

# Statistical Models of Reconstructed Phase Spaces for Signal Classification

Richard J. Povinelli, *Senior Member, IEEE*, Michael T. Johnson, *Senior Member, IEEE*,  
Andrew C. Lindgren, *Student Member, IEEE*, Felice M. Roberts, *Student Member, IEEE*, and  
Jinjin Ye, *Student Member, IEEE*

**Abstract**—This paper introduces a novel approach to the analysis and classification of time series signals using statistical models of reconstructed phase spaces. With sufficient dimension, such reconstructed phase spaces are, with probability one, guaranteed to be topologically equivalent to the state dynamics of the generating system, and, therefore, may contain information that is absent in analysis and classification methods rooted in linear assumptions. Parametric and nonparametric distributions are introduced as statistical representations over the multidimensional reconstructed phase space, with classification accomplished through methods such as Bayes maximum likelihood and artificial neural networks (ANNs). The technique is demonstrated on heart arrhythmia classification and speech recognition. This new approach is shown to be a viable and effective alternative to traditional signal classification approaches, particularly for signals with strong nonlinear characteristics.

**Index Terms**—Reconstructed phase spaces (RPSs), signal classification, statistical models.

## I. INTRODUCTION

WE present this new approach based on reconstructed phase spaces (RPSs) as an alternative to traditional linear approaches to signal analysis and classification, which are typically based on frequency domain characteristics. The underlying assumption of traditional linear methods is that the salient information about signal characteristics is contained in the frequency power spectrum. From a stochastic process perspective, the first- and second-order statistics of the signal are represented by this power spectral information.

However, there are many types of signals, both theoretical and experimental, for which a frequency domain representation is insufficient, because it is generally not possible to distinguish between signals that have the same power spectra but differing phase and/or higher order spectra. For example, signals generated through nonlinear differential or difference equations typ-

Manuscript received July 22, 2004; revised July 11, 2005. This work was supported in part by the National Science Foundation by Grant IIS-0113508, by the Department of Education GAANN Fellowship, by an American Heart Association Fellowship, and by the Milwaukee Foundation Frank Rogers Bacon Research Assistantship. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Tulay Adali.

R. J. Povinelli, M. T. Johnson, and F. M. Roberts are with the Electrical and Computer Engineering Department, Marquette University, Milwaukee, WI 53233 USA (e-mail: richard.povinelli@mu.edu; mike.johnson@mu.edu; felice.roberts@mu.edu).

A. C. Lindgren is with the ATK Mission Research Corporation, Dayton, OH 45430 USA (e-mail: andrew.lindgren@atk.com).

J. Ye is with the Electrical Engineering Department, University of California, Los Angeles, CA 90095 USA (e-mail: jjye@icsl.ucla.edu).

Digital Object Identifier 10.1109/TSP.2006.873479

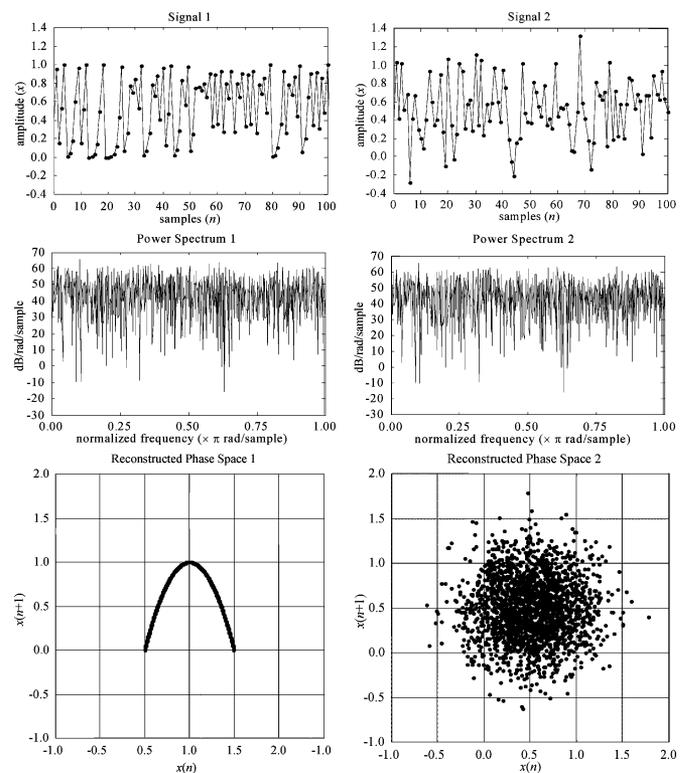


Fig. 1. Signals, broadband power spectra, and reconstructed phase spaces for two examples.

ically exhibit broadband power spectral characteristics that are difficult to interpret, because the spectra do not contain sharp resonant peaks. An illustrative example of two signals that are indistinguishable using power spectral classification methods are seen in Fig. 1. The first signal in Fig. 1 is the logistic map  $x_{n+1} = kx_n(1 - x_n)$ , with  $k = 4$ . The second is the logistic map's FT surrogate, which maintains the signal's Fourier transform amplitude, but randomizes the phase [1].

Fig. 1 provides a graphical motivation for the signal analysis and classification approach presented in this paper. Signal classification techniques based on power spectral information can not distinguish between different signals that have the same power spectrum, but such signals may be distinguishable in a RPS. Additionally, the underlying theory, outlined in more detail in Section III, guarantees that the dynamics of any system are fully described by a RPS generated from any single state variable, provided that the dimension of the RPS is greater than twice the box counting dimension of the original system [2]. From a practical perspective for discrete-time signals, this

means that a RPS, which is a multidimensional plot of the time series signal against time-delayed versions of itself, contains all the information of the underlying system [3].

The particular models used here are statistical distributions that can be learned over RPSs and then used to classify unseen signals. Both nonparametric distributions based on binning and occurrence counts and parametric distributions based on Gaussian mixture model (GMM) distributions are illustrated. Two types of classifiers are applied: a Bayes maximum likelihood classifier and an artificial neural network (ANN) classifier.

Thus, the proposed approach is particularly suited for differentiating between signals where the phase of the signal contains important differentiating information, because such differentiating phase information is captured by the RPS, or where the state structure captured by the RPS reveals state variables and relationships between state variables that provide greater differentiability across classes than the original state variable by itself.

The new approach is applied to two domains—distinguishing heart arrhythmia and speech phoneme classification. The results from these two applications show that this new approach can be applied to a variety of signal classification problems.

Previous work in signal classification is discussed in Section II. Background of the underlying dynamical system's theory is given in Section III, with detailed descriptions of the distribution models, learning algorithms, and classification techniques in Section IV. Section V discusses the experimental setup and results, with conclusions in Section VI.

## II. PREVIOUS WORK

In addition to the traditional signal analysis models based on spectrum amplitude such as autoregressive modeling [4] or cepstral analysis [5], there is also a large body of work on signal detection and classification in the field of communications [6], based on statistical decision theory. These methods are founded on the existence of an *a priori* underlying transmission or symbol model that can be used to analytically determine closed form conditional distributions for each signal class and derive a maximum likelihood detector. In the case of arbitrary signal classification, with complex underlying systems such as heart rhythms and speech, such complete time-domain signal models do not exist. The general approach to classification for these systems is based on machine learning principles, using preprocessing to identify relevant features for classifier models, rather than a symbol-based time-domain representation.

Most research in using RPSs focuses on estimating dynamical invariants, which are not sensitive to initial conditions or smooth transformations of the space. These invariants may be classified into three categories: metric (Lyapunov exponents and dimension), natural measure (density) [7], and topology. Various methods of estimating dimension have been proposed, such as correlation dimension [8], minimum phase space volume [9], and box counting [7]. Density estimation techniques are typically histogram [10], [11] or GMM [12] based. Topological analysis techniques include templates [13] and global vector field reconstruction [14].

Dynamical systems methods have been used in many applications. Lyapunov exponents have been used for classifying

signals as chaotic [14] and as additional features for speech recognition [15]. Fractal dimension, Richardson dimension, Lempel-Ziv complexity, and Hurst exponent have been used as features for classifying simulated and beta emission signals [16]. Estimates of dimensions have been used to analyze speech [17] and heart rate variability [18]. However, estimation of metric invariants is highly sensitive to noise and sample size. Without large sample sizes and the use of nonlinear filtering techniques, their estimations are suspect [19].

Topological analysis techniques such as templates [13] have been applied to simulated chaotic systems [20], voltage [20], and laser [21] time series. Global modeling techniques have been applied to the computation of Lyapunov exponents [14] and biological signal classification [22]. The topological analysis approach fits a functional model to the attractor. This approach is seen in [22].

However, there is relatively little literature directly applying RPSs to classification. Our approach differs from metric-based approaches by modeling the density of the RPS directly instead of calculating a measure of an average trajectory divergence (Lyapunov exponents) or estimating the dimension. The work presented here develops statistical features of RPSs, whereas topological analysis approach builds global vector reconstructions.

## III. PHASE SPACE RECONSTRUCTION THEORY

The basis of this approach is that given access to the state structure of a system, a classification of such systems can be developed. We start by presenting a theoretical construct of the problem. Given a finite-dimensional system state space  $M$  and  $\varphi : M \rightarrow M$ , the dynamics of the system, a system is described by the pair  $\langle M, \varphi \rangle$ . We then define a set  $\Phi$  of all possible dynamics on  $M$  with a topology  $\mathfrak{T}$ . Without loss of generality, we assume  $M$  to be  $d$ -dimensional, because given any  $M' \subset M$ ,  $M$  can be replaced by  $M \cup M'$ . The system classification then becomes one of partitioning  $\Phi$  according to the requirements of the classification problem with a particular dynamics  $\varphi$  identified with a particular partition  $P_i$  such that  $\Phi = \cup P_i$ , where  $P_i \cap P_j = \emptyset, i \neq j$ .

The problem for real world systems is how to gain access to and represent  $\varphi$  for a particular system. The approach used here is phase space reconstruction, also known as phase space embedding, and was first proposed in [23]. The methods for representing  $\varphi$ , which are the contributions of this work, are presented in the following section.

The central premise is that a space and its associated dynamics, which are topologically equivalent to the original system space  $M$  and its dynamics  $\varphi$ , can be recovered or unfolded from a time series of observations of a single state variable for the original system  $\langle M, \varphi \rangle$ .

Whitney showed that an  $d$ -dimensional topological space can be embedded in  $\mathbb{R}^{2d+1}$  [24], where an embedding is a homeomorphic mapping from one topological space to another. Takens [3] showed that it is a generic property that the map  $\Phi_{(\varphi, x)} : M \rightarrow \mathbb{R}^{2d+1}$  defined by

$$\Phi_{(\varphi, x)}(y) = (x(y), x(\varphi(y)), \dots, x(\varphi^{2d}(y))) \quad (1)$$

is an embedding, where  $M$  is an  $d$ -dimensional state space,  $\varphi : M \rightarrow M$  is a twice continuously differentiable diffeomorphism that describes the dynamics of the system, and  $x : M \rightarrow \mathbb{R}$  is a twice continuously differentiable function representing the observation of a single state variable.

Working from these original theorems, Sauer and Yorke [2] extended Takens' work by showing that almost every time-delay map  $\Phi_{(\varphi, x)}$  is an embedding, indicating that except for a set of degenerate cases with measure zero, the topological equivalence property is guaranteed. Therefore, these theorems guarantee that for almost every time delay embedding, the reconstructed dynamics of such a map are topologically identical to the true dynamics of the system [2]. In addition, they found a tighter bound on the required dimension as  $d > 2d_0$ , where  $d_0$  is the boxcounting dimension of the attractor of the underlying system.

In other words, we have a mechanism for obtaining a continuous, one-to-one, and onto transformation from  $\langle M, \varphi \rangle$  to  $\langle \mathbb{R}^d, \mathbf{X} \rangle$ , where  $d > 2d_0$  and  $\mathbf{X}$  is the trajectory matrix defined as follows. Given a time series  $x = x_n, n = 1, \dots, N$ , a sequence of state variable observations, a trajectory matrix  $\mathbf{X}$  of dimension  $d$  and time lag  $\tau$  is defined as

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_{1+(d-1)\tau} \\ \mathbf{x}_{2+(d-1)\tau} \\ \vdots \\ \mathbf{x}_N \end{bmatrix} = \begin{bmatrix} x_{1+(d-1)\tau} & \cdots & x_{1+\tau} & x_1 \\ x_{2+(d-1)\tau} & \cdots & x_{2+\tau} & x_2 \\ \vdots & & \ddots & \\ x_N & \cdots & x_{N-(d-2)\tau} & x_{N-(d-1)\tau} \end{bmatrix} \quad (2)$$

where each row vector in the matrix represents a single point in the space;

$$\mathbf{x}_n = [x_{n-(d-1)\tau} \quad \cdots \quad x_{n-\tau} \quad x_n] \quad (3)$$

where  $n = (1 + (d-1)\tau) \dots N$ . A row vector  $\mathbf{x}_n$  is a point in the RPS. The pattern traced out by  $\mathbf{X}$  in  $\mathbb{R}^d$  is typically referred to as an attractor, even when the technical definition of an attractor is not formally met. We adopt this terminology.

Recall, that the system/signal classification problem in this work is addressed by transforming a signal from a particular system into a RPS, which has a mathematical correspondence with the underlying system. Therefore, given  $\mathfrak{X}$ , the collection of all possible RPSs  $\mathbf{X}$ , the system/signal classification problem is to define a partition of  $\mathfrak{X}$  such that  $\mathfrak{X} = \cup P_i$  and a mechanism for identifying a particular  $\mathbf{X}$  with a particular  $P_i$ .

As is discussed next, a statistical machine learning approach is taken for defining the partition and for identifying an  $\mathbf{X}$  with a  $P_i$ . In general, the classification accuracy will depend on how well the model of a partition  $P_i$  of  $\mathfrak{X}$  characterizes the signals that are labeled as belonging to that partition and how different the model of  $P_i$  is from the models of other partitions.

#### IV. METHODS

As aforementioned, the approach adopted here is based on direct statistical modeling of the RPS, through the estimation of a joint probability density function over that space. Both discrete nonparametric and continuous parametric distributions are

implemented. There are three main steps to applying the new methods. The first step, data analysis, is to determine the time lag and embedding dimension of the RPS and compensate for any nonstationarity of the observation function. The second step is to generate statistical models of the attractors. This is done using both discrete and continuous models. The final step is to build classifiers.

##### A. Data Analysis

To construct a RPS from a signal, the dimension of the RPS and the time lag at which to sample the signal must be selected. Although proper selection of dimension and time lag may appear to be critical to the success of the methods presented in this paper, in practice the methods presented here are effective across a range of dimensions and time lags. In addition, existing methods for choosing time lag, such as the first minimum of the automutual information function or the first zero crossing of the autocorrelation [25], dimension, such as false nearest neighbor [26] or Cao's method [27], or both [28], [29] do not optimize to classification accuracy. Hence, they provide only a first estimate of appropriate time lags and dimensions. Thus, as proposed in [12], the mode of distribution of the first minimum of the automutual information function across all signals is used as an initial estimate of the time lag. Similarly, the mean plus two standard deviations of the distribution of false nearest neighbor dimensions across all signals is used as an initial estimate of the dimension. We show that the best empirical time lag and dimension may differ from the initial estimate in Section V.

Recall from (1) that  $x : M \rightarrow \mathbb{R}$  is a twice continuously differentiable function representing the observation of a single state variable. For many applications, the gain of this function is not controlled across and sometime within signals. For the heart arrhythmia example presented in Section V, baseline wandering, where the mean of the signal changes when the electrical sensor is physically disturbed, is a problem. For both datasets, the gain varies across signals. Thus, a mechanism is needed to compensate for a time varying observation function. The type of compensation will depend on the nature of the observation function. For example, the electrocardiogram (ECG) signals can be low pass filtered to remove baseline wandering and standardized in the time domain as follows:  $x'_n = (x_n - \mu_{\mathbf{x}}) / \sigma_{\mathbf{x}}$ .

The next step is to form the RPSs as specified by (2) for statistical modeling. When a Bayes classifier is used, the training signals are used to form RPSs for each class by appending in a column fashion the  $\mathbf{X}$ 's formed from the signals belonging to that class. This enables statistical models for each class to be learned. When the ANN classifier is used, all training signals are used to form a single RPS. This enables a consistent set of features to be extracted for training the ANN.

##### B. Statistical Models

The use of histograms as an estimate of the discrete probability mass function (pmf) of the attractor is straightforward. The space is divided into bins and occurrence counts in each bin over the training examples, divided by the total number

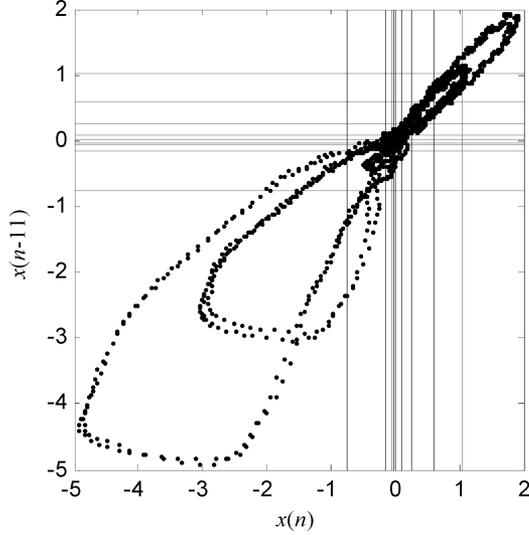


Fig. 2. Bin-based distribution modeling (100 bins).

of points, provide a direct estimate of the posterior probability within each bin

$$\hat{p}(b) = \frac{\text{Number of example points in bin } b}{\text{Total number of points}}. \quad (4)$$

Given  $\mathbf{x}_n$  in bin  $b$ , the posterior probability for  $\mathbf{x}_n$  is  $\hat{p}(b)$ .

In order to improve the reliability of the pmf estimate, it is desirable to establish a binning system that more accurately reflects the underlying distribution. The distribution of points throughout the phase space is nonuniform; therefore, nonuniformly spaced bins are used. Fig. 2 illustrates the nonuniformly spaced bins. A two-step process is used to form the bins. First, along each dimension a set of intercepts is computed such that in that dimension the histogram formed by the intercepts is uniform. The outlying bins of this one-dimensional (1-D) histogram extend to infinity. Second, the higher-dimensional bins are formed as hypercubes whose boundaries are formed by the intercepts determined in the first step. In forming a RPS for a signal of reasonable length, the intercepts are approximately the same in each dimension. Fig. 2 illustrates, as an example, the case of a 2-D RPS with nine intercept values per dimension, which forms a 10 by 10 bin mass function over the entire space.

Disadvantages of the bin-based system include an exponentially increasing number of bins as RPS dimension increases and fixed bin boundaries. Thus, a GMM, which has soft boundaries and does not necessarily require an exponential number of mixtures with respect to RPS dimension, is studied. The probability of  $\mathbf{x}_n$  is

$$p(\mathbf{x}_n) = \sum_{m=1}^M w_m p_m(\mathbf{x}_n) = \sum_{m=1}^M w_m \mathcal{N}(\mathbf{x}_n; \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \quad (5)$$

where  $M$  is the number of mixtures,  $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$  is a normal distribution with mean  $\boldsymbol{\mu}_m$  and covariance matrix  $\boldsymbol{\Sigma}_m$ , and  $w_m$  is the mixture weight. With the constraint that  $\sum w_m = 1$ , a GMM is a probability model that can accurately reflect a wide range of distributions with arbitrary precision given enough mixture components.

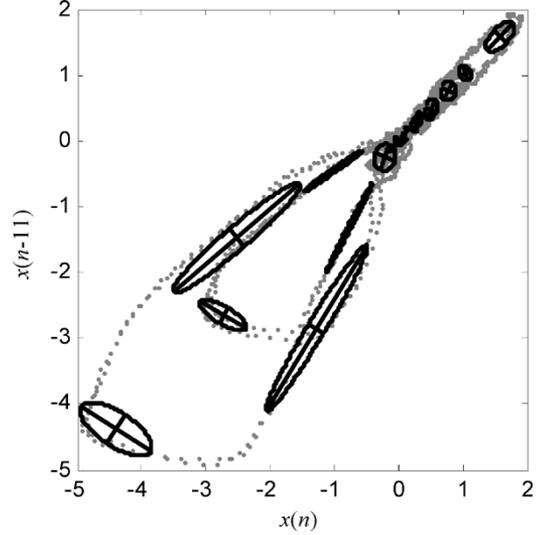


Fig. 3. GMM-based distribution modeling (16 mixtures).

Given training data, the parameters for the GMM can be estimated using the well-known Expectation-Maximization (EM) algorithm [30]. The number of mixtures  $M$  is determined empirically, but is robust across a range of mixture components as shown in Section V. A visualization of a GMM is shown in Fig. 3, where the principle axes of the ellipses indicate the one standard deviation of each mixture in the model.

Both the GMM and bin-based approaches require exponential increases in model complexity to effectively model arbitrary feature spaces as the dimensionality of such feature spaces is increased. Because of the structure and hard bound boundaries of the bin-based model, it suffers from these scalability issues when applied to a RPS, because it models the whole space. The GMM approach is not completely immune to scalability problems, but it does provide two features in dealing with scalability that the bin-based approach does not. First, through the use of EM, the GMM model more accurately models the distribution of points in the RPS, i.e., it models the attractor. An example of this is illustrated in Fig. 3. Second, if the number of mixtures is held constant, a GMM increases linearly in complexity as the RPS dimension is increased.

### C. Classifiers

To use either the binning or GMM models for signal classification, training data from several different types of signals are normalized, embedded, and the selected statistical model is learned. We have used two classifiers: a Bayesian maximum likelihood and an ANN [31].

Bayes classification of a new test signal is accomplished by computing the conditional likelihoods of the signal under each learned model and selecting the model with the highest likelihood. The likelihoods are computed on a point-by-point basis from the learned attractor models

$$p(\mathbf{X} | c_i) = \prod_{n=1}^N p(\mathbf{x}_n | c_i) \quad (6)$$

$$\hat{c} = \arg \max_i p(\mathbf{X} | c_i)$$

where  $c_i$  is the  $i$ th class. As aforementioned, attractor models are learned for each class individually. Thus, for the bin-based method each class has its own set of intercepts with corresponding probability mass functions, while for the GMM each class has unique mixture means and variances with corresponding cluster weights.

The use of ANNs as a classifier requires that a set of features be extracted from the learned statistical model. This is done by using all training signals to form a single RPS. This enables a consistent set of features to be extracted for training so that there is a common set of intercepts in the case of bins or a common set of mixtures in the case of GMMs. For a GMM, this arrangement is often referred to as a tied-mixture model. Hence, the bin-based system with global intercepts is referred to as a tied-bin model. The features are the GMM or bin weights and are calculated, respectively, for a particular signal by determining the weights from the global GMM or the weights for the global bins for that signal.

A set of ANNs is trained, one for each class, by using the bin counts or mixture weights as inputs. The training output is set to one when the signal belongs to that class and to zero when it does not. Testing is accomplished by computing the features over a new signal sample, and selecting the class label corresponding to the ANN with the greatest output.

## V. EXPERIMENTS AND RESULTS

We now present the application of the RPS-based signal classification methods to two different domains. The first application is to distinguishing heart arrhythmias. This application shows how the new approach can accurately classify heart arrhythmias with only a 2-s signal. The second application is speech phoneme classification. Here we give initial results showing that for small datasets the new approach is similar to traditional spectral methods.

### A. Heart Arrhythmia Classification

The goal of this application is to rapidly and accurately classify four different heart rhythms using signals generated by lead II of an electrocardiogram (ECG) [10]. These rhythms are sinus rhythm (SR) and the three arrhythmias: monomorphic ventricular tachycardia (MVT), polymorphic ventricular tachycardia (PVT), and ventricular fibrillation (VF). This is a clinically relevant problem because different therapies are applied depending on the type of rhythm. No therapeutic action is taken for SR. For VF and PVT, electronic shock is the most prevalent therapy. In the case of VF, collapse occurs within seconds and death within minutes unless the VF is corrected with the passage of a large electrical current through the heart muscle. However, shocking SR can sometimes induce VF.

The data for these experiments were obtained from six patients during intercardiac defibrillator implantation. Data was collected from lead II of a 12 lead ECG. The signals were antialias filtered with a cutoff frequency of 200 Hz and subsequently digitized at 1200 Hz. The dataset includes 306 s of SR, 126 s of MVT, 116 s of PVT, and 114 s of VF. Because the data was collected during surgery and the chest was open, the lead placement was not ideal. This shows that the proposed methods

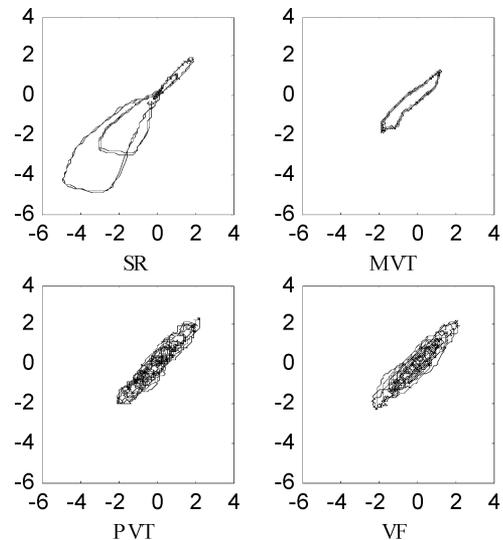


Fig. 4. Reconstructed 2-D phase space for examples of SR, MVT, PVT, and VF with  $\tau = 11$ .

are somewhat robust to variances in the exact state variable that is measured. Data were examined by two experts, whose classification initially agreed on only 80% of four-second epochs. After consultation, they concurred on the remaining 20%. Fig. 4 provides an example of the RPSs for the four rhythm types.

The signals are segmented into 2 s intervals and zero meaned and unit normalized. The initial estimates of time lag and dimension are 11 and 9, respectively, using the method proposed in [12]. Figs. 5 and 6, respectively, show classification accuracy versus dimension with lag held constant and accuracy versus time lag with dimension held constant using a 16 mixture full covariance GMM and a Bayes classifier (GMM:Bayes). A 10-fold cross validation is performed for all experiments. Data segments for each fold are randomly selected with the constraint that the proportion of classes be the same within each. The experiments are thus not patient independent, in that a patient's data may appear in more than one fold. Classification accuracy is the total number of correctly classified signals divided by the total number of signals.

As can be seen in Figs. 5 and 6, the initial estimates of time lag and dimension are reasonable, but not optimal. Additionally, it can be seen that the GMM:Bayes method is relatively robust for this problem across a range of time lags from 11 to 15 and dimensions from 13 to 19. Using the best empirical time lag and dimension, the number of mixtures is varied from 1 to 64. The accuracy increases from 75.8% to 92.2% as the number of mixtures increases from 1 to 8. From 16 to 64 mixtures, the accuracy is relatively stable in the range from 94.5% to 95.5%. Table I shows the best results for the two methods, bin-based statistical model and Bayesian classifier (bin:Bayes) and GMM:Bayes, with the structure of each RPS and number of mixtures/bins. The sensitivity results for the GMM:Bayes classifier are 100.0% for SR, 95.2% for MVT, 82.7% for PVT, and 96.5% for VF.

Also seen in Table I are the results of two baseline techniques commonly used in automatic cardioverter defibrillators—a heart rate-based method and a gradient pdf-based method—and frequency-based approach. Details of the first two methods can

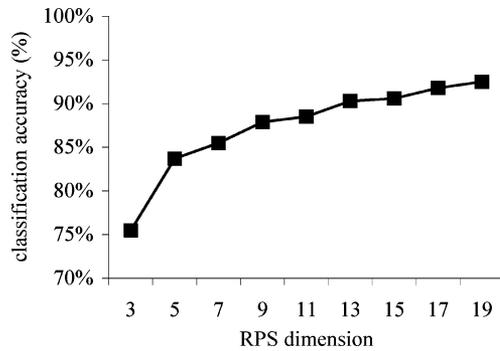


Fig. 5. ECG classification accuracy versus dimension with  $\tau = 11$  for a 16 mixture GMM and Bayes classifier.

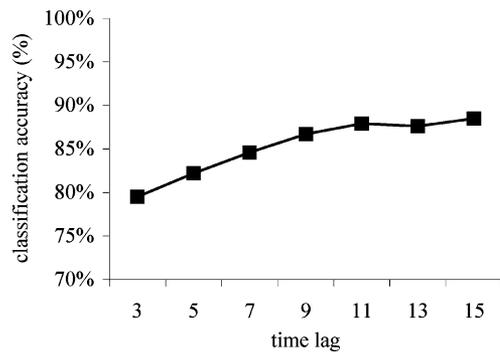


Fig. 6. ECG classification accuracy versus time lag with  $d = 9$  for a 16 mixture GMM and Bayes classifier.

TABLE I  
ECG ACCURACY RESULTS FOR EACH METHOD

Classifier	Accuracy	Dimension	Lag	Mixtures/Bins
Gradient PDF	71.0%			
Heart Rate	53.5%			
Frequency Based	43.8%			1
bin:Bayes	83.7%	3	15	1000
GMM:Bayes	95.5%	21	15	32

be found in [32] and [33]. In short, the heart rate method estimates the heart rate for each of the four rhythms and classifies according to heart rate bands. The gradient pdf approach builds models of the gradient of the ECG signal and classifies using a Bayesian approach. The frequency method uses the centroid frequency of the power spectrum as a feature [34]. The best results were obtained when using a single mixture to model this feature across each class. A maximum likelihood classifier is used on the test signals. The GMM:Bayes method outperforms all other tested approaches including the bin:Bayes. The difficulty in adjusting the granularity of the bin:Bayes method in comparison to the GMM:Bayes method is also seen. The GMM:Bayes is able to successfully model a 21-dimensional space. To apply the bin-based method to such a space with only one division per dimension would require over one million bins. We also see that the initial estimates for time lag and dimension are reasonable, but not optimal, estimates.

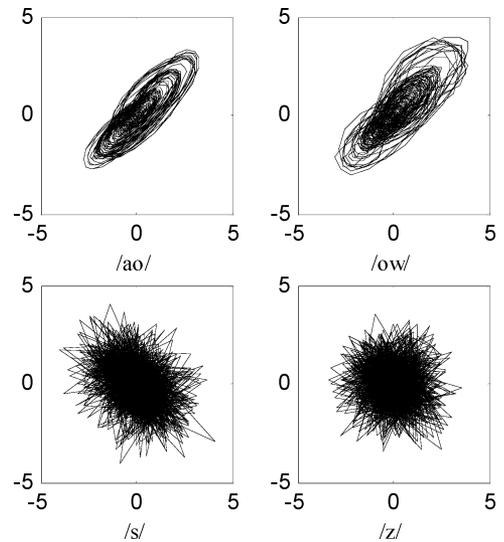


Fig. 7. 2-D RPS for examples of normalized phonemes: /ao/, /ow/, /s/, /z/ with  $\tau = 2$ .

## B. Speech Recognition

In this set of experiments, we apply our RPS-based classification methods to automatic speech recognition, specifically speaker dependent isolated phoneme classification [35]. The classification of isolated phonemes is directly related to the problem of continuous speech recognition, a necessary element of such important tasks as automated transcription and machine translation. We compare a traditional cepstral-based method to the RPS approaches. Examples of four speech phoneme 2-D RPS are shown in Fig. 7. The speech signals, which are sampled at 16 KHz, are taken from the TIMIT corpus [36]. The speech signals in the TIMIT corpus contain expertly labeled time-stamped phoneme boundaries, which can be used to extract the isolated phoneme data.

The illustrated phonemes are of two different types, vowels, /ao/ and /ow/, and fricatives, /s/ and /z/. The vowels have smooth locally correlated trajectories, whereas the fricatives contain random locally uncorrelated trajectories, as would be expected from the associated differences in speech production mechanisms.

We use a single speaker's data and implement several different RPS models. For this speaker, there are 417 total phoneme utterances belonging to 47 classes. One class of the standard 48 is not present in this data set. For each method, a model is learned for each of the 47 classes, yielding 47 models. These 47 classes are folded into 39 classes as is consistent with the literature [37]. A leave-one-out cross-validation testing is used for the MFCC approach and bin-based statistical model combined with an ANN classifier (bin:ANN), while a class balanced 10-fold cross validation is used for all other experiments.

The initial estimates of time lag and dimension are 2 and 12, respectively. In Fig. 8, we can see that these initial estimates are reasonable, but not optimal. We can also see that the method's accuracy is stable across the range of dimensions from 10 to 18 for the GMM:Bayes method. The best results for each approach are given in Table II.

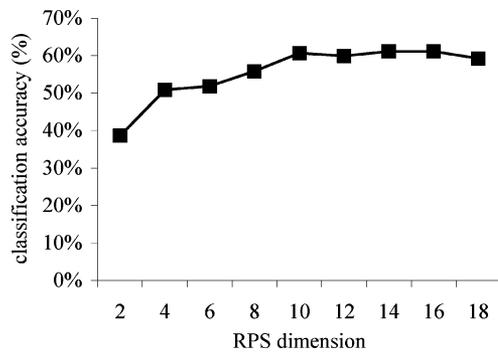


Fig. 8. GMM:Bayes phoneme classification accuracy versus dimension with  $\tau = 2$  for a 16 mixtures.

TABLE II  
SPEAKER DEPENDENT CLASSIFICATION RESULTS

Classifier	Accuracy	Dimension	Lag	Mixtures/Bins
MFCC	51.6%	12		8
bin:Bayes	24.0%	6	2	64
bin:ANN	32.4%	2	6	400
GMM:Bayes	58.0%	12	1	8
GMM:Bayes	62.6%	16	1	32

The baseline Mel-frequency cepstral coefficients (MFCC) approach uses eight mixtures to model the distribution of 12 MFCCs for each phoneme. This is compared to three RPS-based methods. The bin:Bayes method has the lowest accuracy of 24.0%. The bin:ANN with a tied-bin statistical model with 400 ( $20 \times 20$ ) bins in combination with an ANN classifier outperforms the bin:Bayes approach. The ANN uses bin counts as inputs to an ANN with a 400-10-3-1 (400 input neurons, 10 sigmoid neurons in the first hidden layer, three sigmoid neurons in the second hidden layer, and one linear output neuron) architecture. Two versions of the GMM:Bayes approach are shown. The first is directly comparable to the MFCC result. The second uses a larger dimension and number of mixtures to achieve an accuracy of 62.6%. The sensitivity results for this GMM:Bayes classifier are: vowels 61.8%, fricatives 71.9%, nasals 57.1%, semivowels 40.0%, stops 46.3%, and silence 82.5%.

The results indicate that RPS methods are capable of discriminating between phonemes. From the results of the last experiment, we can see that for this speaker dependent task, a GMM:Bayes approach outperforms the MFCC approach. In contrast to the ECG classification, where the best RPS approach outperforms the best traditional approach by 24.5%, only the GMM:Bayes approach outperforms the MFCC approach and by only 11%. These results give an indication of how well RPS approaches will perform on tasks that are well characterized by linear models such as speech production.

## VI. CONCLUSION

We have presented the use of RPS representations as a novel and theoretically well founded method for signal classification, and shown that statistical models of the RPSs are viable for capturing the information in such a space. The approach is applied to two signal classification tasks.

A key advantage of such phase space signal models is that this representation is capable of capturing the full dynamic structure of any finite-dimensional generating system. In the limit as the amount of data and corresponding model dimension increases, the attractor structure can fully describe complex behaviors of a system of arbitrarily large order. Another advantage of this approach is its ability to distinguish signals of very short duration, where the frequency resolution needed for accurate classification is unattainable. This is useful in applications such as heart arrhythmia classification, where such rapid classification enables previously inconceivable therapy options.

It is important to note the types of signal classification problems for which RPS-based methods will not work better than power spectral-based methods. Assuming the chosen modeling approach is of high enough order; power spectral-based methods will outperform RPS-based methods when the phase of the signal is not important for differentiating between classes, as would be the case for a linear system. This can be seen in the phoneme recognition task discussed above, as there is an ongoing debate as to how important phase is in speech recognition [38]. Additionally, RPS-based methods would underperform methods based on a single state variable when the state structure captured by the RPS fails to provide additional signal class differentiating information.

## REFERENCES

- [1] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, "Testing for nonlinearity in time series: The method of surrogate data," *Physica D*, vol. 58, pp. 77–94, 1992.
- [2] T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology," *J. Stat. Phys.*, vol. 65, pp. 579–616, 1991.
- [3] F. Takens, "Detecting strange attractors in turbulence," in *Proc. Dynam. Syst. Turbulence*, Warwick, 1980, pp. 366–381.
- [4] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Rev. ed. San Francisco, CA: Holden-Day, 1976.
- [5] B. Gold and N. Morgan, *Speech and Audio Signal Processing*. New York: Wiley, 2000.
- [6] J. G. Proakis, *Digital Communications*, 4th ed. Boston, MA: McGraw-Hill, 2001.
- [7] E. Ott, *Chaos in Dynamical Systems*. Cambridge, U.K.: Cambridge Univ. Press, 1993.
- [8] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D*, vol. 9, pp. 189–208, 1983.
- [9] H. Leung, "System identification using chaos with application to equalization of a chaotic modulation system," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 45, pp. 314–320, 1998.
- [10] R. J. Povinelli, F. M. Roberts, K. M. Ropella, and M. T. Johnson, "Are nonlinear ventricular arrhythmia characteristics lost, as signal duration decreases?," in *Proc. Comput. Cardiol.*, Memphis, TN, 2002, pp. 221–224.
- [11] D. M. Tumeay, P. E. Morton, D. F. Ingle, C. W. Downey, and J. H. Schnurer, "Neural network classification of EEG using chaotic preprocessing and phase space reconstruction," in *Proc. 1991 IEEE 17th Ann. Northeast Bioeng. Conf.*, 1991, pp. 51–52.
- [12] R. J. Povinelli, M. T. Johnson, A. C. Lindgren, and J. Ye, "Time series classification using Gaussian mixture models of reconstructed phase spaces," *IEEE Trans. Knowl. Data Eng.*, vol. 16, pp. 779–783, 2004.
- [13] D. Sciamarella and G. B. Mindlin, "Unveiling the topological structure of chaotic flows from data," *Phys. Rev. E*, vol. 64, pp. 036 209:1–7, 2001.
- [14] R. Brown, "Calculating Lyapunov exponents for short and/or noisy data sets," *Phys. Rev. E*, vol. 47, pp. 3962–3969, 1993.
- [15] A. Petry, D. Augusto, and C. Barone, "Speaker identification using nonlinear dynamical features," *Chaos Solitons Fractals*, vol. 13, pp. 221–231, 2002.
- [16] P. E. Rapp, T. A. A. Watanabe, P. Faure, and C. J. Cellucci, "Nonlinear signal classification," *Int. J. Bifurc. Chaos in Appl. Sci. Eng.*, vol. 12, pp. 1273–93, 2002.

- [17] N. Tishby, "A dynamical systems approach to speech processing," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 1990, pp. 365–368.
- [18] Y. Ashkenazy, P. C. Ivanov, S. Havlin, C.-K. Peng, A. L. Goldberger, and H. E. Stanley, "Magnitude and sign correlations in heartbeat fluctuations," *Phys. Rev. Lett.*, vol. 86, pp. 1900–3, 2001.
- [19] M. Banbrook, S. McLaughlin, and I. Mann, "Speech characterization and synthesis by nonlinear methods," *IEEE Trans. Speech Audio Process.*, vol. 7, pp. 1–17, 1999.
- [20] H. D. I. Abarbanel, T. A. Carroll, L. M. Pecora, J. J. Sidorowich, and L. S. Tsimring, "Predicting physical variables in time-delay embedding," *Phys. Rev. E*, vol. 49, pp. 1840–1853, 1994.
- [21] T. Sauer, "Time series prediction using delay coordinate embedding," in *Time Series Prediction: Forecasting the Future and Understanding the Past*, A. S. Weigend and N. A. Gershenfeld, Eds. Reading, MA: Addison-Wesley, 1994, pp. 175–194.
- [22] J. Kadtko, "Classification of highly noisy signals using global dynamical models," *Phys. Lett. A*, vol. 203, pp. 196–202, 1995.
- [23] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a time series," *Phys. Rev. Lett.*, vol. 45, pp. 712–716, 1980.
- [24] H. Whitney, "Differentiable manifolds," *Ann. Math.*, ser. 2nd, vol. 37, pp. 645–680, 1936.
- [25] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1997.
- [26] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data*. New York: Springer, 1996.
- [27] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D*, vol. 110, pp. 43–50, 1997.
- [28] T. Gautama, D. P. Mandic, and M. M. V. Hulle, "A differential entropy based method for determining the optimal embedding parameters of a signal," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Hong Kong, 2003, pp. 29–32.
- [29] J. B. Vitrano and R. J. Povinelli, "Selecting dimensions and delay values for time-delay embedding using a genetic algorithm," in *Proc. Genetic Evolut. Computat. Conf.*, San Francisco, CA, 2001, pp. 1423–1430.
- [30] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Statist. Soc.*, ser. B, vol. 39, pp. 1–38, 1977.
- [31] T. M. Mitchell, *Machine Learning*. New York: McGraw-Hill, 1997.
- [32] I. Singer, *Implantable Cardioverter-Defibrillator*. Amonk, NY: Futura, 1994.
- [33] C. Cabo and D. S. Rosenbaum, *Quantitative Cardiac Electrophysiology*. New York: Marcel Dekker, 2002.
- [34] I. Romero and L. Serrano, "ECG frequency domain features extraction: A new characteristic for arrhythmias classification," in *Proc. 23rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Istanbul, Turkey, 2001, pp. 2006–2008.
- [35] M. T. Johnson, R. J. Povinelli, A. C. Lindgren, J. Ye, X. Liu, and K. M. Indrebo, "Time-domain isolated phoneme classification using reconstructed phase spaces," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 4, pp. 458–466, Jul. 2005.
- [36] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. Dahlgren, and N. L. Dahlgren, *Darpa timit acoustic-phonetic continuous speech corpus*. [CD-ROM], 1993.
- [37] K.-F. Lee and H.-W. Hon, "Speaker-independent phone recognition using hidden Markov models," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 1641–1648, 1989.
- [38] L. D. Alsteris and K. K. Paliwal, "Importance of window shape for phase-only reconstruction of speech," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Montreal, QC, Canada, 2004, pp. I573–I576.



**Richard J. Povinelli** (S'85–M'97–SM'01) received the B.S. degree in electrical engineering and the B.A. degree in psychology from the University of Illinois, Champaign-Urbana, in 1987, the M.S. degree in computer and systems engineering from Rensselaer Polytechnic Institute, Troy, NY, in 1989, and the Ph.D. degree in electrical and computer engineering from Marquette University, Milwaukee, WI, in 1999.

From 1987 to 1990, he was a software engineer with General Electric (GE) Corporate Research and Development. From 1990 to 1994, he was with GE Medical Systems, where he served as a Program Manager and then as a Global Project Leader. From 1995 to 1998, he held the positions of Lecturer and

Adjunct Assistant Professor with the Department of Electrical and Computer Engineering, Marquette University, where, since 1999, he has been an Assistant Professor. His research interests include data mining of time series, chaos and dynamical systems, computational intelligence, and financial engineering.

Dr. Povinelli is a member of the Association for Computing Machinery, American Society of Engineering Education, Tau Beta Pi, Phi Beta Kappa, Sigma Xi, Eta Kappa Nu, Upsilon Pi Epsilon, and the Golden Key. He was voted Young Engineer of the Year for 2003 by the Engineers and Scientists of Milwaukee, Inc.



**Michael T. Johnson** (S'93–M'93–SM'02) received the B.S. degree in computer science engineering and the B.S. degree in engineering and electrical concentration, both from LeTourneau University, Long View, TX, in 1989 and 1990, respectively. He received the M.S.E.E. degree from the University of Texas, San Antonio, in 1994 and the Ph.D. degree from Purdue University, West Lafayette, IN, in 2000.

He was a design engineer from 1990 to 1991 for Micronyx, Inc. and Microtechnology Services, Inc., Dallas, TX, and from 1991 to 1993 for Datapoint

Corp., San Antonio. He was a Senior Engineer and Engineering Manager for SNC Manufacturing, Oshkosh, WI, from 1993 to 1996. Since 2000, he has been an Assistant Professor of Electrical and Computer Engineering with Marquette University, Milwaukee, WI. His primary research area is speech and signal processing, with an emphasis in speech recognition systems. Other areas of research include machine learning, statistical pattern recognition, and nonlinear signal processing.

Dr. Johnson is a registered professional engineer in the state of Wisconsin. He is a member of the Association for Computing Machinery, the Acoustical Society of America, and the International Speech Communications Association. His honor society memberships include Sigma Xi, Eta Kappa Nu, and Upsilon Pi Epsilon.



**Andrew C. Lindgren** (S'02) received the B.S. degree in physics in 2001 from DePaul University, Chicago, IL, and the M.S. degree in electrical and computer engineering in 2003 from Marquette University, Milwaukee, WI.

His past research experiences include working in the area of automatic speech recognition and enhancement. He has conducted research at Argonne National Laboratory in the area of nondestructive evaluation techniques for industrial ceramics (1999–2001), and also carried out research at DePaul

University (2000–2001) in the area of condensed matter physics that involved modeling magnetic reversal properties on ultrathin films. In 2003, he joined ATK Mission Research, and he has worked on space-time adaptive processing for phased-array radar systems, bistatic radar geolocation algorithms, automatic target recognition, and hyperspectral data exploitation. His current research interests include array signal processing, radar signal processing, automatic speech recognition and enhancement, and machine learning.



**Felice M. Roberts** (S'96) received the B.S. degree in electrical engineering and computer science and the M.S. degree in electrical and computer engineering from Marquette University, Milwaukee, WI, in 1987 and 2000, respectively. She is currently working towards the Ph.D. degree in electrical and computer engineering at Marquette University.

Her current research interests include signal processing and nonlinear analysis of biomedical signals, particularly ECGs. She worked for GE Medical Systems from 1986 to 1988 as an Applications Engineer in magnetic resonance imaging. She was a systems engineer with EDS, Detroit, MI, and Ruesselsheim, Germany, from 1988 to 1996, developing and maintaining CGS, a proprietary CAD/CAM/CAE system.



**Jinjin Ye** (S'02) received the B.S. degree in electronic science and engineering from Nanjing University, Nanjing, China, in 1999, and the M.S. degree in electrical and computer engineering from Marquette University, Milwaukee, WI, in 2004, respectively.

He was with the Institute of Human-Computer Interaction and Media Integration, Tsinghua University, from 1999 to 2001, and with the Speech and Signal Processing Lab, Marquette University, from 2002 to 2004. Currently, he is pursuing the Ph.D. degree in electrical engineering at the University of California, Los Angeles (UCLA). His current research interests include signal processing, speech analysis, and robust speech recognition and enhancement.