Characterization of Multiple Transient Acoustical Sources From Time-Transform Representations

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

***Abstract*—This paper introduces a new framework for detecting, classifying, and estimating the signatures of multiple transient acoustical sources from a time-transform representation (TTR) of an audio waveform. A TTR is a vector observation sequence containing the coefficients of consecutive windows of data with respect to known sampled basis waveforms. A set of likelihood ratio tests is hierarchically applied to each time slice of a TTR to detect and classify signals in the presence of interference. Since the signatures of each acoustical event typically span several adjacent dependent observations, a Kalman filter is used to generate the parameters necessary for computing the likelihood values. The experimental results of applying the proposed method to a problem of detecting and classifying man-made and natural transient acoustical events in national park soundscape recordings attest to its effectiveness at performing the aforementioned tasks.**

***Index Terms*—Kalman filtering, multivariate analysis, sequence estimation, signal classification, transient signal detection.**

I. Introduction

THIS paper considers the detection and classification of highly variable transient acoustical events that contribute to the soundscapes of national parks. Acoustical monitoring ultimately advances the overreaching purpose of national parks, which is to conserve the natural state of resources therein, by providing a scientific basis for identifying trends in acoustical conditions, quantifying impacts of certain sources, and assessing consistency with park management objectives. The goal is to prevent manual post observation and evaluation of large volumes of data, which is the current approach due to the complexity of the soundscapes. Of particular interest is determining the composition of the acoustical scene in terms of commonly occurring man-made (e.g. aircraft) and natural (e.g. weather effects) sources, which are often present simultaneously. We associate these two categories with the more general terms of "signal" and "interference", respectively, whereas the term source means anything that produces an acoustical event, i.e., signal or interference. An event refers to the signatures of a single source that often span several observations, e.g., the sounds produced by a jet flyover. There are typically only a few signal types of interest for a given park, while interference sources are those that disrupt proper characterization of signals, yet do not fit the ambient noise model.

To accomplish this source characterization task, monitoring stations are typically deployed in certain locations in parks for months at a time to constantly record the soundscapes. Historically, a particular type of lossy data reduction has been used for efficient storage. This data reduction is done by transforming consecutive windows of the streaming audio data by using a bank of filters and forming a data vector based on the energy in the filtered waveforms. In general, any time-transform representation (TTR) can be employed for this lossy data reduction. Here, a TTR refers to any transformation with real-valued basis [1] (e.g., discrete cosine or discrete wavelets [2]) applied to time-windowed streaming data. Therefore, the goal is to develop a unified approach for analyzing data collected in the past, as well as new data to be collected using other TTR, for on-line implementation on the monitoring station. This means that many existing approaches [3]–[9] developed for similar applications are not appropriate solutions as they presume access to the raw audio data.

In general, automated detection and classification (discriminating between different source types) of acoustical transient events has a multitude of applications including speech recognition [7], [10], habitat monitoring [8], [11], and battlefield surveillance [9], [12]. These tasks are complicated by many factors that often depend on the particular application. These factors include time-varying source signatures, a large number of possible event types, variable number of sources that may be simultaneously present leading to superimposed signatures, variable event structure and duration even among those events associated with a single source type, unknown arrival times and Doppler shift, and the presence of ambient noise as well as environmental and operational variations.

Detection and classification of acoustical events from TTR in particular carries a unique set of challenges including lack of optimality of the representation if the transform vectors are not matched to source signatures of interest [1] and potential reduction of discriminatory information relative to the original audio data when using a lossy transform. One common approach to such problems is to identify novel portions of a TTR based on certain features or statistical properties. The work in [10] identifies speech patterns using 2-D patches that are extracted from spectrograms and projected onto a 2-D discrete cosine basis. Another study [13] exploits the distribution of individual time-frequency components in a spectrogram, which is then transformed to a space appropriate for segmenting time-frequency slices containing noise only from those containing noise plus an unknown signal.

Manuscript received October 30, 2012; revised February 08, 2013 and April 25, 2013; accepted May 05, 2013. Date of publication May 14, 2013; date of current version July 12, 2013. This work was supported under a cooperative agreement with the National Park Service, Fort Collins, CO, under Contract #H2370094000. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Laurent Daudet.

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/TASL.2013.2263141

Template-based methods are also frequently used for acoustical source characterization from TTR. In [14], use of the spectral-band matched filter is explored for those source types that do not generate highly variable acoustical events. If the dependencies between successive observations are ignored and sources emit deterministic signatures (unrealistic assumption) then the matched subspace detector [15] can be used to detect the presence of the signal in each observation separately, and a uniformly most powerful detector is formed even in the presence of known subspace interference. Since template based methods typically fail to account for inconsistent signatures that are emitted by most acoustical source types, Hidden Markov Models (HMM) [16] may be used as a viable alternative. In [11], a framework is presented for automated detection and classification of sounds from birds, crickets, and frogs using a HMM to represent sequences of statistical features extracted from spectrograms. In [17], a transient detection method was introduced that implements Page's test [18], which minimizes the worst case delay to detection, with dependent observations using HMMs. Methods also exist that take a more targeted approach to environmental sound classification, such as the work in [19], which uses a self-organizing map, along with a locally-excitatory/globally-inhibitory oscillator network, to identify co-occurring sound features extracted from a spectral representation and group sound fragments, respectively.

While the aforementioned methods excel at their intended applications, they all have shortcomings that prevent them from achieving acceptable performance on the data considered in this study. In particular, these methods in their original form either (1) are incapable of recognizing any general source type [10], [11], or (2) cannot handle significant within class diversity [14], [15], or (3) cannot not perform classification (only detection) [13], [17], or (4) are not robust to the presence of structured interference [10], [11], [13], [14], [17], [19]; all of which are essential for the problem at hand.

To address the shortcomings of existing methods for characterization of multiple source types from TTR, this paper develops a sequential random coefficient tracking (SRCT) framework that applies a hierarchy of log likelihood ratio tests (LLRT) to individual observations, each of which may contain the signatures of a subspace signal and/or a subspace interference source, both of which can be one of multiple types. Source signatures are modeled as random to capture variability between different events associated with the same source type. A Kalman filter that exploits known source subspace and coefficient dependency models is used to generate the parameters of the conditional densities necessary for calculating the test statistics, where dependence is on previous observations and a specific source model. It is assumed that at most one type of signal and one type of interference are present at a given time, though this SRCT method may be continuously applied to incoming streaming data in order to detect and classify new transient sources, possibly of different types. To this end, the method developed in this paper also performs separation of the signal and interference portions of the measurement and produces estimates of their signatures in isolation; a property that is not shared by other methods mentioned before.

This paper is organized as follows. Section II introduces the observation model used as the basis for developing the SRCT method. Section III formally introduces the problem and discusses the entire SRCT source characterization framework, including the general form of the LLRT, the procedure for calculating the parameters necessary to construct a LLRT, and the explicit form of each of the LLRTs that are hierarchically applied to individual observations. Section IV provides the experimental results of applying the SRCT method to real acoustical vector observation sequences, that represent national park soundscape recordings. Finally, Section V provides concluding remarks.

II. Observation Model

The kth time slice of a TTR, which contains N transform coefficients, is referred to as an observation vector and is denoted by $\mathbf{z}_k = [z_{kN+1} \cdots z_{(k+1)N}] = \mathbf{T}\mathbf{u}_k$, where $\mathbf{u}_k \in \mathbb{R}^L$ is the kth time interval (contains samples $[kL+1, (k+1)L]$) of the original sampled audio recording, and $\mathbf{T} \in \mathbb{R}^{N \times L}$, with $N \leq L$, represents the time-transform mapping, as it has linearly independent rows that are sampled basis waveforms, e.g., discrete cosine or wavelet basis vectors. Since the soundscape is being constantly recorded, $\mathbf{z}_k$'s are continually arriving, and can be generally represented as

$$\mathbf{z}_k = \alpha_k \mathbf{s}_k + \beta_k \mathbf{h}_k + \boldsymbol{\omega}_k$$

where $\mathbf{s}_k \in \mathbb{R}^N$ and $\mathbf{h}_k \in \mathbb{R}^N$ are random vectors defined similar to $\mathbf{z}_k$, but represent the signatures produced by the unknown signal and interference sources to be characterized, respectively; α_k and β_k are binary variables that indicate the presence $(\alpha_k, \beta_k = 1)$ or absence $(\alpha_k, \beta_k = 0)$ of a signal and interference source, respectively; and $\boldsymbol{\omega}_k$ is an i.i.d. Gaussian measurement noise vector with known statistics, $E[\boldsymbol{\omega}_k] = \mathbf{0}$ and $E[\boldsymbol{\omega}_k \boldsymbol{\omega}_{k-j}^T] = \mathbf{R}_{\boldsymbol{\omega}} \delta(j)$, where $\delta(j)$ represents the Kronecker delta function and $\mathbf{R}_{\boldsymbol{\omega}} \in \mathbb{R}^{N \times N}$ is a full rank noise covariance matrix. The distribution of $\boldsymbol{\omega}_k$ stems from the assumption that noise in $\mathbf{u}_k$ is i.i.d. Gaussian with zero mean and variance σ^2, and hence, we have $\boldsymbol{\omega}_k \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{T}\mathbf{T}^T)$. Although, for the present application, signal and interference are associated with man-made and natural sources, respectively, in general they are user-defined categories that should be complementary, such that one source from each category may be simultaneously present.

For reasons discussed in the next section, it is more convenient to operate on transformed observations $\mathbf{y}_k = \mathbf{R}_{\boldsymbol{\omega}}^{-(1/2)} \mathbf{z}_k$ with white observation noise $\mathbf{w}_k = \mathbf{R}_{\boldsymbol{\omega}}^{-(1/2)} \boldsymbol{\omega}_k$, i.e. $E[\mathbf{w}_k \mathbf{w}_{k-j}^T] = \mathbf{I}_N \delta(j)$, where $\mathbf{I}_N$ is the $N \times N$ identity matrix. Here, it is assumed that the transformed signal vector $\mathbf{R}_{\boldsymbol{\omega}}^{-(1/2)} \mathbf{s}_k$ and transformed interference vector $\mathbf{R}_{\boldsymbol{\omega}}^{-(1/2)} \mathbf{h}_k$ lie in known low dimensional subspaces, $\langle \mathbf{S}_k \rangle$ and $\langle \mathbf{H}_k \rangle$, respectively, that are spanned by the columns of $\mathbf{S}_k \in \mathbb{R}^{N \times M}$ and $\mathbf{H}_k \in \mathbb{R}^{N \times M}$, respectively, with $M \ll N$. These subspaces are time-dependent since the source types may change. Specifically, $\mathbf{S}_k \in \{\mathbf{S}^{(p)}\}_{p=1}^P$ and $\mathbf{H}_k \in \{\mathbf{H}^{(q)}\}_{q=1}^Q$, meaning there are P possible signal types and Q possible interference types. All source subspaces, $\langle \mathbf{S}^{(p)} \rangle$'s and $\langle \mathbf{H}^{(q)} \rangle$'s, are assumed to be linearly independent of each other (not necessarily orthogonal)

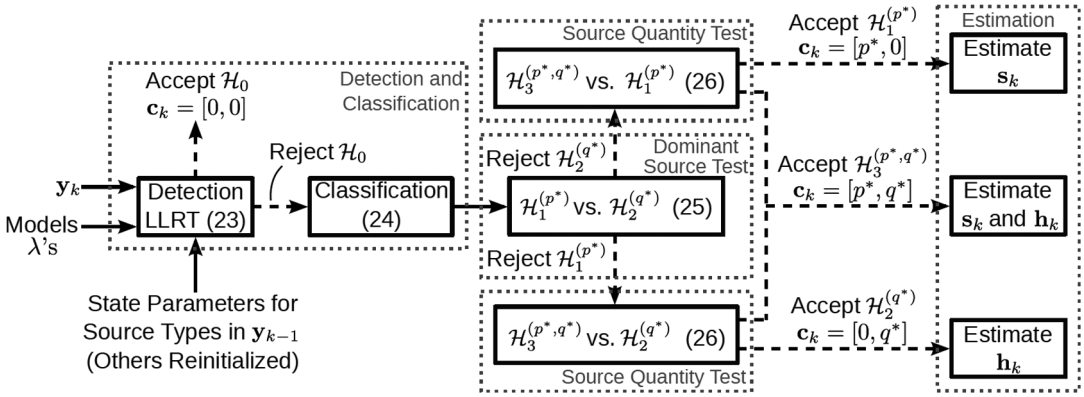


Fig. 1. Block diagram showing application of the proposed SRCT method to the observation $\mathbf{y}_k$. Dashed lines indicate that a path is followed only when the corresponding decision to accept or reject a given hypothesis is made. The dotted boxes indicate the stage in the hierarchical process.

and exactly M-dimensional, where the latter requirement is necessary for subspace classifiers [20] in order to prevent bias in favor of those source types with larger associated subspace dimensions. The model for the transformed observation can therefore be written as

$$\mathbf{y}_k = \alpha_k \mathbf{S}_k \mathbf{a}_k + \beta_k \mathbf{H}_k \mathbf{b}_k + \mathbf{w}_k \tag{1}$$

where $\mathbf{a}_k \in \mathbb{R}^M$ and $\mathbf{b}_k \in \mathbb{R}^M$ are random signal and interference coefficient vectors. Hereafter, $\mathbf{y}_k$ (rather than $\mathbf{z}_k$) is referred to as the observation at time k.

A subspace model is useful for acoustical source classification from TTR since different linear combinations of basis vectors may be used to capture various nonstationarities that are often present in the source's signatures. For instance, when using the discrete cosine transform to form the TTR, the time-frequency signatures of a Doppler shifted waveform may change from narrowband to broadband during the time the most rapid frequency shifts occur, and different linear combinations of basis vectors can be used to model this time-varying behavior. Thus, to represent variations among different source events, it is assumed that $\mathbf{a}_k$ and $\mathbf{b}_k$ obey the following respective vector linear autoregressive (AR) models

$$\mathbf{a}_k = \sum_{j=1}^{J} \boldsymbol{\Phi}_{j,k} \mathbf{a}_{k-j} + \boldsymbol{\xi}_k, \quad \mathbf{b}_k = \sum_{j=1}^{J} \boldsymbol{\Psi}_{j,k} \mathbf{b}_{k-j} + \boldsymbol{\nu}_k \tag{2}$$

where $\boldsymbol{\Phi}_{j,k}$ and $\boldsymbol{\Psi}_{j,k}$ are the jth AR parameter matrices for the signal and interference sources at time k, respectively. The vectors $\boldsymbol{\xi}_k$ and $\boldsymbol{\nu}_k$ are zero mean temporally uncorrelated (but potentially colored element-wise) Gaussian driving processes for the signal and interference, respectively, with known covariance matrices $\mathbf{R}_{\boldsymbol{\xi},k}$ and $\mathbf{R}_{\boldsymbol{\nu},k}$, and are assumed to be independent of each other and with $\mathbf{w}_k$. The time index k on the AR parameters again indicates that they may change over time when a new source type becomes extant. Although in (2) all source coefficient AR models are assumed to be of order J for notational simplicity, the proposed method can still apply for different AR model orders for each source type.

III. Source Characterization

The SRCT method introduced in this section performs transient detection, classification, and estimation of multiple source types by sequentially evaluating $\mathbf{y}_k$'s on-line, i.e., $\mathbf{y}_{k+j}$, $j > 0$ is not accessible at time k. This is done by applying a hierarchy of LLRTs to each $\mathbf{y}_k$ to test between several hypotheses that account for each possible observation composition in terms of signal and interference according to the model in (1). Specifically, these hypotheses are given by

$$\begin{aligned}
\mathcal{H}_0 &: \mathbf{y}_k = \mathbf{w}_k \ \text{ or } \ \alpha_k = \beta_k = 0 \\
\mathcal{H}_1^{(p)} &: \mathbf{y}_k = \mathbf{S}^{(p)} \mathbf{a}_k + \mathbf{w}_k \ \text{ or } \ \alpha_k = 1, \ \ \beta_k = 0 \\
\mathcal{H}_2^{(q)} &: \mathbf{y}_k = \mathbf{H}^{(q)} \mathbf{b}_k + \mathbf{w}_k \ \text{ or } \ \alpha_k = 0, \ \ \beta_k = 1 \\
\mathcal{H}_3^{(p,q)} &: \mathbf{y}_k = \mathbf{S}^{(p)} \mathbf{a}_k + \mathbf{H}^{(q)} \mathbf{b}_k + \mathbf{w}_k \ \text{ or } \ \alpha_k = \beta_k = 1
\end{aligned} \tag{3}$$

where superscripts $p \in \{1, \ldots, P\}$ and/or $q \in \{1, \ldots, Q\}$ indicate dependencies of hypotheses on source models. Thus, there are a total of $P + Q + PQ$ different possible hypotheses excluding $\mathcal{H}_0$. The models associated with the above hypotheses are specified as

$$\begin{aligned}
\mathcal{H}_0 &: \lambda_0 \\
\mathcal{H}_1^{(p)} &: \lambda_1^{(p)} = \left\{ \mathbf{S}^{(p)}, \boldsymbol{\Phi}_1^{(p)}, \ldots, \boldsymbol{\Phi}_J^{(p)}, \mathbf{R}_{\boldsymbol{\xi}}^{(p)} \right\} \\
\mathcal{H}_2^{(q)} &: \lambda_2^{(q)} = \left\{ \mathbf{H}^{(q)}, \boldsymbol{\Psi}_1^{(q)}, \ldots, \boldsymbol{\Psi}_J^{(q)}, \mathbf{R}_{\boldsymbol{\nu}}^{(q)} \right\} \\
\mathcal{H}_3^{(p,q)} &: \lambda_3^{(p,q)} = \lambda_1^{(p)} \cup \lambda_2^{(q)}
\end{aligned} \tag{4}$$

which are formed using training data prior to the application of the SRCT method.

The flow of the SRCT method is shown in the block diagram of Fig. 1. As can be seen, the inputs to the system are $\mathbf{y}_k$, the models defined in (4), and a set of state parameters that allows the system to exploit the dependency structure of each source (see Section III-B). Ultimately, the system produces one signal and/or one interference label for $\mathbf{y}_k$, in addition to estimates of $\mathbf{s}_k$ and/or $\mathbf{h}_k$, if desired. The test statistics that implement the SRCT method are log likelihood ratios (LLR) formed using the conditional probabilities of observing $\mathbf{y}_k$ given relevant previous observations and specific models in (4). The associated LLRTs are applied hierarchically to a given $\mathbf{y}_k$ as follows

1) **Detection and Classification**: using (23), determine whether $\mathbf{y}_k$ consists of noise alone or contains the signatures of one or two sources; if the latter, then reject $\mathcal{H}_0$ and proceed. The source types, p^* and q^*, that are most likely present in $\mathbf{y}_k$ are also estimated by (24).
2) **Dominant Source Test**: use (25) to determine whether signal or interference is dominant and, consequently,

which remaining single source hypothesis ($\mathcal{H}_1^{(p^*)}$ or $\mathcal{H}_2^{(q^*)}$) is rejected.

3) **Source Quantity Test**: use (26) to test the hypothesis corresponding to the dominant source type against the two source hypothesis $\mathcal{H}_3^{(p^*,q^*)}$ in order to finally accept either $\mathcal{H}_1^{(p^*)}$, $\mathcal{H}_2^{(q^*)}$, or $\mathcal{H}_3^{(p^*,q^*)}$.
4) **Estimation**: Use estimates of $\mathbf{a}_k$ and $\mathbf{b}_k$, that were generated for the above LLRTs, to form estimates of the actual source signatures. This step is optional and does not involve a LLRT.

A hierarchy of tests is applied here rather than directly finding the most likely hypothesis in (4). This is done due to the fact that the complexity of the observation model (number of parameters) varies between hypotheses, which would lead to a bias in LLRTs in favor of more complex hypotheses [21]. A hierarchical test circumvents these issues through the use of different thresholds to compare hypotheses with different relative complexities.

Since each main step in the SRCT process evaluates different hypotheses using the same general form of the LLRT, this form is briefly discussed in the next subsection. Sections III-B and III-C discuss calculating distribution parameters for each hypothesis, that are used to form the LLRTs. Explicit forms of the LLRTs are then developed in Section III-D. To conclude this section, the procedure for obtaining estimates of the TTR signatures of detected signal and interference sources is briefly discussed.

A. General Form of a LLR

In this paper, a hierarchy of tests is applied to $\mathbf{y}_k$, each of which is based on a likelihood ratio having the general form

$$L_k\left(\boldsymbol{\theta}_k,\boldsymbol{\theta}_k'\right)=\frac{l(\boldsymbol{\theta}_k;\mathbf{y}_k)}{l\left(\boldsymbol{\theta}_k';\mathbf{y}_k\right)},\quad \boldsymbol{\theta}_k\neq\boldsymbol{\theta}_k' =\frac{f\left(\mathbf{y}_k|\mathbf{Y}_{k_0}^{k-1},\lambda\right)}{f\left(\mathbf{y}_k|\mathbf{Y}_{k_0'}^{k-1},\lambda'\right)},\quad \lambda\neq\lambda' \tag{5}$$

where $l(\boldsymbol{\theta}_k;\mathbf{y}_k)$ is the likelihood function of the *distribution parameter* $\boldsymbol{\theta}_k$ (describes conditional PDFs), f is a conditional density function, λ is one of the source models in (4), and $\mathbf{Y}_{k_0}^{k-1}=\{\mathbf{y}_{k-1},\ldots,\mathbf{y}_{k_0}\}$ is the set observations that $\mathbf{y}_k$ is dependent on under the hypothesis associated with λ. In other words, if a particular source type has been present since time $k_0\leq k-1$, then $\mathbf{y}_k$ will be dependent on $\mathbf{Y}_{k_0}^{k-1}$ under any hypothesis that assumes the presence of this source, owing to dependence between source coefficient vectors at different times, as shown in (2).

The statistic in (5) may be used for the test

$$L_k\left(\boldsymbol{\theta}_k,\boldsymbol{\theta}_k'\right)\begin{array}{c}\text{reject }\mathcal{H}'\\ \gtrless\\ \text{reject }\mathcal{H}\end{array}\gamma' \tag{6}$$

where $\mathcal{H}$ and $\mathcal{H}'$ are hypotheses from (3) associated with the models λ and λ', respectively, and γ' is a predetermined threshold. Equations (5) and (6) show that λ implicitly defines the hypothesis a likelihood function is dependent on, as well as the distribution parameter $\boldsymbol{\theta}_k$. Such tests are used to assign class labels to each $\mathbf{y}_k$ separately, as opposed to accumulating evidence to make a decision based on the likelihood of observing the entire sequence $\mathbf{Y}_{k_0}^{k}$ under a given λ, as in a traditional sequential LLRT [18]. Although the latter approach is possible under the proposed framework, it is avoided since both likelihoods in a LLRT must be found given the same set of observations [22]. On the other hand, since the presence and types of signal and interference change independently as new observations arrive, the set of observations that can reasonably occur under each hypothesis is often not identical. Therefore, using a common $\mathbf{Y}_{k_0}^{k}$, $k_0<k$ to evaluate all likelihoods can lead to low likelihoods for hypotheses associated with sources that are actually present.

The next subsection shows that $\mathbf{y}_k$ is conditionally multivariate Gaussian with mean vector $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$, given $\mathbf{Y}_{k_0}^{k-1}$ and λ, owing to the fact that all observations follow the general model introduced in the previous section. Therefore, defining $\boldsymbol{\theta}_k=\{\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k\}$ and taking the natural log of (5) yields the LLR as

$$\Lambda_k\left(\boldsymbol{\theta}_k,\boldsymbol{\theta}_k'\right)=\ln\frac{l(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k;\mathbf{y}_k)}{l\left(\boldsymbol{\mu}_k',\boldsymbol{\Sigma}_k';\mathbf{y}_k\right)}=\boldsymbol{\zeta}_k\left(\boldsymbol{\theta}_k'\right)-\boldsymbol{\zeta}_k(\boldsymbol{\theta}_k) \tag{7}$$

where

$$\zeta_k(\boldsymbol{\theta}_k)=\frac{1}{2}\ln\det(\boldsymbol{\Sigma}_k)+\frac{1}{2}(\mathbf{y}_k-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}(\mathbf{y}_k-\boldsymbol{\mu}_k) \tag{8}$$

and similarly for $\boldsymbol{\zeta}_k(\boldsymbol{\theta}_k')$. Since all tests performed by the SRCT method use a LLR that assumes the general form in (7), generating $\boldsymbol{\theta}_k$ under different λ (i.e. different hypotheses) is an integral step of this process, and is discussed next.

B. Generating Parameter Sets

This subsection discusses calculating $\boldsymbol{\zeta}_k(\boldsymbol{\theta}_k)$ for a given λ, which involves calculating the associated $\boldsymbol{\theta}_k$, i.e., the parameters of the conditional distribution. Beginning with the simple case of $\mathcal{H}_0$ (i.e., $\lambda=\lambda_0$), where $\mathbf{y}_k=\mathbf{w}_k$, we have

$$\begin{aligned}\mathcal{H}_0:\boldsymbol{\mu}_k&=E\left[\mathbf{y}_k|\mathbf{Y}_{k_0}^{k-1},\lambda_0\right]=E[\mathbf{w}_k]=\mathbf{0}\\ \boldsymbol{\Sigma}_k&=E\left[(\mathbf{y}_k-\boldsymbol{\mu}_k)(\mathbf{y}_k-\boldsymbol{\mu}_k)^T|\mathbf{Y}_{k_0}^{k-1},\lambda_0\right]\\ &=E\left[\mathbf{w}_k\mathbf{w}_k^T\right]=\mathbf{I}_N\end{aligned}$$

due to the fact that $\mathbf{w}_k$'s are independent at different times. From (8), it follows that $\boldsymbol{\zeta}_k(\boldsymbol{\theta}_k)=(1/2)\mathbf{y}_k^T\mathbf{y}_k$ under $\mathcal{H}_0$.

Finding the conditional probability of observing $\mathbf{y}_k$ given $\lambda\neq\lambda_0$ is accomplished by using such a model to establish a series of state equations and applying a Kalman filter to obtain the estimates of the time dependent source signatures. More specifically, the idea is to obtain an estimate of the state vector $\mathbf{x}_k$ that represents the relevant past and present source basis coefficient vectors ($\mathbf{a}_{k-j}$'s and/or $\mathbf{b}_{k-j}$'s), under a given hypothesis in (3). See Table I for explicit definitions of the state variables for each model in (4). For a specific λ, the general forms of the state equations are given by

$$\mathbf{x}_k=\mathbf{F}\mathbf{x}_{k-1}+\mathbf{D}\mathbf{v}_k \tag{9a}$$

$$\mathbf{y}_k=\mathbf{C}\mathbf{x}_k+\mathbf{w}_k \tag{9b}$$

where $\mathbf{F}$, $\mathbf{D}$, and $\mathbf{C}$ are matrices of appropriate forms and dimensions (see Table I), and $\mathbf{v}_k$ is the process noise vector with

TABLE I
STRUCTURE OF THE STATE VARIABLES FOR EACH MODEL WHERE $\mathbf{0}_M$ AND $\mathbf{0}_{M_1\times M_2}$ ARE $M\times M$ AND $M_1\times M_2$ ZERO MATRICES, RESPECTIVELY

Model	$\mathbf{x}_k$	$\mathbf{F}$	$\mathbf{D}$	$\mathbf{C}$	$\mathbf{v}_k$	$\mathbf{R_v}$
$\lambda_1^{(p)}$	$\mathbf{x}_{1,k}=\begin{bmatrix}\mathbf{a}_k\\ \vdots\\ \mathbf{a}_{k-J+1}\end{bmatrix}$	$\mathbf{F}_1^{(p)}=\begin{bmatrix}\boldsymbol{\Phi}_1^{(p)} & \cdots & \boldsymbol{\Phi}_J^{(p)}\\ \mathbf{I}_{(J-1)M} & & \mathbf{0}_{(J-1)M\times M}\end{bmatrix}$	$\begin{bmatrix}\mathbf{I}_M\\ \mathbf{0}_{(J-1)M\times M}\end{bmatrix}$	$\mathbf{S}^{(p)}\mathbf{D}^T$	$\boldsymbol{\xi}_k^{(p)}$	$\mathbf{R}_{\boldsymbol{\xi}}^{(p)}$
$\lambda_2^{(q)}$	$\mathbf{x}_{2,k}=\begin{bmatrix}\mathbf{b}_k\\ \vdots\\ \mathbf{b}_{k-J+1}\end{bmatrix}$	$\mathbf{F}_2^{(q)}=\begin{bmatrix}\boldsymbol{\Psi}_1^{(q)} & \cdots & \boldsymbol{\Psi}_J^{(q)}\\ \mathbf{I}_{(J-1)M} & & \mathbf{0}_{(J-1)M\times M}\end{bmatrix}$	$\begin{bmatrix}\mathbf{I}_M\\ \mathbf{0}_{(J-1)M\times M}\end{bmatrix}$	$\mathbf{H}^{(q)}\mathbf{D}^T$	$\boldsymbol{\nu}_k^{(q)}$	$\mathbf{R}_{\boldsymbol{\nu}}^{(q)}$
$\lambda_3^{(p,q)}$	$\begin{bmatrix}\mathbf{x}_{1,k}\\ \mathbf{x}_{2,k}\end{bmatrix}$	$\begin{bmatrix}\mathbf{F}_1^{(p)} & \mathbf{0}_{JM}\\ \mathbf{0}_{JM} & \mathbf{F}_2^{(q)}\end{bmatrix}$	$\begin{bmatrix}\mathbf{I}_M & \mathbf{0}_{JM\times M}\\ & I_M\\ \mathbf{0}_{(2J-1)M\times M} & \mathbf{0}_{(J-1)M\times M}\end{bmatrix}$	$\left[\mathbf{S}^{(p)},\mathbf{H}^{(q)}\right]\mathbf{D}^T$	$\begin{bmatrix}\boldsymbol{\xi}_k^{(p)}\\ \boldsymbol{\nu}_k^{(q)}\end{bmatrix}$	$\begin{bmatrix}\mathbf{R}_{\boldsymbol{\xi}}^{(p)} & \mathbf{0}_M\\ \mathbf{0}_M & \mathbf{R}_{\boldsymbol{\nu}}^{(q)}\end{bmatrix}$

known covariance matrix $\mathbf{R_v}$. These equations are obtained by arranging the observation model equations in (1) and (2) in state space form, and replacing the parameters in the these equations with those from one of the models $\lambda_1^{(p)}$, $\lambda_2^{(q)}$, or $\lambda_3^{(p,q)}$. Using a given set of state equations, a standard Kalman filter can be applied to obtain the *a priori* and *a posteriori* estimates of the source coefficients (states), denoted by $\hat{\mathbf{x}}_{k|k-1}$ and $\hat{\mathbf{x}}_{k|k}$, respectively. The associated error covariance matrices are denoted by

$$\mathbf{Q}_{k|k-1}=E\left[\boldsymbol{\epsilon}_{k|k-1}\boldsymbol{\epsilon}_{k|k-1}^T\right],\quad \mathbf{Q}_{k|k}=E\left[\boldsymbol{\epsilon}_{k|k}\boldsymbol{\epsilon}_{k|k}^T\right] \tag{10}$$

where $\boldsymbol{\epsilon}_{k|k-1}=\mathbf{x}_k-\hat{\mathbf{x}}_{k|k-1}$ and $\boldsymbol{\epsilon}_{k|k}=\mathbf{x}_k-\hat{\mathbf{x}}_{k|k}$ are the *a priori* and *a posteriori* state error vectors, respectively.

To find $\boldsymbol{\theta}_k=\{\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k\}$ for a given λ, the definitions of $\boldsymbol{\epsilon}_{k|k-1}$ and the state variables in Table I may be used to write (9b) as

$$\mathbf{y}_k=\mathbf{C}(\hat{\mathbf{x}}_{k|k-1}+\boldsymbol{\epsilon}_{k|k-1})+\mathbf{w}_k. \tag{11}$$

Standard Kalman filter theory [23] dictates that, given the past observations $\mathbf{Y}_{k_0}^{k-1}$, we have $\boldsymbol{\epsilon}_{k|k-1}\sim\mathcal{N}(\mathbf{0},\mathbf{Q}_{k|k-1})$. Since $\boldsymbol{\epsilon}_{k|k-1}$ and $\mathbf{w}_k$ are both zero mean Gaussian, and together represent the random part of (11), $\mathbf{y}_k$ is also conditionally multivariate Gaussian, given $\mathbf{Y}_{k_0}^{k-1}$ and λ. In particular, $f(\mathbf{y}_k|\mathbf{Y}_{k_0}^{k-1},\lambda)$ is parameterized by the mean vector [23]

$$\begin{aligned}\boldsymbol{\mu}_k&=\mathbf{C}E\left[\mathbf{x}_k|\mathbf{Y}_{k_0}^{k-1},\lambda\right]\\&=\mathbf{C}\hat{\mathbf{x}}_{k|k-1}=\mathbf{C}\mathbf{F}\hat{\mathbf{x}}_{k-1|k-1}\end{aligned} \tag{12}$$

and covariance matrix

$$\begin{aligned}\boldsymbol{\Sigma}_k&=E\left[(\mathbf{C}\boldsymbol{\epsilon}_{k|k-1}+\mathbf{w}_k)(\mathbf{C}\boldsymbol{\epsilon}_{k|k-1}+\mathbf{w}_k)^T|\mathbf{Y}_{k_0}^{k-1},\lambda\right]\\&=\mathbf{C}E\left[\boldsymbol{\epsilon}_{k|k-1}\boldsymbol{\epsilon}_{k|k-1}^T|\mathbf{Y}_{k_0}^{k-1},\lambda\right]\mathbf{C}^T+E\left[\mathbf{w}_k\mathbf{w}_k^T\right]\\&=\mathbf{C}\mathbf{Q}_{k|k-1}\mathbf{C}^T+\mathbf{I}_N\end{aligned} \tag{13}$$

where $\hat{\mathbf{x}}_{k-1|k-1}$ is the *a posteriori* state estimate found at time $k-1$, and the expectation in each case is over an ensemble set of source event realizations with a similar structure. Equation (12) uses the fact that $\mathbf{w}_k$ is independent of the source model and previous observations (regardless of their composition), while (13) exploits $E[\boldsymbol{\epsilon}_{k|k-1}\mathbf{w}_k^T|\mathbf{Y}_{k_0}^{k-1},\lambda]=\mathbf{0}$ due to conditional independence of $\mathbf{w}_k$ and source signatures.

It is clear that a separate Kalman filter for each hypothesis in (3), except for $\mathcal{H}_0$, is needed to generate a corresponding $\boldsymbol{\zeta}_k(\boldsymbol{\theta}_k)$, as they all assume different observation compositions. Therefore, *a posteriori* values, $\hat{\mathbf{x}}_{k-1|k-1}$ and $\mathbf{Q}_{k-1|k-1}$, should only be used for estimating $\mathbf{x}_{k|k}$ for hypotheses associated with source types that were determined to be present at time $k-1$ (also see Fig. 1). The state vector for every other hypothesis should be estimated using $\hat{\mathbf{x}}_{k_0-1|k_0-1}$ and $\mathbf{Q}_{k_0-1|k_0-1}$, that are reinitialized values of the state parameters, and are derived in the next subsection. The reason being such hypotheses essentially assume the onset of a new source type, and hence, there are no valid past states to use. Thus, reinitialization sets k_0 for a given hypothesis, and correspondingly determines the past observations used to compute a likelihood, as in (5). This implies that, when there are many source types, most $\boldsymbol{\zeta}_k(\boldsymbol{\theta}_k)$ will be generated using reinitialized state parameters. However, it is shown below that this process is simple, as the initial state error covariance and state vector estimator have a closed form for each hypothesis.

C. Initializing State Parameters

This subsection derives explicit forms of $\hat{\mathbf{x}}_{k_0-1|k_0-1}$ and $\mathbf{Q}_{k_0-1|k_0-1}$, that are initial values of the state vector estimate and error covariance matrix, respectively. As mentioned before, these initial parameters are needed for generating $\boldsymbol{\zeta}_k(\boldsymbol{\theta}_k)$ for any hypothesis that assumes the presence of a given source type in $\mathbf{y}_k$ that was absent in $\mathbf{y}_{k-1}$, according to the results of applying the SRCT method to this prior observation.

Proposition 1: The error covariance matrices under the single source hypotheses are initialized as

$$\begin{aligned}\mathcal{H}_1^{(p)}:\mathbf{Q}_{k_0-1|k_0-1}&=\mathbf{I}_J\otimes\left(\mathbf{S}^{(p)T}\mathbf{S}^{(p)}\right)^{-1}\\\mathcal{H}_2^{(q)}:\mathbf{Q}_{k_0-1|k_0-1}&=\mathbf{I}_J\otimes\left(\mathbf{H}^{(q)T}\mathbf{H}^{(q)}\right)^{-1}\end{aligned} \tag{14}$$

where $\otimes$ denotes the Kronecker product. Corresponding state vectors are initialized using (15) and (16).

Proof: The state vector estimate under a single source hypothesis can be initialized with the linear least squares estimates (or maximum likelihood estimates (MLE) since the observation noise is Gaussian) of the coefficients of $\mathbf{y}_{k_0-j}$, $j=1,\ldots,J$ relative to the appropriate basis vectors, i.e.

$$\begin{aligned}\mathcal{H}_1^{(p)}:\hat{\mathbf{a}}_{k_0-j}&=\left(\mathbf{S}^{(p)T}\mathbf{S}^{(p)}\right)^{-1}\mathbf{S}^{(p)T}\mathbf{y}_{k_0-j}\\&=\mathbf{S}^{(p)\dagger}\mathbf{y}_{k_0-j}\\\mathcal{H}_2^{(q)}:\hat{\mathbf{b}}_{k_0-j}&=\left(\mathbf{H}^{(q)T}\mathbf{H}^{(q)}\right)^{-1}\mathbf{H}^{(q)T}\mathbf{y}_{k_0-j}\\&=\mathbf{H}^{(q)\dagger}\mathbf{y}_{k_0-j}\end{aligned} \tag{15}$$

where $\mathbf{S}^{(p)\dagger}$ and $\mathbf{H}^{(q)\dagger}$ denote the Moore-Penrose inverses [24] of $\mathbf{S}^{(p)}$ and $\mathbf{H}^{(q)}$, respectively, which always exist and

may be calculated using the given explicit forms owing to the columns of $\mathbf{S}^{(p)}$ and $\mathbf{H}^{(q)}$ being linearly independent, as stated in Section II. Now, the state vector (see Table I) estimates are initialized using the coefficient vector estimates in (15) as

$$\begin{aligned}\mathcal{H}_1^{(p)} &: \hat{\mathbf{x}}_{k_0-1|k_0-1} = \left[\hat{\mathbf{a}}_{k_0-1}^T \cdots \hat{\mathbf{a}}_{k_0-J}^T\right]^T \\ \mathcal{H}_2^{(q)} &: \hat{\mathbf{x}}_{k_0-1|k_0-1} = \left[\hat{\mathbf{b}}_{k_0-1}^T \cdots \hat{\mathbf{b}}_{k_0-J}^T\right]^T. \end{aligned} \tag{16}$$

To derive $\mathbf{Q}_{k_0-1|k_0-1} = E[\boldsymbol{\epsilon}_{k_0-1|k_0-1}\boldsymbol{\epsilon}_{k_0-1|k_0-1}^T]$, under $\mathcal{H}_1^{(p)}$, first note that in this case

$$\begin{aligned}\hat{\mathbf{x}}_{k_0-1|k_0-1} - \mathbf{x}_{k_0-1} &= \begin{bmatrix}\mathbf{S}^{(p)\dagger}\mathbf{y}_{k_0-1}\\ \vdots \\ \mathbf{S}^{(p)\dagger}\mathbf{y}_{k_0-J}\end{bmatrix} - \begin{bmatrix}\mathbf{a}_{k_0-1}\\ \vdots \\ \mathbf{a}_{k_0-J}\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{S}^{(p)\dagger}\mathbf{w}_{k_0-1}\\ \vdots \\ \mathbf{S}^{(p)\dagger}\mathbf{w}_{k_0-J}\end{bmatrix}.\end{aligned}$$

Now, since $E[\mathbf{w}_k\mathbf{w}_{k-j}^T] = \mathbf{I}_N\delta(j)$, from (10) we have

$$\mathcal{H}_1^{(p)} : \mathbf{Q}_{k_0-1|k_0-1} = \mathbf{I}_J \otimes \mathbf{S}^{(p)\dagger}\mathbf{I}_N\left(\mathbf{S}^{(p)\dagger}\right)^T$$

which can be reduced to (14). Similar steps are used to derive $\mathbf{Q}_{k_0-1|k_0-1}$ under $\mathcal{H}_2^{(q)}$. ■

Note that (15) motivates transforming the original observation to $\mathbf{y}_k = \mathbf{R}_{\boldsymbol{\omega}}^{-(1/2)}\mathbf{z}_k$, as this ensures the elements of the transformed observation noise $\mathbf{w}_k$ are uncorrelated and the coefficient estimators in (15) have minimum variance [25].

Proposition 2: The error covariance matrix under a dual source hypothesis is initialized as

$$\begin{aligned}&\mathcal{H}_3^{(p,q)} : \mathbf{Q}_{k_0-1|k_0-1} \\ &= \begin{bmatrix}\mathbf{I}_J \otimes \left(\mathbf{S}^{(p)T}\mathbf{P}_{\mathbf{H}}^{(q)\perp}\mathbf{S}^{(p)}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_J \otimes \left(\mathbf{H}^{(q)T}\mathbf{P}_{\mathbf{S}}^{(p)\perp}\mathbf{H}^{(q)}\right)^{-1}\end{bmatrix}\end{aligned} \tag{17}$$

where $\mathbf{P}_{\mathbf{S}}^{(p)\perp} = \mathbf{I}_N - \mathbf{S}^{(p)}(\mathbf{S}^{(p)T}\mathbf{S}^{(p)})^{-1}\mathbf{S}^{(p)T}$ and $\mathbf{P}_{\mathbf{H}}^{(q)\perp} = \mathbf{I}_N - \mathbf{H}^{(q)}(\mathbf{H}^{(q)T}\mathbf{H}^{(q)})^{-1}\mathbf{H}^{(q)T}$ project onto the orthogonal complements of the subspaces $\langle\mathbf{S}^{(p)}\rangle$ and $\langle\mathbf{H}^{(q)}\rangle$, respectively. The corresponding state vector is initialized using (19) and (20).

Proof: Ideally, under $\mathcal{H}_3^{(p,q)}$, $\hat{\mathbf{x}}_{k_0-1|k_0-1}$ should contain signal and interference coefficients that are free from the effects of each other. The MLEs of $\mathbf{a}_{k_0-j}$'s and $\mathbf{b}_{k_0-j}$'s under this two source assumption can be obtained using oblique projection [25] matrices

$$\begin{aligned}\mathbf{E}_{\mathbf{S}}^{(p,q)} &= \mathbf{S}^{(p)}\left(\mathbf{S}^{(p)T}\mathbf{P}_{\mathbf{H}}^{(q)\perp}\mathbf{S}^{(p)}\right)^{-1}\mathbf{S}^{(p)T}\mathbf{P}_{\mathbf{H}}^{(q)\perp} \\ \mathbf{E}_{\mathbf{H}}^{(p,q)} &= \mathbf{H}^{(q)}\left(\mathbf{H}^{(q)T}\mathbf{P}_{\mathbf{S}}^{(p)\perp}\mathbf{H}^{(q)}\right)^{-1}\mathbf{H}^{(q)T}\mathbf{P}_{\mathbf{S}}^{(p)\perp}.\end{aligned}$$

The matrices $\mathbf{E}_{\mathbf{S}}^{(p,q)}$ and $\mathbf{E}_{\mathbf{H}}^{(p,q)}$ have respective range spaces $\langle\mathbf{S}^{(p)}\rangle$ and $\langle\mathbf{H}^{(q)}\rangle$ and respective null spaces $\langle\mathbf{H}^{(q)}\rangle$ and $\langle\mathbf{S}^{(p)}\rangle$, and hence, we have

$$\begin{aligned}\mathcal{H}_3^{(p,q)} &: \mathbf{E}_{\mathbf{S}}^{(p,q)}\mathbf{y}_{k_0-j} = \mathbf{S}^{(p)}\mathbf{a}_{k_0-j} + \mathbf{E}_{\mathbf{S}}^{(p,q)}\mathbf{w}_{k_0-j} \\ \mathcal{H}_3^{(p,q)} &: \mathbf{E}_{\mathbf{H}}^{(p,q)}\mathbf{y}_{k_0-j} = \mathbf{H}^{(q)}\mathbf{b}_{k_0-j} + \mathbf{E}_{\mathbf{H}}^{(p,q)}\mathbf{w}_{k_0-j}.\end{aligned} \tag{18}$$

Now, the MLEs of the signal and interference coefficients under $\mathcal{H}_3^{(p,q)}$ are

$$\begin{aligned}\mathcal{H}_3^{(p,q)} &: \hat{\mathbf{a}}_{k_0-j} = \mathbf{S}^{(p)\dagger}\mathbf{E}_{\mathbf{S}}^{(p,q)}\mathbf{y}_{k_0-j} \\ \mathcal{H}_3^{(p,q)} &: \hat{\mathbf{b}}_{k_0-j} = \mathbf{H}^{(q)\dagger}\mathbf{E}_{\mathbf{H}}^{(p,q)}\mathbf{y}_{k_0-j}\end{aligned} \tag{19}$$

and the state vector estimate is initialized using these coefficients as

$$\mathcal{H}_3^{(p,q)} : \hat{\mathbf{x}}_{k_0-1|k_0-1} = \left[\hat{\mathbf{a}}_{k_0-1}^T \cdots \hat{\mathbf{a}}_{k_0-J}^T\ \hat{\mathbf{b}}_{k_0-1}^T \cdots \hat{\mathbf{b}}_{k_0-J}^T\right]^T. \tag{20}$$

The steps to find $\mathbf{Q}_{k_0-1|k_0-1}$ under $\mathcal{H}_3^{(p,q)}$ are similar to those under $\mathcal{H}_1^{(p)}$, and hence, from (18) we have

$$\begin{aligned}\hat{\mathbf{x}}_{k_0-1|k_0-1} - \mathbf{x}_{k_0-1} &= \begin{bmatrix}\hat{\mathbf{a}}_{k_0-1}\\ \vdots \\ \hat{\mathbf{a}}_{k_0-J}\\ \hat{\mathbf{b}}_{k_0-1}\\ \vdots \\ \hat{\mathbf{b}}_{k_0-J}\end{bmatrix} - \begin{bmatrix}\mathbf{a}_{k_0-1}\\ \vdots \\ \mathbf{a}_{k_0-J}\\ \mathbf{b}_{k_0-1}\\ \vdots \\ \mathbf{b}_{k_0-J}\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{S}^{(p)\dagger}\mathbf{E}_{\mathbf{S}}^{(p,q)}\mathbf{w}_{k_0-1}\\ \vdots \\ \mathbf{S}^{(p)\dagger}\mathbf{E}_{\mathbf{S}}^{(p,q)}\mathbf{w}_{k_0-J}\\ \mathbf{H}^{(q)\dagger}\mathbf{E}_{\mathbf{H}}^{(p,q)}\mathbf{w}_{k_0-1}\\ \vdots \\ \mathbf{H}^{(q)\dagger}\mathbf{E}_{\mathbf{H}}^{(p,q)}\mathbf{w}_{k_0-J}\end{bmatrix}.\end{aligned} \tag{21}$$

Using the fact that

$$\begin{aligned}\mathbf{S}^{(p)\dagger}\mathbf{E}_{\mathbf{S}}^{(p,q)}\mathbf{E}_{\mathbf{S}}^{(p,q)T}\left(\mathbf{S}^{(p)\dagger}\right)^T &= \left(\mathbf{S}^{(p)T}\mathbf{P}_{\mathbf{H}}^{(q)\perp}\mathbf{S}^{(p)}\right)^{-1} \\ \mathbf{H}^{(q)\dagger}\mathbf{E}_{\mathbf{H}}^{(p,q)}\mathbf{E}_{\mathbf{H}}^{(p,q)T}\left(\mathbf{H}^{(q)\dagger}\right)^T &= \left(\mathbf{H}^{(q)T}\mathbf{P}_{\mathbf{S}}^{(p)\perp}\mathbf{H}^{(q)}\right)^{-1}\end{aligned}$$

the covariance of the error term in (21) is found to be as shown in (17). ■

Since $\mathbf{y}_k$'s are processed sequentially, the initial state estimates in (16) and (20) are formed using the J observations prior to a given $\mathbf{y}_{k_0}$. While these observations might not contain strong signatures of the sources associated with these estimates, these are the best linear estimates of the coefficients that can be obtained.

All components necessary for generating $\boldsymbol{\theta}_k$ for a given λ have now been provided, which may subsequently be used to generate a corresponding $\boldsymbol{\zeta}_k(\boldsymbol{\theta}_k)$ using (8). The variables $\boldsymbol{\zeta}_k(\boldsymbol{\theta}_k)$'s may then be used to form test statistics for the hypothesis test in (3), as described next.

D. Determining Source Composition Using LLRTs

This subsection introduces the LLRTs needed to assign a class label vector, $\mathbf{c}_k = [\hat{\alpha}_k p, \hat{\beta}_k q]$, with $p \in \{1, \ldots, P\}$ and $q \in \{1, \ldots, Q\}$, consisting of one signal and one interference label, to $\mathbf{y}_k$. Here " $\hat{}$ " denotes estimate of the indicator variables in (1) that determine the presence of signal and/or interference.

1) Detection and Classification: As indicated by Fig. 1, the first steps for processing $\mathbf{y}_k$ are source detection, where it is decided whether to accept or reject $\mathcal{H}_0$, and source classification, where estimates of the most likely signal and interference source

TABLE II
CLASS LABELS AND INDICATOR ESTIMATES UNDER EACH HYPOTHESIS

	Accepted Hypothesis			
	$\mathcal{H}_0$	$\mathcal{H}_1^{(p^*)}$	$\mathcal{H}_2^{(q^*)}$	$\mathcal{H}_3^{(p^*,q^*)}$
$\left(\hat{\alpha}_k,\ \hat{\beta}_k\right)$	$(0,\ 0)$	$(1,\ 0)$	$(0,\ 1)$	$(1,\ 1)$
$\mathbf{c}_k$	$[0,\ 0]$	$[p^*,\ 0]$	$[0,\ q^*]$	$[p^*,\ q^*]$

types, p^* and q^*, are found. This is accomplished by forming a set of LLR, each of which may be used to test the hypothesis that $\mathbf{y}_k$ consists of noise alone versus the hypothesis that $\mathbf{y}_k$ contains noise plus a type p signal source and/or a type q interference source. Using the general form in (7) with $\boldsymbol{\theta}_k$'s generated using the appropriate models, this LLR is given by

$$\Lambda_k\left(\boldsymbol{\theta}_{3,k}^{(p,q)},\boldsymbol{\theta}_{0,k}\right)=\boldsymbol{\zeta}_k(\boldsymbol{\theta}_{0,k})-\boldsymbol{\zeta}_k\left(\boldsymbol{\theta}_{3,k}^{(p,q)}\right) \tag{22}$$

where $\boldsymbol{\theta}_{0,k}$ and $\boldsymbol{\theta}_{3,k}^{(p,q)}$ are the distribution parameters corresponding to models λ_0 and $\lambda_3^{(p,q)}$, respectively. Detection of sources in $\mathbf{y}_k$ is then performed using the LLRT

$$\max_{(p,q)}\Lambda_k\left(\boldsymbol{\theta}_{3,k}^{(p,q)},\boldsymbol{\theta}_{0,k}\right)\begin{array}{c}\text{reject } \mathcal{H}_0\\ \geq\\ <\\ \text{accept } \mathcal{H}_0\end{array}\gamma \tag{23}$$

where γ is a predetermined threshold that adjusts the sensitivity of the detector. Estimates of the signal and interference source types (classification) are then found using the test statistics already calculated for detection as

$$(p^*,q^*)=\arg\max_{(p,q)}\Lambda_k\left(\boldsymbol{\theta}_{3,k}^{(p,q)},\boldsymbol{\theta}_{0,k}\right)\geq\gamma. \tag{24}$$

Assuming a transient source has been detected at time k, the dominant source and source quantity tests outlined below can subsequently be initiated. In this case, (3) is reduced to testing between $\mathcal{H}_1^{(p^*)}$, $\mathcal{H}_2^{(q^*)}$, and $\mathcal{H}_3^{(p^*,q^*)}$. When hypothesis $\mathcal{H}_0$ is accepted, no further processing is required for $\mathbf{y}_k$ and the next observation $\mathbf{y}_{k+1}$ is evaluated. In this case, we have $\hat{\alpha}_k=\hat{\beta}_k=0$ so that $\mathbf{c}_k=[0,0]$. Justification for using (24) to find p^* and q^* is based on the fact that, for source types that were determined to be absent at time $k-1$, the corresponding $\boldsymbol{\zeta}_k(\boldsymbol{\theta}_{3,k}^{(p,q)})$'s are found using reinitialized state variables, generated using oblique projections to remove the effects of type q interference from the type p signal estimate and vice versa (see (19) in Section III-C). For sources types that were present at time $k-1$, $\boldsymbol{\zeta}_k(\boldsymbol{\theta}_{3,k}^{(p,q)})$ is formed using the best (in the mean squared error sense) linear estimates of the source signatures.

2) Dominant Source Test: When $\mathcal{H}_0$ is rejected for $\mathbf{y}_k$ the next step involves determining whether the dominant source is a type p^* signal or a type q^* interference (see Fig. 1). Note that this step does not exclude the possibility that both sources are simultaneously present, but rather it defines the most likely source in the single source case. To this end, the LLRT for determining the dominant source is given by

$$\Lambda_k\left(\boldsymbol{\theta}_{1,k}^{(p^*)},\boldsymbol{\theta}_{2,k}^{(q^*)}\right)=\boldsymbol{\zeta}_k\left(\boldsymbol{\theta}_{2,k}^{(q^*)}\right)-\boldsymbol{\zeta}_k\left(\boldsymbol{\theta}_{1,k}^{(p^*)}\right)\begin{array}{c}\text{reject } \mathcal{H}_2^{(q^*)}\\ \geq\\ <\\ \text{reject } \mathcal{H}_1^{(p^*)}\end{array}\tau \tag{25}$$

where $\boldsymbol{\theta}_{1,k}^{(p^*)}$ and $\boldsymbol{\theta}_{2,k}^{(q^*)}$ are the distribution parameters corresponding to models $\lambda_1^{(p^*)}$ and $\lambda_2^{(q^*)}$, respectively, and τ is a predetermined threshold that may be adjusted based on the risk associated with missing signals in a given application.

3) Source Quantity Test: Following the dominant source test, the remaining single source hypothesis that has not been rejected ($\mathcal{H}_1^{(p^*)}$ or $\mathcal{H}_2^{(q^*)}$) is tested against the remaining dual source hypothesis, $\mathcal{H}_3^{(p^*,q^*)}$. The LLRT for making this decision is

$$F_k\begin{array}{c}\text{accept } \mathcal{H}_3^{(p^*,q^*)}\\ \geq\\ <\\ \text{reject } \mathcal{H}_3^{(p^*,q^*)}\end{array}\eta \quad \left(\text{accept } \mathcal{H}_1^{(p^*)} \text{ or } \mathcal{H}_2^{(q^*)}\right) \tag{26}$$

where η is a predetermined threshold that should also be set based on the risk associated with missing (weaker) signals and

$$F_k=\begin{cases}\Lambda_k\left(\boldsymbol{\theta}_{3,k}^{(p^*,q^*)},\boldsymbol{\theta}_{1,k}^{(p^*)}\right), & \text{when } \mathcal{H}_2^{(q^*)} \text{ rejected}\\ \Lambda_k\left(\boldsymbol{\theta}_{3,k}^{(p^*,q^*)},\boldsymbol{\theta}_{2,k}^{(q^*)}\right), & \text{when } \mathcal{H}_1^{(p^*)} \text{ rejected}\end{cases}. \tag{27}$$

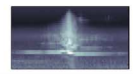
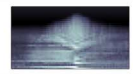
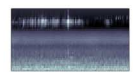
This means that the single source log-likelihood value used to form F_k depends on the results of the dominant source test in (25) (also see Fig. 1). Note that, at this stage, all $\boldsymbol{\zeta}_k(\boldsymbol{\theta}_k)$'s have already been calculated for use in other LLRTs, and hence, these values may be used in (27) to find F_k.

Upon completion of the source quantity test (or when no sources are detected in $\mathbf{y}_k$), a class label $\mathbf{c}_k$ may be assigned to this observation that depends on the accepted hypothesis according to Table II, which also provides the values of the estimated indicator variables, $\hat{\alpha}_k$ and $\hat{\beta}_k$, in each case. The entire process in Fig. 1 may then be continually applied to subsequent observations $\mathbf{y}_{k+j}$'s indefinitely or until all available observations have been processed.

4) Source Estimation: Once $\mathbf{c}_k$ has been determined for $\mathbf{y}_k$, estimates of the identified source signatures may then be obtained from the state estimates for the accepted hypothesis, if desired. Specifically, the coefficient estimates, $\hat{\mathbf{a}}_k$ and/or $\hat{\mathbf{b}}_k$, for time k, can be extracted from $\hat{\mathbf{x}}_{k|k}$ and used in conjunction with their respective basis matrices, $\mathbf{S}^{(p^*)}$ and/or $\mathbf{H}^{(q^*)}$, to yield estimates of the vectors representing the signal and interference components, denoted by $\hat{\mathbf{s}}_k$ and $\hat{\mathbf{h}}_k$, respectively. In particular, these estimates are found (see Table I) as

$$\begin{aligned}\mathcal{H}_1^{(p^*)}:\hat{\mathbf{s}}_k&=\mathbf{R}_{\boldsymbol{\omega}}^{\frac{1}{2}}\mathbf{S}^{(p^*)}\mathbf{D}^T\hat{\mathbf{x}}_{k|k}\\ \mathcal{H}_2^{(q^*)}:\hat{\mathbf{h}}_k&=\mathbf{R}_{\boldsymbol{\omega}}^{\frac{1}{2}}\mathbf{H}^{(q^*)}\mathbf{D}^T\hat{\mathbf{x}}_{k|k}\\ \mathcal{H}_3^{(p^*,q^*)}:\hat{\mathbf{s}}_k&=\left[\mathbf{R}_{\boldsymbol{\omega}}^{\frac{1}{2}}\mathbf{S}^{(p^*)}\ \mathbf{0}_{N\times M}\right]\mathbf{D}^T\hat{\mathbf{x}}_{k|k}\\ \hat{\mathbf{h}}_k&=\left[\mathbf{0}_{N\times M}\mathbf{R}_{\boldsymbol{\omega}}^{\frac{1}{2}}\mathbf{H}^{(q^*)}\right]\mathbf{D}^T\hat{\mathbf{x}}_{k|k}\end{aligned}$$

TABLE III
DESCRIPTIONS FOR DIFFERENT SOURCE TYPES IN THE KENAI FJORDS 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	
	Helicopter	Signatures evolve from low-frequency narrowband to very wideband, and revert back to narrowband.	40–400 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 days	

where each basis matrix is premultiplied by $\mathbf{R}_{\boldsymbol{\omega}}^{1/2}$ since their columns are basis vectors for the transformed observation space. In the case $\mathcal{H}_3^{(p^*,q^*)}$ is accepted, a type of source separation is performed by producing estimates, $\hat{\mathbf{s}}_k$ and $\hat{\mathbf{h}}_k$, that represent the separated signal and interference components of $\mathbf{y}_k$, respectively.

IV. EXPERIMENTAL RESULTS

This section presents the results of applying the SRCT method to a real acoustical source characterization problem. The data used to perform these experiments is first introduced, followed by a brief description of the Gaussian Mixture Model (GMM)-based method [5], [8], [26] that is used to benchmark the performance of the proposed SRCT method. The experimental test setup and associated results are then presented in terms of individual performance evaluations for each test outlined in the previous section, and confusion matrices for the overall detection/classification performance.

A. Data Description

The data used for this study represents acoustical recordings of a soundscape associated with a relatively remote site within Kenai Fjords National Park, Alaska, where a variety of signal and interference sources are intermittently present. The original audio waveforms were converted to sequences of one-third octave data vectors by the acoustical monitoring station that recorded the soundscape, where a vector was extracted from every non-overlapping one-second time segment and has $N = 33$ elements that represent the average energy in different one-third octave frequency bands for the corresponding one second interval.

In total, 19 full days of recordings from a single monitoring station are available and used for this study, where each day consists of 86,400 observations. The types of signal and interference sources that were frequently captured by this monitoring station are listed in Table III, along with brief descriptions of the general structures of their corresponding one-third octave signatures, durations of typical associated events, and example events for each source type. It is important to note that there is typically a large amount of variation between events associated with the same source type due to factors such as source distance and trajectory w.r.t. the recording station, Doppler, etc. This source signature diversity is perhaps the most challenging aspect of the present problem. Other significant challenges include constant water flow noise due to the monitoring station's close proximity to a river, the presence of rain of varying intensity throughout a large portion of the recordings, and mildly nonstationary noise statistics due to various phenomena associated with day/night cycles as well as other factors.

As mentioned, due to the complexity of the soundscape, manual annotation of the data was previously the only available approach for locating and labeling sources. Therefore, such annotations existed before the development of the SRCT method, and serve as the truth that is used to generate results. In particular, two well-trained operators visually inspected the data to identify acoustical events associated with signal sources of interest, which are those listed in Table III, as they occur most frequently and prominently in this particular site. Only the presence of interference (not its type) was annotated since it was present a large portion of the time and such sources are viewed as a nuisance for the present application. Occasionally, there were novel source types that contributed to the soundscape, e.g., wildlife activity or human speech, that are generally not of interest due to their rarity and consequent low impact on the soundscape.

B. Source Characterization Using GMMs

A GMM-based framework is adopted as the benchmark method in this study since this approach is commonly used in applications involving environmental sound [5], wildlife call [8], and speech [26] recognition. It is the ability of GMMs to model arbitrary distributions with multiple modes that makes them suitable for recognizing features extracted from inconsistent acoustical signatures that are a mainstay of these applications. A GMM is a weighted sum of K Gaussian densities, and is specified by $\lambda_{\mathbf{s}}^{(p)} = \{w_{\mathbf{s},i}^{(p)}, \boldsymbol{\mu}_{\mathbf{s},i}^{(p)}, \boldsymbol{\Sigma}_{\mathbf{s},i}^{(p)}\}_{i=1}^{K}$ for signal type p, where $w_{\mathbf{s},i}^{(p)}$, $\boldsymbol{\mu}_{\mathbf{s},i}^{(p)}$, $\boldsymbol{\Sigma}_{\mathbf{s},i}^{(p)}$ are the ith component weight, mean vector, and covariance matrix, respectively. Interference and dual source GMMs many be similarly defined as $\lambda_{\mathbf{h}}^{(q)}$ and $\lambda_{\mathbf{s},\mathbf{h}}^{(p,q)}$, respectively. The likelihood of a given set of signal

GMM parameters, given a vector observation $\mathbf{y}_k$, may be found as [5]

$$l\left(\lambda_{\mathbf{s}}^{(p)};\mathbf{y}_k\right)=\sum_{i=1}^{K} w_{\mathbf{s},i}^{(p)} f\left(\mathbf{y}_k;\boldsymbol{\mu}_{\mathbf{s},i}^{(p)},\boldsymbol{\Sigma}_{\mathbf{s},i}^{(p)}\right) \tag{28}$$

where $f(\cdot)$ represents a multivariate Gaussian density function.

In order to detect the simultaneous presence of signal and interference sources, a separate GMM is needed for modeling one-third octave vectors that correspond to different hypotheses in (3), besides $\mathcal{H}_0$. The likelihood under $\mathcal{H}_0$ may be found using a standard unimodal multivariate Gaussian likelihood function for parameters that are the fixed and known noise statistics introduced in Section II. In each case, a mixture of Gaussian densities can represent observations that contain both a random noise component, as well as the component(s) associated with one or two random sources. A hierarchical scheme, similar to that implemented by the SRCT method, may then be applied to each $\mathbf{y}_k$ using likelihoods calculated as in (28) for a set of LLRTs. While the dominant source and source quantity LLRTs for the GMM-based method have the same forms as those used for the SRCT method, i.e., (25) and (26), respectively, detection and classification are performed using single source GMMs. More specifically, GMM-based detection is performed using

$$\max_{\lambda}\frac{l(\lambda;\mathbf{y}_k)}{l(\lambda_0;\mathbf{y}_k)},\quad \lambda\in\left\{\lambda_{\mathbf{s}}^{(p)}\right\}_{p=1}^{P}\cup\left\{\lambda_{\mathbf{h}}^{(q)}\right\}_{q=1}^{Q}$$

as the test statistic, while classification is performed using

$$p^*=\arg\max_{p}\frac{l\left(\lambda_{\mathbf{s}}^{(p)};\mathbf{y}_k\right)}{l\left(\lambda_0;\mathbf{y}_k\right)}\quad q^*=\arg\max_{q}\frac{l\left(\lambda_{\mathbf{h}}^{(q)};\mathbf{y}_k\right)}{l(\lambda_0;\mathbf{y}_k)}.$$

The reason for this discrepancy between the SRCT and GMM-based methods is that, in the former, the dual sources models and likelihood values can accommodate the presence of a single source, as coefficient estimates for absent sources will be negligible (see Section III-D-1) whereas a dual source GMM represents observations that necessarily contain both signal and interference components.

C. Test Setup

In order to apply the SRCT method to the Kenai Fjords data, the subspace and AR parameter matrix dictionaries for each source type, as well as the observation noise statistics, need to be extracted from a set of "clean" training events that contain only the signatures of interest for a particular model. These training events were chosen to ensure a given source model adequately captures the diversity of the associated signatures. For example, in the Kenai Fjords site, a significant amount of within-class diversity exists for propeller plane events, not only because this is the most commonly occurring type of signal, but also owing to the presence of different types of planes. Consequently, more plane training events are required when compared to other signal types. To extract a source's subspace, eigenvalue decomposition [27] was applied to a covariance matrix representing the data from several clean events corresponding to that source type. The $M=3$ eigenvectors corresponding to the largest eigenvalues were then used as basis vectors for that source model. Using the parameter identification procedure in [28], vector AR model parameters (with $J=2$) for a given source were also extracted from the aforementioned set of training events. Noise statistics were estimated using observation segments that did not contain any structured signal or interference. 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 was subtracted from each observation as it arrived, before being processed by a given method. The experimental distribution of the noise vectors was also found to be approximately multivariate Gaussian, thus justifying the use of the model in (1).

Single source GMMs were formed for each signal and interference type by applying the expectation maximization (EM) algorithm [29] to the same training events used to form the SRCT models. Dual source GMMs were extracted from observations formed by superimposing different combinations of the clean signal and interference observations used to form the single source GMMs. The number of components in a given GMM was selected by applying the silhouette method [30] to the corresponding set of observations the GMM was extracted from.

Disjoint training and testing sets were then formed using a collection of two-hour-long segments found throughout the 19 days of data. 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 interferences sources, had a good mix of hypotheses and hypothesis transitions, and contained events with highly variable signatures. In particular, among all the observations used in this study (training and testing), 36.8% represented $\mathcal{H}_0$, 4.4% represented $\mathcal{H}_1^{(p)}$, 53.1% represented $\mathcal{H}_2^{(q)}$, and 5.7% represented $\mathcal{H}_3^{(p,q)}$. Observations containing interference alone were more common due to rather relentless rain and thunder, which is also the reason interference was present for the majority of observations where a signal was present. The training set consisted of 10 data segments and was used to determine the detection, dominant source, and source quantity thresholds, which were found to be $\gamma=-1.5$, $\tau=3.6$, and $\eta=54.2$, respectively for the SRCT method, and $\gamma=-4.7$, $\tau=20.5$, and $\eta=-38.7$, respectively for the GMM method. These thresholds were selected such that no signals in the training segments were missed. The observations in the training segments that contained the signatures of only one source were also used to estimate the source models, as described above. The testing set consisted of 40 segments and was used to evaluate the performance of the each method. No testing segments contained an event used to estimate source models, and at least one segment from each of the 19 available days of data was used.

Finally, an HMM-based post-processing was applied to the signal detection/classification results produced by each method. This process finds the likelihood that a certain signal type is present, given a sequence of assigned labels, using a procedure that is similar to that proposed in [17]. In particular, a separate three-state HMM was learned for each signal type using the Baum-Welch algorithm. The observations in this case are signal class labels assigned by the SRCT method (or GMM when evaluating these decision sequences). The HMM states do not have any specific parametric meaning, but may correspond

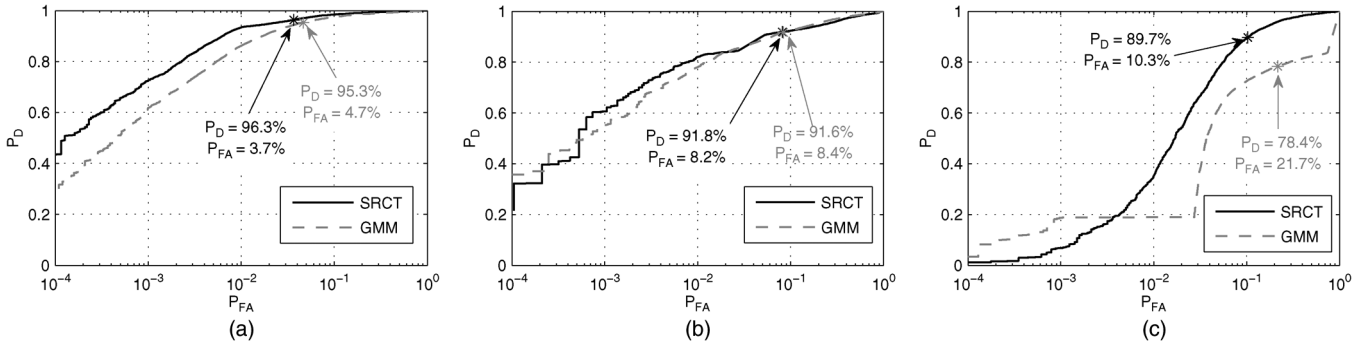


Fig. 2. ROC curves produced by the SRCT and GMM methods for each test statistic. (a) Detection test. (b) Dominant source test. (c) Source quantity test.

to different signal label distributions resulting from the time-varying signatures of signals considered in this application. This means that state transitions typically occur when the structure of an acoustical event being evaluated changes (see examples in Table III), as a system's decisions may change slightly. Different HMMs were used for the SRCT and GMM methods, and each HMM was formed using decision sequences generated by a given method for segments in the training set. This approach is appropriate for on-line processing and is done both to smooth out the detection/classification results, and to associate a cluster of detections that have the same label with a single event for more concise and meaningful classification results. In the case of the GMM method, this HMM-based smoothing provides a means to incorporate temporal dependencies between decisions, that would otherwise be independent.

D. Test Statistic Performance Evaluation

To measure the performance of each test in Section III-D when applied to the testing segments, receiver operator characteristic (ROC) curves were generated and are presented in Figs. 2(a)–2(c). Each ROC curve shows how the probability of detection (P_D) and probability of false alarm (P_{FA}) change as the appropriate threshold (γ, τ, or η) is modified. For each test, P_D indicates the probability of correctly accepting the hypothesis associated with the numerator of a given LLRT, i.e., $\mathcal{H}_3^{(p,q)}$, $\mathcal{H}_1^{(p^*)}$, and $\mathcal{H}_3^{(p^*,q^*)}$ for (23), (25), and (26), respectively. On the other hand, P_{FA} indicates the probability of falsely accepting the hypothesis associated with the denominator of a given LLRT, i.e., $\mathcal{H}_0$, $\mathcal{H}_2^{(q^*)}$, and $\mathcal{H}_1^{(p^*)}$ *or* $\mathcal{H}_2^{(q^*)}$ for (23), (25), and (26), respectively. The axis showing P_{FA} is log-scale in each case to highlight the differences between each method. 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 decision threshold where $P_D + P_{FA} = 1$.

As can be seen from Fig. 2(a), the knee-points of the ROC curves for the SRCT and GMM detectors (see Section III-D-1) are $(P_D = 96.3\%, P_{FA} = 3.7\%)$ and $(P_D = 95.3\%, P_{FA} = 4.7\%)$, respectively, while the AUCs are 0.992 and 0.987, respectively. Although both detectors operate similarly, the SRCT detector performs slightly better since the source estimates used to form the likelihoods can accommodate arbitrary combinations of signal and interference. Nonetheless, the single source GMMs still produce sufficiently high likelihoods for observations containing one or more sources. In both cases missed detections are primarily caused by samples associated with mild interference, such as very light rain, as well as the presence of novel source signatures. The latter is due to the inability of source basis vectors for the SRCT method, or source signature distributions for the GMM method, to capture signatures that rarely occur during training. False alarms are mostly caused by ambient noise statistics changing slightly over time, owing to the properties of natural soundscape data. However, this deviation is small enough such that the detectors remain fairly robust to this temporal change.

The ROC curves in Fig. 2(b) for the SRCT and GMM dominant source tests (see Section III-D-2) exhibit knee-points at $(P_D = 91.8\%, P_{FA} = 8.2\%)$ and $(P_D = 91.6\%, P_{FA} = 8.4\%)$, respectively, while the AUCs are 0.965 and 0.968, respectively. The GMM and SRCT methods perform very similarly in the dominant source stage, as the models used by each are appropriate for classification in these single source scenarios. This result reinforces the suitability of GMM-based methods for the interference-free acoustical source recognition tasks that they are typically applied to, owing to the ability of GMMs to represent complicated and erratic signatures. For both methods, dominant source errors can mostly be attributed to closeness between some realizations of rain and thunder interference to the models associated with helicopter and jet signal types, which occasionally leads to $\mathcal{H}_1^{(p^*)}$ being accepted when $\mathcal{H}_2^{(q^*)}$ is true. This confusion is typically only the case when unusually strong thunder is encountered.

Finally, Fig. 2(c) shows the ROC curves for the SRCT and GMM source quantity test statistics (see Section III-D-3), where the knee-points are $(P_D = 89.7\%, P_{FA} = 10.3\%)$ and $(P_D = 78.3\%, P_{FA} = 21.7\%)$, respectively, and the AUCs are 0.953 and 0.806, respectively. The GMM source quantity ROC curve is smoothest from $P_{FA} = 19\%$ to $P_{FA} = 84\%$ since this is the region where the dual source GMMs produce numerically significant likelihoods, and hence, the LLRT is not affected by round off error. These results show that the main difference between the SRCT and GMM-based methods is their respective performance in determining source quantity, where the SRCT vastly outperforms the GMM-based method. This is due in part to the SRCT exploiting the dependency structure of each source's signatures when forming the LLRT. More importantly, the dual source GMMs have difficulty representing the extreme variations associated with superimposed signatures from two

different sources. For instance, not only is there a significant amount of diversity between events associated with a given source type, but both signal and interference sources present in an observation can be at different relative stages in their evolution (i.e., temporal position within an event), and can have arbitrary amplitudes. A given dual source GMM must capture all combinations of these variations, leading to a lower likelihood for the GMM given any specific observation, whereas the SRCT method directly estimates the source signatures to account for this diversity. For the SRCT method, in some cases a weak signal superimposed with strong interference triggers false acceptance of $\mathcal{H}_2^{(q^*)}$ over $\mathcal{H}_3^{(p^*,q^*)}$, thus missing the signal. Other errors for this test are again due to novel source signatures and their closeness to subspaces of other models. Specifically, strong plane and helicopter signals occasionally falsely trigger $\mathcal{H}_3^{(p^*,q^*)}$ over $\mathcal{H}_1^{(p^*)}$, and very strong thunder also causes false acceptance of $\mathcal{H}_3^{(p^*,q^*)}$. Clearly, performance issues arising from close subspaces are more prevalent when the system must contend with more source types, and these issues are exacerbated the closer M is to N.

TABLE IV
SRCT CONFUSION MATRIX SHOWING TOTAL NUMBER OF INSTANCES EACH SIGNAL TYPE WAS ASSIGNED A GIVEN LABEL (INTERFERENCE ABSENT/INTERFERENCE PRESENT)

		Assigned			
		None	Plane	Heli	Jet
Truth	None	-	1 (0/1)	0 (0/0)	7 (2/5)
	Plane	4 (2/2)	**120 (61/59)**	0 (0/0)	2 (0/2)
	Heli	9 (5/4)	2 (0/2)	**27 (11/16)**	0 (0/0)
	Jet	2 (1/1)	0 (0/0)	0 (0/0)	**17 (11/6)**

TABLE V
GMM CONFUSION MATRIX SHOWING TOTAL NUMBER OF INSTANCES EACH SIGNAL TYPE WAS ASSIGNED A GIVEN LABEL (INTERFERENCE ABSENT/INTERFERENCE PRESENT)

		Assigned			
		None	Plane	Heli	Jet
Truth	None	-	18 (4/14)	7 (1/6)	1 (0/1)
	Plane	13 (3/10)	**112 (53/59)**	1 (0/1)	0 (0/0)
	Heli	10 (5/5)	4 (3/1)	**24 (8/16)**	0 (0/0)
	Jet	2 (1/1)	6 (1/5)	1 (0/1)	**10 (10/0)**

E. Overall Classification Performance Evaluation

In order to provide an indication of the overall performance (involving all tests) achieved by two different approaches for correctly classifying signal events contained in the 40 testing sequences, Tables IV and V show confusion matrices for the SRCT and GMM methods, respectively. Each main entry in these tables indicates the number of times a certain type of signal event was assigned a given label. To demonstrate the effects of interference on signal detection/classification, the numbers in parentheses separate the overall confusion matrix into cases where interference was absent and present (e.g., $c(a/b)$ in an entry means there were a and b instances where interference was absent and present, respectively, with $c = a + b$). Interference events are also classified and estimated by each method, but results are not reported since these source types were not labeled for reasons mentioned above. An event is considered correctly classified if the correct label is assigned to at least half of the observations it is present. Since "none" means no source of interest (either present or assigned), the first column in each confusion matrix indicates instances where signal events are missed ($\mathcal{H}_0$ or $\mathcal{H}_2^{(q^*)}$ falsely accepted), whereas the first row indicates false alarms ($\mathcal{H}_1^{(p^*)}$ falsely accepted). The diagonal elements indicate the number of events of each signal type that are assigned a correct label, which show overall correct signal classification rates of 89.62% and 79.78% for the SRCT and GMM, respectively. These classification rates exclude false alarms, as there are no distinct "signal-free" events that may be labeled as such, but rather a false alarm arises any time a signal label is assigned to an area where there are none present. Therefore, false alarm rates are reported in terms of the percentage of all event detections (i.e., entries in the last three columns in each confusion matrix) that are false, which are 4.55% and 14.13% for the SRCT and GMM, respectively. This is also the reason "$-$" appears for the "none" diagonal entry in each table.

As can be seen from Tables IV and V, the overall classification results produced by the SRCT method are significantly better than those produced by the GMM, which is mainly due to the relative differences in performance of the associated source quantity tests. For example, Table V shows that the GMM-based method missed more signals when interference was present. The SRCT missed fewer signal events overall, since it was able to adapt to dual source scenarios due to the increased flexibility offered by the estimation process used to find the likelihood values. Confusion between signal types was also more common for the GMM method. The fact that such confusion occurred most often when interference was also present shows that the GMMs were not quite as effective at estimating the signal types in the presence of two sources. Finally, the GMM-based method produced significantly more false alarms owing to a less powerful source quantity test and the fact that the threshold for this test was set to detect all signals in the training segments. Most false alarms produced by the GMM-based method were associated with planes since strong thunder that is frequently present (interference alone) tends to resemble superimposed rain and plane signatures (dual source). Most false alarms produced by the SRCT method were for jets since the subspace for this source is close to weak thunder signatures. In total, 35% and 73% of all errors (false alarms and missed signals) produced by the SRCT and GMM methods, respectively, were caused by the presence of strong interference.

F. Computational Cost

As a final note, the computational cost of the SRCT method for processing a single $\mathbf{y}_k$ is considered. This cost is driven by the processes with the largest associated growth rates, that are required to find $\boldsymbol{\zeta}_k(\boldsymbol{\theta}_k)$'s for the PQ dual source hypotheses (out of all $P + Q + PQ$ hypotheses). In particular, the most costly processes are implementation of the Kalman filters and calculating $\boldsymbol{\zeta}_k(\boldsymbol{\theta}_k)$'s directly using (8), which require $O(PQM^{2.373})$ and $O(PQN^{2.373})$ operations, respectively, if efficient implementations are used. Therefore, since $N > M$ the latter represents the SRCT algorithm cost. For the GMM method, the matrix inversions and determinants required to find the likelihoods

may be calculated off-line, but a likelihood must be found for each component in a GMM. Therefore, assuming an average of K components in each GMM, the cost of this method is $O(KPQN^2)$ operations, which is very comparable to that of the SRCT method for our application.

V. Conclusions

This paper considers the complex problem of detecting, classifying, and estimating the signatures of random transient acoustical sources from TTR where a signal and an interference source, both of unknown type, may be simultaneously present. In such problems, it can be difficult to quantify the effects of the presence of one source on the detection of the other since all sources have unknown scaling and times of arrival, and their signatures may overlap to varying degrees. The method introduced in this paper deals with these complications using a series of LLRT, designed specifically to discover the composition of observations that follow proposed model. Each test uses the log-likelihoods of distribution parameters under specific source composition hypotheses, which incorporate known information about the dependencies between basis coefficients for each source type, thus yielding a framework that is appropriate for characterizing transient sources from sequential data. The experimental results on data representing the soundscape of a national park demonstrates that the SRCT method excels at performing each of the aforementioned tasks with high precision.

Acknowledgment

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

References

[1] B. Friedlander and B. Porat, "Performance analysis of transient detectors based on a class of linear data transforms," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 665–673, Mar. 1992.

[2] 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.

[3] E. D. Chesmore, "Application of time domain signal coding and artificial neural networks to passive acoustical identification of animals," *Appl. Acoust.*, vol. 62, pp. 1359–1374, Dec. 2001.

[4] M. Ramona, G. Richard, and B. David, "Vocal detection in music with support vector machines," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 2008, pp. 1885–1888.

[5] 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.

[6] J. D. Krijnders, M. E. Niessen, and T. C. Andringa, "Sound event recognition through expectancy-based evaluation of signal-driven hypotheses," *Pattern Recogn. Lett.*, vol. 31, pp. 1552–1559, Aug. 2010.

[7] B. Gajić and K. K. Paliwal, "Robust speech recognition in noisy environments based on subband spectral centroid histograms," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 2, pp. 600–608, Mar. 2006.

[8] P. Somervuo, A. Härmä, and S. Fagerlund, "Parametric representation of bird sounds for automatic species recognition," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 6, pp. 2252–2263, Nov. 2006.

[9] H. Wu and J. Mendel, "Classification of battlefield ground vehicles using acoustic features and fuzzy logic rule-based classifiers," *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 1, pp. 56–72, Feb. 2007.

[10] J. Bouvrie, T. Ezzat, and T. Poggio, "Localized spectro-temporal cepstral analysis of speech," in *Proc. IEEE ICASSP*, May 2008, pp. 4733–4736.

[11] T. S. Brandes, "Feature vector selection and use with hidden Markov models to identify frequency-modulated bioacoustic signals amidst noise," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 8, pp. 1173–1180, Aug. 2008.

[12] D. E. Lake, "Harmonic phase coupling for battlefield acoustic target identification," in *Proc. IEEE ICASSP*, May 1998, vol. 4, pp. 2049–2052.

[13] C. Hory, N. Martin, and A. Chehikian, "Spectrogram segmentation by means of statistical features for non-stationary signal interpretation," *IEEE Trans. Signal Process.*, vol. 50, pp. 2915–2925, Dec. 2002.

[14] R. A. Altes, "Detection, estimation, and classification with spectrograms," *J. Acoust. Soc. Amer.*, vol. 67, pp. 1232–1246, Apr. 1980.

[15] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.

[16] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. the IEEE*, vol. 77, pp. 257–286, February 1989.

[17] 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.

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

[19] 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.*, Jul. 2010, pp. 1–8.

[20] A. B. Salberg, A. Hanssen, and L. L. Scharf, "Robust multidimensional matched subspace classifiers based on weighted least-squares," *IEEE Trans. Signal Process.*, vol. 55, pp. 873–880, Mar. 2007.

[21] D. Luengo, C. Pantaleon, I. Santamaria, L. Vielva, and J. Ibañez, "Multiple composite hypothesis testing: A competitive approach," *J. VLSI Signal Process. Syst.*, vol. 37, pp. 319–331, Jun. 2004.

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

[23] A. P. Sage and J. L. Melsa, *Estimation Theory With Applications to Communications and Control*. New York, NY, USA: McGraw-Hill, 1971.

[24] A. Ben-Israel and T. Greville, *Generalized Inverses: Theory and Applications*, 2nd ed. New York, NY, USA: Springer, 2003.

[25] R. T. Behrens and L. L. Scharf, "Signal processing applications of oblique projection operators," *IEEE Trans. Signal Process.*, vol. 42, no. 6, pp. 1413–1424, Jun. 1994.

[26] 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.

[27] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: Johns Hopkins, 1996.

[28] M. R. Azimi-Sadjadi, "New results in strip Kalman filtering," *IEEE Trans. Circuits Syst.*, vol. 36, no. 6, pp. 893–897, Jun. 1989.

[29] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.

[30] P. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987.

Neil Wachowski (S'09) received the B.S. degree in electrical engineering from Michigan Technological University, Houghton, in 2002, and the M.S. degree in electrical engineering from Colorado State University (CSU), Fort Collins, in 2009, with specialization in signal processing. Currently, he is working towards the Ph.D. degree at CSU 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 theory, 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.