be used in a variety of problems associated with transient signal processing, and it can provide robustness without performance degradation relative to conventional methods as the GLRT. Structural signal information is lost when using ordered measurements, and therefore, if such information is available, the OS approach is beneficial only where its advantages cover for the information loss. The OS approach achieves these advantages in the cost of the increasing computational complexity of the processors. Modern computational capabilities enable the implementation of algorithms based on this approach and the benefit of its advantages.

**APPENDIX**

**PROOF OF THEOREM**

Assume that (8) has a local maximum at \( \hat{L} \). In order to prove that this local maximum is also a global maximum, we have to prove that the likelihood function is a monotonic increasing function for every \( L < \hat{L} \) and monotonic decreasing function for every \( L > \hat{L} \). We prove that the likelihood function is a monotonic decreasing function for \( L > \hat{L} \). Proving the other part is similar.

Since \( \hat{L} \) is a local maximum

\[
\frac{f_{X_m:N}(x_m \mid \hat{\theta}, \hat{L})}{f_{X_m:N}(x_m \mid \hat{\theta}, L + 1)} \geq 1. 
\]

We plug in (8) in (13), and we use basic algebraic operations that finally yield

\[
1 - F_N(x_m \mid \hat{\theta}) \leq \frac{\hat{L} - m + 1}{\hat{L} + 1}. \tag{14}
\]

We now prove that

\[
\frac{f_{X_m:N}(x_m \mid \hat{\theta}, \hat{L} + 1)}{f_{X_m:N}(x_m \mid \hat{\theta}, \hat{L} + 2)} \geq 1. \tag{15}
\]

Assume that (15) does not hold so that \( f_{X_m:N}(x_m \mid \hat{\theta}, \hat{L} + 1)/f_{X_m:N}(x_m \mid \hat{\theta}, \hat{L} + 2) \leq 1 \). Using the same method that was used to simplify (13), we get

\[
1 - F_N(x_m \mid \hat{\theta}) \geq \frac{\hat{L} - m + 2}{\hat{L} + 2}. \tag{16}
\]

The result of combining (14) and (16) is

\[
\frac{\hat{L} - m + 2}{\hat{L} + 2} \leq 1 - F_N(x_m \mid \hat{\theta}) \leq \frac{\hat{L} - m + 1}{\hat{L} + 1}. \tag{17}
\]

It is easy to verify that \( (\hat{L} - m + 1)/(\hat{L} + 1) \) is smaller than \( (\hat{L} - m + 2)/(\hat{L} + 2) \). Since \( 0 \leq 1 - F_N(x_m \mid \hat{\theta}) \leq 1 \), (17) is false, and therefore, (15) is true.

Using the same method, we prove, by induction, that for every \( L > \hat{L} \), if \( f_{X_m:N}(x_m \mid \hat{\theta}, L)/(f_{X_m:N}(x_m \mid \hat{\theta}, L + 1)) > 1 \), then \( f_{X_m:N}(x_m \mid \hat{\theta}, L + 1)/(f_{X_m:N}(x_m \mid \hat{\theta}, L + 2)) > 1 \). This completes proof.

**REFERENCES**


**Detection of Transients in 1/f Noise with the Undecimated Discrete Wavelet Transform**

Thomas T. Liu and Antony C. Fraser-Smith

**Abstract**—A method based on pattern matching in the undecimated discrete wavelet transform domain is introduced for the detection of a known transient with unknown parameters in the presence of 1/f noise. Maxima tracking techniques are used to reduce the computational complexity of the matching procedure by an order of magnitude, with minimal performance impact.

**Index Terms**—Signal detection, wavelet transforms.

**I. INTRODUCTION**

Wavelet transforms have been widely applied to the problem of transient detection and processing, primarily because the transform basis functions provide good time localization. For a transient with known parameters in 1/f Gaussian noise, Wornell [1] described a shift-variant, matched filter detector in the discrete wavelet transform (DWT) domain. Detectors based on maxima tracking in either the undecimated discrete wavelet transform (UDWT) domain [2] or the analytic wavelet transform domain [3] have also been proposed. These techniques rely on the observation that the evolution of the transform maxima across scales provides a measure of the local regularity of the signal [2]. Maxima tracking makes intuitive sense, but the connection to standard detection theory has not been clear.

In this correspondence, we extend the work of [1] and introduce a shift-invariant, generalized likelihood ratio test (GLRT) detector for a known transient signal of unknown amplitude, scale, and delay parameters.

Manuscript received December 3, 1998; revised August 27, 1999. This work was supported in part by the United States Geological Survey under Grants 1434-HQ-97-GR-03124 and 1434-HQ-96-GR-027015. The associate editor coordinating the review of this paper and approving it for publication was Dr. Lal C. Godara.

T. T. Liu is with the Department of Radiology, University of California, San Diego Medical Center/Thornton Hospital, La Jolla, CA 92037 USA (e-mail: tliu@ucsd.edu).

A.C. Fraser-Smith is with the Space, Telecommunications, and Radioscience Laboratory, Stanford University, Stanford, CA 94305 USA.

Publisher Item Identifier S 1053-587X/00/03294-3.
eters in 1/f Gaussian noise. The proposed detection method can be implemented in a straightforward manner as a pattern-matching procedure in the UDWT domain. Furthermore, the computational complexity of the matching process can be reduced, with negligible impact on performance, by an order of magnitude through the use of the transform local maxima to estimate the delays and scales that are most likely to maximize the detection statistic. This result provides a means of interpreting maxima-tracking-based techniques in the context of detection theory.

II. THEORY

We use the subscript notation $\psi_m(t) = 2^{-m/2}\psi(2^{-m}t)$ and $\psi_{m,n}(t) = 2^{-m/2}\psi(2^{-m}t - n)$, where $\psi(t)$ is an orthonormal wavelet. The notation $\psi_m$ refers to the Coiflet (parameter 2) family of wavelets and scaling functions [4]. The inner product of two functions is defined as $\langle f(t), g(t) \rangle = \int_{-\infty}^{\infty} f(t)g^*(t)dt$, where $g^*$ is the complex conjugate of $g$. The DWT of a signal $x(t)$ is $X_{m,n} = \langle x(t), \psi_{m,n}(t) \rangle$. The choice of time origin for the basis functions $\psi_{m,n}(t)$ is arbitrary, and we define other DWT’s with basis functions $\psi_{m,n}(t-J)$ and the notation $X_{m,n}^{[j]} = \langle x(t), \psi_{m,n}(t-J) \rangle$, where $J$ is an arbitrary integer shift. If $M$ denotes the largest analysis scale of interest, then the $X_{m,n}^{[j]}$ are invariant to shifts by integer multiples of $2^M$, i.e., $X_{m,n}^{[j]} = X_{m,n-2^j}^{[j]}$. As a result, there are $2^{J_{\max}}$ unique DWT shifts, with each shift giving rise to a different decomposition of the signal $x(t)$. The UDWT is defined as $X_{m,n} = \langle x(t), \psi_{m,n}(t-n) \rangle$.

A. Detection Model

Consider the detection of a transient signal with unknown amplitude, scale, and delay. We have the standard choice between two hypotheses $H_0: x(t) = v(t)$ and $H_1: x(t) = As_k(t) + v(t)$, where $A \neq 0$ is the unknown amplitude, and $k$ and $l$ are integers that represent the quantized scale and delay, respectively. We define $s_k(t) = 2^{-k/2}\xi_k(t-l)$, where $\xi_k(t) = 2^{-k/2}\xi(2^{-k}t)$, and $\xi(t)$ is the signal model. The reason for the leading $2^{-k/2}$ factor is given in Section II-C. The additive noise $v(t)$ is assumed to be a 1/f Gaussian random process.

A standard scheme for detection with unknown parameters is the generalized likelihood ratio test (GLRT) [5]. It has the following form: Choose $H_1$ if the likelihood ratio $r(x(t)) = r_1$, where

$$r(x(t)) = \frac{\max\{A, k, l, H_1\}}{\max\{A, k, l, H_0\}}$$

and $r_1$ is a threshold value chosen to achieve a desired probability of false alarm (PFA), and choose $H_0$ otherwise. In most cases, the likelihood ratio can be replaced with a sufficient test statistic that has a simpler expression [5].

Detection in 1/f noise is based on the observation that the DWT acts as an approximate whitening transform for such processes [1]. If $v(t)$ represents the noise process with power spectrum $\sigma^2/f^\alpha$ over a range of frequencies $[f_{\min}, f_{\max}]$, then the DWT coefficients $V_{m,n}$ are approximately uncorrelated, and the variance at each analysis scale $m$ is $\sigma^2_2 = \sigma^2_2 2^{-m}$, where $\sigma^2_2 = (2\pi\gamma)^2\sigma^2_2$, and $\gamma$ is a parameter that depends on the choice of wavelet and $\gamma$. We also assume that $v(t)$ has a finite mean so that $V_{m,n}$ has zero mean.

B. Detection with Known Parameters

We first consider the detection problem with known signal parameters in order to develop concepts that will be useful in understanding the unknown parameter case. Because the DWT acts as an approximate whitening transform, the equivalent hypotheses (see [1] and [6]) in the DWT domain are $H_0: X_{m,n} = V_{m,n}$ and $H_1: X_{m,n} = A S_{m,n}^{[k]} + V_{m,n}$, where $A, k$, and $l$ are known parameters, $X_{m,n}^{[k]}$ was previously defined, and

$$S_{m,n}^{[k]} = \langle s_k(t), \psi_{m,n}(t-l) \rangle = \text{DWT of } s_k(t)$$

with shift $J$. The value $J = 0$ corresponds to the standard, unshifted DWT. The likelihood ratio $r(X^{[k]}) = f(X^{[k]}|H_1)/f(X^{[k]}|H_0)$, with $X^{[k]} = \{X_{m,n}^{[k]}, m, n \in \mathbb{Z}\}$ is the vector of DWT observations. The ratio simplifies to yield the sufficient test statistic

$$\lambda(X) = \sum_{m,n} 2^{-m} X_{m,n}^{[k]} S_{m,n}^{[k]}$$

where the superscript on $\lambda$ has been omitted for notational convenience.

Because the DWT is shift variant, the performance of the detector is also shift variant. Shift variance in detectors that use a whitening transform arises when the signal energy in the transformed coordinates is measured for use in the remainder of this correspondence. For a given known signal $s_k(t)$, we are free to choose the DWT shift that maximizes the performance index. For the signal $s_k(t)$ with delay $l = 0$, we define a detector with a test statistic of the form

$$\lambda(X) = \sum_{m,n} 2^{-m} X_{m,n}^{[k]} S_{m,n}^{[k]}$$

where $J_k = \arg\max_{j} \lambda(X|H_1) = \arg\max_{j} \lambda(X|H_1) = \lambda(X|H_1)$. The subscript $k$ on $J_k$ indicates that the optimum DWT shift is a function of the scale $k$ of the input signal. The performance index of this detector is equal to the maximum performance index of the shift variant detector since $\lambda(X|H_1) = \sum_{m,n} 2^{-m} (S_{m,n}^{[k]} S_{m,n}^{[k]})^2$. Using the previously stated identity, we note that $\sum_{m,n} 2^{-m} (S_{m,n}^{[k]} S_{m,n}^{[k]})^2$
In the UDWT domain, the statistic is written as
\[ \lambda(X, l) = \sum_{m, n} 2^{-\gamma m} X_{[0]^m, [k]^n} S_{m, n}^{[k], [l]} \]
and used the relation \( S_{m, n}^{[k], [l]} = \langle s_{m, n}, (l \cdot \psi_{m, n}(t - n)) \rangle \) for the UDWT of \( s_{m, n}(t) \) and used the relation \( S_{m, n}^{[k], [l]} = S_{m, n}^{[k], [l]} \).

The test statistic \( \lambda(X, l) \) conditional means \( E[\lambda(X, l) | H_0] = 0 \) and \( E[\lambda(X, l) | H_0'] = \varepsilon_{k,l} A \) and conditional variances \( \text{var}(\lambda(X, l) | H_0) = \text{var}(\lambda(X, l) | H_0') = \sigma^2 \varepsilon_{k,l} \).

As a result, the performance index is \( d = (A \sqrt{E[\varepsilon_{k,l}]}) \), and the detector is shift invariant with respect to the delay parameter \( l \). The likelihood ratio that corresponds to this test statistic is
\[ r(X) = \frac{f(X^{[k], [l]} | H_1)}{f(X^{[k], [l]} | H_0)} \]
where the form of the ratio emphasizes the fact that the detector uses the known parameter values \( k \) and \( l \) to choose the DWT shift \( j_k + l \) that maximizes its performance.

C. Detection with Unknown Parameters

It can be shown that a straightforward extension of the shift-variant detector described in the previous section to the case of unknown parameters yields a GLRT detector that is also shift variant [7]. In order to obtain shift invariance, we extend (1) to the unknown parameter case and define a likelihood ratio \( r(\hat{X}) \) equal to
\[
\max_{\{A, k, l\}} \frac{f(X^{[k], [l]} | A, k, l, H_1)}{f(X^{[k], [l]} | H_0)} \prod_{m, n} \exp \left( -\frac{(X_{[0]^m, [k]^n})^2}{2 \sigma^2 \varepsilon_{k,l}} \right)
\]
where \( \hat{X} = \{\hat{X}_{m,n}, m, n \in \mathbb{Z}\} \) is the vector of UDWT observations and contains all possible DWT coefficients \( X_{m,n}, l \in \mathbb{Z} \).

Note that the DWT shift \( j_k + l \) is chosen to maximize the detector performance conditioned on the unknown parameters \( k \) and \( l \). With the substitution of the maximum likelihood estimate for \( A \), the ratio in (2) simplifies to yield a sufficient test statistic \( \lambda(\hat{X}) = \max_{\{A, k, l\}} (1/\varepsilon_{k,l}) \sum_{m, n} 2^{-\gamma m} X_{[0]^m, [k]^n} S_{m, n}^{[k], [l]} \).

To demonstrate shift invariance, we define \( g_{k,l} = 2^{-\gamma m} X_{[0]^m, [k]^n} S_{m, n}^{[k], [l]} \) such that \( \lambda(\hat{X}) = \max_{\{k,l\}} g_{k,l} \).

The test statistic \( \lambda(\hat{X}) \) is shift invariant if the set of random variables \( G = \{g_{k,l} : k, l \in \mathbb{Z}\} \) is invariant with respect to the assumed signal delay for hypothesis \( H_1 \). Consider two signals \( A_0 s_{k_0, l_0}(t) \) and \( A_0 s_{k_0, l_1}(t) \) with signal delays of \( l_0 \) and \( l_1 \), respectively. For delay \( l_0 \), \( g_{k,l_0} = (1/\sqrt{\varepsilon_{k,l}}) \sum_{m, n} 2^{-\gamma m} (A_0 S_{m, n}^{[k], [l_0]} + V_{m, n}^{[k], [l_0]}) S_{m, n}^{[k], [l_0]} \).

For delay \( l_1 \), the corresponding \( g_{k,l} = (1/\sqrt{\varepsilon_{k,l}}) \sum_{m, n} 2^{-\gamma m} (A_0 S_{m, n}^{[k], [l_1]} + V_{m, n}^{[k], [l_1]}) S_{m, n}^{[k], [l_1]} \).

The sets \( G = \{g_{k,l_0}\} \) and \( G' = \{g_{k,l_1}\} \) are identical if for each \( l \) there exists a unique \( l' \) such that \( g_{k,l} = g_{k,l'} \). The choice \( l' = l + l_0 - l_1 \) succeeds.

Recall that we have assumed \( A \neq 0 \) in our detection model. With the more restrictive assumption that \( A > 0 \), the absolute value operation may be omitted from the test statistic [7].

III. IMPLEMENTATION

Equations (3) and (4) represent two equivalent ways of computing the detection statistic. We implement (4), which can be viewed as a pattern-matching operation in the UDWT domain, where for each unknown scale the pattern is \( S_{m, n}^{[k], [l]} \). The pattern matching procedure in (4) requires \( O(2N \sum_{k,l} P_k) \) operations, whereas \( \mathbb{K} \) is the set of unknown scales, and \( P_k \) is the number of nonzero coefficients in \( S_{m, n}^{[k], [l]} \).

The contribution of most of these coefficients to the detection process is negligible, and we can reduce the number of required coefficients by ranking them according to their contribution to \( \varepsilon_{k,l} \) and selecting only the largest \( T_k \) coefficients. In the case where \( \varepsilon(t) \) is a C2 scaling function and \( \gamma = 1 \), the largest \( T_k = 20 \) coefficients account for 99.4% of the value of \( \varepsilon_{k,l} \).

We can further reduce the number of computations required by observing that the most significant coefficients of the pattern \( S_{m, n}^{[k], [l]} \) tend to be located near peaks in the UDWT domain. Fig. 2 shows an example for both the C2 scaling function and the Gaussian function. This result reflects the fact that the shift \( j_k \) was chosen to maximize \( \varepsilon_{k,l} \). In particular, we note that a number of the coefficients lie somewhere on the maximum peaks at each analysis scale \( m \). We expect, therefore, that “maxima ridges” in the UDWT domain, where the local maxima propagate across scales, are likely to correspond to scales and delays for which the detection statistic is maximized. Instead of computing \( \lambda(\hat{X}, l, k) \) for all possible values of \( k \) and \( l \), we constrain the computation to a subset of values that correspond to maxima ridges. If we constrain the search for transform ridges to analysis scales that
Fig. 3. ROC curves for methods a) (solid line, square), b) (dash-dot line, circle), c) (dashed line, triangle), and d) (dotted line, asterisk). (a) $k = 7$, (b) $k = 8$. Error bars are plus or minus one median absolute deviation. Transient model is a $C_2$ scaling function with $k = 7$ or 8, and $d = 5$. The ROC for detector with known parameters is shown in bold.

are much coarser than the finest analysis scales, e.g., $m > 5$, then the number of maxima that need to be considered is typically 1 to 2 orders of magnitude less than the length $N$ of the original signal. As a result, this preselection of locations can typically reduce the computational complexity of the pattern matching procedure by at least an order of magnitude.

The preceding discussion leads us to propose the following detection methods.

a) **Baseline GLRT Method**: Compute $\lambda(\hat{\mathbf{x}}, k, l)$ for all values of $k$ and $l$, and find the maximum.

b) **Preselection of Scales and Delays Using Transform Ridges**: Use a ridge-finding algorithm to generate estimates of both the scale and delay of the transient. The $g$th estimate is referred to as an ordered pair $(\hat{k}_g, \hat{l}_g)$. Compute the detection statistic $\lambda(\mathbf{X}) = \max_{k, l} \lambda(\mathbf{X}, \hat{k}_g, \hat{l}_g)$.

c) **Preselection of Delays Using Transform Ridges**: Use a ridge-finding algorithm to estimate the delays but not the scales. The detection statistic is $\lambda(\mathbf{X}) = \max_{k} \lambda(\mathbf{X}, \hat{k}_g, \hat{l}_g)$.

d) **Preselection of Delays Using Local Maxima at Selected Scales**: Use the positions of the local maxima over a range of analysis scales as delay estimates. The form of the detection statistic is the same as for Method C.

The ridge-finding algorithm is described in detail in [7], where it is also shown that the computational complexities of methods b), c), and d) are typically an order of magnitude less than that of method a). Variations of the proposed methods are also described in [7].

We used Monte Carlo simulations to obtain the receiver operating characteristics (ROC) of the detectors described above. Each simulation consisted of 200 independent trials. Using the method described in [8], we generated $1/f$ noise sample paths of length $N = 16384$ and with $\gamma = 1$. We employed three different transient signal models.

1) $C_2$ scaling function;

2) Gaussian function $g(t) = e^{-at^2}$;

3) two-sided exponential function $e(t) = \beta e^{-|t|}$.

The parameters $a, b, \alpha, \beta$ were chosen so that the temporal widths and parameter $E_s$ were equivalent for all signal models. For each signal type, $P_k = 20$. The performance index $d = (A \sqrt{E_s} / \sigma_w)$ is used as a measure of the signal to noise ratio of each simulation, with the caveat that, for a GLRT detector, $d$ no longer represents the normalized distance between the distributions under the two detection hypotheses. We computed the UDWT using a $C_2$ wavelet for analysis scales $m = 1$–10. For methods a), c) and d), the maximization over $k$ was performed over the set of unknown scales $k^* = \{6, 7, 8\}$.

Fig. 3 shows ROC curves for methods a)–d) with the following parameters:

- $C_2$ scaling function transient model;
- $k = 7$ and 8;
- $d = 5$.

Curves for other parameter values can be found in [7]. The performance loss of the GLRT detector [method a)] is due primarily to the large range of unknown delay values. The performance of methods c) and d) is equivalent to that of a) for all transient models and input scales, thus showing that maxima-tracking-based methods are effective for estimating the parameter values that maximize the proposed GLRT statistic. The performance of method b) is slightly worse for some cases, indicating errors in the estimation of the scale of the underlying transient from the scales of the transform maxima. The proposed detector methods have also been shown to be robust with respect to small errors in the assumed transient model and scale while providing good discrimination against transients with scales outside the desired detection range. Details are provided in [7].

**IV. CONCLUSION**

We have described a GLRT detector for transients in $1/f$ noise by making use of the approximate whitening properties of the DWT. The
Median-Based Cyclic Polyspectrum Estimation
Antonio Napolitano and Chad M. Spooner

Abstract—In this paper, median-based estimation methods for the cyclic polyspectrum are proposed. The algorithms do not require a priori knowledge of the $\beta$ submanifolds, that is, they do not require the knowledge of all of the lower order cycle frequencies of the time-series available for the estimation of the cyclic polyspectrum. Therefore, such methods are particularly useful when the cyclostationarity of the signals under consideration is not completely known. The proposed estimators converge to the theoretical values of the cyclic polyspectrum when the collect time approaches infinity and the spectral resolution becomes infinitesimal. Furthermore, their accuracy is very nearly the same as that of the usual time- and frequency-smoothed cyclic periodogram methods that use a priori knowledge of lower-order cycle frequencies to avoid the $\beta$ submanifolds.

Index Terms—Cyclic polyspectrum estimation, cyclostationarity, higher order statistics.

I. INTRODUCTION

In recent years, the theory of higher order cyclostationary signals has been developed in both the stochastic and fraction-of-time (FOT) probability frameworks [1], [3]–[5], [7], [9], [10]. Such signals, when processed by homogeneous nonlinear time-invariant transformations, regenerate spectral lines, whose frequencies are called cycle frequencies, which originate from hidden periodicities due to periodic signal processing operations such as modulation, sampling, coding, and multiplexing. The cycle frequencies are related to signal parameters such as carrier frequency, baud rate, sampling frequency, etc. Signal processing algorithms based on the regenerated spectral lines turn out to be signal selective in many cases and, hence, are very tolerant to noise and interference suitable for signal detection and classification purposes [9], [10].

A central statistical parameter in the study and exploitation of higher order cyclostationarity properties is the $N$th-order cyclic polyspectrum (CP) [3]. Estimators for the CP have been proposed in [10] and in [1] for continuous-time and discrete-time signals, respectively. In [10], the measurement problem has been formulated in the FOT probability framework in which the statistical functions, assuming they exist, are defined in terms of a single infinite-length time series and hence, estimators defined by using a single finite-length data record converge by definition to the theoretical values when the measurement collect time becomes infinite. Alternatively, in [1], the estimation problem has been addressed in the stochastic process framework, and mixing conditions involving the cumulant of the stochastic process have been presented to provide the convergence of the estimators in the statistical mean-square sense.

All the estimation methods presented in [1] and [10] require the knowledge of the $\beta$-submanifolds, which is equivalent to knowledge of all cycle frequencies of orders lower than that of the CP to be estimated. Such a priori knowledge is not always available. For example, when a signal of interest is embedded in additive interference and noise, even if the interference does not exhibit cyclostationarity at the considered order $N$, cycle frequency, and conjugation configuration [3], the $\beta$ submanifolds of the composite signal depend not only on the lower order cycle frequencies of the signal of interest but also on those of the interference.

New estimation algorithms for the CP are proposed in this paper. These methods are based on the use of the median as a measure of the average value of a set of quantities rather than the arithmetic mean, which is used in all previously proposed estimators. The application of the median to estimation of the CP is suggested by the results in [6], in which various robust measures of the average, including the median, are applied to estimation of the third-order cumulant of stationary time series with some success. In the present work, the median operation is applied to the $N$th-order cyclic periodogram or to a time- or frequency-smoothed version of it, and this application obviates the need for a priori knowledge of the $\beta$ submanifold of the time-series available for the estimation. Therefore, the methods are thought to be particularly useful for estimation scenarios in which a complete characterization of the signal of interest and/or of the interference is not available.

The signal analysis framework adopted in this correspondence is that of the FOT probability [2], [3]. The proposed algorithms work under the mild assumption that time-shifted versions of the time series are asymptotically independent (in the FOT probability sense) so that the cyclic polyspectrum is a well-behaved function (contains no Dirac delta functions). Moreover, the proposed estimators are shown to converge to their ideal CP values when the collect time approaches infinity and the spectral resolution becomes infinitesimal. Furthermore, they exhibit very nearly the same accuracy as the previously proposed time- and frequency-smoothed cyclic periodogram methods that use a priori knowledge of lower order cyclostationarity to avoid the $\beta$ submanifolds.

[Manuscript received September 1998; revised October 5, 1999. This work was supported in part by the Consiglio Nazionale delle Ricerche (CNR), Italy. The associate editor coordinating the review of this paper and approving it for publication was Prof. Dimitrios Hatzinakos. A. Napolitano is with the Dipartimento di Ingegneria Elettronica e delle Telecomunicazioni, Università di Napoli Federico II, Napoli, Italy (e-mail: antnapoli@unina.it). C. M. Spooner is with Mission Research Corporation, Monterey, CA, 93940-5776 USA (e-mail: cmspooner@mrcmry.com).]