^{-\frac{(v-1)A}{2v}} + e^{\frac{(v-1)A}{2v}}\right)}{2(2\gamma)^{L-1}\sqrt{v}^{L+1}\Gamma(L)}, A \geqslant 0. \quad (22)$$

Using (19), we can rewrite (18) as follows:

$$G(l) = \frac{|A|}{|A| + C_1}. \quad (23)$$

Define that $G(l) = z$, then the p.d.f. of $G(l)$ can be obtained from $p(|A|)$ in (22) directly with the help of [20, (5-5)], given by:

$$p(z) = \begin{cases} \int_1^{+\infty} e^{-\frac{(v+1)C_1 zB}{2v(1-z)}}\sqrt{B^2 - 1}^{L-1} I_{L-1}\left(\frac{\gamma C_1 z}{1-z}\sqrt{\frac{B^2 - 1}{v}}\right)dB \\ \times \frac{\delta^L C_1^{L+1}}{2(2\gamma)^{L-1}\sqrt{v}^{L+1}\Gamma(L)}\frac{z^L\left(e^{-\frac{(v-1)C_1 z}{2v(1-z)}} + e^{\frac{(v-1)C_1 z}{2v(1-z)}}\right)}{(1-z)^{L+2}}, & \text{if } 0 \leqslant z \leqslant 1 \\ 0, & \text{else} \end{cases} \quad (24)$$

where $0 < \gamma < 1$.

Eq. (24) is the p.d.f. of the PLD-DCPF estimator $G(l)$, which will be used to analyze both the NR and the SD of this post-filter estimator in the following sections.

4. Simulations and discussion

By using Monte Carlo simulation results [21], this section verifies the validation of the p.d.f. of the PLD-DCPF estimator $G(l)$ in (24), which is derived in Section 3. The Monte Carlo simulation methods are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results; typically one is used to obtain the distribution of an random process by running repeated simulations many times. In this paper we use the Monte Carlo simulation method to obtain the empirical p.d.f. of $G(l)$, as the method can provide the following two benefits: First, the empirical p.d.f. obtained by the Monte Carlo simulation method can approach the theoretical p.d.f. of $G(l)$ when the amount of simulation data is large enough; Second, the input signals are generated by computer, so we can manually set all the parameters that we need, which makes the analysis and the verification quit simple.

For the simulation, the stationary noise, non-stationary noise and the desired speech are assumed to follow the Gaussian distributions. We want to emphasize that this assumption is only an approximation since the distribution the desired speech is generally super-Gaussian. However, the quantitative results of this analysis can give us a deep insight into the performance of the

traditional PLD-DCPF estimator. We apply this assumption to verify the validation of the p.d.f. of $G(l)$ in (24). Under the Gaussian assumptions, we generate the two-microphone signals as follows:

$$y_1(n) = (w_1(n) + \mu w_2(n))u(n) + (w_4(n) + \mu w_5(n))u(n - N_0)$$
$$+ \sqrt{\xi_1}w_7(n)u(n - 2N_0), \tag{25}$$

and

$$y_2(n) = \sqrt{v_d}((w_1(n) + \mu w_3(n))u(n) + (w_4(n) + \mu w_6(n))u(n - N_0))$$
$$+ \sqrt{v_s\xi_1}w_7(n)u(n - 2N_0), \tag{26}$$

where $w_i(n)$, with $i = 1,\ldots,7$, are uncorrelated white Gaussian noise (WGN) processes with the variance $\sigma_{w_1}^2 = \sigma_{w_2}^2 = \sigma_{w_3}^2$, $\sigma_{w_4}^2 = \sigma_{w_5}^2 = \sigma_{w_6}^2$ and $\sigma_{w_7}^2 = (1 + \mu^2)(\sigma_{w_1}^2 + \sigma_{w_4}^2)$. N_0 is a constant positive integer that should be large enough to get sufficient amount of data for the Monte Carlo simulation. The 7 WGN processes can combine to generate the stationary noise component, the non-stationary noise component and the desired speech component for the two-microphone signals.

It should be pointed out that in the original PLD-DCPF estimator, the SNPSD $P_{D_s}(l)$ is estimated simply by a first order recursive smoothing from the first several frames in silent interval, thus it cannot cope with any change of the noise subsequently [13, (18)]. For the simulation signals generated by (25) and (26), the first N_0 samples is stationary Gaussian series with the variance $(1 + \mu^2)\sigma_{w_1}^2$ and can be regarded as only containing the stationary noise component, since it could be tracked by the SNPSD estimator of the PLD-DCPF estimator. The second N_0 samples is also stationary Gaussian series with the variance $(1 + \mu^2)(\sigma_{w_1}^2 + \sigma_{w_4}^2)$, which can be seen as two uncorrelated Gaussian series added together, one with the variance $(1 + \mu^2)\sigma_{w_1}^2$ and the other $(1 + \mu^2)\sigma_{w_4}^2$. Considering that the SNPSD estimator of the PLD-DCPF estimator only estimate the NPSD using the first several frames, $P_{D_s}(l) \approx (1 + \mu^2)\sigma_{w_1}^2$ holds true for all samples even after the time index N_0. Thus for the second N_0 samples, the Gaussian series with the variance $(1 + \mu^2)\sigma_{w_1}^2$ could be regarded as stationary noise component for the SNPSD estimator of the PLD-DCPF estimator, and the Gaussian series with the variance $(1 + \mu^2)\sigma_{w_4}^2$ could be regarded as the non-stationary noise component for this estimator, because it cannot be tracked by the SNPSD estimator. Therefore, the second N_0 samples can be regarded as containing both the stationary and the non-stationary noise components as the increase of the powers of the signals at the time index N_0 could not be tracked by the SNPSD estimator. For $n > 2N_0$, there are three types of signals at the two microphones, which includes the desired speech component, the stationary and the non-stationary noise components.

Obviously, the parameters v_s, v_d and ξ_1 in (25) and (26) satisfy the definition in (15). From (25) and (26), we obtain:

$$C_{d_1 d_2} = 1/(1 + \mu^2), \tag{27}$$

$$R = \sigma_{w_1}^2/(\sigma_{w_1}^2 + \sigma_{w_4}^2), \tag{28}$$

$$\gamma = \frac{\sqrt{v_s}\xi_1 + C_{d_1 d_2}\sqrt{v_d}}{\sqrt{(\xi_1 + 1)(v_s\xi_1 + v_d)}} = \frac{\sqrt{v_s}\xi_1 + \sqrt{v_d}/(1 + \mu^2)}{\sqrt{(\xi_1 + 1)(v_s\xi_1 + v_d)}}, \tag{29}$$

where $C_{d_1 d_2}$ represents the square root of the MSC of the noise. Using the two-microphone signals generated by (25) and (26) by properly choosing the values of all the parameters, the theoretical p.d.f. of $G(l)$ can be verified by Monte Carlo simulation results. $|H_{12}(l)|^2$ is replaced by v_s when calculating the theoretical p.d.f.

and the empirical p.d.f. of $G(l)$. In all simulations, the signals are sampled at 8 kHz and a non-overlapped 256-point STFT is used. L is set to be 9 and $C_{d_1 d_2}$ is set to be 0.5. As can be seen from (27), $C_{d_1 d_2}$ is only determined by μ, so we can set $0 \leqslant C_{d_1 d_2} \leqslant 1$ to any values that we need by properly choosing μ. In this paper, $C_{d_1 d_2}$ is set to a fixed value since we find that $C_{d_1 d_2}$ does not have significant impacts on the performance of the traditional PLD-DCPF estimator if $C_{d_1 d_2} \neq 1$.

Fig. 1 shows the comparison results of the theoretical p.d.f. and the empirical p.d.f. of $G(l)$. As can be seen from Fig. 1, the empirical results fit quite well with the theoretical results, which proves that (24) provides a good fit for the p.d.f. of $G(l)$. After verifying the theoretical p.d.f. of $G(l)$, we will use it to further study the impacts of the parameters in (15) on the NR and the SD in the next section.

5. Applications

It can be seen from Sections 3 and 4 that there are total seven parameters that have impacts on the distribution of the p.d.f. of $G(l)$, which are $C_{d_1 d_2}$, L, R, v_s, v_d, ξ_1 and the estimated ATF of the desired speech between the two microphones. Obviously, these seven parameters are all involved with the NR and the SD. The parameter R can be regarded as the noise estimation accuracy since the original PLD-DCPF estimator only estimates the stationary noise. Therefore, $R \leq 1$ holds true in most cases, because the non-stationary noise is often underestimated.

We want to emphasize that two assumptions have been made in Sections 3 and 4. First, both the speech and the noise are assumed to follow the Gaussian distributions. Second, the noise and the speech are assumed to be independent identically distributed (i.i.d). It is well-known that the super-Gaussian distribution is more suitable to model the speech. However, although the first assumption is just an approximation, the analysis results could also reveal the behavior of the PLD-DCPF estimator. The main reason is that the modeling error is only a minor factor on the qualitative results. Based on these two assumptions, we use the WGN processes in (25) and (26) to generate the empirical data to verify the theoretical results.

5.1. Analysis of noise reduction

For noise only segments, we have $v_s = 0$ and $\xi_1 = 0$. The theoretical amount of NR can be given by:

$$\text{NR} = -20\lg\left(\int_0^1 zp(z)dz\right)\text{dB} \tag{30}$$

where $z = G(l)$ is the PLD-DCPF estimator. Through the theoretical analysis and the experiments, we find that the parameters $C_{d_1 d_2}$ and L do not have significant impacts on the amount of NR. The corresponding results will no longer be presented in this paper due to the space limit. $C_{d_1 d_2}$ and L are set to fixed values in the following sections, where we set $C_{d_1 d_2} = 0.5$ and $L = 9$ to study the impacts of the two parameters R and v_d on the amount of NR. Notice should be given that the tiny estimation error of the ATF $|H_{12}(l)|^2$ also has a weak impact on the amount of NR, hence $|H_{12}(l)|^2$ is replaced by v_s in all the latter analysis.

Fig. 2(a) plots the NR versus different values of v_d, and Fig. 2(b) plots the NR versus different values of R. As can be seen from Fig. 2, the theoretical results fit quit well with the empirical results. Fig. 2 clearly show the impacts of R and v_d on the amount of NR. First, the amount of NR has the highest value when v_d is close to 1, and decreases significantly as v_d goes far away from 1. Second, the amount of NR decreases significantly as R decreases. From the results in Fig. 2, two remarks can be made. First, if the second assumption, $\sigma_{D_1}^2 = \sigma_{D_2}^2$, does not hold true, the PLD-DCPF estimator reduces its amount of NR significantly. Second, only estimating the
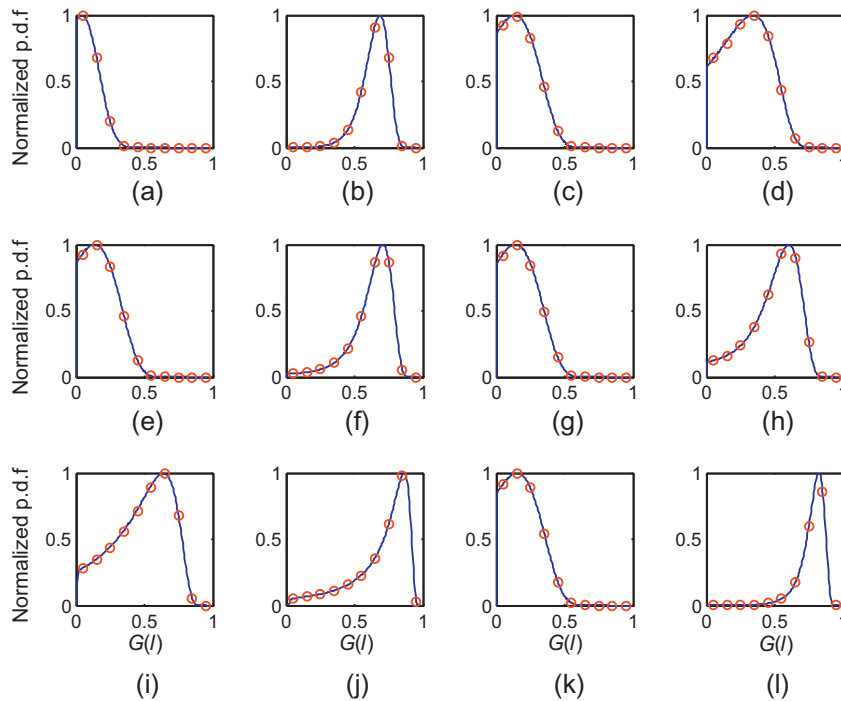
Fig. 1. Comparison of the theoretical p.d.f. of $G(l)$ with the empirical results. (a) $R = 1$, (e) $R = 0.5$, (i) $R = 0.1$; Set $v_s = 0$, $v_d = 1$, $\xi_1 = 0$ for the three cases; (b) $v_s = 0.1$, (f) $v_s = 0.5$, (j) $v_s = 0.9$; Set $R = 1$, $v_d = 1$, $\xi_1 = 5$ for the three cases; (c) $v_d = 1$, (g) $v_d = 0.8$, (k) $v_d = 1.2$; Set $R = 0.5$, $v_s = 0$, $\xi_1 = 0$ for the three cases; (d) $\xi_1 = 1$, (h) $\xi_1 = 3$, (l) $\xi_1 = 10$; Set $R = 1$, $v_s = 0.5$, $v_d = 1$ for the three cases; Blue solid line: theoretical p.d.f. of $G(l)$; Red circle: empirical p.d.f. of $G(l)$. For all the cases $C_{d_1 d_2} = 0.5$, $L = 9$ and $|H_{12}(l)|^2 = v_s$ is used. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
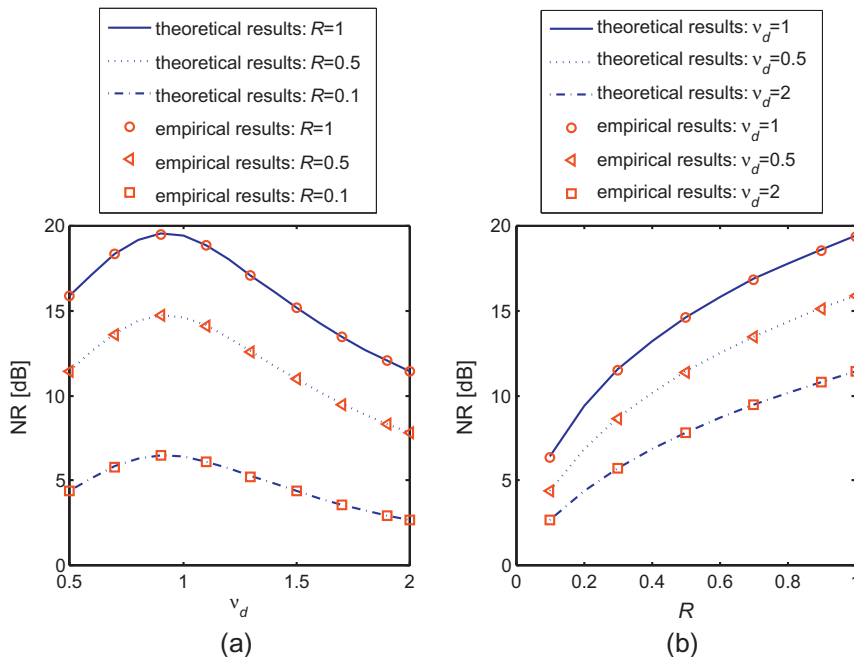

Fig. 2. (a) Noise reduction (NR) of $G(l)$ versus v_d with $R = 1$, $R = 0.5$ and $R = 0.1$; (b) noise reduction (NR) of $G(l)$ versus R with $v_d = 1$, $v_d = 0.5$ and $v_d = 2$. The solid line, dotted line and dash dotted line are the theoretical results of the three cases in (a) and (b), while the circle, triangle and square are the empirical results of them. $C_{d_1 d_2} = 0.5$, $L = 9$, $v_s = 0$ and $\xi_1 = 0$ are used for both (a) and (b).

SNPSD could not suppress non-stationary noise components completely. Therefore, two schemes should be used in practice to improve the performance of the PLD-DCPF estimator. The first scheme of this paper is that the NPSD mismatch at the two microphones should be calibrated to make sure that $v_d \approx 1$ holds true. The second scheme of this paper is that the non-stationary noise component should be estimated.

5.2. Analysis of speech distortion

The theoretical amount of SD over frequency can also be calculated by (30). $C_{d_1 d_2} = 0.5$ and $L = 9$ are used unaltered for analyzing the SD, since both of them have not significant impacts on the amount of SD. Besides, $|H_{12}(l)|^2$ is replaced by v_s due to the weak impact of its estimation error on the amount of SD. Since
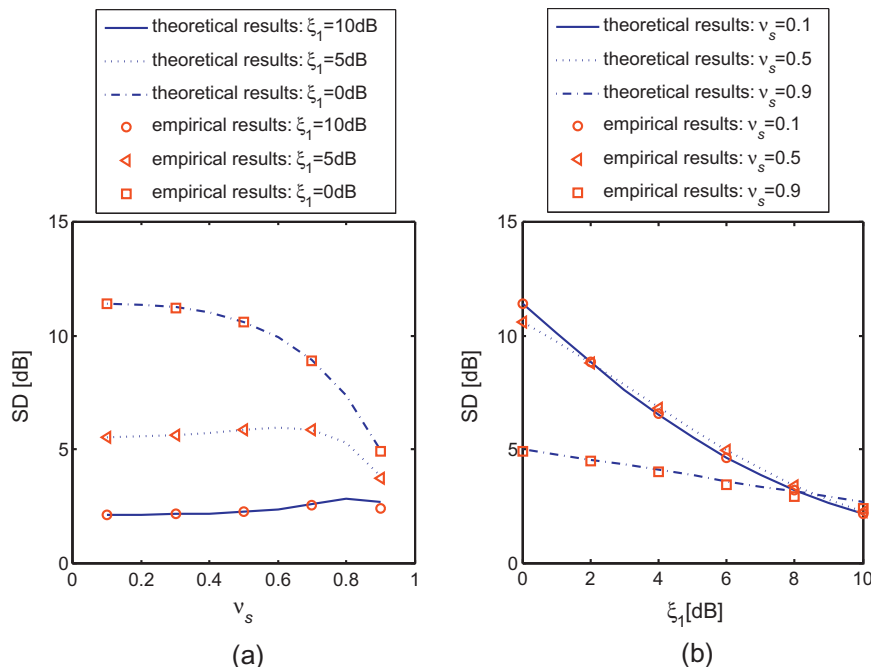
Fig. 3. (a) Speech distortion (SD) of $G(l)$ versus v_s with $\xi_1 = 10$ dB, $\xi_1 = 5$ dB and $\xi_1 = 0$ dB; (b) speech distortion (SD) of $G(l)$ versus ξ_1 with $v_s = 0.1$, $v_s = 0.5$ and $v_s = 0.9$. The solid line, dotted line and dash dotted line are the theoretical results of the three cases in (a) and (b), while the circle, triangle and square are the empirical results of them. $C_{d_1 d_2} = 0.5$, $L = 9$, $R = 1$ and $v_d = 1$ are used for both (a) and (b).

Section 5.1 have already shown that the NPSD mismatch should be calibrated beforehand and the non-stationary noise should be estimated accurately, so $v_d = 1$ and $R = 1$ are directly used in analyzing the SD. Thus, we only need to analyze the two parameters v_s and ξ_1 on the impacts of the amount of SD.

Fig. 3(a) plots the SD versus different values of v_s, and Fig. 3(b) plots the SD versus different values of ξ_1. As can be seen from Fig. 3, the amount of the SD decreases as ξ_1 increases, while the amount of the SD increases as v_s decreases, especially when ξ_1 is not a large value. Moreover, ξ_1 has more impacts than v_s on the SD. Based on these analysis results, two additional schemes can be used to reduce the SD of the PLD-DCPF estimator. The third scheme of this paper is that $y_2(n)$ can be applied to suppress the coherent noise component in $y_1(n)$ adaptively to increase ξ_1. Notice should be given that $v_d = 1$ does not still hold true after applying this adaptive filtering scheme, so we need to calibrate the NPSD mismatch before applying the PLD-DCPF estimator. After calibrating the NPSD mismatch, v_s may become much smaller than before. The fourth scheme of this paper is that we can use $y_1(n)$ to eliminate the speech component in $y_2(n)$ to make sure that v_s is much smaller than 1 to reduce the speech distortion.

This section studies the impacts of several parameters on the amount of NR and that of the SD. After that, four schemes are proposed to improve the performance of the traditional PLD-DCPF estimator.

6. Conclusion

This paper studies statistical properties of the traditional PLD-DCPF estimator. We derives the p.d.f. of this post-filter estimator under Gaussian distributions of both the speech and the noise. After that, we discuss the relationship between the parameters and the amount of the NR and that of the SD. Based on these analysis results, four schemes are proposed to improve the performance of the traditional PLD-DCPF estimator. Further work should concentrate on proposing practical and effective algorithms

to improve its performance, such as a simple NPSD calibration algorithm, an efficient non-stationary NPSD estimation algorithm and a robust adaptive filtering scheme.

References

[1] Brandstein M, Ward D. Microphone arrays: signal processing techniques and applications. Berlin: Springer-Verlag; 2008.
[2] Loizou PC. Speech enhancement: theory and practice. New York: CRC Press; 2007.
[3] Benesty J, Makino S, Chen J. Speech enhancement. Berlin: Springer-Verlag; 2005.
[4] Zheng C, Zhou Y, Hu X, Li X. Speech enhancement based on the structure of noise power spectral density. In: Proc Eur Signal Process Conf (EUSIPCO). Aalborg, Denmark; August, 2010.
[5] Allen JB, Berkley DA, Blauert J. Multimicrophone signal-processing technique to remove room reverberation from speech signals. J Acoust Soc Am 1977;62:912–5.
[6] Bouquin RL, Faucon G. Using the coherence function for noise reduction. IEE Proc Commun Speech Vis 1992;139:276–80.
[7] Bouquin RL, Azirani AA, Faucon G. Enhancement of speech degraded by coherent and incoherent noise using a cross-spectral estimator. IEEE Trans Acoust Speech Signal Process 1997;5:484–7.
[8] Yousefian N, Loizou PC. A dual-microphone speech enhancement algorithm based on the coherence function. IEEE Trans Acoust Speech Signal Process 2012;20:599–609.
[9] Yousefian N, Loizou PC. A dual-microphone algorithm that can cope with competing-talker scenarios. IEEE Trans Acoust Speech Signal Process 2013;21:143–53.
[10] Liu C, Wheeler BC, O'Brien WD, Bilger RC, Lansing CR, Feng AS. Localization of multiple sound sources with two microphones. J Acoust Soc Am 2000;108:1888–905.
[11] Shi G, Arabi P, Jiang H. Phase-based dual-microphone speech enhancement using a prior speech model. IEEE Trans Acoust Speech Signal Process 2007;15:109–18.

[12] Cohen I, Berdugo B. Two-channel signal detection and speech enhancement based on the transient beam-to-reference radio. In: Proc IEEE int conf on acoust, speech, signal process (ICASSP). Hong Kong, China; May, 2003.

[13] Yousefian N, Akbari A, Rahmani M. Using power level difference for near field dual-microphone speech enhancement. Appl Acoust 2009;70:1412–21.

[14] Jeub M, Herglotz C, Nelke C, Beaugeant C, Vary P. Noise reduction for dual-micropone mobile phones exploiting power level difference. In: Proc IEEE int conf on acoust, speech, signal process (ICASSP). Kyoto, Japan; March, 2012.

[15] Marro C, Mahieux Y, Simmer KU. Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering. IEEE Trans Acoust Speech Signal Process 1998;6:240–59.

[16] Zheng C, Zhou Y, Hu X, Li X. Two-channel post-filtering based on adaptive smoothing and noise properties. In: Proc IEEE int conf on acoust, speech, signal process (ICASSP). Prague, Czech Republic; May, 2011.

[17] Zheng C, Liu H, Peng R, Li X. A statistical analysis of two-channel post-filter estimators in isotropic noise fields. IEEE Trans Acoust Speech Signal Process 2013;21:336–42.

[18] Goodman NR. On the joint estimation of the spectra, cospectrum and quadrature spectrum of a two-dimensional stationary Gaussian process. Ph D diss. Princeton Univ, Princeton, USA; 1957.

[19] Martin R. Bias compensation methods for minimum statistics noise power spectral density estimation. Signal Process 2006;86:1215–29.

[20] Papoulis A. Probability, random variables, and stochastic processes. 3rd ed. New York: McGraw-Hill, Inc; 1991.

[21] Landau DP, Binder K. A guide to Monte Carlo simulations in statistical physics. Cambridge: Cambridge University Press; 2009.