DETECTION OF UNKNOWN SIGNALS BASED ON SPECTRAL CORRELATION MEASUREMENTS

Omar A. Yeste Ojeda and Jesús Grajal

Universidad Politécnica de Madrid
Departamento de Señales, Sistemas y Radiocomunicaciones
Madrid, Spain

ABSTRACT

The problem of detecting an unknown signal embedded in white Gaussian noise is addressed. A CFAR detector based on combining the information of the whole frequency-cyclefrequency plane is proposed. An analytical characterization of the detector is provided, and its detection capability evaluated. The FFT-Accumulation Method (FAM) is used to measure the SCF. Approximate analytic expressions of the probability of false alarm are provided.

1. INTRODUCTION

The radiometer is commonly considered as the most appropriate solution to the unknown-signal detection problem, when the signal is stationary and the noise is white [1]. However, most of man-made signals obey an (almost) cyclostationary model rather than a stationary one [2]. This means that the statistics of the signal are (almost) periodic functions of time or, in other words, that all man-made signals exhibit periodicities such as the carrier frequency, the bit rate, the pulse repetition interval, etc. [3]. The objective of this work is to exploit the (second order) cyclostationary properties of man-made signals in order to develop detection schemes outperforming the radiometric approach. Moreover, an approximate analytical expression for the probability of false alarm is provided, and its accuracy discussed.

The cyclostationary approach to the signal detection problem has been formally addressed in [4]. The spectral correlation function (SCF), also known as cyclic spectral density function [2], represents the correlation between two spectral signal components at frequencies \( f + \alpha /2 \) and \( f - \alpha /2 \) [3]. \( f \), which is called the frequency, is the mean frequency of the two components, and \( \alpha \), the cyclefrequency, is their frequency separation. In [4], the SCF of the signal is known except for the signal phase. The knowledge of the signal SCF allows the optimum coherent integration of the SCF along the frequency dimension, for each signal cyclic spectral component. Moreover, the values of cyclefrequency where the signal is present are also known and therefore, the SCF is computed only for these values simplifying the estimation process.

However, the work presented herein deals with the signal detection problem when no assumptions are made about the signal. This produces some differences with the work mentioned previously: First, all the frequency-cyclefrequency plane must be computed, since the cyclefrequencies exhibiting signal components are unknown. This suggests the use of efficient algorithms for measuring the SCF. Second, it is impossible to integrate coherently the signal SCF along the frequency dimension without some previous knowledge about the SCF. Thus, the detection statistics must process the information incoherently.

In the following, we develop two detection statistics based on the spectral correlation measurement (Section 2). These statistics result from decoupling the information acquired through the SCF: That concerning to the conventional power spectral density (PSD), i.e. the SCF for null cyclefrequency, from that concerning to the rest of the SCF. We shall see that this decoupling is adequate because of the different statistical properties of the SCF within these two regions. In Section 3 we study the statistical properties of these statistics and provide an approximate analytical expression for their probability of false alarm (\( P_{FA} \)). The accuracy and range of application are discussed and supported with simulation results. Finally, a new detector based on combining these two statistics for improving radiometric detection is presented in Section 5. Also, the performance results provided shall show the advantages of the cyclostationary approach.

2. DESCRIPTION OF THE DETECTION STATISTICS

Let \( x \) be the sequence of collected data, the detection problem can be formulated through the hypothesis test:

\[
H_0 : x = n \\
H_1 : x = s + n
\]  

where \( s \) is the unknown signal to be detected and \( n \) is an additive white noise. Hereinafter, the noise will be Gaussian, with zero mean and unknown variance \( \sigma_n^2 \).

For solving the previous test, we will use an approach based on measuring the SCF of the collected data. Then, two statistics (\( \gamma_1 \) and \( \gamma_2 \)) are extracted from the measured data and used to detect the signal. The SCF represents the correlation between two spectral signal components at frequencies \( f_1 = f + \alpha /2 \) and \( f_2 = f - \alpha /2 \) [3] where \( f \), the frequency, is the mean of \( f_1 \) and \( f_2 \) and \( \alpha \), the cyclefrequency, is their frequency separation \( \alpha = f_1 - f_2 \).

Since we have made no assumptions about the signal, we cannot know a priori either the spectral or the cyclic spectral components of the signal. Therefore, the SCF must be measured throughout the frequency-cyclefrequency plane. Then, after appropriate normalizations of the SCF for the two different regions in which it is divided, we calculate and analyze two disjoint statistics which will be combined after for a global detector.

---

This work has been supported by the National Grant Program FPU and the project TEC2005-07010-C02-01/TCM.
The SCF information on the whole frequency-cyclefrequency plane has been decoupled in order to analyze the problem. The key reason for this decoupling is the different statistical properties, under hypothesis $H_0$, of the SCF within two regions: $\alpha = 0$ and $\alpha \neq 0$. In the first case, the SCF is always real and positive since it represents the autocorrelation of the signal spectral components, i.e., for $\alpha = 0$ (the frequency axis) the expected SCF matches the conventional power spectral density (PSD) of the signal. On the contrary, for $\alpha \neq 0$, the SCF is not zero (as ideal), but complex valued with zero mean. The reason is that, in practice, the SCF cannot be known, only estimated (what is a noise-dependent operation) [5].

On the other hand, measuring the SCF is computationally very expensive and then, an efficient algorithm should be used. We have chosen the FFT Accumulation Method (FAM) [6]. This algorithm is more efficient than those based on frequency smoothing [6]. Besides, it provides SCF estimate distributed over a rectangular grid, an important feature for detection purposes since many signals exhibit their SCF maxima along straight lines of constant frequency [7]. Nevertheless, the use of FAM introduces non-constant power of the measurement noise, which decreases as the reliability (defined in the next section) increases [8]. Thus, a normalization procedure is required due to the fluctuation of the measurement noise power.

Then, let $\gamma_1$ be the detection statistic concerning the PSD region ($\alpha = 0$):

$$\gamma_1 = \max_f \left[ \tilde{S}_x^2(f) \right] $$

(2)

where $\tilde{S}_x^2(f)$ is the SCF measurement conveniently (mean and variance) normalized with the aim to obtain a set of identically distributed RV. For $\alpha = 0$, we obtain:

$$\tilde{S}_x^2(f) = \frac{\hat{S}_x^2(f) - \hat{\sigma}_x^2}{K(f)\hat{\sigma}_x^2}$$

(3)

where $\hat{S}_x^2(f)$ is the SCF measured through FAM. $\hat{\sigma}_x$ is the mean of $\hat{S}_x^2(f)$ and the noise MLE under the $H_0$ hypothesis:

$$\hat{\sigma}_x = \frac{||x||^2}{N}$$

(4)

where $N$ is the length of the collected data. In (3), $K(f)$ is a function which compensates the variance fluctuation due to FAM [8]:

$$K(f) = \sqrt{1 + \frac{N \sum |h(n;f)|^2}{\sum |h(n;0)|^2}}$$

$$h(n;f) = \sum a(m)a(m+nL)e^{-j\Delta f m}$$

(5)

where, in the FAM implementation, $a(n)$ is the window of the first FFT and $L$ is the decimation factor [6].

The inclusion of $\hat{\sigma}_x^2$ in the denominator of (3) makes $\gamma_1$ invariant to scales and therefore, $\gamma_1$ exhibits CFAR properties with respect to the input noise power.

Now, let $\gamma_2$ be the detection statistic involving the rest of the frequency-cyclefrequency plane, i.e. $\alpha \neq 0$:

$$\gamma_2 = \max_{f: \alpha \neq 0} \left[ \tilde{S}_x^2(f) \right]^2$$

(6)

In this region, the mean of the SCF is zero under $H_0$ and thus, only the variance fluctuations are compensated in the normalization:

$$\tilde{S}_x^2(f) = \frac{\hat{S}_x^2(f)}{K(f,\alpha)\hat{\sigma}_x^2}, \alpha \neq 0$$

(7)

where $K(f,\alpha)$ is the function compensating the variance fluctuations due to FAM, which also depends on $\alpha$ [8]:

$$K(f,\alpha) = \frac{1}{2} \sqrt{\frac{1}{P} \text{Re} \left[ \frac{H_k \left( j2\pi q(\alpha)\frac{\Delta f}{P} ; f \right) }{h_n(n;0)^2 + \alpha N_r |\Delta f|^{-1}} \right] }$$

(8)

where $H_k(j\omega;f)$ is the Fourier Transform of $h_n(n;f)$, $\Delta f$ is the frequency resolution of FAM, $\Delta f$ is the inverse of the cyclefrequency resolution (approximately the input length $N$), $P$ is the length of the second FFT in FAM, and $q$ represents the cyclefrequency offset of the estimate within each channel-pair region. For a better understanding of these FAM parameters, the reader is encouraged to visit [6]. Similarly to $\gamma_1$, the detection statistic $\gamma_2$ is invariant to scales and then presents CFAR properties too.

3. ANALYTIC $P_{FA}$ AND VALIDITY OF THE APPROXIMATION

In this section we will focus on finding an approximate $P_{FA}$ analytical expression for detection statistics $\gamma_1$ and $\gamma_2$. The discussion on the best way of combining $\gamma_1$ and $\gamma_2$ will be deferred until Section 4.

Let $\Gamma_1$ be the set of identically distributed RV maximized by $\gamma_1$, i.e. the set of all SCF measures for $\alpha = 0$, $\tilde{S}_x^2(f)$.

By using the central limit theorem (CLT) for $m$-dependent variables [9], each of the RV in $\Gamma_1$ can be approximated as Gaussian with zero mean and variance $\hat{\sigma}_x^2$. The accuracy of the Gaussian approximation mainly depends on the so-called reliability product $\Delta f/\Delta$, which represents a quality measure of the SCF estimate [5]. In fact, the higher this product is, the more accurate the Gaussian approximation is.

Hereinafter, let us assume the RV in $\Gamma_1$ to be Gaussian. The distribution of $\gamma_1$ is then the distribution of the maximum of a set of correlated Gaussian RV. Indeed, in order to avoid estimation gaps in the frequency-cyclefrequency plane (when FAM is used), they cannot be independent [6]. Nevertheless, assuming the first FFT window exhibits low sidelobes, the approximate $P_{FA}$ curve can be obtained by continuing as if they were independent:

$$P_{FA,\gamma_1}(Th) = 1 - \left[ Q \left( \frac{Th}{\sigma_{\gamma_1}} \right) \right]^{K_1}$$

$$\sigma_{\gamma_1}^2 = \frac{1}{P} \sum |h(n;0)|^2$$

(9)

$$K_1 = \frac{1}{2\Delta} + 1$$

where $Th$ is the detection threshold, $Q(z)$ is the cumulative distribution function of a normal RV $N(0,1)$, and $K_1$ is the
cardinal of the set $\Gamma$, i.e. the number of frequency channels in the range $f = 0$ to $0.5$

Now, let us focus on the detection statistic $\gamma_1$. In this case, and making use of the CLT for m-dependent variables again, it can be seen that each of the RV maximized in (6) exhibit an exponential distribution [8]. Analogously to results for statistic $\gamma_1$, the exactitude of the exponential assumption is intimately related to $\Delta f/\Delta t$. Then, the asymptotic distribution of the extreme value of a set of exponential RV has been studied in [10]. Using the results therein, the $P_{FA}$ of the detection statistic $\gamma_1$ can be approximated by the analytic expression:

$$P_{FA}(Th) = 1 - \exp(-\exp(-Th - \log(K_2)))$$

$$K_2 = \left(\frac{2}{\Delta f} + 1\right)^{-1} \Delta t / \Delta f - 1$$

where $K_2$ is the cardinal of the set maximized by $\gamma_1$.

In Figure 1, the simulated (100,000 trials, thicker trace) and analytical approximate $P_{FA}$ (thinner trace) for both detection statistics $\gamma_1$ (on the left) and $\gamma_2$ (on the right) are plotted. For each statistic, three cases are represented: constant reliability product, constant frequency resolution and constant data length.

However, the approximations made for both statistics are no longer accurate for small values of $\Delta f/\Delta t$. On the other hand, for a constant $\Delta f/\Delta t$, the higher $\Delta f$ is, the more precise analytical $P_{FA}$. As a result, in order to preserve the accuracy, a decrease in $\Delta f$ (better frequency resolution) requires a proportionally greater increase in $\Delta t$ (the collected data length).

4. THE DETECTOR

Then, both statistics are combined in order to obtain a single detector. The best solution is to find out the optimum detector in the Neyman-Pearson sense. The optimum detector results from applying the GLRT for statistics $\gamma_1$ and $\gamma_2$:

$$\frac{f_{H_1}(\gamma_1, \gamma_2)}{f_{H_0}(\gamma_1, \gamma_2)} \geq Th$$

where $f_{H_1}(\gamma_1, \gamma_2)$ and $f_{H_0}(\gamma_1, \gamma_2)$ are the joint pdf of $\gamma_1$ and $\gamma_2$ under hypothesis $H_1$ and $H_0$, respectively. $Th$ is the detection threshold, which is set to attain the desired $P_{FA}$. Unfortunately, this detector is hard to implement in practice since $f_{H_1}(\gamma_1, \gamma_2)$ is highly variable with the SNR and many signal features such as its modulation, bandwidth, etc.

For this reason, we use a simple detector consisting of the logical OR of the two single detectors which use, separately, statistics $\gamma_1$ and $\gamma_2$, that is, a detection is declared if any of the single detectors declare a detection. The motivation of this detector makes sense after checking the results shown in the next section, where it can be appreciated that in some cases the best detection probability is achieved with the single detector employing $\gamma_1$, and in other cases the single detector using $\gamma_2$ is better. Thus, the OR detector shall approach the best of them. Moreover, the probability of false alarm of the OR detector can be expressed analytically by using (9) and (10) in the following equation:

$$P_{FA OR}(Th_1, Th_2) = P_{FA H_1}(Th_1) + P_{FA H_2}(Th_2) - P_{FA H_1}(Th_1)P_{FA H_2}(Th_2)$$

where $Th_1$ and $Th_2$ are the detection thresholds of the single detectors. Furthermore, we set $P_{FA H_1}(Th_1) = P_{FA H_2}(Th_2)$ and, assuming they are low, the last term in (12) can be obviated. Then, $Th_1$ and $Th_2$ are set to achieve in each single detector, respectively, half the probability of false alarm of the OR detector. On the other hand, more complex (and ad hoc) detection schemes than the OR detector have been used too, with no significant improve in the detection.

Finally, (12) is valid only if statistics $\gamma_1$ and $\gamma_2$ are independent from each other, which we discuss in the following. The results obtained by simulation are in concordance with the independence assertion, and although it does not represent a formal justification, it is useful to provide an idea of the validity of the independence supposition. We can see in Figure 2 the contour map of the simulated joint pdf (thicker trace) and the analytical joint pdf (thinner trace) obtained by multiplying the marginal pdf, which can be easily obtained from (9) and (10). It can be seen that as the reliability product increases, the simulated curves fit to the analytical curves better.

Additionally, an independence chi-square test [11] has been applied to each of the cases represented. As a figure of merit, it can be used the maximum significance level for which the test declares that statistics $\gamma_1$ and $\gamma_2$ are independent. The significance level represents the probability that
γ1 and γ2 are declared dependent, given that they are independent. Thus, the higher the significance level is, the likelier the independence assumption is [11]. The results show that statistics γ1 and γ2 are independent with a significance level 0.09 for ∆f/∆ = 128, 0.77 for ∆f/∆ = 256 and 0.73 for ∆f/∆ = 512. Thus, the independence assumption of (12) is less clear in the first case, although a commonly used value of the significance level is 0.05 [11].

5. RESULTS

In Figure 3, it is plotted the probability of detection (P_D) resulting for the three detectors described: That using only statistic γ1 (PSD region: α = 0); that using only statistic γ2 (α ≠ 0); and the third one, proposed herein, which combines both statistics by the OR scheme. The examples include three different signal modulations, BPSK, QPSK and MSK, all with a symbol duration Tb = 32 samples, and the detection threshold has been fixed to attain, in each case, P_FA = 10^{-3} (on the left) and P_FA = 10^{-6} (on the right). The results shown have been computed for a reliability product ∆f/∆ = 256, frequency resolution ∆f = 1/64 (and therefore, ∆t = 2^{14}), and decimation factor L = 8 (and thus, P = ∆t/L = 2^{11}). It is noteworthy that the single detector based on statistic γ1 is equivalent to a channelized energy detector and can be used as a reference too. This detector is the result of applying a bank of filters with bandwidth ∆f, and then, in each branch, an amplitude-squaring device followed by a non-coherent integrator. Then, the maximum of all these outputs normalized according to (3), i.e. γ1, is compared with a detection threshold. In addition, just for comparison purposes, the plots also show the probability of detection of three more detectors. The first one is an FFT-based detector, which compares the squared magnitude of the outputs of the FFT with a threshold and assumes that the noise power is known. The other two detectors are detectors which employ statistics γ1 and γ2, respectively, but supposing that the locations of the SCF maximum amplitude in both regions are known, and therefore, the detection is always made by using these single points of the SCF. These last two detectors represent an upper P_F bound for the single detectors in the unknown maximum location case.

In the figures, it can be appreciated that the OR detector practically equals the best of both single detectors. Maximum-location-known detectors need some signal knowledge (just used as reference).

Other important result is that, when decreasing the P_FA, the detector based on γ2 requires a smaller increase in SNR than the detector based on γ1 in order to preserve their P_D. The reason is their different statistics which result in an intersection of the ROC (receiver operating characteristic) curves of both detectors. Thus, for a high P_FA, the detector based on γ1 is better than that based on γ2, but as P_FA decreases,
the detector based on $\gamma_2$ gets better. A similar result was also obtained in [4] for the known signal case.

### 6. SUMMARY

In this paper, we have used a cyclostationary approach to the problem of detecting an unknown signal embedded in white Gaussian noise. The proposed detection scheme first decouples, and after recombines, the information of the spectral correlation function, one concerning $\alpha = 0$ (the PSD) and the other concerning $\alpha \neq 0$. This scheme results in the OR detector described above. Since the SCF should be computed throughout the frequency-cyclefrequency plane, an efficient algorithm (FAM) is used for measuring the SCF. Besides, the OR detector exhibits CFAR properties and equals the best of the single detectors, improving the channelized energy detector sensitivity up to 1 dB. This improvement is better when the $P_{FA}$ decreases, due to an intersection of the ROC curves, and when the signal exhibits a high maximum of the SCF amplitude for $\alpha \neq 0$. Of course the cyclostationary approach only has sense when the signals we want to detect present a SCF with a moderate maximum amplitude for $\alpha \neq 0$. Otherwise, a conventional radiometer will provide similar detection performance.

An interesting future work could be the extension of the results herein to higher order cyclostationarity, in order to improve the sensitivity for signals with a low level of second-order cyclostationarity. On the other hand, the statistical analysis of the detector leads to an approximate analytic expression for the $P_{FA}$. The accuracy of the formula depends on the reliability product $\Delta f/\Delta \alpha$, which also represents a quality measure of the SCF estimator.

### REFERENCES


