On Generalized Auto-Spectral Coherence Function and Its Applications to Signal Detection

Chengshi Zheng, *Member, IEEE*, Hefei Yang, and Xiaodong Li

***Abstract*—Considering that spectral components of one random process are not necessarily independent for all types of signals, this paper defines a generalized auto-spectral coherence function (GAS-CF) to measure this spectral correlation. The GAS-CF is a generalization of the temporal coherence function and the spectral coherence function, where they have already been successfully applied to detect howling components and transient noise components, respectively. After defining the GAS-CF, this paper studies its statistical properties in detail. Simulation results show that the proposed GAS-CF can be applied to detect different types of signals, including transient noise, howling frequency and chirp signal, in a simple way.**

***Index Terms*—Auto-spectral coherence, chirp signal, signal detection, temporal coherence, transient noise.**

I. Introduction

SINCE Goodman defined the coherence function between two wide-sense stationary random processes [1], it has already been widely studied due to its wide applications [2]–[10], such as signal detection, signal estimation and system identification.

In [2], the fast Fourier transform (FFT) was first introduced to calculate the coherence function and the overlapped FFT was further proposed in [3], where these FFT-based coherence functions will be referred as FFT-CF. To improve the frequency resolution, Benesty *et al.* [5] proposed to estimate the coherence function by the minimum variance distortionless response (MVDR) approach, which will be referred as MVDR-CF. It is well-known that the MVDR approach suffers from the serious signal mismatch problem. To solve this problem, the canonical correlation analysis (CCA) method was proposed, which will be referred as CCA-CF [6]. In [7], Jakobsson *et al.* considered the 2-D coherence function estimation and proposed a time-updating estimation method. The coherence function with nonuniformly sampled sequences was first considered in [10]. The three nonparametric coherence functions, including the FFT-CF, the MVDR-CF and the CCA-CF, were first treated in a unified way in [8], which makes it easier to understand these existing coherence functions and their properties. All the above mentioned coherence functions are used to measure some properties of two wide-sense stationary random processes, such as linear and nonlinear system analysis, signal-to-noise-ratio (SNR) and time delay [11].

Unlike the coherence function defined before, Zheng *et al.* defined a temporal coherence function (T-CF) and a spectral coherence function (S-CF) for only one random process to detect howling components and transient noise components in [12] and [13], respectively. This is based on the fact that the independence assumption of spectral components is not necessarily true for all types of signals. For example, the coherence times of howling components are nearly infinite. Although numerous experimental results have verified the validity of the T-CF and the S-CF in detecting the howling and/or the transient noise components, their relationship is still uncovered. Moreover, their applications are not well discussed. In this paper, we will treat the T-CF and the S-CF in a unified way, which will be referred as a generalized auto-spectral coherence function (GAS-CF). Notice should be given that the GAS-CF measures the coherence of different frequency components for only one random process, which is unlike the FFT-CF, the MVDR-CF and the CCA-CF that measure the cross-coherence of two random processes. Furthermore, statistical properties of the GAS-CF will be studied in detail. Finally, some applications in signal detection are discussed and simulations are given to show the validity of the GAS-CF.

The remainder of this paper is organized as follows. The T-CF and the S-CF are reviewed briefly in Section II, and the GAS-CF is also introduced in this section. Statistical properties of the GAS-CF are given in Section III. Practical implementation and some applications of the GAS-CF are presented in Section IV. Section V gives some conclusions.

II. Background and Generalized Auto-Spectral Coherence Function

Before giving the definition of the GAS-CF, we assume that $x(n)$ is a short-time wide-sense stationary random process having zero mean with $n = 0, 1, 2, \cdots$. Define $x_1(n) = x(n)$ and $x_2(n) = x(n - D)$, where D is a non-negative integer number. Assume that $X_1^l(k)$ and $X_2^l(k)$ are the N-point FFTs of $x_1(n)$ and $x_2(n)$, respectively, where $k = 0, 1, \cdots, N-1$ is the frequency index and l is the frame index.

To rewrite the following defined coherence functions in a matrix way, we further define the N Fourier vectors:

$$\mathbf{f}_k = [\,1 \quad e^{-jw_k} \quad \cdots \quad e^{-jw_k(N-1)}\,]^T/\sqrt{N}, \tag{1}$$

where the superscript T is the transpose of a vector and $w_k = 2\pi k/N$, with $k = 0, 1, \cdots, N-1$. The covariance matrices of $x_i(n)$ and $x_j(n)$ in the lth frame can be given by:

$$\mathbf{R}_{x_i x_j}^l = E\left\{\mathbf{x}_i^l (\mathbf{x}_j^l)^H\right\}, i, j = 1, 2, \tag{2}$$

Manuscript received November 08, 2013; revised February 19, 2014; accepted March 01, 2014. Date of publication March 11, 2014; date of current version March 17, 2014. This work was supported by the National Science Fund of China (NSFC) under Grants 61201403 and 61302126 , and was also supported in part by the Tri-Networks Integration Grant KGZD-EW-103-5(3). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xiao-Ping Zhang.

The authors are with the Communication Acoustics Laboratory, Institute of Acoustics, Chinese Academy of Science, Beijing 100190, Beijing, and also with Acoustics and Information Technology Laboratory, Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, Shanghai.. (e-mail: cszheng@mail.ioa.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2014.2310772

where $\mathbf{x}_i^l = [\,x_i(lM) \quad \cdots \quad x_i(lM+N-1)\,]^T$, with M the frame shift and $l = 0, 1, \cdots$. The superscript H denotes Hermitian conjugate of a vector or a matrix. $E\{\bullet\}$ is the expectation function.

A. Definition of the T-CF

When $D = QM$, with Q the positive integer number, holds true, the T-CF can be given by (3), shown at the bottom of the page [12], where the superscript $*$ is the complex conjugate function.

Eq. (3) can be rewritten in a matrix form, which is shown in (4) at the bottom of the page.

B. Definition of the S-CF

In [13, (11)], the S-CF is only calculated for each frame. In this paper, we propose to estimate the S-CF in each bin for each frame, which can be given by:

$$^{\mathrm{S-CF}}C_{xx}^{l}(k,k_0) = \frac{\left|E\left\{X_1^l(k)\left(X_1^l(k+k_0)\right)^*\right\}\right|^2}{E\left\{\left|X_1^l(k)\right|^2\right\}E\left\{\left|X_1^l(k+k_0)\right|^2\right\}}, \tag{5}$$

where $0 < k_0 < N$ is a positive integer.

Eq. (5) can also be rewritten in a matrix form, which is:

$$^{\mathrm{S-CF}}C_{xx}^{l}(k,k_0) = \frac{\left|\mathbf{f}_k^H\mathbf{R}_{x_1x_1}^l\mathbf{f}_{k+k_0}\right|^2}{\left(\mathbf{f}_k^H\mathbf{R}_{x_1x_1}^l\mathbf{f}_k\right)\left(\mathbf{f}_{k+k_0}^H\mathbf{R}_{x_1x_1}^l\mathbf{f}_{k+k_0}\right)}. \tag{6}$$

C. Definition of the GAS-CF

A generalization of the T-CF and the S-CF, which is referred as the GAS-CF, can be given by:

$$C_{xx}^{l}(k,k_0) = \frac{\left|E\left\{X_1^l(k)\left(X_2^l(k+k_0)\right)^*\right\}\right|^2}{E\left\{\left|X_1^l(k)\right|^2\right\}E\left\{\left|X_2^l(k+k_0)\right|^2\right\}}. \tag{7}$$

where $0 \le k_0 < N$. Notice should be given that $k_0 = 0$ is included in (7), which is different from (5).

Obviously, (7) reduces to the T-CF defined in (3) as $k_0 = 0$ and $D = QM$. (7) turns to the S-CF defined in (5) as $D = 0$ and $k_0 = 1, 2, \cdots, N-1$. For the general case, both D and k_0 could be arbitrary integers.

The matrix form of (7) can be given by:

$$C_{xx}^{l}(k,k_0) = \frac{\left|\mathbf{f}_k^H\mathbf{R}_{x_1x_2}^l\mathbf{f}_{k+k_0}\right|^2}{\left(\mathbf{f}_k^H\mathbf{R}_{x_1x_1}^l\mathbf{f}_k\right)\left(\mathbf{f}_{k+k_0}^H\mathbf{R}_{x_2x_2}^l\mathbf{f}_{k+k_0}\right)}. \tag{8}$$

As can be seen from (7) and (8), the T-CF measures the coherence of the spectra of $x(n)$ over time, while the S-CF measures the coherence of the spectra of $x(n)$ over frequency. The GAS-CF can measure the coherence of the spectra of $x(n)$ over both time and frequency. In the following section, statistical properties of the GAS-CF are studied.

III. Statistical Properties of the GAS-CF

The GAS-CF is a generalization of the T-CF and the S-CF, we study the statistical properties of the GAS-CF in this session to give an insight into the mechanism of the proposed GAS-CF in signal detection.

Property 1: $0 \le C_{xx}^{l}(k,k_0) \le 1$ *for all values of* k_0

Proof: Since $\mathbf{R}_{x_ix_j}$, with $i, j = 1, 2$, are assumed to be non-negative definite, it is clear that $C_{xx}^{l}(k,k_0) \ge 0$, where the result can also be derived by (7) directly, since both the numerator and the denominator of (7) are non-negative.

Define $\mathbf{g}_{1,k}^l = (\mathbf{R}_{x_1x_1}^l)^{1/2}\mathbf{f}_k$ and $\mathbf{g}_{2,k}^l = (\mathbf{R}_{x_2x_2}^l)^{1/2}\mathbf{f}_{k+k_0}$, (8) can be rewritten as:

$$C_{xx}^{l}(k,k_0) = \frac{\left|\mathbf{g}_{1,k}^H(\mathbf{R}_{x_1x_1}^l)^{-1/2}\mathbf{R}_{x_1x_2}^l(\mathbf{R}_{x_2x_2}^l)^{-1/2}\mathbf{g}_{2,k}\right|^2}{\left(\mathbf{g}_{1,k}^H\mathbf{g}_{1,k}\right)\left(\mathbf{g}_{2,k}^H\mathbf{g}_{2,k}\right)}. \tag{9}$$

Eq. (9) has the same form with ([5, (31)]), where $C_{xx}^{l}(k,k_0)$ can be easily proved to be not larger than one using the same method presented in [5, (32-37)].

By now, we can conclude that $0 \le C_{xx}^{l}(k,k_0) \le 1$ holds true for all values of k_0 and D. Property 1 reveals that the GAS-CF is a normalized measure of similarity between two spectra of $x(n)$, where its value ranges from 0 to 1.

When $x(n)$ can be modeled by an ARMA process, it can be whitened by applying a whitening filter. Thus, we only study statistical properties of the GAS-CF for a wide-sense stationary white Gaussian process in the following two parts.

Property 2: For the wide-sense stationary white Gaussian noise, $C_{xx}^{l}(k,k_0)$ *decreases as* $D \in [0\ N]$ *increases from 0 to* N *for* $k_0 = 0$

Proof: For the case $D \in [0\ N]$, we have:

$$\mathbf{R}_{x_1x_2}^l = \sigma^2\begin{bmatrix}\mathbf{0}_{D\times(N-D)} & \mathbf{0}_{D\times D}\\ \mathbf{I}_{N-D} & \mathbf{0}_{(N-D)\times D}\end{bmatrix}. \tag{10}$$

where $\mathbf{I}_{N-D}$ is a $(N-D)\times(N-D)$ identity matrix. $\mathbf{0}_{D\times(N-D)}$ is a $D\times(N-D)$ zero matrix.

Substituting (10) into (8) for $k_0 = 0$, we have:

$$C_{xx}^{l}(k,k_0) = \left((N-D)/N\right)^2. \tag{11}$$

Property 2 holds true due to that $C_{xx}^{l}(k,k_0)$ reduces from 1 to 0 as D increases from 0 to N as can be seen from (11). Property 2 shows that the more overlap there is, the higher the GAS-CF is, where this result can also be found in [14].

Notice should be given that $C_{xx}^{l}(k,k_0) \equiv 0$ for all k_0 and k when $D \ge N$, since $\mathbf{R}_{x_1x_2}^l \equiv \mathbf{0}$ in this case.

Property 3: For the wide-sense stationary white Gaussian noise, $C_{xx}^{l}(k,k_0) \le 1/(N\sin(\pi/N))^2$ *for* $D \in [0\ N)$ *and* $k_0 \ne 0$

$$^{\mathrm{T-CF}}C_{xx}^{l}(k,k_0)|_{k_0=0} = \left.\frac{|E\left\{X_1^l(k)\left(X_2^l(k+k_0)\right)^*\right\}|^2}{E\left\{|X_1^l(k)|^2\right\}E\left\{|X_2^l(k+k_0)|^2\right\}}\right|_{k_0=0}, \tag{3}$$

$$^{\mathrm{T-CF}}C_{xx}^{l}(k,k_0)\big|_{k_0=0} = \left.\frac{\left|\mathbf{f}_k^H\mathbf{R}_{x_1x_2}^l\mathbf{f}_{k+k_0}\right|^2}{\left(\mathbf{f}_k^H\mathbf{R}_{x_1x_1}^l\mathbf{f}_k\right)\left(\mathbf{f}_{k+k_0}^H\mathbf{R}_{x_2x_2}^l\mathbf{f}_{k+k_0}\right)}\right|_{k_0=0}. \tag{4}$$

Proof: We can substitute (10) into (8) for $k_0 \neq 0$ directly with $D \in [0\ N)$, then we have:

$$\begin{aligned} C_{xx}^{l}(k,k_0) &= \frac{1}{N^2}\left(\frac{\sin(\pi k_0(N-D)/N)}{\sin(\pi k_0/N)}\right)^2, \\ &\leq \frac{1}{N^2}\left(\frac{1}{\sin(\pi k_0/N)}\right)^2, \\ &\leq \frac{1}{N^2}\left(\frac{1}{\sin(\pi/N)}\right)^2, \end{aligned} \tag{12}$$

where if and only if $k_0 = 1$ and $\pi(N-D)/N = \pi/2$, $C_{xx}^{l}(k,k_0)$ has the maximum value, which equals to $1/(N\sin(\pi/N))^2$. For $N \gg 1$, we have $C_{xx}^{l}(k,k_0) \approx 1/\pi^2$ since $\sin(\pi/N) \approx \pi/N$.

The above two properties reveal that, for the wide-sense stationary white Gaussian process, the maximum value of $C_{xx}^{l}(k,k_0)$ is only about $1/\pi^2$ for all $k_0 \neq 0$. Moreover, the maximum value of $C_{xx}^{l}(k,k_0)$ reduces dramatically with the increasing of D for $k_0 = 0$.

Property 4: If $X_2^l(k+k_0) = X_1^l(k)\{P_{12}\exp(-j\theta_{12})\}$ holds true for some values of D and k_0, where both $P_{12} > 0$ and θ_{12} are constant real values, then $C_{xx}^{l}(k,k_0) = 1$

Proof: When $X_2^l(k+k_0) = X_1^l(k)\{P_{12}\exp(-j\theta_{12})\}$, (7) can be given by:

$$\begin{aligned} C_{xx}^{l}(k,k_0) &= \frac{\left|E\left\{\left|X_1^l(k)\right|^2 P_{12}\exp(-j\theta_{12})\right\}\right|^2}{E\left\{\left|X_1^l(k)\right|^2\right\}E\left\{\left|X_1^l(k)\right|^2 P_{12}^2\right\}} \\ &= \frac{P_{12}^2\left(E\left\{\left|X_1^l(k)\right|^2\right\}\right)^2}{\left(E\left\{\left|X_1^l(k)\right|^2\right\}\right)^2 P_{12}^2} = 1. \end{aligned} \tag{13}$$

Property 4 expresses the two conditions when $C_{xx}^{l}(k,k_0) = 1$ holds true for some values of D and k_0. The first condition, $P_{12} = |X_2^l(k+k_0)|/|X_1^l(k)|$, indicates that $x_1(n)$ and $x_2(n)$ have approximately a fixed ratio of the power spectral density in the kth bin of the lth frame for some values of D and k_0. The second condition, $\theta_{12} = \angle(X_1^l(k)(X_2^l(k+k_0))^*)$, means that $x_1(n)$ and $x_2(n)$ should have a fixed phase difference in the kth bin of the lth frame for some values of D and k_0. In other words, if $X_2^l(k+k_0)$ can be linear predicted by $X_1^l(k)$, then $C_{xx}^{l}(k,k_0) = 1$. There are at least three types of signals including the transient noise components, the howling components and the chirp signal that satisfy the two conditions approximately, where we will give a brief proof in this part.

1) For the Transient Noise Components, $C_{xx}^{l}(k,k_0) \approx 1$ for $D = 0$ and a Small Value of k_0, where $k_0 = 1, 2, 3$: By using the transient noise model presented in [15, (2.1)], the two conditions of property 4 can be derived from [15, (2.12)] directly. Notice should be given that [15, (2.12)] holds true only in the region defined by:

$$\mathbb{B} = \{(w_k, w_{k+k_0}) : |w_k - \alpha_0| < B, |w_{k+k_0} - \alpha_0| < B\}, \tag{14}$$

where α_0 is the center frequency of a stable, unit gain time-invariant bandpass filter as defined in [15] and B is the bandwidth. Generally, the bandwidth B is not a large value, so k_0 should not be too large to ensure that (14) holds true. We propose to use $k_0 \in [1\ 3]$ in practice [13]. Moreover, since the transient noise is highly non-stationary and its duration is extremely short, $D = 0$ should be chosen.

2) For the Howling Components, $C_{xx}^{l}(k,k_0) \approx 1$ for $k_0 = 0$ and Arbitrary Values of D: It is well-known that the closed-loop system becomes unstable if and only if the following two conditions occur [16]:

$$\begin{cases} |G(w_k,l)F(w_k,l)| \geq 1 \\ \angle(G(w_k,l)F(w_k,l)) = m2\pi, \quad m = 1,2,\cdots \end{cases} \tag{15}$$

where $G(w_k,l)$ and $F(w_k,l)$ are, respectively, the short-term frequency responses of the forward and feedback path at w_k in the lth frame. When (15) holds true, an oscillation at w_k will occur, especially when the forward path and the feedback path are time-invariant [16]. In other words, the howling components are always pure oscillations, which makes it have an infinite coherence times for these howling components. Therefore, D could be an arbitrary large value if we ignore its impact on the performance of the detection lag, while $k_0 = 0$ should be selected.

3) For the Chirp Signal, $x_{\text{chirp}}(n) = A_c\cos(2\pi(v_0 n + \kappa n^2/2) + \phi)$, where A_c is the Amplitude, ϕ is a Random Phase, κ is the Chirp Rate and $v_0 + \kappa n$ is the Instantaneous Frequency, $C_{xx}^{l}(k,k_0) \approx 1$, at $k = (v_0 + \kappa n)N$ and $k_0 = \kappa ND$, Holds True After Compensating the Phase Difference for Each Frame: Obviously, $P_{12} = 1$ always holds true for this chirp signal when $k = (v_0 + \kappa n)N$ and $k_0 = \kappa ND$. Unfortunately, the phase difference, $\theta_{12} = v_0 D + \kappa D^2/2 + \kappa Dn$, is not a fixed value since it varies with n. However, κDn is not a random variable, so we can compensate this definite phase difference before calculating the GAS-CF. In other words, the second condition in property 4 holds true after compensating the definite phase difference for each frame.

Based on these statistical properties of the GAS-CF, we can find that the GAS-CF can be easily applied in signal detection, such as detecting the transient noise components [13], the howling components [12] and the chirp signal buried in not only the wide-sense stationary white Gaussian process but also some wide-sense stationary ARMA process of which spectral components only have low spectral correlation. In next section, we will show that the GAS-CF can detect these three types of signals from the noisy/clean speech effectively, where the clean/noisy speech is a typical short-time wide-sense stationary ARMA process.

IV. Practical Implementation and Simulations

A. Practical Implementation

To obtain $X_1^l(k)$ and $X_2^l(k)$ simultaneously, the value of $D = MQ$ should be used in practice. The GAS-CF can be estimated by (16), shown at the bottom of the page, where (16) averages

$$C_{xx}^{l}(k,k_0) = \frac{\left|\sum_{k_\alpha=-K_1}^{K_2}\sum_{l_\alpha=-L_1}^{L_2}\left(X_1^{l+l_\alpha}(k+k_\alpha)\left(X_2^{l+l_\alpha}(k+k_0+k_\alpha)\right)^*\right)\right|^2}{\left(\left(\sum_{k_\alpha=-K_1}^{K_2}\sum_{l_\alpha=-L_1}^{L_2}\left|X_1^{l+l_\alpha}(k+k_\alpha)\right|^2\right)\times\left(\sum_{k_\alpha=-K_1}^{K_2}\sum_{l_\alpha=-L_1}^{L_2}\left|X_2^{l+l_\alpha}(k+k_0+k_\alpha)\right|^2\right)\right)}, \tag{16}$$

K_1+K_2+1 frequency bins and L_1+L_2+1 frames to approximate the expectation operator in (7). In general, $K = K_1 = K_2$ can be used in practice. For the real-time implementation, $L_2 = 0$ should be chosen.

There are some general guidelines to choose K and L_1:

1) For the transient noise detection, $L_1 = 0$ should be chosen since the duration of the transient noise is extremely short. The value of K could be large enough to reduce the variance of the estimated GAS-CF.
2) For the howling frequency detection, the value of K could not be too large to ensure that the frequency resolution is high enough. The value of L_1 could have a large value to reduce the variance of the GAS-CF.
3) For the chirp detection, the value of K should not be too large to reduce the influence of the noise and the value of L_1 should not be too large to track the variation of the chirp rate κ.
4) In all, the parameters K and L_1 should be chosen manually according to the time-frequency properties of the input signal.

B. Simulation Examples

In this part, we give some simulation examples to show the validity of the proposed GAS-CF in signal detection. In all the following three simulation examples, $N = 512$ and $M = 256$ are used. Before calculating the FFT, a Hann window is applied in each frame to reduce spectral leakage.

In the first example, the signal is the microphone output signal used in [12], where the time domain and the spectrogram of this signal can be found in [12, Fig. 4(a) and (b)], respectively. The close-loop sound reinforcement system in this example contains a forward path and a backward path, where the forward path is provided by Waterschoot and Moonen [16], while the backward path is set to be a constant value, i.e. 1.79. The source signal is the clean speech taken from the TIMIT database [17]. The parameters $k_0 = 0$, $D = 5M$, $K = 0$, and $L_1 = 9$ are set in this example. We assume that $H_0^l(k)$ and $H_1^l(k)$ are the null and the alternative hypotheses, respectively, where $H_1^l(k)$ indicates that the howling dose occur at the kth bin of the lth frame, while $H_0^l(k)$ indicates that the howling is absent. The input signal-to-noise-ratio (SNR) is about 0 dB.

In the second example, the transient noise is the mouse clicking noise used in [13], which can be downloaded from the website [18]. The same as [19], the clean speech taken from the TIMIT database [17] is degraded by the mouse clicking noise. $k_0 = 3$, $D = 0$, $K = 32$ and $L_1 = 0$ are used in this example. $H_1^l(k)$ indicates that the transient noise component dose exist, while $H_0^l(k)$ indicates that the transient noise component is absent. The input SNR is 0 dB.

In the third example, the chirp signal is generated as follows:

$$x(n) = x_{\text{chirp}}(n) + w(n) + s(n), \tag{17}$$

where $w(n)$ is a zero-mean white Gaussian noise process with unit variance and $s(n)$ is the clean speech that is also taken from the TIMIT database [17]. $k_0 = \kappa ND$, $D = 5M$, $K = 0$ and $L_1 = 9$ are used. $H_1^l(k)$ indicates that the chirp signal exists, while $H_0^l(k)$ indicates that the chirp signal is absent. The input SNR is 0 dB.

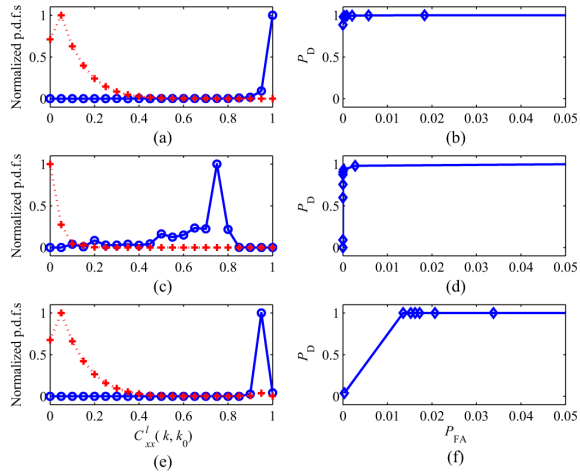

Fig. 1. Empirical results of $f_{C_{xx}^l(k,k_0)|H_i^l(k)}(x)$, with $i = 0, 1$, for (a) howling frequency detection; (c) transient noise detection; (e) chirp signal detection. (b), (d) and (f) are the ROC curves of (a), (c) and (e), respectively. Red dashed line with plus: normalized $f_{C_{xx}^l(k,k_0)|H_0^l(k)}(x)$; blue solid line with circle: normalized $f_{C_{xx}^l(k,k_0)|H_1^l(k)}(x)$.

Denote $f_{C_{xx}^l(k,k_0)|H_1^l(k)}(x)$ as the probability density function (p.d.f) of $C_{xx}^l(k,k_0)$ under $H_1^l(k)$. $f_{C_{xx}^l(k,k_0)|H_0^l(k)}(x)$ is the p.d.f of $C_{xx}^l(k,k_0)$ under $H_0^l(k)$. Fig. 1 plots the empirical results of $f_{C_{xx}^l(k,k_0)|H_i^l(k)}(x)$ for the above mentioned three examples with $i = 0, 1$. The receiver operating characteristic (ROC) curves of these three examples are also plotted in this figure, where P_{D} indicates the probability of detection and P_{FA} indicates the false-alarm rate.

As can be seen from Fig. 1(a), the GAS-CF is close to 1 at the howling frequency, while its value is close to 0 at other frequencies. For the transient noise, the GAS-CF is close to 1 when the transient noise occurs, while it is close to 0 at other segments, where the results can be found from Fig. 1(c). For the chirp signal, the GAS-CF has a large value that is close to 1 only at the instantaneous frequency, while it is close to 0 at other frequencies, where these results can be found from Fig. 1(e). Note that the two p.d.f.s $f_{C_{xx}^l(k,k_0)|H_i^l(k)}(x)$, with $i = 0, 1$, have good separation, implying that we can detect the expected signal efficiently by using the proposed GAS-CF. The ROC curves in Fig. 1 further show that the proposed GAS-CF can detect these signals correctly at very low false-alarm rate.

These results reveal that the proposed GAS-CF can be applied to detect signal in a simple way. Because the GAS-CF is a normalized measure of similarity between two spectra of the input signal, it is not difficult to determine the detection threshold in most cases, where the main reason is that the value of the GAS-CF is not influenced by the relative value of the power spectral density of the input signal.

V. Conclusion

The coherence function between two random processes is very popular and well-known in signal processing. In this paper, we define a generalized auto-spectral coherence function for only one random process to measure the similarity between two spectra of this random process. By studying statistical properties of the GAS-CF, we show how to apply it to detect signal. Simulation results show the validity of the GAS-CF.

REFERENCES

[1] N. R. Goodman, "On the joint estimation of the spectra, cospectrum and quadrature spectrum of a two-dimensional stationary Gaussian process," Ph.D dissertation, Princeton Univ., Princeton, NJ, USA, 1957.

[2] V. A. Benignus, "Estimation of the coherence spectrum and its confidence interval using the fast Fourier transform," *IEEE Trans. Audio Electroacoust.*, vol. AU-17, pp. 145–150, 1969.

[3] G. C. Carter, C. H. Knapp, and A. H. Nuttall, "Estimation of the magnitude-squared coherence function via overlapped fast Fourier transform processing," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 337–344, Aug. 1973.

[4] D. J. Thomson, "Spectrum estimation, and harmonic analysis," *Proc. IEEE*, vol. 70, pp. 1055–1096, 1982.

[5] J. Benesty, J. Chen, and Y. Huang, "A generalized MVDR spectrum," *IEEE Signal Process. Lett.*, vol. 12, pp. 827–830, Dec. 2005.

[6] I. Santamaria and J. Via, "Estimation of the magnitude squared coherence spectrum based on reduced-rank canonical coordinates," in *IEEE Int. Conf. Acoust., Speech and Signal Process.*, Honolulu, HI, USA, Apr. 15–20, 2007.

[7] A. Jakobsson, S. R. Alty, and J. Benesty, "Estimating and time-updating the 2-D coherence spectrum," *IEEE Trans. Signal Process.*, vol. 55, pp. 2350–2354, 2007.

[8] C. Zheng, M. Zhou, and X. Li, "On the relationship of non-parameter methods for coherence function estimation," *Signal Process.*, vol. 88, pp. 2863–2867, 2008.

[9] C. Zheng, Y. Zhou, and X. Li, "Generalised framework for nonparametric coherence function estimation," *Electron. Lett.*, vol. 46, pp. 150–152, 2010.

[10] N. R. Butt and A. Jakobsson, "Coherence spectrum estimation from nonuniformly sampled sequences," *IEEE Signal Process. Lett.*, vol. 17, pp. 339–342, 2010.

[11] G. Carter and C. Knapp, "Coherence and its estimation via the partitioned modified chirp-z transform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 23, pp. 257–264, 1975.

[12] C. Zheng, H. Liu, R. Peng, and X. Li, "Temporal coherence-based howling detection for speech applications," in *133rd Audio Eng. Soc. Conv.*, San Francisco, CA, USA, Oct. 26–29, 2012.

[13] C. Zheng, X. Chen, S. Wang, R. Peng, and X. Li, "Delayless method to suppress transient noise using speech properties and spectral coherence," in *135th Audio Eng. Soc. Conv.*, New York, NY, USA, Oct. 17–20, 2013.

[14] R. Martin, "Bias compensation methods for minimum statistics noise power spectral density estimation," *Signal Process.*, vol. 86, pp. 1215–1229, 2006.

[15] K. W. Baugh and K. R. Hardwicke, "On the detection of transient signals using spectral correlation," *Circuits Syst. Signal Process.*, vol. 13, no. 4, pp. 467–479, 1994.

[16] T. V. Waterschoot and M. Moonen, "Fifty years of acoustic feedback control: State of the art, and future challenges," *Proc. IEEE*, vol. 99, no. 2, pp. 288–327, 2010.

[17] J. S. Garofolo, "Getting started with the DARPA TIMIT CD-ROM: An acoustic-phonetic continuous speech database," in *Nat. Inst. Stand. Technol. (NIST)*, Gaithersburg, MD, 1993.

[18] [Online]. Available: http://www.freesound.org

[19] R. Talmon, I. Cohen, and S. Gannot, "Single-channel transient interference suppression with diffusion maps," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 1, pp. 132–144, Jan. 2013.