Detection and Classification of Nonstationary Transient Signals Using Sparse Approximations and Bayesian Networks

Neil Wachowski, *Student Member, IEEE*, and Mahmood R. Azimi-Sadjadi, *Senior Member, IEEE*

***Abstract*—This paper considers sequential detection and classification of multiple transient signals from vector observations corrupted with additive noise and multiple types of structured interference. Sparse approximations of observations are found to facilitate computation of the likelihood of each signal model without relying on restrictive assumptions concerning the distribution of observations. Robustness to interference may be incorporated by virtue of the inherent separation capabilities of sparse coding. Each signal model is characterized by a Bayesian Network, which captures the temporal dependency structure among coefficients in successive sparse approximations under the associated hypothesis. Generalized likelihood ratios tests may then be used to perform signal detection and classification during quiescent periods, and quiescent detection whenever a signal is present. The results of applying the proposed method to a national park soundscape analysis problem demonstrate its practical utility for detecting and classifying real acoustical sources present in complex sonic environments.**

***Index Terms*—Multivariate analysis, signal classification, sparse representations, transient detection.**

I. Introduction

THE problem of detecting and classifying multiple nonstationary transient signals from sequential multivariate data, in the presence of variable interference and noise, is considered here. This process involves detecting the onset of each new signal event, estimating its duration, and assigning a class label as new multivariate observations arrive. These capabilities are useful for a wide variety of applications such as speech recognition [1], [2], habitat monitoring [3], [4], medical diagnosis [5], and soundscape characterization [6], [7] in National Parks, which is the application considered here. Therefore, it is crucial to develop a system that can achieve high performance even when multiple types of structured interference, which obstruct signal components, are simultaneously present, and when significant between-class similarities exist.

Manuscript received December 12, 2013; revised April 15, 2014; accepted August 06, 2014. Date of publication August 18, 2014; date of current version August 26, 2014. This work was supported under a cooperative agreement with the National Park Service. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Rongshan Yu.

The authors are with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523 USA (e-mail: nswachow@engr.colostate.edu; azimi@engr.colostate.edu).

Digital Object Identifier 10.1109/TASLP.2014.2348913

An extensive amount of research has already focused on transient detection [8], mainly estimating an unknown signal onset time from independent scalar observations, as in the classical Page's Test [9]. In [10], a version of Page's test that can operate on dependent observations was implemented using Hidden Markov Models (HMM). This method allows for less restrictive assumptions on the structure of the transient signal and noise. Unfortunately, many existing methods are unable to continually detect multiple transient signals from a sequential data stream of indeterminate length. Additionally, they are best suited to scalar observations, not designed to perform classification, nor consider the presence of structured interference.

Sparse representations [11] have recently seen widespread use for detection and/or classification from multivariate observations by using only a few atoms from an overcomplete dictionary to represent signals of interest [1], [12]–[14]. In [14], separate dictionaries are learned using K-SVD [15], that are capable of sparsely representing different types of audio signals, and a support vector machine is used to directly classify sparse coefficient vectors. A related approach is to use a dictionary that consists of training templates for different classes and assign a class label based on which sparse subset of templates provides the smallest reconstruction error. This approach was adopted in [12] for face recognition and extended to handle multiple observations in [13]. However, these methods process either a single observation or an ensemble of observations simultaneously, and hence, may be inadequate for continually detecting and classifying multiple signals using sequential data. This shortcoming may be addressed by modeling the dependencies between atom coefficients [16]–[19] extracted from different observations. For instance, [18] introduces an approach for detection of arbitrary transients in audio waveforms by modeling them as chains of atom coefficients in the time-scale domain with energy that monotonically decays over time. This coefficient modeling concept is central to the method proposed in this paper, though most existing work only applies to time series data, does not consider simultaneous detection and classification, and/or the presence of interference.

One approach that directly addresses the same problem as that in this paper was developed in [6]. This method assumes different types of source signatures lie in linearly independent subspaces and have random basis coefficients that obey vector linear autoregressive models. This model allows for estimating source signatures under various hypotheses involving the presence of different pairs of signal and interference sources. The

main issues with the approach in [6] are: (a) only one type of interference may be present at a time, which is impractical for some applications, and (b) the autoregressive model fails to capture novel variations in acoustical events, leading to less accurate estimates.

The sparse coefficient state tracking (SCST) method introduced in this paper draws from the concepts of inference in a sparse domain and modeling of sparse atom coefficients to yield a cohesive framework that is applicable to data containing signal, interference, and noise components that may be difficult to model using convenient distributions, e.g., multivariate Gaussian. To simplify the data representation, sparse coding and quantization are first applied to each incoming observation. This allows for using a Bayesian network (BN) [20] to model the temporal evolution of a given class of signal events. The likelihoods of BNs for different signal types and noise may then be used to form a set of cumulative test statistics for detection and classification of multiple transient signal events. The SCST method was designed to be applicable to many different types of sequential multivariate data. However, here we illustrate its performance on a soundscape characterization application [21], which involves determining soundscape compositions in terms of recurrent extrinsic sources (e.g., aircraft), that are often simultaneously present with competing intrinsic sources (e.g., weather effects). The results indicate excellent performance for detecting and classifying the extrinsic sources when compared to the approaches in [6]. Additionally, the proposed method prevents prohibitively time-consuming manual post observation and evaluation of large volumes of acoustical recordings, which is the current exercise.

This paper is organized as follows. Section II describes the problem formulation in the original data space, including the observation model and GLRTs used for signal detection and classification. Section III introduces the process for obtaining sparse coefficient state data representations, as well as the associated reformulations of the GLRTs, for practical application of the SCST method. Section IV presents the results of applying the SCST method to data sequences collected as part of National Park Service acoustical monitoring efforts. Finally, Section V provides concluding remarks.

II. Detection and Classification of Transient Events - Original Data Space

This section develops the underlying framework for implementing transient characterization in the original sequential multivariate data space. We first present the observation model and the basic mechanics used for detecting transient signals. The general forms of the tests required for detection and classification are then provided.

A. Observation Model and Detection and Classification Hypotheses

Let $\mathbf{Y}_1^n = \{\mathbf{y}_k\}_{k=1}^n, n = 1, 2, \ldots$ be the observation sequence recorded as of the current time n, where $\mathbf{y}_k \in \mathbb{R}^N$ is the observation at time k. Data arrives continually, meaning n is increasing. Detecting and classifying multiple transient signals requires two distinct phases: 1) signal detection to look for the presence of a signal while it is assumed that none are present, and 2) quiescent detection to look for observations that contain no signal while it is assumed that one is present. The idea is to alternate between these two phases as new $\mathbf{y}_n$'s arrive, while performing classification by exploiting all available information within a given detected signal event. To facilitate the classification framework (discussed at the end of Section II-B) it is assumed that a maximum of one transient signal can be present in a given $\mathbf{y}_n$, meaning the signatures of two transients will never be superimposed. Furthermore, it is assumed that a quiescent period will always separate any given pair of signal events. If analyzing data that contains multiple overlapping signal events, typically only a single event will be detected, thereby missing the others.

Since signals are continually detected and classified, it is helpful to adopt notation associated with the onset of various detection periods, relative to the current time n. Let k_0 and k_1 denote the unknown onset times of the next quiescent and signal periods, respectively, and let $\hat{k}_0$ and $\hat{k}_1$ denote the estimated (known) onset times for the most recently detected quiescent and signal periods, respectively. Fig. 1 demonstrates the two-phase concept using a 1/3 octave [22] observation sequence (see Section IV), which is a type of time-frequency representation, and in this case includes the acoustical signatures of two propeller planes. This figure shows the circumstances for implementing each phase, as well as the most recent estimated onset times relative to the current time n. The test statistics for each phase are also displayed, which are discussed in Section II-B.

When the data has been in a quiescent period since time $\hat{k}_0$, signal detection and classification are performed on each $\mathbf{y}_n$ according to the following multiple hypotheses test

$$\mathcal{H}_0 : \mathbf{y}_k = \beta_k \sum_{q=1}^{Q} \mathbf{h}_k^{(q)} + \mathbf{w}_k, \hat{k}_0 \le k \le n$$

$$\mathcal{H}_1^{(p)} : \mathbf{y}_k = \begin{cases} \beta_k \sum\limits_{q=1}^{Q} \mathbf{h}_k^{(q)} + \mathbf{w}_k, & \hat{k}_0 \le k < k_1 \\ \mathbf{s}_k^{(p)} + \beta_k \sum\limits_{q=1}^{Q} \mathbf{h}_k^{(q)} + \mathbf{w}_k, & k_1 \le k \le n \end{cases} \quad (1)$$

where $\mathbf{s}_k^{(p)}$ is a random class $p \in [1, P]$ signal vector, $\mathbf{h}_k^{(q)}$ is a random class $q \in [1, q]$ interference vector, β_k is a binary variable indicating the presence ($\beta_k = 1$) or absence ($\beta_k = 0$) of interference, and $\mathbf{w}_k$ is an independent and identically distributed (IID) ambient noise vector with $E[\mathbf{w}_k] = \mathbf{0}$. The null hypothesis $\mathcal{H}_0$ indicates signal components have been absent since time $\hat{k}_0$, while under the alternative hypothesis $\mathcal{H}_1^{(p)}$ the onset of signal components $\mathbf{s}_k^{(p)}, k \in [k_1, n]$ occurs at the unknown time k_1. The goal is then to find the estimate $\hat{k}_1$, as well as the class of the newly onset signal. The summation over $\mathbf{h}_k^{(q)}$'s indicates that multiple types of interference may be simultaneously present, where $\mathbf{h}_k^{(q)} = \mathbf{0}$ if class q interference is absent from $\mathbf{y}_k$. Interference differs from ambient noise in several ways, namely it is 1) typically not IID or zero mean, 2) associated with a specific set of sources that are usually not of interest, and 3) not necessarily always present.

Since transient signals have finite extent, the next quiescent period must be detected before the process of detecting the next

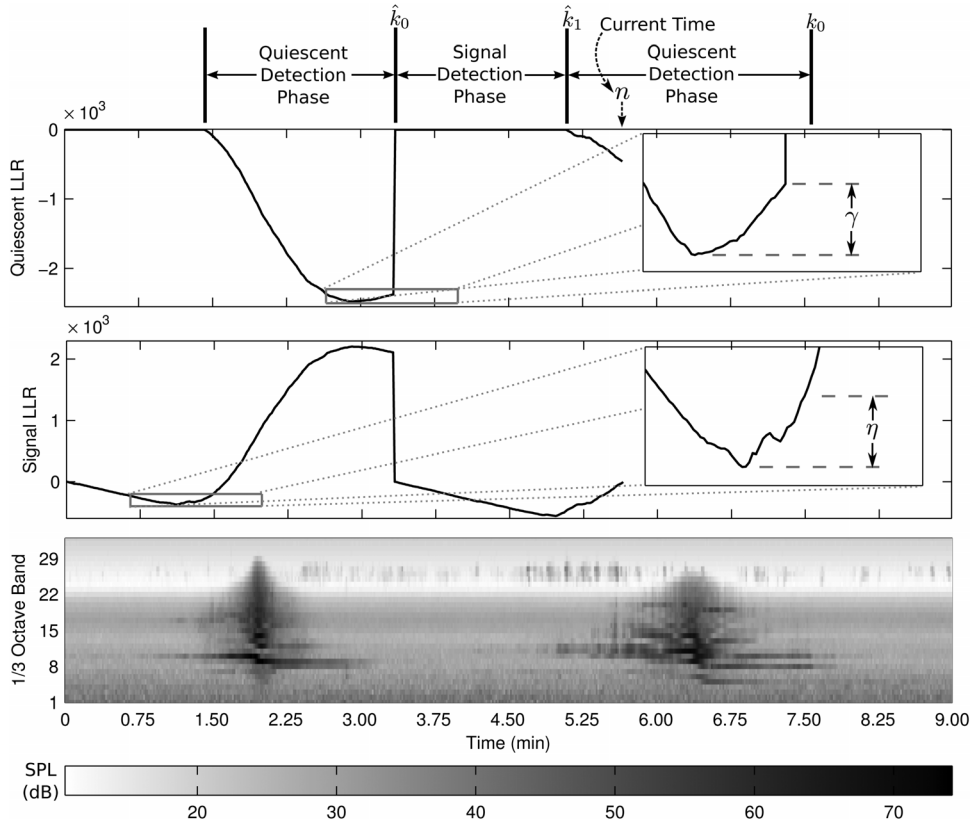


Fig. 1. Illustration of the two phase detection approach, where the durations of several phases are shown above a 1/3 octave observation sequence (bottom), and the corresponding test statistics used to detect signal (middle) and quiescent (top) periods. The times where the onset of a quiescent period and a signal event were last detected, denoted $\hat{k}_0$ and $\hat{k}_1$, respectively, are shown relative to the current time n. A signal and quiescent period are detected when their associated LLRs increase by at least η and γ, respectively.

signal can begin. Therefore, when a signal has been present since time $\hat{k}_1$, the following hypothesis test is used in place of (1) to perform quiescent detection

$$\mathcal{H}_1^{(p)} : \mathbf{y}_k = \mathbf{s}_k^{(p)} + \beta_k \sum_{q=1}^{Q} \mathbf{h}_k^{(q)} + \mathbf{w}_k, \hat{k}_1 \leq k \leq n$$

$$\mathcal{H}_0 : \mathbf{y}_k = \begin{cases} \mathbf{s}_k^{(p)} + \beta_k \sum\limits_{q=1}^{Q} \mathbf{h}_k^{(q)} + \mathbf{w}_k, & \hat{k}_1 \leq k < k_0 \\ \beta_k \sum\limits_{q=1}^{Q} \mathbf{h}_k^{(q)} + \mathbf{w}_k, & k_0 \leq k \leq n \end{cases} \quad (2)$$

i.e., $\mathbf{s}_k^{(p)}$'s cease to be extant at the unknown time k_0 under $\mathcal{H}_0$. In summary, signal and quiescent detection are performed when $\mathbf{y}_{n-1} \in \mathcal{H}_0$ and $\mathbf{y}_{n-1} \in \mathcal{H}_1^{(p)}$, respectively.

B. GLRTs for Hypothesis Testing

1) Signal Detection: Throughout the remainder of this section $\beta_k = 0, \forall k$ in (1) and (2), i.e., interference is not considered. This stems from the fact that the SCST method addresses interference through the use of an alternate data representation, which is presented in Section III. To implement the hypothesis test in (1) consider the log-likelihood ratio (LLR) for the general null and alternative hypothesis parameter sets, denoted by λ_0 and λ_p, respectively, given the data $\mathbf{Y}_{\hat{k}_0}^n$

$$L_{\hat{k}_0}^n(\lambda_p, k_1) = \ln\left(\frac{\ell\left(\lambda_p, k_1; \mathbf{Y}_{\hat{k}_0}^n\right)}{\ell\left(\lambda_0, k_1; \mathbf{Y}_{\hat{k}_0}^n\right)}\right) = \ln\left(\frac{f_{\lambda_0}\left(\mathbf{Y}_{\hat{k}_0}^{k_1-1}\right) f_{\lambda_p}\left(\mathbf{Y}_{k_1}^n\right)}{f_{\lambda_0}\left(\mathbf{Y}_{\hat{k}_0}^{k_1-1}\right) f_{\lambda_0}\left(\mathbf{Y}_{k_1}^n\right)}\right) = \ln\left(\frac{f_{\lambda_p}\left(\mathbf{Y}_{k_1}^n\right)}{f_{\lambda_0}\left(\mathbf{Y}_{k_1}^n\right)}\right) \quad (3)$$

where $f_\lambda(\mathbf{Y}_{\hat{k}_0}^n)$ is a general probability distribution modeled by the parameter set $\lambda \in \{\lambda_0, \lambda_p\}$. This LLR is a function of two unknowns, namely the change time k_1 and the signal parameter set λ_p. The second equality in (3) follows from independence of $\mathbf{y}_n$'s before and after the unknown change time k_1 under all hypotheses. This was the motivation for setting $\beta_k = 0$ in this section, namely since interference is typically not IID, thus invalidating the second equality when $\beta_k = 1$.

To implement (1), consider the GLRT for change detection with an unknown signal parameter set after the hypothesis change [8]

$$\max_{\hat{k}_0 \leq k \leq n} \max_p L_{\hat{k}_0}^n(\lambda_p, k) \begin{array}{l} \mathbf{y}_n \notin \mathcal{H}_0 \\ \geq \\ < \\ \mathbf{y}_n \in \mathcal{H}_0 \end{array} \eta \quad (4)$$

where $\eta > 0$ is a predetermined signal detection threshold. Double maximization makes this test generalized and states that a signal is detected when any $L^n_{\hat{k}_0}(\lambda_p, \hat{k}_0)$ (i.e., for any $p \in [1, p]$) increases by at least η from its lowest point [8], and the earliest time this level of increase is witnessed marks the estimated signal onset time $\hat{k}_1$. This concept is illustrated by the plot of the signal detection statistic in the middle of Fig. 1, which shows $L^n_{\hat{k}_0}(\lambda_p, k)$, for one λ_p (associated with the plane signal type), increasing as new observations containing signatures that fit this model arrive, but decreasing otherwise.

2) Quiescent Detection: Recall that a complete solution must account for the inevitably that a detected signal will cease to be extant. This process is simplified by the previously stated assumption that immediate switching from one signal class to another will not be encountered. This involves the test in (2), which uses the LLR

$$F^n_{\hat{k}_1}(\lambda_{p^*}, k_0) = \ln\left(\frac{\ell\left(\lambda_0, k_0; \mathbf{Y}^n_{\hat{k}_1}\right)}{\ell\left(\lambda_{p^*}, k_0; \mathbf{Y}^n_{\hat{k}_1}\right)}\right) = \ln\left(\frac{f_{\lambda_0}\left(\mathbf{Y}^n_{k_0}\right)}{f_{\lambda_{p^*}}\left(\mathbf{Y}^n_{k_0}|\mathbf{Y}^{k_0-1}_{\hat{k}_1}\right)}\right) \quad (5)$$

where

$$p^* = \arg\max_p \max_{\hat{k}_0 \le k \le n} L^n_{\hat{k}_0}(\lambda_p, k) \quad (6)$$

is the maximum likelihood (ML) signal model at time n, i.e., λ_{p^*} satisfies (4). This means that quiescent detection is performed relative to the most likely signal class, though classification is only performed when the most likely signal is no longer extant.

The test used to implement (2) is

$$\max_{\hat{k}_1 \le k \le n} F^n_{\hat{k}_1}(\lambda_{p^*}, k) \begin{array}{c} \mathbf{y}_n \in \mathcal{H}_0 \\ \geq \\ < \\ \mathbf{y}_n \notin \mathcal{H}_0 \end{array} \gamma \quad (7)$$

where $\gamma > 0$ is a predetermined quiescent detection threshold and maximization is performed with respect to the unknown onset time of the next quiescent period k_0. Equation (7) states that $\mathcal{H}_0$ is again accepted for samples starting at time n (i.e. $\hat{k}_0 = n$) if $F^n_{\hat{k}_1}(\lambda_{p^*}, \hat{k}_1)$ has increased by at least γ at this time. This concept is illustrated by the top plot in Fig. 1, which shows the quiescent LLR decreasing when a signal is present, but increasing during times leading up to detection of a quiescent period, where signal components are absent. Note that this LLR is zero during signal detection phases since it is not used during these times.

3) Signal Classification: The class label assigned to $\mathbf{y}_n$ is denoted by $c_n \in [0, P]$, where $c_n = 0$ means $\mathbf{y}_n \in \mathcal{H}_0$ and $c_n = p$ means $\mathbf{y}_n \in \mathcal{H}^{(p)}_1$. Event-wide classification is performed, meaning a unified label is assigned to the set of observations associated with the most recently detected event only after using (7) to again accept $\mathcal{H}_0$, i.e., end of extant. The reason being that, due to the random and time-varying nature of signals, some events associated with different signal types may appear similar for subsets of their observations. Therefore, more accurate labels are assigned when taking into account the likelihood of each signal model over the course of all observations associated with an event (signatures of one signal). The assigned label p^* corresponds to the ML model parameter set λ_{p^*} (as in (6)) at time $\hat{k}_0 - 1$, i.e., the time step immediately preceding the start of the newly detected quiescent period. More formally $\{c_k\}^{\hat{k}_0-1}_{k=\hat{k}_1} = p^*$.

C. Practical Considerations

Calculating the likelihoods $\ell(\lambda_p, k; \mathbf{Y}^n_{\hat{k}_0}), \forall p$, for use in (3) and (5), requires knowledge of the probability distributions $f_{\lambda_p}(\mathbf{Y}^n_{\hat{k}_0})$, parameterized by λ_p, which is generally infeasible without assuming independence of $\mathbf{y}_n$'s under each $\mathcal{H}^{(p)}_1$. In the absence of interference, it is possible to fit, e.g., an HMM to $\mathbf{y}_n$'s under $\mathcal{H}^{(p)}_1$, and use an approach similar to that in [10] for detecting and classifying signals. However, as mentioned before, the intermittent presence of multiple types of interference in our soundscape characterization problem leads to difficulties when using HMMs. In particular, using a separate HMM for each unique combination of signal and interference would lead to an abundance of models, and frequent switching between these models even when the signal type does not change. Alternatively, interference could be incorporated into each signal HMM, but this often results in extreme variations in the data, which are too difficult to model using a set of multivariate probability distributions. In the next section, we use sparse coding to simplify the modeling of temporal dependencies between $\mathbf{y}_n$'s, as well as to remove the effects of interference as much as possible. This allows for efficient likelihood calculation without making extensive assumptions about the structure of signals to be detected.

III. Detection and Classification of Transient Events - Sparse Coefficient State Space

The SCST method is implemented according to the block diagram in Fig. 2. As can be seen, each incoming data vector $\mathbf{y}_n$ is first transformed to $\mathbf{z}_n$ using sparse coding and coefficient quantization, in order to simplify the relationships between observations and their distributions, respectively, as discussed in Sections III-A and III-B. These steps provide a realistic and flexible means for calculating the likelihoods of λ's given representative data, which may then be updated as detailed in Section III-D. In this section $\beta_k = 1$ in (1) and (2), meaning multiple types of interference may be present at any time. Robustness to this interference is inherently handled during the sparse coding stage, as the signal and interference components of the observation can be mostly separated and associated with different atoms in the dictionary. The process then proceeds as in Section II, where LLRs are used to perform signal and quiescent detection, though here $\mathbf{z}_n$'s are used in place of $\mathbf{y}_n$'s. Note that, $\mathbf{y}_n$'s are typically raw data vectors, e.g., 1/3 octave vectors for the soundscape monitoring application [21], or Mel-frequency cepstral coefficients [2] for speech recognition. The reason is that $\mathbf{z}_n$'s can be viewed as a type of feature vector extracted specifically for ease of modeling using a BN.

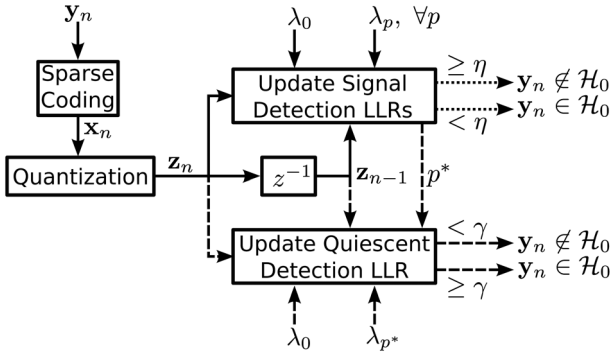


Fig. 2. Block diagram of the proposed signal detection and classification framework. The dashed and dotted lines indicate that the connected processes are only executed during the quiescent detection (when $\mathbf{y}_{n-1} \in \mathcal{H}_1^{(p)}$) and signal detection (when $\mathbf{y}_{n-1} \in \mathcal{H}_0$) phases, respectively.

A. Sparse Coding

To simplify the dependencies between consecutive observation vectors and make the structure of nonstationary events more tractable, the SCST method first finds a sparse approximation of a newly arrived $\mathbf{y}_n$, denoted by $\mathbf{x}_n = [x_{1,n} \cdots x_{i,n} \cdots x_{M,n}]^T \in \mathbb{R}^M$ and using one of several existing sparse coding methods [11], [23], [24]. An underlying assumption is that any $\mathbf{s}_n^{(p)}$ or $\mathbf{h}_n^{(q)}$ contained in $\mathbf{y}_n$ admits a sparse representation over some rank N dictionary matrix $\mathbf{A} = [\mathbf{A}_s, \mathbf{A}_h] \in \mathbb{R}^{N \times M}$, with $N \ll M$ and normalized columns (atoms). Furthermore, it is assumed that the atoms typically used to provide sparse representations of $\mathbf{s}_n^{(p)}$'s are mostly disjoint from those used to represent $\mathbf{h}_n^{(q)}$'s, due to the separability assumption. This implies that signal and interference components can be represented in terms of two dictionaries, i.e., $\mathbf{A}_s$ and $\mathbf{A}_h$, respectively, that are relatively incoherent [25]. Note that some overlap between the atoms used to represent these two components is inevitable in many practical cases, but reasonably small overlap will typically not degrade performance (see Section IV-E). Further details concerning the recommended structure for $\mathbf{A}$ are discussed below. Apart from signal and interference separability, the merit of using $\mathbf{x}_n$'s is that they will contain many coefficients close or equal to zero. This means $x_{i,n}$ will be dependent on a relatively small set of other $x_{i',n-j}$ with time lag $j \geq 0$, and hence, the temporal evolution of the sequence $\mathbf{X}_1^n = \{\mathbf{x}_k\}_{k=1}^n$ will be easier to model and track than that of the original data sequence $\mathbf{Y}_1^n$.

To generate $\mathbf{x}_n$, consider the underdetermined linear system $\mathbf{A}\mathbf{v} = \mathbf{y}_n$, which has infinitely many solutions $\mathbf{v}$, meaning constraints are required to find a unique solution. Since sparsity is desired and observations are noisy, an intuitive approach is to find $\mathbf{x}_n$ using the following optimization problem [11]

$$\mathbf{x}_n = \min_{\mathbf{v}} \|\mathbf{v}\|_0 \text{ subject to} \|\mathbf{y}_n - \mathbf{A}\mathbf{v}\|_2 \leq \delta \tag{8}$$

where $\delta \geq 0$ is an error tolerance proportional to the noise energy $\|\mathbf{w}_n\|_2$ [11], and $\|\cdot\|_0$ is the ℓ_0-norm. The motivation for permitting a discrepancy of δ between $\mathbf{y}_n$ and $\mathbf{A}\mathbf{v}$ is to extract a $\mathbf{x}_n$ that contains fewer components representing $\mathbf{w}_n$ when compared to the case of using an equality constraint. Since (8) is NP-hard due to non-convexity of the ℓ_0-norm [11], approximate solutions to (8) are required. The SCST method is flexible in that any pursuit method can be used to obtain $\mathbf{x}_n$. Common choices are matching pursuit algorithms [23] (greedy approaches), and basis pursuit algorithms [11], [24], which transform (8) into a convex problem by replacing the ℓ_0-norm with the ℓ_1-norm. Therefore, a proper value of δ should be selected based on criteria established for the chosen sparse coding algorithm.

To obtain consistently sparse $\mathbf{x}_n$'s and maximize signal discrimination, $\mathbf{A}$ must be intelligently designed relative to any signal and interference vectors that may be observed. In this paper we use

$$\mathbf{A} = [\mathbf{S}_1 \cdots \mathbf{S}_P \mathbf{H}_1 \cdots \mathbf{H}_Q] \tag{9}$$

where $\mathbf{S}_p$ and $\mathbf{H}_q$ are dedicated subdictionaries capable of providing sparse representations of class p signal vectors $\mathbf{s}_n^{(p)}$'s and class q interference vectors $\mathbf{h}_n^{(q)}$'s, respectively. Such subdictionaries may be extracted by applying, e.g., K-SVD [15] or any other sparse dictionary learning algorithm (e.g., [26], [27]) to a training data set in the original data space containing the associated signal/interference types. Note that generic dictionaries (e.g., wavelets) may also be used to represent a broad category of signal and interference types. Without loss of generality, it is assumed that the first M_s and last M_h columns of $\mathbf{A}$ are associated with the composite signal and interference dictionaries, i.e., $\mathbf{A}_s = [\mathbf{S}_1 \cdots \mathbf{S}_P] \in \mathbb{R}^{N \times M_s}$ and $\mathbf{A}_h = [\mathbf{H}_1 \cdots \mathbf{H}_Q] \in \mathbb{R}^{N \times M_h}$, respectively, with $M = M_s + M_h$. Consequently, the sparse coding process incorporates robustness to interference by encoding the majority of signal and interference energy using the first M_s and last M_h coefficients in $\mathbf{x}_n$, respectively. Likelihoods used for signal detection may then be based only on the M_s signal coefficients, while the M_h interference coefficients are ignored. As will be explained in Section III-D, the separation between these components need not be perfect [25], as the learned signal models can account for the fact that some signal energy will be present in $\{x_{M_s+i,n}\}_{i=1}^{M_h}$, while all learned models can account for the fact that some interference energy will be present in $\{x_{i,n}\}_{i=1}^{M_s}$. On the other hand, encoding most of the interference energy in $\{x_{M_s+i,n}\}_{i=1}^{M_h}$ allows for improved discrimination between different signal types and noise by discarding information that is a nuisance to detection and classification.

B. Quantization of Sparse Coefficients

Just as $\mathbf{x}_n$ is extracted from $\mathbf{y}_n$ to simplify the dependencies between consecutive observation vectors, the *sparse coefficient state vector* $\mathbf{z}_n = [z_{1,n} \cdots z_{i,n} \cdots z_{M_s,n}]^T \in \mathbb{R}^{M_s}$ is in turn generated by quantizing the coefficients in $\mathbf{x}_n$ corresponding to signal atoms. Instead of assuming $x_{i,n}$'s obey a convenient but unlikely distribution (e.g., Gaussian [16]), quantization ensures they may be parameterized in a simple but accurate manner using a collection of categorical (i.e., L-level discrete) distributions, while still retaining sufficient information for signal detection and classification. More explicitly, "sparse coefficient states" are obtained as

$$z_{i,n} = \begin{cases} 0, & \|x_{i,n}\| \leq \epsilon \\ H(x_{i,n}), & \text{otherwise} \end{cases}, \quad i \in [1, M_s] \tag{10}$$

where

$$H(x) = l \text{ if } x \in (t_{l-1}, t_l],\ l \in [1, L-1] \tag{11}$$

is a $(L-1)$-level quantization function (the zero-state represents the remaining level) dependent on the distribution of x under different hypotheses (defined below), and ϵ is a predetermined threshold used to determine those coefficients that are inactive ($x_{i,n} \approx 0$). The purpose of ϵ is to give coefficients close or equal to zero their own state in order to exploit the sparsity of $\mathbf{X}_1^n$ and simplify parameterization, as an overwhelming percentage of $x_{i,n}$'s will be near zero if the matrix $\mathbf{A}$ is appropriately designed. Setting ϵ too low can lead to $\mathbf{z}_n$'s that lack sparsity if $\mathbf{y}_n$'s contain noise and an error tolerant version of (8) is used, while setting ϵ too high can lead to discarding important discriminatory features. Practically speaking, a suitable value of ϵ can be the one that produces $z_{i,n} = 0$ for some large (SNR dependent) percentage of $x_{i,n}$'s extracted from observations in the training set containing noise alone.

The quantization function $H(x)$ is characterized by transition levels $\mathbf{t} = [t_0, t_1, \ldots, t_{L-1}]$, with $-\infty = t_0 < t_1 < \ldots < t_{L-1} = \infty$ and $L \geq 2$, and uses reconstruction levels $\mathbf{r} = [1, \ldots, L-1]$, though the latter is chosen for simplicity as the actual values used for reconstruction are irrelevant to detection and classification performance [28]. Clearly, smaller L leads to simpler parameterization of the data but a greater loss of discriminatory information. In general, L should be set as large as possible while avoiding an abundance of sample-poor cases when forming categorical distributions (used in Section III-D) representing $z_{i,n}$'s from training data. In other words, since the true probability distributions for $z_{i,n}$'s are rarely if ever available, quantization may be viewed as a necessary step for dealing with realities of limited training data in real-world applications, while refraining from making assumptions about these distributions. Note that, when $L = 2$, no quantizer is used and $z_{i,n} \in \{0, 1\}$.

On the other hand, it is important to ensure that $z_{i,n}$'s contain as much information useful for signal detection and classification as possible for a given L. To this end, the maximum J-divergence quantizer [29] is used that, in the case of multiple hypotheses, specifies $\mathbf{t}$ to maximize the sum of the pairwise divergences between sets of distributions corresponding to $z_{i,n}$'s representing different classes of signals. The importance of J-divergence is largely attributed to results [30], [31] linking a maximum of this measure to minimum error probabilities when discriminating between two hypotheses, i.e., bounds on the latter can be expressed in terms of the former [28]. In general, SCST discriminates between different hypotheses by finding the likelihood of a given *pattern* of sparse coefficient states. However, the goal of quantization is to use a single function to generate states with marginal distributions that are optimal (in the sense of J-divergence) for this signal discrimination.

Define $X_i^{(p)}$ and $X_i^{(0)}$ as random variables representing atom coefficients under $\mathcal{H}_1^{(p)}$ and $\mathcal{H}_0$, respectively, with realizations that are sparse coefficients $x_{i,n}$'s. The quantizer function $H(\cdot)$ in (11) is characterized by the transition vector $\mathbf{t}$ that maximizes [29]

$$D(\mathbf{t}) = \sum_{i=1}^{M} \sum_{\substack{p,p'=0 \\ p \neq p'}}^{P} d_i^{(p,p')}(\mathbf{t}) \tag{12}$$

where

$$d_i^{(p,p')}(\mathbf{t}) = \sum_{l=1}^{L-1} \left(r_{i,l}^{(p')}(\mathbf{t}) - r_{i,l}^{(p)}(\mathbf{t}) \right) \ln \left(\frac{r_{i,l}^{(p')}(\mathbf{t})}{r_{i,l}^{(p)}(\mathbf{t})} \right)$$

is the J-divergence between two distributions of quantized coefficients belonging to different classes, p and p', and

$$r_{i,l}^{(p)}(\mathbf{t}) = \Pr\left[t_{l-1} < X_i^{(p)} \leq t_l \right] = \int_{t_{l-1}}^{t_l} f_{X_i^{(p)}}(x) dx \tag{13}$$

is the probability that $X_i^{(p)}$, with probability density function $f_{X_i^{(p)}}(x)$, lies in the interval $(t_{l-1}, t_l]$. As can be seen, $D(\mathbf{t})$ is the sum of the distances between distributions for each quantized atom coefficient under each pair of hypotheses. The use of separate distributions for each coefficient in (12) is unique to this work and is done to exploit the fact that different signals often have sparse representations in terms of different atoms, especially for dictionaries constructed as in (9). Consequently, each J-divergence term $d_i^{(p,p')}(\mathbf{t})$ is typically larger when using separate distributions for each coefficient and class, rather than just one distribution for each class, leading to a quantizer that generates $z_{i,n}$'s with superior class discrimination.

C. Illustrative Example

To illustrate the steps used by the SCST method to extract $\mathbf{z}_n$'s, an alternative application is considered where the goal is to distinguish different animal vocalizations. The audio waveforms for these (swine and horse) vocalizations were concatenated and superimposed with wind interference to create a realistic data sequence. Fig. 3(a) shows a coarse spectrogram (large frequency bins) that was extracted from the resulting time series in order to create a vector sequence $\mathbf{Y}_1^n$. The time segments where swine and horse vocalizations are present are noted at the top of Fig. 3, while the superimposed wind is responsible for a large portion of the intense signatures in the lower two frequency subbands.

Fig. 3(b) shows the sparse representation $\mathbf{X}_1^n$ of the spectrogram in Fig. 3(a). The first and second sets of five atoms in the dictionary represent horse and swine signatures, respectively, while the last two atoms represent wind signatures. As can be seen, the coefficients for the wind atoms have intermittently high values for the entire data sequence, while the coefficients for swine and horse atoms typically only have high energy when their corresponding signatures are present. Fig. 3(c) shows the sparse coefficient state sequence $\mathbf{Z}_1^n$ extracted from $\mathbf{X}_1^n$ in Fig. 3(b), where $L = 2$ was used for the quantizer

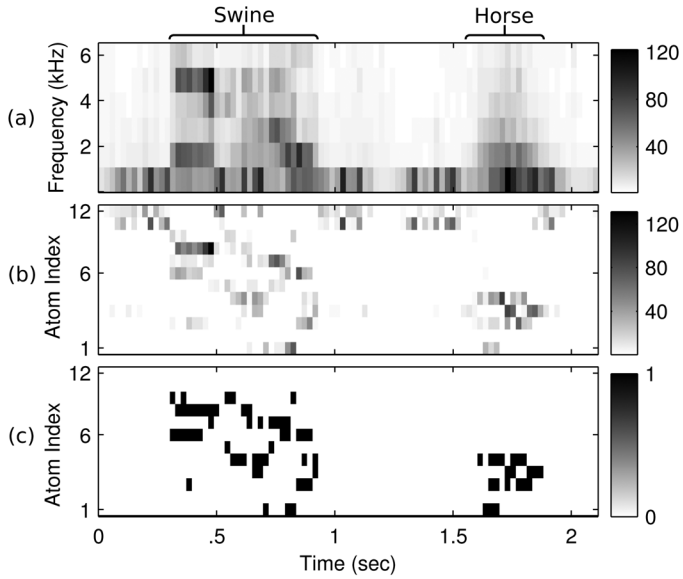


Fig. 3. SCST data transformation applied to a spectrogram representing swine and horse vocalizations superimposed with wind interference. (a) Original data sequence $\mathbf{Y}_1^n$. (b) Sparse representation $\mathbf{X}_1^n$. (c) Sparse coefficient state representation $\mathbf{Z}_1^n$ with $L = 2$.

in Section III-B for simplicity. Since the coefficients associated with interference were ignored (set to zero in this figure), nonzero coefficient states are only witnessed when animal vocalizations are present, thus providing robustness to the wind signatures. Furthermore, the coefficient state sequences for the horse and swine signatures are much different, both in terms of the active coefficients and their temporal evolutions. The state patterns for each signal type can be modeled and exploited for detection and classification, as discussed next.

D. Probability of Coefficient State Sequences

In this subsection, we will show how the proposed coefficient state representation facilitates realistic formation of the hypothesis tests, even for long data sequences of high dimension. More specifically, we explicitly define the probability $f_\lambda(\mathbf{Z}_{k_1}^n)$ of the sparse coefficient state sequence $\mathbf{Z}_{k_1}^n = \{\mathbf{z}_k\}_{k=k_1}^n$ (or $\mathbf{Z}_{k_0}^n$), given the parameter set $\lambda \in \{\lambda_p, \lambda_0\}$, used to form LLRs equivalent to those in (3) and (5). Each model parameter set defines a BN [20], denoted by $\lambda_p = \{G_p, \Theta_p\}$. Here, G_p is a *directed acyclic graph* [20], [32] with nodes $Z_{i,k-j}^{(p)}, i \in [1, M_s]$, that are categorical random variables with time delay $j \in \{0, 1\}$ and corresponding realizations that are sparse coefficient states $z_{i,k-j} \in [0, L-1]$ from the quantizer output as in (10). Edges in G_p describe the dependencies between $Z_{i,k-j}^{(p)}$'s, i.e., the "parents" of each coefficient state. The parameters of the conditional distributions associated with the random variables $Z_{i,k-j}^{(p)}$'s are elements of the set Θ_p, and are described in more detail below. A BN allows for efficiently calculating a complicated joint probability $f_{\lambda_p}(\mathbf{Z}_{k_1}^n)$ (in place of those used in (3) and (5)) by decomposing it into a product of conditional probabilities of $z_{i,k}$'s given other dependent states, which is much simpler to evaluate in practice than $f_{\lambda_p}(\mathbf{Y}_{k_1}^n)$. BNs are appropriate for transient detection as the graph G_p is well-suited for describing causal temporal relationships owing to its directed structure and the sequential processing of nodes that occurs as a result [20].

We first show how to decompose the probability distribution used to form the numerator of the SCST test statistic equivalent to that in (3), i.e.,

$$f_{\lambda_p}\left(\mathbf{Z}_{k_1}^n\right) = f_{\lambda_p}\left(\mathbf{z}_{k_1}\right) \prod_{k=k_1+1}^{n} f_{\lambda_p}\left(\mathbf{z}_k | \mathbf{Z}_{k_1}^{k-1}\right) \tag{14}$$

where $f_{\lambda_p}(\mathbf{z}_{k_1})$ is the prior probability of $\mathbf{z}_{k_1}$ under $\mathcal{H}_1^{(p)}$. Assuming G_p imposes a first order dependency structure, $\mathbf{z}_k$ is only dependent on $\mathbf{z}_{k-1}$, meaning

$$\begin{aligned} f_{\lambda_p}\left(\mathbf{z}_k | \mathbf{Z}_{k_1}^{k-1}\right) &= f_{\lambda_p}\left(\mathbf{z}_k | \mathbf{z}_{k-1}\right) \\ &= f_{\lambda_p}\left(z_{1,k} | \mathbf{z}_{k-1}\right) \prod_{i=2}^{M_s} f_{\lambda_p}\left(z_{i,k} | \{z_{i',k}\}_{i'=1}^{i-1}, \mathbf{z}_{k-1}\right) \end{aligned} \tag{15}$$

where the second equality is a result of using the chain rule to decompose $f_{\lambda_p}(\mathbf{z}_k|\mathbf{z}_{k-1})$ into a product of conditional probabilities of $z_{i,k}$ given $\mathbf{z}_{k-1}$ and previous elements in $\mathbf{z}_k$. Note that the first order assumption in (15) is used for simplicity of derivations, and may be dropped if it is invalid for a particular application.

We now exploit the fact that the dependency structure of $z_{i,k}$'s can be simplified according to the BN λ_p being evaluated. More specifically, owing to sparsity of $\mathbf{z}_k$'s, many $z_{i',k-j}, i' \neq i$ associated with class $p' \neq p$ signal atoms will be zero when a type p signal is present, meaning the corresponding random variables $Z_{i',k-j}^{(p)}$'s for λ_p carry little to no information about $Z_{i,k}^{(p)}$'s associated with atoms in the same-class subdictionary $\mathbf{S}_p$. While sparsity is the predominant factor that enables a simplified dependency structure, there may be other application-specific attributes that allow for independence of $Z_{i,k}^{(p)}$'s. For instance, for the data in Section IV, certain broadband components of plane signatures are typically only present after the onset of specific narrowband mid-frequency components, meaning a node representing the former may be considered conditionally independent of other nodes besides that representing the latter. Regardless, the idea is to measure the dependence between pairs of coefficients during training to determine the edges that connect nodes in a given G_p.

The above justification means (15) may be reduced to

$$f_{\lambda_p}\left(\mathbf{z}_k | \mathbf{z}_{k-1}\right) = \prod_{i=1}^{M_s} f_{\boldsymbol{\theta}_{i|\mathcal{S}}^{(p)}}\left(z_{i,k} | \mathcal{S}_i^{(p)}\right) \tag{16}$$

where $\mathcal{S}_i^{(p)} = \{z_{i',k-j} : J(Z_{i,k}^{(p)}, Z_{i',k-j}^{(p)}) > \mu\}, j \in \{0, 1\}$ with $i' < i$ when $j = 0$ (owing to (15)), and $J(Z, Z')$ is a measure of dependence between random variables Z and Z', e.g., mutual information is used for the results in Section IV. Therefore, $\boldsymbol{\theta}_{i|\mathcal{S}}^{(p)} \in \Theta_p$ is a length L categorical parameter vector for the conditional probability distribution associated with the ith coefficient state under $\mathcal{H}_1^{(p)}$, given $\mathcal{S}_i^{(p)}$. In other words, $\boldsymbol{\theta}_{i|\mathcal{S}}^{(p)}$ encodes the probability that $Z_{i,k}^{(p)} = l$ for $l \in [0, L-1]$, given that surrounding coefficient states $Z_{i',k-j}^{(p)}$, that $Z_{i,k}^{(p)}$ is found to be dependent on, are equal to specific values $z_{i',k-j} \in \mathcal{S}_i^{(p)}$. Clearly, there is a separate $\boldsymbol{\theta}_{i|\mathcal{S}}^{(p)}$ for each i, p, and possible set $\mathcal{S}_i^{(p)}$.

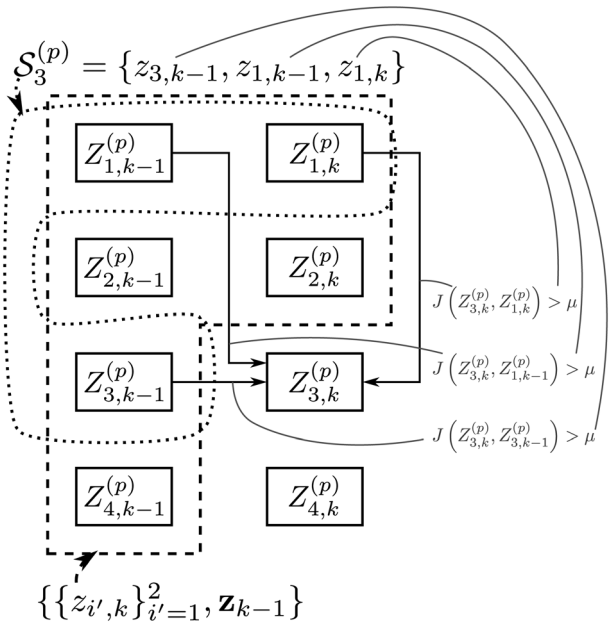


Fig. 4. Example dependency structure imposed on $Z_{3,k}^{(p)}$ by λ_p for $\mathbf{z}_k \in \mathbb{R}^4$. Dashed lines enclose variables $Z_{3,k}^{(p)}$ is dependent on using the decomposition in (15). Dotted lines enclose the reduced set of variables $Z_{3,k}^{(p)}$ is dependent on according to the measure $J(Z_{3,k}^{(p)}, Z_{i',k-j}^{(p)})$ (e.g., mutual information), thus defining edges in G_p (represented by arrows) and the set $\mathcal{S}_3^{(p)}$.

In general, $J(Z, Z')$ can be any measure that best captures the dependence of Z on Z', so long as it is easily calculated from training data. In terms of the graph G_p, if $J(Z_{i,k}^{(p)}, Z_{i',k-j}^{(p)})$ exceeds a predetermined threshold μ then $Z_{i',k-j}^{(p)}$ is a parent of $Z_{i,k}^{(p)}$ (an edge in G_p connects them) and $z_{i',k-j} \in \mathcal{S}_i^{(p)}$. This concept is illustrated by the example in Fig. 4, which shows the dependency structure used to calculate a single term in (15), and the simplified dependency structure imposed by the BN for calculating a single term in (16). This figure also shows that $\mathcal{S}_i^{(p)} \subseteq \{\{z_{i',k}\}_{i'=1}^{i-1}, \mathbf{z}_{k-1}\}$, i.e., $\mathcal{S}_i^{(p)}$ always contains a subset of the coefficient states used in the full decomposition in (15). Selection of the threshold μ can be based on examining the empirical probability density function of $J(Z, Z')$ for training data to look for statistically significant values. Typically, there is a high probability associated with $J(Z, Z') \approx 0$ due to sparsity. Setting μ too high results in ignoring potentially useful discriminatory information. Setting μ too low can lead to large sets $\mathcal{S}_i^{(p)}$, an abundance of conditional distributions, and generally poor sampling of these distributions, thus creating a poor fit of the model λ_p to the training data.

To complete the decomposition of (14), the prior probability of observing $\mathbf{z}_{k_1}$ for λ_p is required, and may be written as

$$f_{\lambda_p}(\mathbf{z}_{k_1}) = f_{\lambda_p}(z_{1,k_1}) \prod_{i=2}^{M_s} f_{\lambda_p}\left(z_{i,k_1} | \{z_{i',k_1}\}_{i'=1}^{i-1}\right) = \prod_{i=1}^{M_s} f_{\boldsymbol{\phi}_{i|\mathcal{T}}^{(p)}}\left(z_{i,k_1} | \mathcal{T}_i^{(p)}\right) \quad (17)$$

where $\mathcal{T}_i^{(p)} = \{z_{i',k_1} : J(Z_{i,k_1}^{(p)}, Z_{i',k_1}^{(p)}) > \mu\}, i' < i$ and $Z_{i,k_1}^{(p)}$ and $Z_{i',k_1}^{(p)}$ are random variables with corresponding realizations z_{i,k_1} and z_{i',k_1}, respectively, that are states of different coefficients in the first observation associated with a class p signal event. As before, $\boldsymbol{\phi}_{i|\mathcal{T}}^{(p)} \in \Theta_p$ is a length L categorical parameter vector for the prior probability distribution associated with the ith coefficient state under $\mathcal{H}_1^{(p)}$, given $\mathcal{T}_i^{(p)}$. The prior probabilities are defined similar to the conditional probabilities in (16) except they are conditioned on $\mathcal{T}_i^{(p)}$ rather than $\mathcal{S}_i^{(p)}$, where the former does not contain coefficient states z_{i',k_1-1}'s from the previous vector. This is due to the fact that the first vector in a signal event is independent of previous vectors and, consequently, the interelement dependency structure of $\mathbf{z}_{k_1}$ may be different from that of subsequent vectors in the event. The elements of $\mathcal{T}_i^{(p)}$ are still dictated by G_p, and hence, a given λ_p contains all the necessary components for calculating the probability of observing $\mathbf{Z}_{k_1}^n$ under $\mathcal{H}_1^{(p)}$. The full distribution parameter set associated with λ_p can now be written as

$$\Theta_p = \left\{ \left\{ \boldsymbol{\theta}_{i|\mathcal{S}}^{(p)} \right\}_{i,\mathcal{S}}, \left\{ \boldsymbol{\theta}_{i|\mathcal{T}}^{(p)} \right\}_{i,\mathcal{T}} \right\}.$$

We now show how to decompose the denominator of the SCST test statistic equivalent to that in (3), i.e., the probability of $\mathbf{Z}_{k_1}^n$ given λ_0. Since interference terms are mostly nullified by the sparse coding process, and since noise is IID, we can write

$$f_{\lambda_0}(\mathbf{Z}_{k_1}^n) = \prod_{k=k_1}^{n} f_{\lambda_0}(\mathbf{z}_k). \quad (18)$$

Using a similar concept to that in (17), each term on the right side of (18) can be expressed as

$$f_{\lambda_0}(\mathbf{z}_k) = f_{\lambda_0}(z_{1,k}) \prod_{i=2}^{M_s} f_{\lambda_0}\left(z_{i,k} | \{z_{i',k}\}_{i'=1}^{i-1}\right) = \prod_{i=1}^{M_s} f_{\boldsymbol{\phi}_{i|\mathcal{T}}^{(0)}}\left(z_{i,k} | \mathcal{T}_i^{(0)}\right) \quad (19)$$

where $\mathcal{T}_i^{(0)} = \{z_{i',k} : J(Z_{i,k}^{(0)}, Z_{i',k}^{(0)}) > \mu\}, i' < i$ is a set defined similar to $\mathcal{T}_i^{(p)}$, but for $\mathcal{H}_0$ coefficient state sequences. Naturally, $\boldsymbol{\phi}_{i|\mathcal{T}}^{(0)} \in \Theta_0$ is a length L categorical parameter vector defined similar to $\boldsymbol{\phi}_{i|\mathcal{T}}^{(p)}$ and the elements of $\mathcal{T}_i^{(0)}$ are dictated by the edges in G_0.

The required BNs, λ_0 and λ_p's, can be learned [20] using a set of training sequences for each hypothesis. The dependence measure $J(Z, Z')$ between random variables representing the coefficient states is fully observable, meaning G_p has a closed form given a specific set of training data. Each parameter vector in Θ_p can then be found using ML estimation by tabulating the number of times each coefficient is equal to a specific value given the associated set of dependent coefficient states. This training procedure allows for imperfect separation between signal and interference when finding $\mathbf{x}_n$'s since λ_p will model the dependency structure of $\mathbf{Z}_{k_1}^n$ when a class p signal is present, possibly superimposed with multiple types of interference. In other words, even if $\mathbf{Z}_{k_1}^n$ does not fully represent all of the signal components originally present in $\mathbf{Y}_{k_1}^n$, and additionally contains some interference components, training λ_p using such superimposed events accounts for this.

E. Sequential GLRT Implementation using Sparse Coefficient States

This subsection presents the proposed SCST implementation of the GLRTs when using BNs and a coefficient state sequence $\mathbf{Z}_{k_1}^n$, rather than general model parameter sets and the original data sequence $\mathbf{Y}_{k_1}^n$, as in Section II-B. Owing to the decompositions presented in (14)–(19), these GLRTs can be implemented using a set of cumulative test statistics that are updated as new data arrives, similar to the CUSUM procedure [9], [10]. Therefore, to implement the signal detection phase of the SCST method, calculating the GLRTs in (3) is replaced by calculating test statistics given by

$$B_p(n) = \max\{0, B_p(n-1)+b_p(n)\}, n = \hat{k}_0, \hat{k}_0+1, \dots \quad (20)$$

and initialized as $B_p(\hat{k}_0 - 1) = 0, \forall p$. This statistic is updated at time n using the nonlinearity

$$b_p(n) = \begin{cases} \ln\left(\frac{f_{\lambda_p}(\mathbf{z}_n|\mathbf{z}_{n-1})}{f_{\lambda_0}(\mathbf{z}_n)}\right), B_p(n-1) > 0 \\ \ln\left(\frac{f_{\lambda_p}(\mathbf{z}_n)}{f_{\lambda_0}(\mathbf{z}_n)}\right), \quad B_p(n-1) = 0. \end{cases}$$

Alternatively, $b_p(n)$ can be expressed as

$$b_p(n) = \begin{cases} \sum_{i=1}^{M_s} \ln\left(\frac{f_{\boldsymbol{\theta}_{i|\mathcal{S}}^{(p)}}\left(z_{i,n}|\mathcal{S}_i^{(p)}\right)}{f_{\boldsymbol{\theta}_{i|\mathcal{T}}^{(0)}}\left(z_{i,n}|\mathcal{T}_i^{(0)}\right)}\right), & B_p(n-1) > 0 \\ \sum_{i=1}^{M_s} \ln\left(\frac{f_{\boldsymbol{\theta}_{i|\mathcal{T}}^{(p)}}\left(z_{i,n}|\mathcal{T}_i^{(p)}\right)}{f_{\boldsymbol{\theta}_{i|\mathcal{T}}^{(0)}}\left(z_{i,n}|\mathcal{T}_i^{(0)}\right)}\right), & B_p(n-1) = 0. \end{cases}$$

Note that $B_p(n)$ accumulates at the same rate as the LLR in (3) (when given the same data), though it resets whenever $B_p(n) \leq 0$ [10]. Therefore, while in Section II-B a signal is detected when any time segment of (3) increases by η, here a signal is detected at time n whenever

$$\max_p B_p(n) \geq \eta \quad (21)$$

i.e., the cumulative value of $B_p(n)$ simply must exceed the threshold η.

Naturally, to implement a sequential version of the quiescent detection phase of the SCST method, calculating the GLRTs in (7) is replaced by calculating test statistics

$$T_p(n) = \max\{0, T_p(n-1)+t_p(n)\}, n = \hat{k}_1, \hat{k}_1+1, \dots \quad (22)$$

that are initialized as $T_p(\hat{k}_1 - 1) = 0, \forall p$, and updated using the nonlinearity

$$t_p(n) = \ln\left(\frac{f_{\lambda_0}(\mathbf{z}_n)}{f_{\lambda_p}(\mathbf{z}_n|\mathbf{z}_{n-1})}\right) = \sum_{i=1}^{M_s} \ln\left(\frac{f_{\boldsymbol{\theta}_{i|\mathcal{T}}^{(0)}}\left(z_{i,n}|\mathcal{T}_i^{(0)}\right)}{f_{\boldsymbol{\theta}_{i|\mathcal{S}}^{(p)}}\left(z_{i,n}|\mathcal{S}_i^{(p)}\right)}\right).$$

Unlike $b_p(n)$, the value of $t_p(n)$ does not depend on the value of the corresponding test statistic at time $n-1$ since conditional distributions are always used under $\mathcal{H}_1^{(p)}$ in the quiescent detection phase, as shown in (5). As before, $T_p(n)$ accumulates at the same rate as the LLR in (5), meaning the absence of any signal is declared at time n whenever

$$T_{p^*}(n) \geq \gamma \quad (23)$$

where

$$p^* = \arg\max_p B_p(n).$$

As in Section II-B, when $\mathcal{H}_0$ is again accepted a class label is assigned based on the ML signal type at time $\hat{k}_0 - 1$ and the process reverts back to looking for a new signal of unknown type according to (1). This phase switching process can continue indefinitely or until the end of the observation sequence has been reached, if applicable.

IV. Experimental Results

This section presents the results of applying the SCST method to a real acoustical signal detection and classification problem. The sequential multivariate data used to perform these experiments is first introduced, followed by details of the experimental test setup. Finally, results are presented in terms of receiver operator characteristics (ROC) for signal detection, and confusion matrices that indicate the overall performance for detecting and classifying transient signals. The performance of the SCST method is compared with that of random coefficient tracking (RCT) [6] and sparse reconstruction (SR)-based [12] methods.

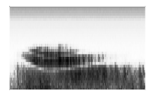
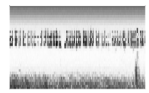
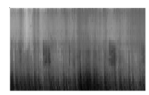
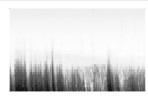
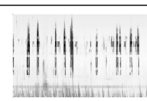
A. Data Description

The data used for this study represents acoustical recordings captured by a monitoring station with a single microphone located in a relatively remote site within Great Sand Dunes National Park, Colorado, where a variety of signal and interference sources are intermittently present[1]. The original audio waveforms were converted to sequences of 1/3 octave vectors [22] by the monitoring station that recorded the soundscape. Each 1/3 octave vector was extracted from a non-overlapping one-second time segment, and has $N = 33$ elements that represent the average energy in different 1/3 octave frequency bands for the corresponding one second interval.

The soundscape was recorded for approximately 17 full days from September 24th to October 10th, 2008, where each day consisted of 86,400 observations. The types of signal and interference sources that were frequently captured by this monitoring station are listed in Table I, along with brief descriptions of their properties. Note that the example events in this table only provide a rough indication of the structure of typical 1/3 octave signatures associated with a given source type, as significant within-class variations are a defining characteristic of most soundscape data. The main challenge associated with this data is the large number of different interference types, which severely diminishes the number of features that can be reliably used for signal detection and classification. In particular, strong wind is dominant throughout a large percentage of the recordings, and

[1]For these experiments, extrinsic and intrinsic acoustical sources are synonymous with terms *signal* and *interference*, respectively, whereas *source* means anything besides ambient noise.

TABLE I
DESCRIPTIONS FOR DIFFERENT SOURCE TYPES IN THE GREAT SAND DUNES DATA SET

	Source	Typical Event Description	Typical Duration	Example
Signal	Propeller Plane	Signatures evolve from mid-frequency narrowband to wideband, and revert back to narrowband.	30–240 s	
	Jet	Starts with low-to-mid frequency signatures, with the mid-frequency signatures slowly fading approximately half-way through the event.	80–260 s	
Interference	Birdsong	Signatures are restricted to high frequency bands and have erratic temporal patterns.	1 s – several hours	
	Rain / Thunder	Rain has mid-to-high frequency broadband signatures, while thunder is often superimposed with rain and adds impulsive, low-to-mid frequency signatures.	few seconds – several hours	
	Strong Wind	Impulsive low-to-mid frequency signatures appearing similar to thunder, but typically more relentless.	few seconds – several days	
	Elk Calls	Similar to birdsong, but having mid-to-high frequency broadband signatures and typically higher magnitude.	1 s – several hours	

its signatures commonly occlude the low-to-mid frequency signatures of signal events to be detected (jets in particular).

Due to the complexity of the soundscape, manual annotation of the data was previously the only available approach for locating and labeling signals. Therefore, such annotations existed before the development of the SCST method, and serve as the ideal that is used to generate results. In particular, two well-trained operators visually inspected the data to identify acoustical events associated with signals of interest, which are those listed in Table I, as they occur most frequently and prominently in this particular site.

B. Test Setup

To apply the SCST method, disjoint training and testing sets were formed using a collection of two-hour-long data segments found during the 17 days of continuous data recording. This segment length was chosen to balance the competing objectives of ensuring diverse conditions were encountered within a given segment (varying interference types and signal arrival times) and minimizing occurrences of long periods of stationary acoustical conditions. In order to provide robust training and a challenging testing environment, segments were selected for both sets that contained a relatively large number of signal and interference sources and events with highly variable signatures. The training set consisted of 10 data segments (about 4.9% of total data) and was used to form the dictionary matrix $\mathbf{A}$ in (9), learn the BNs λ_p's and λ_0 [20], and choose the detection thresholds $\eta = 43.7$ and $\gamma = 20.0$ in (21) and (23), respectively, such that no signals in the training segments were missed. In particular, events within the training segments representing $\mathcal{H}_1^{(p)}$ and containing the signatures of one signal source, often superimposed with one or more types of interference, were used to learn an associated BN λ_p. Similarly, training events representing $\mathcal{H}_0$ and containing only interference and noise were used to learn the BN λ_0.

To form the dictionary matrix $\mathbf{A}$ during the training phase, K-SVD [15] was applied separately to different sets of observations, each representing a single type of signal or interference, to extract $\mathbf{S}_p \in \mathbb{R}^{N\times 25}, \forall p$ and $\mathbf{H}_q \in \mathbb{R}^{N\times 15}, \forall q$. The concatenation of these atoms yields $\mathbf{A}$ as in (9), meaning $M_s = 50$ and $M_h = 60$. Selection of the number of atoms in each dictionary was performed according to the guidelines outlined in [15]. Note that learning a dictionary using K-SVD involves two steps, namely a codebook update stage where each atom is updated to minimize the error between the training observations and their sparse reconstructions, and a sparse coding stage where (8) is used in conjunction with the updated dictionary to extract sparse atom coefficients. Basis pursuit denoising [24] was used to perform the latter step of the K-SVD process since, like the SCST method, the sparse coding strategy is user-defined. The resulting dictionary $\mathbf{A}$ was then used by the SCST method during the testing phase to extract each $\mathbf{x}_n$ in (8), which was also performed using basis pursuit denoising.

Based on the criterion that $z_{i,k} = 0$ for 99% of sparse coefficients representing observations in the training set containing noise alone, $\epsilon = 8.61$ was selected to determine the zero-state in (10). To determine parent-child relationships in the BNs λ_p's, mutual information was used as the dependence measure with a threshold of $\mu = 0.1$, which corresponds to the largest 2% of values witnessed for this measure, thus defining the sets $\mathcal{S}_i^{(p)}$, $\mathcal{T}_i^{(p)}$, and $\mathcal{T}_i^{(0)}$ used in (16), (17), and (19), respectively.

As mentioned before, to maximize class discrimination, we desire L in (11) to be as large as possible while avoiding sample poor distributions used to form (16), (17), and (19). Realistically, even with an abundance of training data, there may be no available samples to form some of the conditional distributions. For instance, if for class p signals $z_{i,k}$ is found to be depen-

dent on $\|\mathcal{S}_i^{(p)}\| = K$ other coefficients, then L^K separate conditional distributions must be formed for $z_{i,k}$, and the structure of class p signal events may dictate that many of the dependent coefficient sets are rarely, if ever, realized in the training data. Therefore, the criterion used in this paper is that at least one of the conditional distributions for each $Z_{i,k}^{(p)}$ (every coefficient and class combination) must be formed using $\geq 4L$ samples, since each requires L estimates. Requiring more than one sample-rich distribution may be unrealistic for some classes of signals for reasons mentioned before. When this criterion is met, the remaining distributions that are considered sample poor are set to be uniform. This procedure led to $L = 4$, as setting $L > 4$ violated the sample-rich criterion.

The testing set consisted of 38 segments (about 18.6% of total data) and was used to evaluate the performance of each method. At least two segments from each of the 17 available days of data was used to form the testing set. Note that, since the data vectors used for this study represent frequency subband acoustical energy at different times they are not zero-mean, and hence, the noise mean (estimated from training segments) was subtracted from each observation, before being processed by a given method.

The proposed SCST method is benchmarked against two other approaches, namely RCT [6] and SR [12]. The former applies a hierarchy of likelihood ratio tests to each $\mathbf{y}_n$ in order to accept either (a) a noise alone hypothesis or (b) a signal alone or interference alone hypothesis or (c) a dual source (signal and interference) hypothesis. The RCT method is used here since its design was also strongly motivated by the National Park soundscape analysis problem and demonstrated the best performance in [6] when compared to a Gaussian mixture model-based approach.

The SR method [12] was originally developed for face recognition from images with potential occlusions (i.e., interference), and assigns class labels based on which class-specific set of atoms can reconstruct the interference-free observation most accurately. Interference is removed using sparse coding to separate the corresponding components from those of signals, and subtracting the reconstructed interference estimate from the original observation. Since the method in [12] does not address the detection problem, it must be extended in order to handle the soundscape data in this section. Denote $\boldsymbol{\alpha}_p(\cdot)$ as the indicator function that selects coefficients associated with the pth signal class, i.e., $\boldsymbol{\alpha}_p(\mathbf{x}_n) \in \mathbb{R}^M$ only has nonzero entries in indices corresponding to the columns of $\mathbf{S}_p$. In this paper, SR-based detection is performed on each observation separately using

$$\max_p \|\mathbf{A}\boldsymbol{\alpha}_p(\mathbf{x}_n)\|_2 \begin{array}{c} \mathbf{y}_n \notin \mathcal{H}_0 \\ \geq \\ < \\ \mathbf{y}_n \in \mathcal{H}_0 \end{array} \eta' \tag{24}$$

i.e., the energy of at least one signal estimate $\mathbf{A}\boldsymbol{\alpha}_p(\mathbf{x}_n)$ must exceed a signal detection threshold η'. Classification is then performed on each observation separately as in [12], i.e.

$$c_n = \min_p \|\mathbf{y}_n - \mathbf{A}\boldsymbol{\omega}(\mathbf{x}_n) - \mathbf{A}\boldsymbol{\alpha}_p(\mathbf{x}_n)\|_2$$

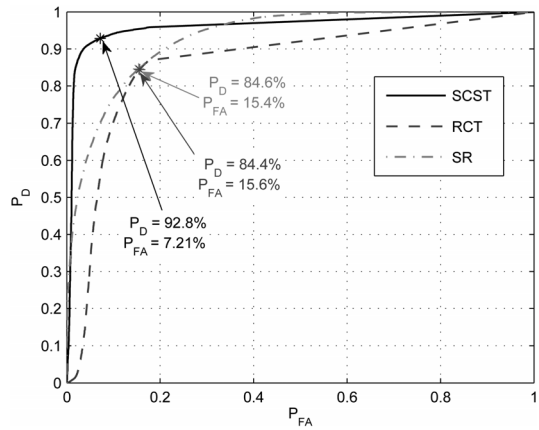


Fig. 5. Signal detection ROC curves.

where $\boldsymbol{\omega}(\cdot)$ is the indicator function that selects coefficients associated with interference atoms in $\mathbf{A}_h$, thus removing the estimated interference from the observation. Clearly, the behavior of $\boldsymbol{\alpha}_p(\cdot)$'s and $\boldsymbol{\omega}(\cdot)$ are dictated by the structure of $\mathbf{A}$ in (9). SR is used for benchmarking since it is one of the few existing methods that is applicable to our problem without significant modifications, and additionally provides an indication of expected performance for methods that do not exploit the temporal dependencies between observations.

Since the RCT and SR methods assign class labels to each observation separately, an HMM-based post-processing was applied to the signal classification results produced by these methods, the details of which are described in [6]. This allows for association of a cluster of detections that have the same label with a single event for more concise and meaningful classification results.

C. Signal Detection Performance

To measure the signal detection performance of each method when applied to testing segments, ROC curves were generated and are shown in Fig. 5. Each ROC curve shows the change in probability of signal detection (P_D) and probability of false alarm (P_{FA}) as the detection threshold is modified, e.g., η in (21) for SCST. In other words, P_D refers to the probability of correctly accepting $\mathcal{H}_1^{(p)}$ over $\mathcal{H}_0$ irrespective of any interference that may be present. Detection performance is measured using individual observations rather than entire events, i.e., the temporal position of an observation is irrelevant and only its associated detection statistic is considered. Though SCST detection is dependent on both (21) and (23), only a single threshold may be modified to generate the ROC, and hence, the threshold for the latter remained fixed. Similarly, the RCT signal detection ROC was generated by only modifying the statistic for the initial detection test [6] to determine the associated impact on signal detection performance, while thresholds for the other tests in the hierarchy remained fixed. For the SR method, (24) can easily be used to generate the ROC by varying η'. The evaluation metrics considered are the area under the ROC curve (AUC) and the P_D and P_{FA} at its "knee-point". The AUC is important since it represents the discrimination ability of a test, while the knee-point corresponds to a threshold where $P_D + P_{FA} = 1$.

As can be seen from Fig. 5, the knee-points of the signal detection ROC curves for the SCST, RCT, and SR methods are ($P_D = 92.8\%$, $P_{FA} = 7.21\%$), ($P_D = 84.4\%$, $P_{FA} = 15.6\%$), and ($P_D = 84.6\%$, $P_{FA} = 15.4\%$), respectively, while the AUCs are 0.962, 0.863, and 0.931, respectively. The SCST achieves a much higher P_D at a given P_{FA} primarily because it exploits the dependencies between the signal components in temporally adjacent observations to yield a cumulative test statistic. Consequently, even when an event contains some observations with weak or novel signal components, a sufficiently high detection statistic is maintained throughout such an event. In contrast, the RCT and SR methods perform detection on each observation independently, leading to more missed detections within events, though the former performs worse (according to AUC) since signal detection is based on three tests, and the thresholds for the second two are fixed. Missed detections for the SCST method are primarily due to delayed signal detections that are inherent with transient detection schemes using cumulative test statistics [9], which cause a small number of samples to be missed at the beginning of each signal event. Similarly, false alarms generated by the SCST method are mainly caused by quiescent detection delays, leading to a few false detections at the end of each signal event. The detection delays are the reason $P_D \neq 100\%$ even for high P_{FA} values for the SCST ROC.

For all methods, some false alarms were also caused by the presence of interference. Occasionally, for the SCST and SR methods, the energy of novel interference was associated with signal atoms, resulting in a state sequence that produced a relatively high signal likelihood for the SCST method and a high SR detection statistic. Similarly, the hierarchy of tests used by the RCT methods sometimes detected a signal when only novel strong interference was present. One factor that reduced the occurrence of false alarms for the SCST and SR methods is their ability to remain robust to the simultaneous presence of multiple types of interference. In contrast, the RCT method assumes a maximum of one type of signal and one type of interference can be simultaneously present. Due to the wide variety of interference source types associated with the Great Sand Dunes soundscape, the presence of multiple types of interference in a given observation would sometimes lead to false alarms for the RCT method, since the superimposed signal and interference model would produce a higher likelihood than the interference alone model.

D. Overall Detection and Classification Performance

While the above ROC analysis shows how well each method is able to detect the presence of a signal in individual observations, here the overall performance of each method for correctly detecting and classifying entire signal events in the 38 testing sequences is evaluated. This provides an indication of how each method performs on a real soundscape analysis problem, where the goal is to tabulate the number of times and when each signal type is present. For instance, in the proposed SCST method, (21) and (23) are used to estimate the time of arrival, duration, and class label of a given signal event. If at least half of the set of observations associated with a manually annotated event (truth) are also in the set of observations associated with detected signal event, and they additionally have the same class label, then the annotated event is considered correctly detected and classified. In other words, performing classification presumes correct detection. Missed detections result when too few or no observations in the annotated event are assigned a label other than $c_k = 0$, and misclassifications occur when the wrong label is assigned a majority of the time. False alarms occur when a signal event is thought to be present where there is none.

The overall detection and classification results are presented in terms of the confusion matrices in Table II. Each entry in this table indicates the number of times a certain type of signal event was assigned a specific label by a given method (SCST / RCT / SR). Since "none" means no signal of interest, the first column in each confusion matrix indicates instances where signal events are missed ("none" assigned), whereas the first row indicates false alarms ("none" true according to annotation). The shaded diagonal entries indicate the number of events of each signal type that are assigned the correct label, which show overall correct signal classification rates of 93.0%, 89.0%, and 87.0% for the SCST, RCT, and SR methods, respectively. False alarm rates are reported in terms of the percentage of all event detections (i.e., entries in the last two columns in each confusion matrix) that are false, which are 3.75%, 11.0%, and 16.1% for the SCST, RCT, and SR methods, respectively. This is also the reason "-" appears for the "none" diagonal entry in Table II. Note that the ROC performance potential of the RCT method was limited due to fixing thresholds for two of its tests. Since no such limitations exist when evaluating classification performance, the RCT method was able to produce better results than the SR method.

As can be seen from Table II, the overall classification results produced by the SCST method are noticeably better than those produced by the RCT and SR methods. The gap in classification performance is caused by the drastically different approaches taken by each method. For instance, the vector linear autoregressive basis coefficient source model used by the RCT method [6] can sometimes fail to estimate subtle or novel variations in acoustical events. In contrast, the SCST method makes no assumptions concerning the distributions of the signals, interference, and noise, but instead simplifies the data representation just enough so that likelihoods can be realistically computed. In other words, simplifying the data representation has provided superior class discrimination when compared to restricting the plausible structure of observation components. Moreover, the SCST method is generally better suited for adapting to sudden changes in the structure of source signatures considered in this study owing to, e.g., Doppler effects. For example, the 1/3 octave bands that contain significant energy can rapidly change if a signal source has a high velocity and becomes relatively close to the receiver. For SCST, such a quick change conveniently manifests itself as a change in the atoms used in the sparse representation, which can easily be modeled by a BN. On the other hand, the autoregressive coefficient model used by the RCT method typically causes greater errors when estimating rapidly changing source signatures.

As expected, the presence of interference was most detrimental to the classification performance of the RCT method, which misclassified a few jets as planes, and vice versa, when strong wind was present, owing to the superposition of plane

TABLE II
CONFUSION MATRICES SHOWING THE TOTAL NUMBER OF INSTANCES EACH SIGNAL TYPE WAS ASSIGNED A GIVEN LABEL BY EACH METHOD (SCST / RCT / SR)

		Assigned		
		None	**Plane**	**Jet**
Truth	**None**	-	0 / 10 / 0	11 / 24 / 53
	Plane	3 / 9 / 13	**38 / 30 / 15**	2 / 4 / 15
	Jet	16 / 16 / 11	0 / 4 / 0	**242 / 238 / 247**

and wind signatures resembling those of a jet. The SCST method was able to reject interference via the sparse coding process, and hence, the majority of the wind signatures were coded using the interference atoms in these cases, thus minimizing confusion between jet and plane events. The SR method, on the other hand, produced a large number of jet false alarms and misclassified a fair number of planes as jets. The latter case was often caused by a subset of observations within a given plane event resembling signatures typically associated with the jet class. Since the SCST method considers the joint likelihood of all observations in a given event, it is typically more robust to the presence of such novel signatures. In contrast, the RCT and SR-based methods make decisions on individual observations, and aggregate results using postprocessing [6].

E. Separation Sensitivity Analysis

As mentioned in Section III-A, the SCST method assumes the sparse coding process provides adequate signal and interference separation, such that $\mathbf{z}_n$'s predominantly represent signal components. Therefore, the goal here is to examine the sensitivity of the performance of the SCST and SR methods relative to varying levels of separation quality. Since some of the signal and interference sources in Table I share many of the same features (e.g., the low frequency signatures present in both wind and jet events), the separation of these two components was not always perfect for the experiments reported above. This allows for conducting a separate experiment where each signal event in the testing set was assigned a separation quality tag of either "excellent", "good", or "fair", based on a visual analysis of sparse coefficient vectors $\mathbf{x}_n$'s extracted from the event, and the corresponding reconstructed signal and interference component estimates, i.e., $[\mathbf{A}_s \mathbf{0}_{N\times M_h}]\mathbf{x}_n$ and $[\mathbf{0}_{N\times M_s}\mathbf{A}_h]\mathbf{x}_n$, respectively, where $\mathbf{0}_{N\times m}$ is an $N \times m$ matrix of zeros. Excellent separation corresponds to little to no perceived signal energy present in the interference estimates and similarly for interference energy relative to signal estimates. Good and fair signal event estimates are noticeably degraded relative to the excellent estimates, and contain observations with less energy. Specifically, the mean energy of good and fair signal observation estimates is 92% and 76% of the mean energy of excellent signal observations estimates, respectively.

Tables III and IV show the results of the sensitivity experiment in the form of confusion matrices for the SCST and SR methods, respectively, for each separation quality level. Results are shown in terms of percentages of acoustical events associated with a given signal type that were assigned a given label, since each quality level has a different number of associated events. False alarms (true label of "None") are not relevant for these results since observations that contain only noise are not

TABLE III
CONFUSION MATRICES SHOWING THE PERCENTAGE OF EVENTS OF EACH SIGNAL TYPE THAT WERE ASSIGNED A GIVEN LABEL BY THE SCST METHOD UNDER DIFFERENT SEPARATION QUALITY SCENARIOS (EXCELLENT/GOOD/FAIR)

		Assigned		
		None	**Plane**	**Jet**
Truth	**Plane**	0% / 0% / 20%	**100% / 100% / 67%**	0% / 0% / 13%
	Jet	2.6% / 5.6% / 9.9%	0% / 0% / 0%	**97% / 94% / 90%**

TABLE IV
CONFUSION MATRICES SHOWING THE PERCENTAGE OF EVENTS OF EACH SIGNAL TYPE THAT WERE ASSIGNED A GIVEN LABEL BY THE SR METHOD UNDER DIFFERENT SEPARATION QUALITY SCENARIOS (EXCELLENT/GOOD/FAIR)

		Assigned		
		None	**Plane**	**Jet**
Truth	**Plane**	0% / 32% / 47%	**56% / 42% / 13%**	44% / 26% / 40%
	Jet	1.4% / 2.6% / 7.7%	0% / 0% / 0%	**99% / 97% / 92%**

considered. As expected, the classification performance of both methods gradually decreases along with the quality of signal and interference separation, as the features representing signal components diminish. However, for the SCST method, the performance difference for excellent and good events is minimal. Indeed, Table III shows that the majority of errors made by the SCST method were caused by events with fair signal separation quality. Plane events with fair separation quality caused the most problems since these events were typically the weakest, and erroneous association of the energy in these events would sometimes leave insufficient signal features in the coefficient state sequence. The SR method, on the other hand, does well as far as classifying jets but does poorly at classifying planes at all separation quality levels, for reasons mentioned before. Essentially, since the SR method does not exploit the structure of entire events, it is less robust to a decrease in feature quality. Overall, these results show that imperfect signal and interference separation does harm the performance of the SCST method, but it does not break its functionality. The main reason for this is the probabilistic nature of the BN used to model each signal event, meaning it can remain fairly robust to scenarios where some coefficient states assume atypical values.

F. Computational Complexity

To conclude this section, the computational complexity of the SCST method for evaluating a single $\mathbf{y}_n$ is considered. This analysis assumes that the quantization function $H(\cdot)$ in (11), the dictionary matrix $\mathbf{A}$ in (9), and each BN are all computed off-line, as was the case for the results reported above. Clearly, the cost of the sparse coding process dominates the overall computational complexity of both the SCST and SR methods. If using basis pursuit denoising [24], this process involves finding the solution to a quadratic programming problem, which can be accomplished with a wide variety of algorithms, each with a different complexity that depends on the error tolerance δ. Orthogonal matching pursuit [23] generally requires fewer computations and is recommended for applications where N and M are large. The cost of computationally efficient implementations of orthogonal matching pursuit is $O(KNM)$ [33], where K is the number of iterations (sparsity level), that depends on δ. In the

absolute worst case scenario, $K = N$, meaning $O(MN^2)$ operations are required for sparse coding. Otherwise, SCST simply requires updating the test statistics in (20) and (22), which is very simple since the distribution parameters are computed offline, requiring only $O(PM_s)$ operations.

The computational cost of the RCT method was shown to be $O(PQN^{2.373})$ [6]. Though the complexity of a given method depends on M (for SCST and SR only) and the number of signal sources P and interference sources Q, it is reasonable to suggest that the cost of SCST is similar to the benchmark methods for many applications, so long as an efficient sparse coding algorithm is used. For the soundscape characterization application considered in this paper, the SCST (using basis pursuit denoising), RCT, and SR [12] methods took an average of 58.7, 5.70, and 48.9 milliseconds, respectively, to process a single observation using MATLAB on a computer with a 3.2 GHz quad-core processor and 8 GB of RAM. Clearly, the SCST method was the slowest in this case, but it still processed the data about 17 times faster than the data sampling rate of one observation per second.

V. Conclusions

This paper introduces a new method for detection and classification of transient events from multivariate observations using the patterns of corresponding coefficient state sequences to determine the likelihood of each known signal model. The motivation behind this approach stems from the fact that coefficient state sequences provide a simple way to represent nonstationary components and facilitate realistic calculation of likelihoods, even for lengthy vector sequences. This is especially important for applications where transient events associated with a given signal type are very erratic and have complex temporal evolutions. Furthermore, few assumptions need to be made concerning the statistics of observation components compared to, e.g., the benchmark method in [6]. Finally, the proposed method inherently provides robustness to multiple competing interference sources, owing to the separation capabilities of sparse coding when using an appropriately designed dictionary. The presented results demonstrate the effectiveness of the proposed SCST method at performing simultaneous detection and classification of extrinsic acoustical sources in a natural landscape. The downsides of the SCST method relative to the RCT method are its increased computational complexity and inability to provide class labels for interference sources.

Acknowledgment

The authors would like to thank E. Lynch and D. Joyce at the National Park Service for providing the data and its associated annotation, that were used to generate experimental results.

References

[1] J. M. K. Kua, E. Ambikairajah, J. Epps, and R. Togneri, "Speaker verification using sparse representation classification," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICA)*, May 2011, pp. 4548–4551.

[2] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 3, no. 1, pp. 72–83, Jan. 1995.

[3] H. Wang, J. Elson, L. Girod, D. Estrin, and K. Yao, "Target classification and localization in habitat monitoring," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICA)*, Apr. 2003, vol. 4, pp. 844–847.

[4] S. Chu, S. Narayanan, and C. C. J. Kou, "Environmental sound recognition with time-frequency audio features," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 6, pp. 1142–1158, Aug. 2009.

[5] S. G. Lingala, Y. Hu, E. DiBella, and M. Jacob, "Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1042–1054, May 2011.

[6] N. Wachowski and M. Azimi-Sadjadi, "Characterization of multiple transient acoustical sources from time-transform representations," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 9, pp. 1966–1978, Sep. 2013.

[7] D. Oldoni, B. De Coensel, M. Rademaker, B. De Baets, and D. Botteldooren, "Context-dependent environmental sound monitoring using SOM coupled with LEGION," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2010, pp. 1–8.

[8] M. Basseville and I. V. Nikiforov, "Detection of Abrupt Changes: Theory and Application," in . Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[9] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 42, pp. 1897–1908, Jun. 1971.

[10] B. Chen and P. Willet, "Detection of hidden Markov model transient signals," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 4, pp. 1253–1268, Oct. 2000.

[11] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, pp. 34–81, Feb. 2009.

[12] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 1–18, Feb. 2009.

[13] H. Zhang, N. M. Nasrabadi, T. S. Huang, and Y. Zhang, "Transient acoustic signal classification using joint sparse representation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICA)*, May 2011, pp. 2220–2223.

[14] S. Zubair and W. Wang, "Audio classification based on sparse coefficients," in *Proc. Sensor Signal Process. for Defence*, Sep. 2011, pp. 1–5.

[15] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[16] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 886–902, Apr. 1998.

[17] L. Daudet, "Sparse and structured decompositions of signals with the molecular matching pursuit," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 5, pp. 1808–1816, Sep. 2006.

[18] V. Bruni, S. Marconi, and D. Vitulano, "Time-scale atoms chains for transients detection in audio signals," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 3, pp. 420–433, Mar. 2010.

[19] S. J. Godsill, A. T. Cemgil, C. Févotte, and P. J. Wolfe, "Bayesian computational methods for sparse audio and music processing," in *Proc. 15th Eur. Signal Process. Conf.*, Sep. 2007, pp. 345–349.

[20] D. E. Holmes and L. C. Jain, "Innovations in Bayesian Networks," in , 1st ed. Berlin/Heidelberg, Germany: Springer, 2008.

[21] E. Lynch, D. Joyce, and K. Fristrup, "An assessment of noise audibility and sound levels in U.S. national parks," *Landscape Ecol.*, vol. 26, pp. 1297–1309, Aug. 2011.

[22] E. H. Berger, L. H. Royster, J. D. Royster, D. P. Driscoll, and M. Layne, *The Noise Manual*. Falls Church, VA, USA: AIHA, 2003.

[23] J. M. Adler, B. D. Rao, and K. Kreutz-Delgado, "Comparison of basis selection methods," *Proc. 30th Asilomar Conf. Signals, Syst. Comput.*, vol. 1, pp. 252–257, Nov. 1996.

[24] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, pp. 6–18, Jan. 2006.

[25] D. L. Donoho and G. Kutyniok, "Analysis of ℓ_1 minimization in the geometric separation problem," in *Proc. 42nd Annu. Conf. Inf. Sci. Syst.*, Mar. 2008, pp. 274–279.

[26] D. Barchiesi and M. D. Plumbley, "Learning incoherent dictionaries for sparse approximation using iterative projections and rotations," *IEEE Trans. Signal Process.*, vol. 61, no. 4, pp. 2055–2065, Apr. 2013.

[27] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 543–550.

[28] H. V. Poor and J. B. Thomas, "Applications of Ali-Silvey distance measures in the design of generalized quantizers for binary decision systems," *IEEE Trans. Commun.*, vol. COM-25, no. 9, pp. 893–900, Sep. 1977.

[29] J. N. Tsitsiklis, "Extremal properties of likelihood-ratio quantizers," *IEEE Trans. Commun.*, vol. 41, no. 4, pp. 550–558, Apr. 1993.

[30] H. Kobayashi and J. B. Thomas, "Distance measures and related criteria," in *Proc. 5th Annu. Allerton Conf. Circuit Syst. Theory*, Oct. 1967, pp. 491–500.
[31] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. R. Statist. Soc. Ser. B*, vol. 28, pp. 131–142, Apr. 1966.
[32] A. C. Davison, *Statistical Models*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[33] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202–6216, Dec. 2012.

Neil Wachowski (S'09) received the B.S. degree in electrical engineering from Michigan Technological University, Houghton, in 2002, and the M.S. and Ph.D. degrees from Colorado State University, Fort Collins, in 2009 and 2014, respectively, both in electrical engineering with specialization in signal processing. He was a Field Applications Engineer at Texas Instruments Inc. from 2002 to 2006. His research interests include parameter estimation, statistical signal processing, and transient signal detection.

Mahmood R. Azimi-Sadjadi (S'81–M'81–SM'89) received the M.S. and Ph.D. degrees from the Imperial College of Science and Technology, University of London, London, U.K., in 1978 and 1982, respectively, both in electrical engineering with specialization in digital signal/image processing.

Currently, he is a Full Professor at the Electrical and Computer Engineering Department, Colorado State University (CSU), Fort Collins. He is also the Director of the Digital Signal/Image Laboratory, CSU. His main areas of interest include digital signal and image processing, wireless sensor networks, target detection, classification and tracking, adaptive filtering, system identification, and neural networks. His research efforts in these areas resulted in over 250 journal and referenced conference publications. He is the coauthor of the book *Digital Filtering in One and Two Dimensions* (New York: Plenum Press, 1989).

Prof. Azimi-Sadjadi served an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE TRANSACTIONS ON NEURAL NETWORKS.