DATA-DRIVEN SIGNAL DETECTION AND CLASSIFICATION

Akbar M. Sayeed*
Department of Electrical and Computer Engineering
Rice University
6100 S. Main Street
Houston, TX 77005, USA
Email: akbar@rice.edu

ABSTRACT

In many practical detection and classification problems, the signals of interest exhibit some uncertain nuisance parameters, such as the unknown delay and Doppler in radar. For optimal performance, the form of such parameters must be known and exploited as is done in the generalized likelihood ratio test (GLRT). In practice, the statistics required for designing the GLRT processors are not available *a priori* and must be estimated from limited training data. Such design is virtually impossible in general due to two major difficulties: *identifying* the appropriate nuisance parameters, and *estimating* the corresponding GLRT statistics. We address this problem by using recent results that relate joint signal representations (JSRs), such as time-frequency and time-scale representations, to quadratic GLRT processors for a wide variety of nuisance parameters. We propose a general data-driven framework that: 1) *identifies* the appropriate nuisance parameters from an arbitrarily chosen finite set, and 2) *estimates* the second-order statistics that characterize the corresponding JSR-based GLRT processors.

1. INTRODUCTION

Optimal detection and classification of signals in the presence of noise requires the knowledge of certain underlying statistics. In most practical problems, those statistics are not known *a priori* and must be estimated from available training data.

In many situations, the signals to be detected are best modeled as stochastic signals exhibiting some uncertain parameters, the so-called *nuisance parameters*. For example, the radar returns from a complex object can be modeled as a random signal exhibiting unknown delay and Doppler parameters. Such detection problems are formulated as binary *composite* hypothesis tests of the form[1]

$$\begin{aligned} H_1 &: \quad x(t) = s^{\theta}(t) + n(t) \\ H_0 &: \quad x(t) = n(t) \end{aligned} \tag{1}$$

where $t \in T$, the observation interval, x is the observed waveform, s^{θ} is a (nonstationary) stochastic signal with nuisance parameters $\theta = (\theta_1, \theta_2, \cdots, \theta_M) \in S \subset \mathbb{R}^M$, and n is the additive noise.

For optimal performance, the form of the nuisance parameters must be known and exploited. The generalized likelihood ratio test (GLRT) is often used in practice in which the maximum likelihood (ML) estimate of the parameters is first formed and then the likelihood ratio (LR) corresponding to the estimate is used as the test statistic[2]:

$$L(x) = L^{\hat{\theta}_{ML}}(x) = \max_{\theta \in S} L^{\theta}(x) \ , \tag{2}$$

where L^{θ} is the LR corresponding to the nuisance parameters θ [1].

The form of L^{θ} depends on the statistics of s^{θ} and n. Since the statistics have to be estimated from training data, we assume a characterization of L^{θ} in terms of second-order statistics. In particular, we assume that n is zero-mean circular Gaussian with correlation function $R_n(t_1, t_2) = E[n(t_1)n^*(t_2)]$, and s^{θ} is a zero-mean second-order (not necessarily Gaussian) signal with correlation function R_s^{θ}.[3] We further assume the low-SNR (signal-to-noise-ratio) regime in which the *locally optimal* test statistic is given by [1]

$$L^{\theta}(x) \equiv \langle \mathbf{R}_s^{\theta} \mathbf{R}_n^{-1} x, \mathbf{R}_n^{-1} x \rangle - \mathrm{Tr}\left(\mathbf{R}_s^{\theta} \mathbf{R}_n^{-1}\right) \ , \tag{3}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product and $\mathrm{Tr}(\cdot)$ denotes the trace of an operator[4] (sum of the eigenvalues). Thus, under our assumptions, the GLRT detector is completely determined by R_s^{θ} and R_n.

When the GLRT processor must be designed from labeled training data, as is often the case in practice, there are two key issues that need to be addressed:

1. **Identification of parameters.** How can the nuisance parameters that underlie a given data set be identified?
2. **Estimation of GLRT statistics.** Given the nuisance parameters, how can the GLRT statistics ($\mathbf{R}_s^{\theta}$ and $\mathbf{R}_n$) be estimated from training data?

The first issue is fairly obvious yet highly nontrivial. The second issue arises because the estimation of R_s^{θ} is not at all straightforward (if not impossible) in general. The reason is that even if noise-free realizations of s^{θ} are available, different realizations $s_i^{\theta_i}$ correspond to different values of θ. Thus, for estimation of R_s^{θ} at a particular value of θ, all the realizations must be "aligned" in some sense to that value. This is not always possible in general and crucially depends on the dependence of R_s^{θ} on θ.

Joint signal representations (JSRs), such as time-frequency representations (TFRs) and time-scale representations (TSRs), can realize GLRT detectors in a broad

*This work was supported by the ONR under grant nos. N00014-95-1-0674 and N00014-95-1-0849, and the NSF under grant no. MIP–9457438.

[1] Such classification problems can be similarly described via an M-ary composite hypothesis test – we restrict our discussion to detection problems in this paper.

[2] which is compared to a threshold to decide whether the signal is present (H_1) or not (H_0)

[3] Extension to nonzero-mean situations is straightforward.

[4] $\mathbf{R}$ denotes the operator defined by the function $R(t_1, t_2)$ as $(\mathbf{R}x)(t) = \int R(t, u)x(u)du$. A white-noise component guarantees that $\mathbf{R}_n^{-1}$ exists.

range of composite hypothesis testing problems of the type discussed above [2, 3]. Such JSR-based GLRT detectors admit a unified formulation in terms of *parameterized unitary operators* which can be used to model a wide variety of nuisance parameters that are quite relevant from a practical viewpoint[5] — radar/sonar, machine fault diagnostics, and biomedical signal classification are some typical applications. For example, TFR-based detectors are appropriate for unknown time- and frequency-offset nuisance parameters and have been successfully applied in machine fault diagnostics [4]. More importantly, JSRs impose a structure on nuisance parameters that provides a natural mechanism for addressing the above-mentioned issues encountered in data-driven detection and classification.

In this paper, we propose a general data-driven detection and classification framework by exploiting the structure and unified formulation of JSR-based detectors. Using *labeled (H_0 vs. H_1) training data*, our framework:

1. *Identifies* the nuisance parameters, from a finite set chosen *a priori*, that "best fit" the data in a precise sense.
2. *Estimates* the statistics characterizing the GLRT processor corresponding to the "best" nuisance parameters.

For example, our framework can determine whether time- and frequency-shifts, or time-shifts and scale-changes are the appropriate nuisance parameters in a given data set, and design the corresponding (TFR or TSR) GLRT processor.

In the next section we provide a brief description of JSR-based detectors and highlight their relevant features. In Section 3, we describe the structure of the JSR-based data-driven framework. Section 4 illustrates the performance of TFR- and TSR-based algorithms on simulated data. Some concluding remarks are presented in Section 5.

2. JSR-BASED DETECTORS

JSRs represent signal characteristics jointly in terms of two or more variables or physical quantities (which define the nuisance parameters) — for example, time and frequency in TFRs, and time and scale in TSRs. From a detection viewpoint, the *covariance-based* JSRs are the appropriate vehicle [3] and we now provide a brief description of such JSRs.

Let $G \subset \mathbb{R}^M$ be an M-parameter Lie group and $\{\mathbf{U}_\theta : \theta \in G\}$ be a family of unitary operators that is a unitary representation of G on $L^2(\mathbb{R})$; that is $\mathbf{U}_\theta : L^2(\mathbb{R}) \to L^2(\mathbb{R})$, $\langle \mathbf{U}_\theta s_1, \mathbf{U}_\theta s_2\rangle = \langle s_1, s_2\rangle$, and $\mathbf{U}_\theta \mathbf{U}'_\theta = \mathbf{U}_{\theta \bullet \theta'}$, for $\theta, \theta' \in G$, where $\bullet$ denotes the group operation [3]. The "coordinates" of $\theta = (\theta_1, \theta_2 \cdots \theta_M)$ represent the variables of interest, and the unitary operators $\mathbf{U}_\theta$ represent the signal transformations of interest (that produce the nuisance parameters), such as time and frequency shifts, scale changes etc.

Consider a given family $\{\mathbf{U}_\theta : \theta \in G\}$. On one hand, $\{\mathbf{U}_\theta\}$ defines a class C of JSRs with respect to the variables $\theta = (\theta_1, \theta_2 \cdots \theta_M)$ via [3]

$$(\mathbf{P}s)(\theta; \mathbf{K}) = \langle \mathbf{K}\mathbf{U}_\theta^{-1} s, \mathbf{U}_\theta^{-1}\rangle \tag{4}$$

where the JSR is denoted by the operator $\mathbf{P}$ that maps $s \in L^2(\mathbb{R})$ into the space of functions defined on G. Different choices of the operator $\mathbf{K} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ yield different JSRs from C. On the other hand, $\{\mathbf{U}_\theta\}$ defines a class of composite hypothesis testing situations of the form (1) characterized by [3]

$$\mathbf{R}_s^\theta = \mathbf{U}_\theta \mathbf{R}_r \mathbf{U}_\theta^{-1} \tag{5}$$

for some "reference" correlation function $\mathbf{R}_r$, where the nuisance parameters correspond to the JSR variables $\theta = (\theta_1, \theta_2, \cdots \theta_M)$. (5) is equivalent to the signal model

$$s^\theta(t) = (\mathbf{U}_\theta s_r)(t) \tag{6}$$

for some reference signal s_r with correlation function $\mathbf{R}_r$. The key point is that the JSRs from C constitute the canonical GLRT statistics L^θ for the hypothesis testing problem characterized by (5) [3]:

$$L^\theta(x) = (\mathbf{P}y)(\theta; \mathbf{R}_r) - \mathrm{Tr}(\mathbf{R}_s^\theta \mathbf{R}_n^{-1}) \ , \quad y = \mathbf{R}_n^{-1} x \ , \tag{7}$$

where note that the operator characterizing the JSR is $\mathbf{K} = \mathbf{R}_r$. The GLRT detector can then be realized via (2) by computing the maximum over $\theta \in S \subset G$ (the range of the nuisance parameters).

Intuition for estimation. Note from (5) that if $\mathbf{U}_\theta$ is known, we need to estimate $\mathbf{R}_s^\theta$ for only *one* value of θ because of the group structure. The reason is that for any $\theta, \theta' \in G$, $\mathbf{R}_s^\theta$ and $\mathbf{R}_s^{\theta'}$ can be transformed into each other: $\mathbf{R}_s^{\theta'} = \mathbf{U}_{\theta' \bullet \theta^{-1}} \mathbf{R}_s^\theta \mathbf{U}_{\theta' \bullet \theta^{-1}}^{-1}$, and, similarly, $s^{\theta'} = \mathbf{U}_{\theta' \bullet \theta^{-1}} s^\theta$, for the same underlying s_r in (6). This suggests a mechanism for estimating $R_s^{\theta_o}$ at a particular value θ_o by "aligning" different realizations of s^θ, corresponding to different values of θ, to θ_o. From a purely detection viewpoint, the actual value of θ_o does not matter because of the group structure: $\mathbf{R}_s^{\theta_o}$ for any θ_o can serve as the reference $\mathbf{R}_r$ in (5) and (7).

Intuition for identification. First of all, from a practical perspective, the choice of nuisance parameters that best describe a data set can only be made from a finite set. Different classes of JSRs, characterized by different unitary representations $\{\mathbf{U}_\theta\}$ of different groups, can model a wide range of practically relevant nuisance parameters [3]. Depending on the particular application at hand, a finite candidate set of such nuisance parameters can be chosen *a priori*. The "best" nuisance parameters from the set[6] are those which result in an $\mathbf{R}_r$ estimate after "alignment" that has the smallest effective rank.[7] The intuition behind the criterion is that, in general, the "unaligned" signal correlation function is of a higher rank than $\mathbf{R}_r$ due to nuisance parameters. Moreover, a mismatch of parameters in "alignment" will also in general lead to a higher rank estimate of $\mathbf{R}_r$. For quadratic detection, *deflection* [1] captures this notion of smallest rank, and is also a measure of SNR. In our case, deflection is given by $H(\mathbf{R}_r) \equiv \mathrm{Tr}((\mathbf{R}_r \mathbf{R}_n^{-1})^2)$.

3. UNIFIED JSR-BASED DATA-DRIVEN FRAMEWORK

Consider the hypothesis testing problem (1) characterized by R_s^θ and R_n assumed to be unknown. Suppose that we have N_j training realizations, x_i^j, $i = 1, 2, \cdots N_j$, under H_j, $j = 1, 2$, available to us. Further suppose that that we have Q classes, C_i, $i = 1, 2, \cdots Q$, of JSRs (defined by a corresponding group G_i and a family of unitary operators $\mathbf{U}_{\theta^i}^i$), that have been chosen *a priori* to reflect the types of likely nuisance parameters (such as time-frequency shifts, scale changes, chirp rate changes etc.). The structure of the JSR-based detectors suggests the following two-part data-driven algorithm to design the "best" GLRT detector.

Estimation Algorithm

[5] The operator parameters define the nuisance parameters.

[6] with a corresponding class of JSRs

[7] Rank is the number of nonzero eigenvalues.

Step 1. Estimate $R_n = R_0$ and R_1:

$$\widehat{R}_n(t_1,t_2) = \frac{1}{N_0}\sum_{i=1}^{N_0} x_i^0(t_1)x_i^{0*}(t_2)\ , \tag{8}$$

$$\widehat{R}_1(t_1,t_2) = \frac{1}{N_1}\sum_{i=1}^{N_1} x_i^1(t_1)x_i^{1*}(t_2)\ . \tag{9}$$

Define the operator $\widehat{\mathbf{D}} = (\widehat{\mathbf{R}}_1 - \widehat{\mathbf{R}}_0)\widehat{\mathbf{R}}_1^{-1}$ which is the (estimated) Wiener filter for estimating the signal component from an H_1 realization.[8]

Step 2. Choose the "reference" template, $\widehat{\mathbf{R}}_a = \widehat{\mathbf{R}}_1 - \widehat{\mathbf{R}}_n$, which serves as an estimate of $\mathbf{R}_r$ for the purpose of alignment.

Step 3. For each JSR class C_j, $j = 1, 2, \cdots Q$, do:

Initialize $\widehat{R}_r(t_1,t_2) = 0$, $(t_1,t_2) \in T \times T$.

For $i = 1$ to N_1

begin

$$\hat{\theta} = \arg\max_{\theta\in S}\left[(\mathbf{P}y_i)(\theta;\widehat{\mathbf{R}}_a) - \mathrm{Tr}\left(\widehat{\mathbf{R}}_s^\theta\widehat{\mathbf{R}}_n^{-1}\right)\right]\ ,$$

$$\widehat{\mathbf{R}}_s^\theta = \mathbf{U}_\theta\widehat{\mathbf{R}}_a\mathbf{U}_\theta^{-1}\ ,\ y_i = \widehat{\mathbf{R}}_n^{-1}x_i^1 \tag{10}$$

$$\hat{s}_i = \widehat{\mathbf{D}}x_i^1 \quad (\text{"unaligned" signal estimate}) \tag{11}$$

$$\hat{s}_{r,i} = \mathbf{U}_{\hat{\theta}}^{-1}\hat{s}_i \quad (\text{"aligned" signal estimate}) \tag{12}$$

$$\widehat{R}_r(t_1,t_2) = \frac{[(i-1)\widehat{R}_r(t_1,t_2) + \hat{s}_{r,i}(t_1)\hat{s}_{r,i}^*(t_2)]}{i} \tag{13}$$

end

$\widehat{\mathbf{R}}_r^j = \widehat{\mathbf{R}}_r$ (corresponding to the JSR class C_j)

Identification Algorithm

Compute the deflections $H(\widehat{\mathbf{R}}_r^j)$, $j = 1, 2, \cdots Q$. The JSR class C_j corresponding to the largest deflection defines the nuisance parameters that best fit the training data, and also defines the corresponding GLRT test statistics (7).

Remarks. The objective of the estimation algorithm is to estimate the underlying $\mathbf{R}_r$ for each JSR class. It does so by aligning (Step 3) all the H_1 realization to the value of θ implicitly defined by the reference template $\widehat{\mathbf{R}}_a$ (Step 2). Note that the estimate of θ in (10) is the ML estimate if $\widehat{\mathbf{R}}_a = \mathbf{R}_r$. Step 3 may be repeated to yield better $\widehat{\mathbf{R}}_r$ estimates by replacing $\widehat{\mathbf{R}}_a$ with $\widehat{\mathbf{R}}_r$. The identification step essentially chooses the nuisance parameters which pack the signal energy in the smallest subspace. We note that different choices for the operators $\widehat{\mathbf{D}}$ and $\widehat{\mathbf{R}}_a$ are possible — due to space limitations we discuss only the above.

4. EXAMPLES: TIME-FREQUENCY AND TIME-SCALE DETECTORS

We illustrate the data-driven framework with time-frequency and time-scale GLRT detector design based on simulated data. We use white Gaussian noise for all the experiments.

Based on the available training data, the simplest quadratic detector is

$$L_{\mathrm{B}1}(x) = \langle\widehat{\mathbf{R}}_s x, x\rangle\ \ ,\ \widehat{\mathbf{R}}_s = \widehat{\mathbf{R}}_1 - \widehat{\mathbf{R}}_n\ , \tag{14}$$

which is an estimate of the (locally optimal) Bayesian detector[9] since $\widehat{\mathbf{R}}_s$ is an estimate of $\mathbf{R}_s = \int \mathbf{R}_s^\theta p_\Theta(\theta)d\theta$. JSR-based detectors, on the other hand, exploit the structure of nuisance parameters via the GLRT detector (2) and (7). Moreover, $L_{\mathrm{B}1}$ also admits a JSR-based realization via

$$L_{\mathrm{B}2}(x) = \int(\mathbf{P}x)(\theta;\widehat{\mathbf{R}}_r)\hat{p}_\Theta(\theta)d\theta \tag{15}$$

where $\hat{p}_\Theta$ is an estimate of p_Θ. Although (14) and (15) are equivalent under exact knowledge of statistics, their designs based on training data can have substantially different performance as we will see. We will compare the performance of L, $L_{\mathrm{B}1}$ and $L_{\mathrm{B}2}$ in the following.[10] Recall that given $\mathbf{U}_\theta$ (which defines the nuisance parameters), $\mathbf{R}_r$ characterizes the detectors since $\mathbf{R}_n = \mathbf{I}$ (white noise).

We used an observation interval of $T = 50$ samples in all the simulations. The underlying signal s_r was $N_s = 16$ samples long and of the form $s_r(n) = a\sum_{k=1}^3 Z_k e^{-\beta n^2}e^{j2\pi f_k n}$, $-N_s/2+1 \le n \le N_s/2$, where $Z_k \sim \mathcal{N}(0,1)$, a is positive constant to control the SNR, β is the (fixed) variance of the Gaussian envelope, and $f_k \in [1/16, 1/4]$ (fixed). This corresponds to a rank-3 $N_s \times N_s$ matrix $\mathbf{R}_r$. Without loss of generality, for the "zero" value of the nuisance parameters, the signal $s(\theta = 0) = s_r$ is *centered* in the observation interval T. Using the estimation algorithm of Section 3, the detectors were designed using $N_d = 25, 100$ realizations each under H_0 and H_1, and were tested using 100 new realizations each under H_1 and H_0.

4.1. Estimation

Time-Frequency Detectors. For Cohen's class of TFRs, the nuisance parameters are time and frequency shifts; that is, $\theta = (t,f) \in \mathbb{R}^2$ and $(\mathbf{U}_{(t,f)}s)(\tau) = s(\tau - t)e^{j2\pi f\tau}$ [2]. $N_f = 32$ point FFTs were used in the computation of the TFRs of the observed signal x. The time and frequency nuisance parameters uniformly took on values so that the signal $s^{(t,f)}$ could be anywhere in the $T \times N_f$ time-frequency plane. In this case, the data were generated at SNR $= 10\log(E[s^H s]/E[n^H n]) = 2.3$dB. ROC curves for the three detectors for different amounts of training data are shown in Figure 1. The time-frequency (TF) GLRT detector L performs the best, followed by the TF Bayes detector $L_{\mathrm{B}2}$, which in turn performs better than the unaligned Bayes detector $L_{\mathrm{B}1}$.

Time-Scale Detectors. For the affine class of TSRs, the nuisance parameters are time-shifts and scale changes; that is, $\theta = (t,c) \in \mathbb{R} \times (0,\infty)$, and $(\mathbf{U}_{(t,c)}s)(\tau) = \frac{1}{\sqrt{c}}s\left(\frac{\tau - t}{c}\right)$ [2]. The TSRs were computed between the scales of $1/4$ and 2 using $N_c = 65$ samples. The nuisance parameters uniformly took on values so that the entire $T \times N_c$ time-scale plane was occupied by the signal $s^{(t,c)}$. The SNR in this case was 1.5dB. Figure 2 shows the comparison of the three detectors for different amounts of training data. Again the time-scale (TS) GLRT detector (L) performs the best, followed by the TS Bayes ($L_{\mathrm{B}2}$) which in turn performs better than the unaligned Bayes ($L_{\mathrm{B}1}$). The performance of the detectors becomes comparable for larger (100) training data.

4.2. Identification

Data with time-scale nuisance parameters were generated as before and the estimation algorithms for both TFRs and TSRs were applied to yield corresponding estimates of $\mathbf{R}_r$, whose eigenvalues are shown in Figure 3, along with those for the unaligned estimate $\widehat{\mathbf{R}}_s$. The eigenvalue profile is most concentrated for the time-scale-aligned data as expected. Correspondingly, the deflections of the three

[8] We assume enough data so that $\widehat{\mathbf{R}}_n$ and $\widehat{\mathbf{R}}_1$ are invertible; regularization or pseudo-inverses can be used otherwise.

[9] Assuming random parameters with probability density $p_\Theta(\theta)$.

[10] Note that the estimation algorithm also implicitly results in an estimate of p_Θ.

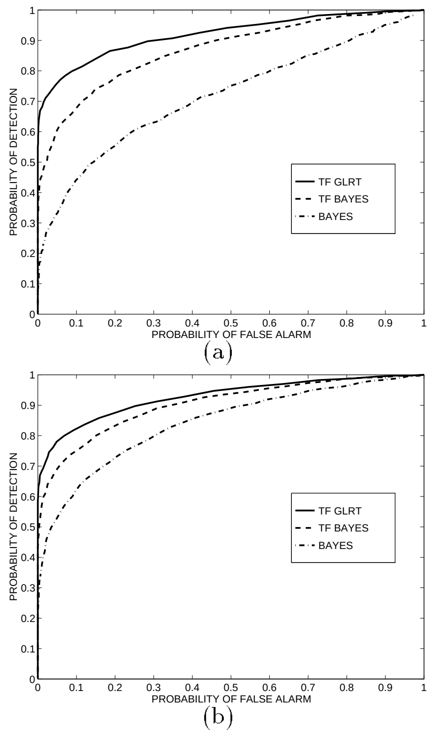

Figure 1. Time-Frequency Detectors: (a) 25 training vectors, (b) 100.

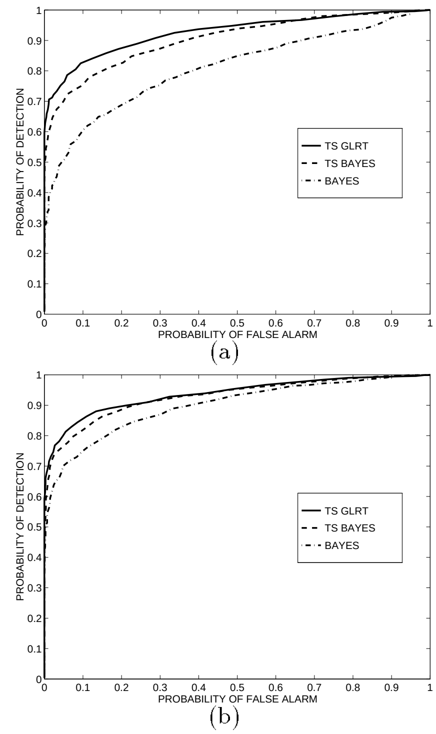

Figure 2. Time-Scale Detectors: (a) 25 training vectors, (b) 100.

correlation function estimates are: 5.7(TS), 4.09(TF) and 1.92(unaligned).[11]

5. DISCUSSION AND CONCLUSIONS

Theoretically, the ML GLRT detector cannot outperform the Bayes detector.[12] However, if the detectors are *estimated* from limited training data, the JSR-based GLRT can substantially outperform the Bayes detector as evident from Figures 1 and 2.[13] An intuitive explanation is that the GLRT detector design effectively reduces the dimensionality of the problem[14] by exploiting the signal structure. The same improvement is evident in the performance of the JSR-based realization of the Bayes detector. Moreover, Figure 3 illustrates that our algorithm has the ability of correctly "identifying" the true underlying nuisance parameters.

In summary, we have proposed a flexible framework for the practical design of GLRT-based detectors and classifiers from training data. In general, such design is rendered impossible due to difficulties in *identifying* the nuisance parameters, and in *estimating* the corresponding GLRT statistics. Our framework overcomes these difficulties by exploiting the structure of JSR-based GLRT detectors that provide a unified formulation for a wide variety of nuisance parameters. By relaxing the requirement of *a priori* statistical information, the proposed framework is suitable for the numerous practical applications of composite hypothesis testing in which only limited training data is available.

[11] The sum of the eigenvalues is the same in the three cases.

[12] In the case of random parameters. It can be shown that the Bayes detector is equivalent to a GLRT but not necessarily the ML GLRT [1].

[13] The performance of all detectors converges for increasing data — detailed analysis appears elsewhere.

[14] thereby improving the small sample performance

REFERENCES

[1] H. Poor, *An Introduction to Signal Detection and Estimation*. Springer-Verlag, 1988.

[2] A. Sayeed and D. Jones, "Optimal quadratic detection using bilinear time-frequency and time-scale representations," *IEEE Trans. Sig. Proc.*, Dec. 1995.

[3] A. Sayeed and D. Jones, "Optimum quadratic detection and estimation using generalized joint signal representations," *IEEE Trans. Sig. Proc.*, Dec. 1996.

[4] B. Samimy, G. Rizzoni, A. Sayeed, and D. Jones, "Design of training data-based quadratic detectors with application to mechanical systems," in *Proc. ICASSP '96*.

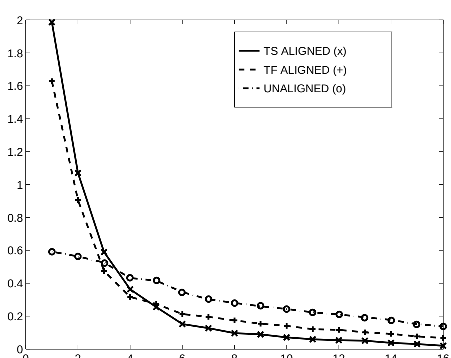

Figure 3. Comparison of time-frequency versus time-scale identification (time-scale data). Eigenvalue profile of the various detectors.