TRAINING-FREE EMPIRICAL DETECTION IN BPSK COMMUNICATIONS

Sumant M. Paranjpe, Victor Stolpman, Geoffrey C. Orsak

Department of Electrical Engineering,
Southern Methodist University, Dallas, TX 75275.
sumant@seas.smu.edu, stolpman@seas.smu.edu, gorsak@seas.smu.edu

ABSTRACT

Wireless communication links suffer from dynamic channels characterized by low SNR, fading, signal multipaths and multiple access interference which can vary significantly from Gaussian models. In model-based receivers, model mismatches can degrade system performance to unacceptable levels. As a solution, empirical receivers have been designed which achieve asymptotic optimality over a wide range of interference and noise models. Standard empirical detection techniques employ labeled training data, which requires transmission of pilot symbols resulting in bandwidth overheads. Moreover, reliable training data might not always be available, especially in dynamic channel conditions as experienced by mobile cellular radio.

In this paper we describe three training-free empirical receivers which perform as well as a reasonably trained universal detectors and nearly as well as the globally optimal receiver. We show through simulations that these universal training-free receivers based on the three estimation algorithms are robust to a wide array of channel statistics and interference models and as such serve as an ideal candidate as a wireless receiver.

1. INTRODUCTION

Typical wireless communications environments suffer from ill-characterized interference, fading and multipath, and often low SNR. One effective method for combating these channel characteristics is to adaptively optimize the communication receiver with respect to training data or a pilot signal. This approach allows for the communication system to "sense" the physical nature of the channel and use this information in constructing good receivers. Unfortunately, requiring the transmission of these additional overhead signals reduces the effective data rate over a fixed bandwidth channel.

To address this shortcoming, researchers have pursued a variety of so-called "blind" approaches to receiver adaptation. These techniques do not require the use of a pilot signal to adapt the receiver. Rather, they often utilize adaptation criteria based entirely on the transmission data (channel equalization using the constant modulus algorithm is a good example of such a blind technique). However, the tradeoff is often a measurable loss in performance over that of trained systems.

Adapting the receiver to multipath, fading, or certain forms of structured interference has been in some respects successfully address in the research literature. This however has not been the case for adaptively optimizing the receiver to statistical changes in the background interference. The difficulty in this case is primarily due to the fact that optimal receivers with respect to most non-Gaussian interference require the use of often severe non-linearities. And as is well appreciated, optimally adapting wide ranging non-linear systems still remains an open and active area of research.

1.1. Universal Receivers

Recently, new asymptotically optimal techniques to signal classification were developed in [1, 2, 3]. These new approaches attempt to classify data sequences, sourced from arbitrary distributions, based entirely upon training data obtained from each of the known sources. The fundamental theoretical approach used in these algorithms is the method of types [4]. Training sequences from each of the sources is collected in length N vectors denoted $\mathbf{T}_i^N$, where i denotes the source. The length n test sequence to be classified is denoted as $\mathbf{X}^n$. It was previously shown[1] that the asymptotically optimal statistic for determining if the test vector $\mathbf{X}^n$ was generated by the same source at the training vector $\mathbf{T}_i^N$ is given by

$$h_i(\mathbf{X}^n, \mathbf{T}_i^N) \quad = \quad d_{KL}(P_{\mathbf{X}_q^n}, P_{(\mathbf{X}^n, \mathbf{T}_i^N)_q})$$

$$+ \quad \frac{N}{n} d_{KL}(P_{(\mathbf{T}_i^N)_q}, P_{(\mathbf{X}^n,\mathbf{T}_i^N)_q}). \quad (1)$$

The quantities $P_{(\mathbf{T}_i^N)_q}$, $P_{\mathbf{X}_q^n}$, and $P_{(\mathbf{X}^n,\mathbf{T}_i^N)_q}$ represent the so-called types of the quantized data[1] vectors $\mathbf{T}_i^N$, $\mathbf{X}^n$, and the concatenated vectors $(\mathbf{X}^n, \mathbf{T}_i^N)$. These types simply represent the empirical (histogram) estimates of the joint statistics of the data vectors. The function d_{KL} is the well known divergence or relative entropy between the probability mass functions in the argument.

It has been shown that this detector is asymptotically optimum over a very wide range of statistical models for the sources[1]; for this reason, it is often referred to as a "Universal Classifier." As discussed in the abstract, this classifier has been extended and applied to problems in communication signal modulation classification with unequalized data [5, 6], face recognition for entry control [7], and as a receiver for wideband CDMA communications [8].

However, in all applications to date the formulation for the universal receiver has assumed the existence of plentiful and reliable training data from each of the sources. For communication systems, this means that either some of the available bandwidth must be used to transmit pilot signals or some of the time on the channel must be reserved for observing the background interference. In both cases, the effective data rate for the communication link will be reduced. To overcome this problem, we present in this paper three approaches for extracting the necessary training information directly from the observed test data. The first two approaches are modifications to the algorithms originally presented in [9, 10]. The new third approach is a hybrid of the first two algorithms that retains the desired robustness properties of each and can achieve performance near globally optimal detector.

2. A TRAINING-FREE UNIVERSAL CLASSIFIER

The Universal Classifier described above merely requires the "type" of the training data from each of the signal models and not the complete raw training data set. To develop a blind version of this asymptotically optimal classifier, we need only construct a methodology for obtaining accurate estimates of the types of the training data for each of the signal models solely from the observed test data $\mathbf{X}^n$.

In [10] and [9], we derived two algorithms for blindly extracting estimates of training types from the test data. In this paper, we extend both of these algorithms and introduce a new hybrid algorithm for obtaining estimates of $\hat{P}_0$ and $\hat{P}_1$ of $P_{(\mathbf{T}_0)_q}$ and $P_{(\mathbf{T}_1)_q}$ from multiple symbol observations of the data $\mathbf{X}^n$. To implement a training-free universal receiver we simply replace $P_{(\mathbf{T}_0)_q}$ and $P_{(\mathbf{T}_1)_q}$ by $\hat{P}_0$ and $\hat{P}_1$ to obtain the following decision statistics:

$$h_i^B(\mathbf{X}^n) = d_{KL}(P_{\mathbf{X}_q^n}, \bar{P}_i) + M d_{KL}(\hat{P}_i, \bar{P}_i) \qquad (2)$$

where M is the number of data symbols used in estimating $\hat{P}_0$ and $\hat{P}_1$ and where

$$\bar{P}_i = \frac{(M-1)\,\hat{P}_i + P_{\mathbf{X}_q^n}}{M} \qquad (3)$$

with in this case $\mathbf{X}^n$ representing the test data for the single bit under consideration. In this case, $\bar{P}_i$ is an estimate of the concatenated[2] type of $(\mathbf{X}^n, \mathbf{T}_i^N)_q$. The final bit decision is determine by selecting the smaller $h_i^B(\mathbf{X}^n)$.

2.1. Recursive Approach

Our first approach to obtaining $\hat{P}_i$ is based on the fact that the type of the test data observed over multiple symbols will converge to the statistical model

$$Q = \frac{P_0 + P_1}{2},$$

where P_i represent the true (and unknown) conditional statistics of the data given bit value i. Our challenge is to determine P_i from Q. To aid in the analysis, let us begin with a few assumptions: firstly because we have assumed that the background noise and interference is additive, P_0 is a shifted version of P_1, and secondly, we impose the additional and reasonable assumption that the noise and interference statistics are symmetrical about the mean of the signal value. Therefore, from an estimate of Q given by the type of the test data observed over M data symbols ($\hat{Q}^{nM}$), together with the constraints on the interference statistics, we attempt to estimate P_0 and P_1.

In developing our estimates of P_i, we begin with the constraint that $\hat{P}_0$ and $\hat{P}_1$ must perfectly reconstruct $\hat{Q}^{nM}$. That is our solution must be such that $\hat{Q}^{nM} = \frac{\hat{P}_0+\hat{P}_1}{2}$. Unfortunately, it can be shown that there are

[1] Here, all of the data sequences are assumed to be first preprocessed by quantization so that the signals can be analyzed and operated on more efficiently on a DSP (thus the subscript q).

[2] Note, it is easily shown that one may compute the type of the concatenated data set $(\mathbf{X}^n, \mathbf{T}_i^N)_q$ by simply using a weighted average of the individual types $P_{\mathbf{X}_q^n}$ and $P_{\mathbf{T}_q^N}$.

an infinite number of estimates of P_i which will in fact perfectly reconstruct $\hat{Q}^{nM}$.

To limit the solution space, consider the following constraint class of estimates:

$$\hat{P}(P_{\mathbf{X}_q^n}, \hat{Q}^{nM}) = P_{\mathbf{X}_q^n} + 2\,\mathrm{diag}(\Gamma)\left[\hat{Q}^{nM} - \frac{P_{\mathbf{X}_q^n} + P^f_{\mathbf{X}_q^n}}{2}\right] \quad (4)$$

where the superscript f represents the operation which flips the type about the center of the quantizer's range. It can be shown that under certain minimal constraints on the matrix diag(Γ), $\hat{P}(P_{\mathbf{X}_q^n}, \hat{Q}^{nM})$ will always perfectly reconstruct $\hat{Q}^{nM}$, and thus forms a parametric family of potential solutions to our estimation problem.

To determine the best such estimate from this family, we find the matrix diag(Γ) such that

$$\Gamma = \arg\min_{\Gamma} d_{KL}(P_{\mathbf{X}_q^n}, \hat{P}(\mathbf{X}^n, \hat{Q}^{nM})), \quad (5)$$

where $\mathbf{X}_q^n$ is observed over any one bit. The solution to this optimization problem will result in estimates $\hat{P}_0$ and $\hat{P}_1$ which will always reconstruct the type $\hat{Q}^{nM}$ and will be nearest to the test data types, which will be estimates of P_0 or P_1.

It can be shown using the method of Lagrange multipliers that the solution to this problem is

$$\Gamma_k = \frac{P_{\mathbf{X}_q^n}(k)}{P_{\mathbf{X}_q^n}(k) + P_{\mathbf{X}_q^n}(K-k+1)} \quad (6)$$

where $P_{\mathbf{X}_q^n}(k)$ and $P_{\mathbf{X}_q^n}(K-k+1)$ represents the k^{th} and $(K-k+1)^{th}$ element of the type of the test data over a single bit with the additional requirement that $\Gamma_k = 1 - \Gamma_{K-k+1}$ where the constant K represents the total number of bins in the quantizer.

Unfortunately, due to the fact that the solution to this optimization problem is determined entirely from data, there can be some analytic and computational difficulties to arriving at good choices for $\hat{P}_i$. Firstly, the resulting analytic solution is only valid if all of the values of $P_{\mathbf{X}_q^n}$ are not equal to zero. This corresponds to having observed test data in every quantizer bin location during each test bit – not a particularly likely scenario if the number of samples per bit is small, and the number of bins in the quantizer is large. Secondly, there is no guarantee that the resulting estimators for P_i will be symmetrical about their means μ_i. As you will recall, this is one of our underlying assumption regarding the nature of the interference statistics.

By forcing symmetry about the an estimated signal mean $\hat{\mu}_i$, we can obtain a symmetric estimate $\hat{P}_S$. However, this solution is not guaranteed to perfectly reconstruct $\hat{Q}^{nM}$. The obvious solution is to replace $P_{\mathbf{X}_q^n}$ in (4) with $\hat{P}_S$ to solve for another perfectly reconstructing solution. By iterating the two procedures, we then have a recursive algorithm which results in a solution that perfectly reconstructs and is nearly symmetric. For this paper, we define one iteration as first forcing symmetry and then optimizing using (4) and (6).

It should be pointed out that the starting point for this iterative process is arbitrary so long the input to the algorithm is a valid type. The choice described in (5) is to use the type of the current symbol $P_{\mathbf{X}_q^n}$. However, other choices are available to us and offer better overall performance.

2.2. BinFit Algorithm

As an alternate approach to obtaining estimates of training types, the BinFit algorithm [9] is a numerical one-shot scheme to *build* training types estimate using the observed joint empirical statistics $\hat{Q}^{nM}$ derived from the raw data. The algorithm assumes that during any limited observation interval, the observed data under each symbol has finite support. We can therefore design a uniform K-bin quantizer which spans the entire support set. The algorithm additionally requires that the sample signal means ($\mu_i, i = 0, 1$) lie on a bin edge. Though this may put constraints on the quantizer design, with sufficiently fine quantizer resolution and enough data, this requirement can be relaxed at the expense of a marginal loss in performance.

Let K^* be the number of bins separating the two signal means. Then based on the assumptions above

$$\hat{P}_0(l) = \hat{P}_0(K-l+1) \doteq \left[\hat{Q}^{nM}(l) + \hat{Q}^{nM}(K-l+1)\right]/2 \quad (7)$$

for $l = 1, 2, \cdots, K^*$. Having derived these values of $\hat{P}_0$, the remaining values are calculated from the following equation: for $k = K^* + 1, \cdots, K/2$

$$\hat{P}_0(k) = \hat{P}_0(K-k+1) \doteq \left[\hat{Q}^{nM}(k) - \hat{P}_0(k-K^*) + \hat{Q}^{nM}(K-k+1) - \hat{P}_0(k-K^*)\right]/2. \quad (8)$$

The corresponding type estimate for P_1 is computed as $\hat{P}_1 = \hat{P}_0^f$ where the f-operator is the flipped version of the argument about the center of the quantizer. Since the algorithm is entirely data-driven, some computational inconsistencies may arise in the solution with a finite data observation. Since the estimate $\hat{P}_i$ is a probability distribution, it can not have negative bin values. The algorithm forces all negative values to zero and then renormalizes the resultant to form a valid type.

The algorithm is simple to implement (complexity $\mathcal{O}(3K/2)$ in accumulates) but is limited by the fact that errors are propagated past their local support along the quantizer. This is most evident for small data sets where $\hat{Q}^{nM}$ is a coarse estimate of the true joint type. Importantly, with sufficient observed data, the BinFit algorithm gives an efficient method for accurately estimating the training types by leveraging the rich structure in the problem.

2.3. Hybrid Approach

In this section, we propose combining the two above approaches into a "hybrid" algorithm that offers a wider range of stable operating conditions than seen in the BinFit algorithm and improved convergence over the recursive algorithm. We know through experimentation that the performance of the blind universal receiver using the estimates obtained from the recursive algorithm depends on the algorithm's initial input. Rather than using $P_{\mathbf{X}_q^n}$ as the input to the recursive algorithm as discussed in section 2.1, experimental results have shown that $\hat{Q}_{nM}$ often serves as a more reliable initial input for the algorithm. Nevertheless, we will show that further improvements in the recursive algorithm can be made by using more informative initial estimates derived from the output of the BinFit algorithm.

Thus, the new hybrid algorithm simply requires the data first be processed by the BinFit algorithm in order to generate an accurate starting point for the recursive algorithm. Subsequent improvements in this solution are obtained through iterations in the recursive algorithm. Examples presented in the following section demonstrate the utility of the new training-free universal receiver over a wide range of operating conditions.

3. EXPERIMENTS

To demonstrate the relative performance of the three proposed approximation techniques, we compare the blind receiver using the different algorithms against (1) a suitably trained version of the universal receiver and (2) the well known Gaussian based receiver structures for each of different background noises.

All test data is formated into $M = 512$ CDMA bits of 64 chips in length and used BPSK modulation similar to the forward link in the IS-95 standard. Although we only use stationary channels in these experiments,

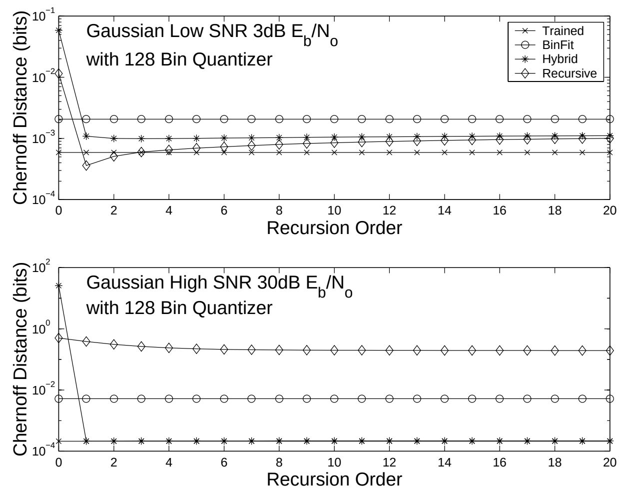

Figure 1: Chernoff distance between various $\hat{P}_0$ estimates and the true P_0 versus the number of recursions for Gaussian background noise at 3dB and 30dB E_b/N_o averaged over 100 frames.

this data framing approach supports quasistationary, slowly fading channels which remain relatively fixed over a frame period.

In the top half of Fig. 1, we compare the Chernoff distances[3] between the estimated training types and the true training types for the three estimation algorithms. As can be seen all perform quite well, with the standard recursive algorithm and the hybrid algorithm resulting in nearly identical performance.

In the bottom half of Fig. 1, we compare the Chernoff distances between the estimates to the true types in an AWGN channel with high SNR (30dB E_b/N_o). Again we see the impact of the initial starting point for the recursive approach. Using the standard input to the recursive algorithm, the estimate converges to a solution near $\hat{Q}^{nM}$ and suffers a significant penalty in detector performance. However, using BinFit's $\hat{P}_0$ as the starting point, the Hybrid algorithm solves the convergence issue by locking onto the clairvoyant, trained estimate.

In Figure 2, we compare detector performances against the globally optimal detector for an AWGN channel. While Fig. 1 shows that the estimates of the training types from each of the algorithms are differ from one another, Fig. 2 establishes that these differences translate into only minor differences in BER performance. Importantly, each of these training-free algo-

[3] As will be seen in Fig. 2, the Chernoff distance serves as an accurate measure of the relative BER performances of these algorithms.

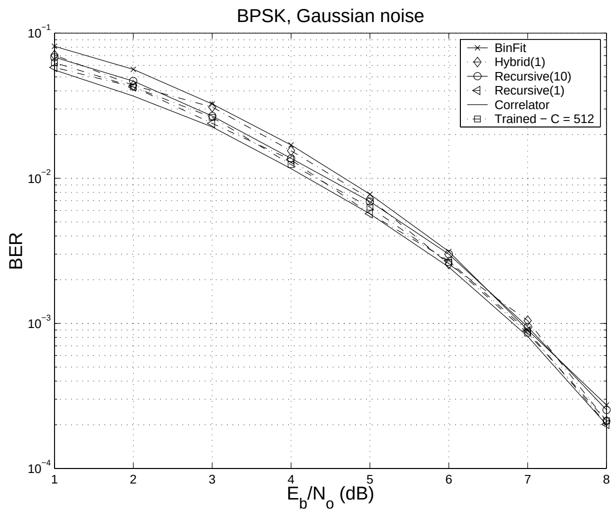

Figure 2: BER versus E_b/No in AWGN channel. The number inside the parenthesis pair for the Hybrid and Recursive implementations indicate the number of iterations.

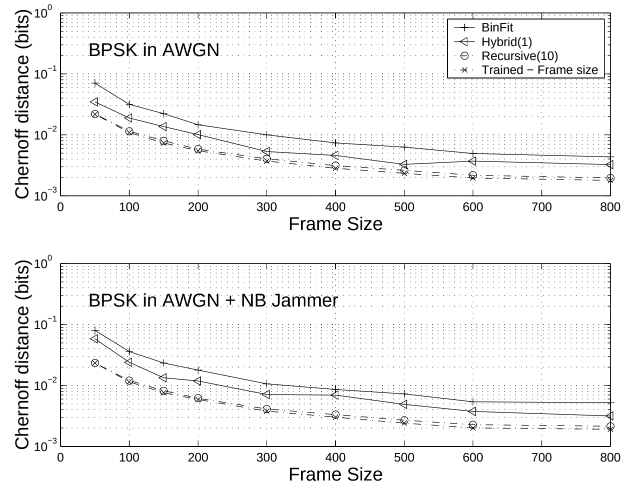

Figure 3: Chernoff distance between various $\hat{P}_0$ estimates and the true P_0 versus the number of data bits in a frame of test data for AWGN channel (top) at 3dB E_b/N_o, (bottom) 3dB E_b/N_o with narrow band (NB) interference. The NB jammer is a pure sinusoid at 4 times the signal amplitude inside the CDMA bandwidth (1.2288MHz) with frequency offset of 240KHz from the carrier (2.4GHz). The number inside the parenthesis pair for the Hybrid and Recursive implementations indicate the number of iterations.

rithms results in performance which nearly matches the performance of the globally optimum receiver and the well-trained universal receiver. Although not shown here, similar experiments using Laplacian noise render nearly identical results.

Let us consider a more complicated channel. In this example, we evaluate the performance of each of the training-free algorithms when the channel suffers from both AWGN and a narrow band interferer. This example is intended to simulate a CDMA system competing in a unlicensed band with existing narrow-band operators. In Fig. 4, we plot the Chernoff distance between the estimates of the channel statistics to the "true" channel statistics derived from exceeding large data sets as we increase the frame size of the packet. Under both channel conditions, all estimates, trained and blind, approach the true source distributions as the frame size increases, and thus both the trained and blind detectors approach the minimum BER.

In Fig. 4, we present the corresponding BER results for the AWGN channel (8dB) with a narrow band interferer precisely at the carrier frequency. Again we used a modulation scheme similar to IS-95 operating at a 2.4GHz carrier. For this experiment, we randomly selected both the user and base station chip signatures of 13 and 128 chips respectively. As can be seen in Fig. 4, the training-free algorithms significantly outperform the Gaussian based receiver (matched filter). Importantly, as the jammer's power increases, the training-free receiver's performance actually gets

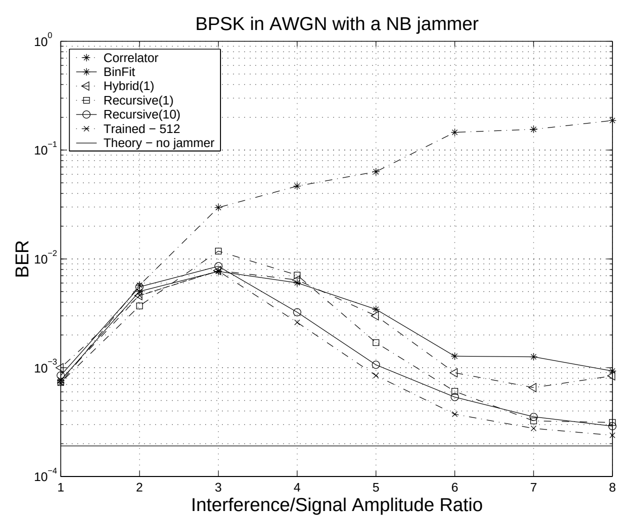

Figure 4: BER versus amplitude of NB jammer relative to signal amplitude. The NB jammer at the carrier frequency (2.4GHz) in an AWGN channel with 8dB E_b/N_o. The number inside the parenthesis pair for the Hybrid and Recursive implementations indicate the number of iterations.

better and tends to approach the performance of the jammer-free channel!

4. CONCLUSION

In this paper we developed a new blind digital communications receiver for channels with ill-characterized background interference. This training-free receiver was shown to perform as well as the trained universal receivers developed in previous research and nearly as well as the globally optimal receiver. In particular, the new "hybrid" blind universal receiver was shown to be computationally efficient, and very robust to a wide array of channel statistics and SNR values.

REFERENCES

[1] M. Gutman, "Asymptotically optimal classification for multiple tests with empirically observed statistics," *IT*, vol. IT-35, no. 2, pp. 401–408, Mar. 1989.

[2] N. Warke and G. C. Orsak, "On the theory and application of universal classification to signal detection," in *Proceedings of the 1994 IEEE/IMS Information Theory Workshop*, 1994.

[3] B.-P. Paris, G. C. Orsak, N. Warke, and H. Chen, "Type-based detection for spread spectrum," in *ICASSP Proc.*, Munich, 1997.

[4] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, 1981.

[5] N. Warke and G. C. Orsak, "A universal methodology for signal classification in non-gaussian environments," in *Proceedings of the 1994 IEEE Digital Signal Processing Workshop*, 1994.

[6] N. Warke, G. C. Orsak, B.-P. Paris, and H. Chen, "Universal modulation classification for unknown non-gaussian dispervie channels," in *Proc. 199 Conference on Information Sciences and Systems*, Johns Hopkins University, Baltimore, Md, March 19–21, 1997.

[7] N. Warke and G. C. Orsak, "An information theoretic methodology for noisy image classification with application to face recognition," in *Proc. 1996 Conference on Information Sciences and Systems*, Princeton University, Princeton, New Jersey, March 20-22, 1996.

[8] V. Stolpman and G. C. Orsak, "Type-based receiver for wideband cdma," in *Proceedings of the 1999 IEEE Wireless Communications and Networking Conference*. September 21-24 1999, New Orleans, LA.

[9] S. M. Paranjpe, V. Stolpman, and G. C. Orsak, "A training-free empirical receiver for qam signals," in *Proc. of 1999 Wireless Communication and Networking Conference*, New Orleans, LA, September 21-24, 1999.

[10] V. Stolpman, S. M. Paranjpe, and G. C. Orsak, "A blind information theoretic approach to automatic signal classification," in *Proc. of 1999 Military Communications International Symposium*, Atlantic City, NJ, October 31 - November 3, 1999.