

Time Series Classification Using Gaussian Mixture Models of Reconstructed Phase Spaces

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

Abstract—A new signal classification approach is presented that is based upon modeling the dynamics of a system as they are captured in a reconstructed phase space. The modeling is done using full covariance Gaussian Mixture Models of time domain signatures, in contrast with current and previous work in signal classification that is typically focused on either linear systems analysis using frequency content or simple nonlinear machine learning models such as artificial neural networks. The proposed approach has strong theoretical foundations based on dynamical systems and topological theorems, resulting in a signal reconstruction, which is asymptotically guaranteed to be a complete representation of the underlying system, given properly chosen parameters. The algorithm automatically calculates these parameters to form appropriate reconstructed phase spaces, requiring only the number of mixtures, the signals, and their class labels as input. Three separate data sets are used for validation, including motor current simulations, electrocardiogram recordings, and speech waveforms. The results show that the proposed method is robust across these diverse domains, significantly outperforming the time delay neural network used as a baseline.

Index Terms—Signal classification, reconstructed phase spaces, Gaussian mixture models.

1 INTRODUCTION

MOST work in signal classification or identification is based on linear systems analysis, using features based on a frequency domain representation. There is also extensive work on signal detection and classification in the field of communications, based on statistical decision theory [1]. Alternatives to these established approaches include nonlinear classifiers such as neural networks or support vector machines, as well as clustering and similarity measurement techniques from the relatively new field of time-series data mining [2]. Many existing time-domain approaches to the task of signal classification are based on the existence of a fairly simple underlying pattern, or template, that is either known a priori or can be learned from the data. In the case of real signals with complex underlying systems, such as cardiac, speech, or electric motor systems, such a simple pattern rarely exists. Frequency-based techniques are based on the existence of spectral patterns, which from a random processes perspective capture only the first and second order characteristics of the system. Recently, studies in dynamical systems and chaos theory have led to new types of signal models based on reconstructed phase spaces (RPSs), and to new signal classification approaches, for example using dynamical invariants as features [3], [4], [5], which can capture information beyond that of a basic spectral representation. However, little work has yet been done in directly modeling

signals in the reconstructed phase space, which is the approach introduced here for classifying time series. A statistical learner is applied to the space, and the resulting maximum likelihood classifier is compared to a baseline time delay neural network (TDNN) approach. Experiments are conducted across three substantially different application areas: electric motor fault detection, heart arrhythmia classification, and speech phoneme recognition.

The importance of accurate signal classification methods can be seen in the breadth of application areas. For example, electric motor fault diagnosis is an important and widely studied industrial problem [6]. Although electric motors are generally reliable, there are currently no effective mechanisms for identifying a wide range of fault types and their corresponding severities, which is essential for detecting faults before they become catastrophic. Another application is the classification of heart arrhythmias. Electronic therapy, which requires the rapid and accurate classification of a heart rhythm, is the preferred method to terminate ventricular fibrillation (VF). There is evidence to suggest that the sooner electronic therapy is delivered following the onset of VF, the greater the success of terminating the arrhythmia and, thus, the greater the chance of survival [7]. The methods used in current practice typically require five or more seconds of data to classify an arrhythmia, so a signal classification method that could accurately classify heart arrhythmias in less than five seconds would be clinically beneficial. In the speech recognition domain, where the signal is the acoustic waveform corresponding to a basic sound unit called a phoneme, improvements in phoneme classification yield corresponding improvements in system recognition accuracy.

Background of the underlying dynamical system's theory and an overview of previous work in this area is given in Section 2. The proposed method is presented in Section 3, with a discussion of the data sets in Section 4. Section 5 reviews experimental results, and conclusions are presented in Section 6.

2 BACKGROUND

The theoretical basis for our new signal classification algorithm comes from the work of Takens [8] and Sauer et al. [9]. This work shows that a time series of observations sampled from a single state variable of a system can be used to reconstruct a space topologically equivalent to the original system. The construction of such an RPS or phase space embedding is straightforward. Given a time series $x = x_n, n = 1 \dots N$, an RPS matrix $\mathbf{X}$ of dimension d and time lag τ is defined by its row vectors:

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

where $n = (1 + (d-1)\tau) \dots N$. A row vector $\mathbf{x}_n$ is a point in the RPS.

The sufficient condition for topological equivalence is that d is greater than twice the box counting dimension of the original system [9]. When d is not known, as is the case for most real systems, it may be estimated using the false nearest-neighbor technique [10], which calculates the percentage of neighboring points which are near because of projection rather than dynamics. In Takens' original work, $\tau = 1$. However, in practice, it has been found that the appropriate selection of the time lag can reduce the required RPS dimension. A common heuristic for determining time lag is to use the first minimum of the automutual information function [10].

Because of the representational capability of RPSs, the proposed classification algorithm is theoretically capable of differentiating between signals generated by topologically different systems. It can differentiate between deterministic nonlinear signals with identical linear characteristics, but different nonlinearities. In

• The authors are with the Electrical and Computer Engineering Department, Marquette University, 1515 W. Wisconsin, WI 53233.
E-mail: {richard.povinelli, mike.johnson, andrew.lindgren, and jinjin.ye}@mu.edu.

Manuscript received 13 Feb. 2003; revised 21 July 2003; accepted 30 Sept. 2003.

For information on obtaining reprints of this article, please send e-mail to: tkde@computer.org, and reference IEEECS Log Number 118292.

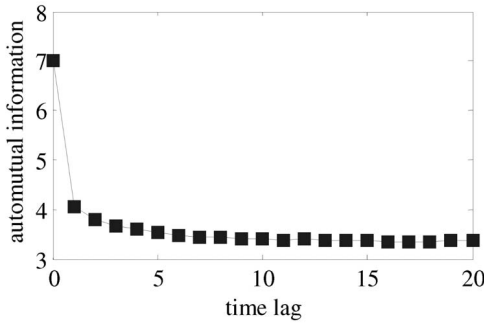


Fig. 1. Automutual information of an ECG signal.

addition, the proposed method, because of the use Gaussian Mixture Models (GMMs), is robust to additive noise components. This theoretical capability is demonstrated empirically across three complex real-world application domains: electric motor fault detection, heart arrhythmia classification, and speech phoneme recognition.

Dynamical systems techniques have been used in many applications. Lyapunov exponents have been used as features for classifying chaotic [3], acoustic [4], and speech [5] signals. Estimates of dimensions have been applied to such areas as speech production [11] and heart rate variability [12], [13]. Topological approaches have been used to analyze a variety of signals and systems, including speech signals [14], chaotic systems [15], voltage waveforms [16], and convection processes [17]. The most similar work to the approach proposed here is Kadtke's [18]. Whereas Kadtke's approach builds global vector reconstructions and differentiates signals in a coefficient space, our approach builds GMMs of signal trajectory densities in an RPS and differentiates between signals using a Bayesian classifier.

3 METHOD

As discussed above, our approach to signal classification is to build GMMs of signal trajectory densities in an RPS and differentiate between signals using a Bayesian classifier. This is done in three steps. The first step, data analysis, includes normalizing the signals and estimating the time lag and dimension of the RPS. The second step is learning the GMMs for each signal class. The final step is signal classification, which is done with a maximum likelihood Bayesian classifier.

3.1 Data Analysis

Each signal is normalized to zero mean and unit standard deviation. The time lag is calculated for each normalized signal using the first minimum of the automutual information function [10]. See Fig. 1 for an example automutual information plot with a first minimum at 11. An overall time lag is selected using the mode of the histogram of the first minima of the automutual information function for all signals. The RPS dimension for each signal is calculated using the global false nearest-neighbor technique [10]. See Fig. 2 for an example plot of the false nearest neighbors by dimension with an indicated dimension of six. Because we want most of the signals to unfold completely in the RPS, the overall RPS dimension is selected as the mean plus two standard deviations of the distribution of individual signal RPS dimensions.

3.2 Gaussian Mixture Models

The second step of the approach is to learn a GMM probability distribution for each signal class. This is done by creating an RPS using the time lag and dimension determined in the previous step

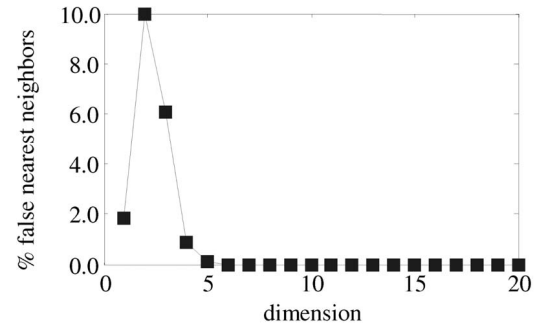


Fig. 2. Global false nearest neighbor of an ECG signal.

and inserting all the signals for a particular class into this space as described by (1) above.

A GMM is defined as:

$$p(\mathbf{x}) = \sum_{m=1}^M w_m p_m(\mathbf{x}) = \sum_{m=1}^M w_m \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m), \quad (2)$$

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 $\sum w_m = 1$. The required number of mixtures is related to the underlying distribution of the RPS density. The classification accuracy tends toward an asymptote as the number of mixtures increases provided there is sufficient training data. The parameters for the GMM are estimated using the well-known Expectation-Maximization (EM) algorithm [19]. This iterative method yields a Maximum Likelihood (ML) estimate, via the estimation formulas:

$$\begin{aligned} \boldsymbol{\mu}'_m &= \frac{\sum_{t=1}^T p_m(x_t) x_t}{\sum_{t=1}^T p_m(x_t)}, \\ \boldsymbol{\Sigma}'_m &= \frac{\sum_{t=1}^T p_m(x_t) (x_t - \boldsymbol{\mu}_m)^T (x_t - \boldsymbol{\mu}_m)}{\sum_{t=1}^T p_m(x_t)}, \\ w'_m &= \frac{\sum_{t=1}^T p_m(x_t)}{\sum_{t=1}^T \sum_{m=1}^M p_m(x_t)}. \end{aligned} \quad (3)$$

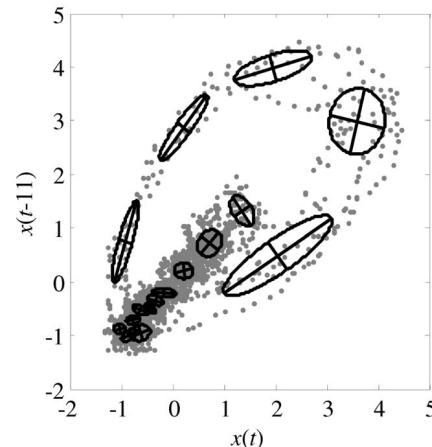


Fig. 3. Gaussian mixture model of an ECG reconstructed phase space.

TABLE 1
Algorithmic Description of Approach

learnModels (<i>signals</i> , <i>classes</i> , <i>M</i>)
<i>normalizedSignals</i> $\leftarrow$ normalize each <i>signal</i> to zero mean, unit variance
<i>timeLag</i> $\leftarrow$ determineTimeLag (<i>normalizedSignals</i>)
<i>dimension</i> $\leftarrow$ determineDimension (<i>normalizedSignals</i> , <i>timeLag</i>)
for each <i>class</i> in <i>classes</i>
form a reconstructed phase space
using EM learn GMM with <i>M</i> mixtures, see (3) above
return the <i>models</i> , <i>timeLag</i> , and <i>dimension</i>
classify (<i>signal</i> , <i>models</i> , <i>classes</i> , <i>timeLag</i> , <i>dimension</i>)
<i>normalizedSignal</i> $\leftarrow$ normalize <i>signal</i> to zero mean, unit variance
form a reconstructed phase space and insert <i>signal</i> , see (1) above
for each <i>class</i> in <i>classes</i>
calculate the likelihood for the corresponding <i>model</i> , see (4) below
determine the classification, see (5) below
return <i>class</i>
determineTimeLag (<i>signals</i>)
for each <i>signal</i> in <i>signals</i>
<i>aif</i> $\leftarrow$ calculate the automutual information function for <i>signal</i> , see [10] for details
<i>lag</i> $\leftarrow$ find first minimum of <i>aif</i>
<i>timeLag</i> $\leftarrow$ mode of <i>lags</i>
return <i>timeLag</i>
determineDimension (<i>signals</i> , <i>timeLag</i>)
for each <i>signal</i> in <i>signals</i>
<i>dim</i> $\leftarrow$ calculate the dimension using the false nearest neighbor technique for <i>signal</i> , see [10] for details
<i>mean</i> , <i>variance</i> $\leftarrow$ calculate the mean and variance of the <i>dims</i> distribution
<i>dimension</i> $\leftarrow$ <i>mean</i> + 2* <i>variance</i>
return <i>dimension</i>

An illustration of a GMM over an RPS is shown in Fig. 3, where the principle axes of the ellipses indicate one standard deviation of each mixture in the model.

3.3 Classification

The last step of the algorithm is to classify test signals. Signals to be classified are first normalized and then embedded in an RPS. Using the GMMs learned for each class of signals as described above, the signal is classified using a Bayesian maximum likelihood classifier.

This 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 models:

$$p(\mathbf{X}|c_i) = \prod_{n=1+(d-1)\tau}^N p(\mathbf{x}_n|c_i), \quad (4)$$

where $\mathbf{X}$ is an RPS matrix of dimension d and time lag τ of the signal, $\mathbf{x}_n$ is a point in the RPS, and $p(\mathbf{x}_n|c_i)$ is the probability of $\mathbf{x}_n$ given the i th class, calculated using (2). The classification is

$$\hat{c} = \arg \max_i p(\mathbf{X}|c_i), \quad (5)$$

where $\hat{c}$ is the maximum likelihood class.

3.4 Algorithms

An algorithmic description of our approach is provided in Table 1. The **learnModels** function generates a GMM for each signal class given a set of labeled signals and the number of mixtures to use for the GMM. The **classify** function classifies a signal using the GMMs

generated by **learnModels**. Two secondary functions are also described. The **determineTimeLag** and **determineDimension** functions calculate the time lag and dimension for the RPS, respectively. A MATLAB implementation of this algorithm is available from <http://povinelli.eece.mu.edu/itr-speech/>.

4 TIME SERIES DATA SETS

We apply our technique to three data sets one simulated and two real. The first data set is generated from a sophisticated simulation of electric motor current signals. The second data set is a collection of electrocardiogram (ECG) signals. The third data set is from the TIMIT speech corpus [20].

4.1 Simulated Motor Current

The first data set consists of simulated electric motor current signals. Because the field collection of electric motor fault data for a wide range faults is very labor intensive and time consuming, we use advanced motor simulations. The signals are current waveforms generated using the time stepping coupled finite element state space (TSCFE-SS) method [21]. The TSCFE-SS approach is a coupling of a finite element model of the magnetic circuits with a circuit network model. The simulations are of the motor dynamics and include the nonlinear effects of magnetic saturation. In this case, the A phase current was generated for a 208-volt, 60-Hz, 2-pole, 1.2-hp, squirrel cage 3-phase induction motor using a finite element grid with 2,295 nodes and a 37th order state model.

Twenty-one different motor operating conditions, including 1-10 broken bars, 1-10 broken end-ring connectors, and a healthy operating mode, are simulated. For each motor operating condition, 20 time series, each with a length of 1,500 points and a

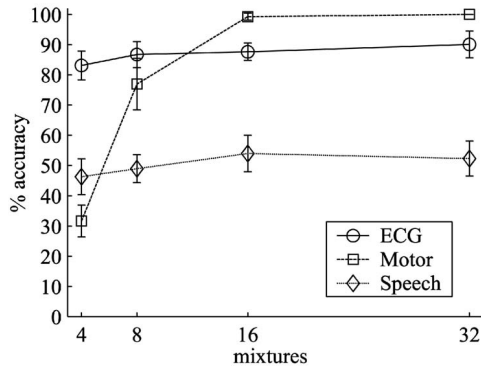


Fig. 4. Classification accuracy of new method with one standard deviation error bars.

sampling rate of 33.3kHz, are generated. This yields a data set of 420 signals divided into 21 classes. The motor data set has been donated to the UCR Time Series Data Mining Archive.

4.2 Electrocardiogram

This second data set was obtained from six patients during intercardiac defibrillator implantation. Data was collected from lead V1 of a 12 lead ECG. The signals were antialias filtered with a cutoff frequency of 200 Hz and, subsequently, digitized at 1,200 Hz. Because the data was collected during surgery and the chest was open, the lead placements were not ideal. Four rhythms were observed: normal sinus rhythm (SR) and three arrhythmias—monomorphic ventricular tachycardia (MVT), polymorphic ventricular tachycardia (PVT), and ventricular fibrillation (VF). Data was labeled by two experts, who initially agreed on 80 percent of the beat-by-beat classifications. After consultation, they concurred on the remaining 20 percent. The data set was divided into two-second (2400 point) segments, yielding 153 SR, 63 MVT, 58 PVT, and 57 VF time series.

4.3 TIMIT Speech Corpus

The last data set is 417 phonemes from the TIMIT speech corpus speaker MJDE0 [20]. The speech signals were sampled at 16KHz. The phoneme signals are of lengths varying from 227 to 5,201 samples, with phoneme boundaries and class labels determined by a group of experts. Of the 417 phonemes, six were spoken only once and one of the standard 48 classes did not occur in this data set, hence, there are 47 classes.

5 EXPERIMENTS AND RESULTS

Our new approach described above in Table 1 is applied to the three data sets also described above. The new method is compared to a time delay neural network (TDNN) [22], which is used as a nonlinear one step predictor. The TDNN classifications are made using minimum prediction error. A 10-fold cross-validation approach is used to compare the methods. The folds are formed in a statistically balanced manner across classes. The same folds are used to train both learners and test both classifiers.

The number of inputs to the TDNN is the dimension of the RPS. Thus, both methods are working with the same number of inputs. The TDNN has two hidden layers. Given d inputs to the TDNN, there are d neurons in the first hidden layer, $\lceil \sqrt{d} \rceil$ neurons in the second hidden layer, and one output neuron. The hidden layers use tansig transfer functions, and a linear transfer function is used for the output.

The signals are normalized to zero mean and unit variance for both methods. The TDNN is trained for 25 epochs. Similarly,

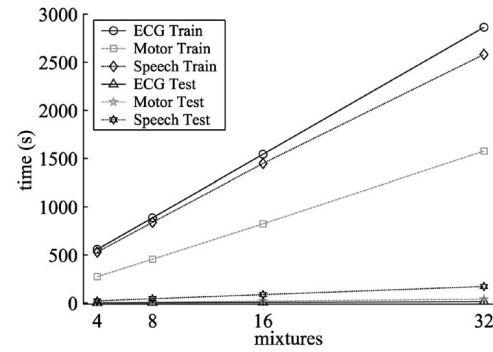


Fig. 5. Computational performance of the new method.

25 iterations of the EM algorithm are used in our new method. The classification accuracy across data sets and number of mixtures is shown in Fig. 4. The expected asymptotic accuracy curves are seen for all data sets with the exception of the 32-mixture speech result, which shows a decrease in mean accuracy. This is most likely due to insufficient training data.

The TDNN classification accuracies including one standard deviation results for the ECG, motor, and speech data sets are $50.8 \pm 8.2\%$, $6.0 \pm 1.7\%$, and $31.9 \pm 7.6\%$, respectively. The new method's accuracy results are statistically greater than the TDNN accuracy results at an $\alpha = 0.001$ level across all data sets and all number of mixtures. The greatest difference between the two approaches is seen in the motor data set. This is most likely due to the sinusoidal nature of the signal. The differences between motor current waveforms are seen across many ac cycles through an envelope that modulates the sinusoid. It appears that the TDNN, which performed barely above chance (4.8 percent), has captured the sinusoidal nature of the signal, which is common across all classes, but not the envelope, which varies across classes.

The mean computational performance of the new method is given in Fig. 5 for both training and testing across all classes and number of mixtures. The experimental platform was a dual 2.0GHz Pentium 4 processor Windows 2000 machine with 1GB of main memory. However, the algorithm runs on only one of the processors. The computational cost is linear in the number of mixtures for both training and testing. The computational cost is comparable to that of the TDNN method as is seen in Table 2. The I/O load is not a significant component of the computational performance.

6 CONCLUSIONS

The results from these three complex and real-world data sets show that this new approach can be successfully applied to a variety of signal classification problems. The new approach, with minimal input tuning (only the number of mixtures), can capture the dynamics of distinctly different categories of signals (motor current, heart ECG, and speech). This is in contrast to the TDNN

TABLE 2
Computational Results for 16 Mixtures and TDNN
(s, mean $\pm$ one standard deviation)

Dataset	New Method		TDNN	
	Train	Test	Train	Test
ECG	1544 $\pm$ 2	8.1 $\pm$ 0.1	776 $\pm$ 12	4.8 $\pm$ 0.4
Motor	823 $\pm$ 35	20.4 $\pm$ 1.1	388 $\pm$ 15	16.5 $\pm$ 0.4
Speech	1448 $\pm$ 45	88.5 $\pm$ 9.9	739 $\pm$ 19	48.2 $\pm$ 2.5

approach which performs significantly worse across all data sets. The advantage of our new approach is that it is able to capture an equivalent to the dynamics of the original system. This advantage is translated into higher classification accuracies in comparison to a TDNN approach across the three data sets tested in this work. Future work will investigate capturing the trajectory of the attractor in addition to its density.

ACKNOWLEDGMENTS

This material is based upon work supported by the US National Science Foundation under Grant No. IIS-0113508 and by the US Department of Education GAANN Fellowship. The authors would like to thank Felice Roberts and Xiaolin Liu for their comments and input during the development of this algorithm.

REFERENCES

- [1] J.G. Proakis, *Digital Comm.*, fourth ed. Boston: McGraw-Hill, 2001.
- [2] E. Keogh and S. Kasetty, "On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration," *Proc. Eighth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining*, 2002.
- [3] J.R. Buchler, Z. Kollath, T. Serre, and J. Mattei, "Nonlinear Analysis of the Lightcurve of the Variable Star R Scuti," *Astrophysical J.*, pp. 462-489, 1996.
- [4] Q. Ding, Z. Zhuang, L. Zhu, and Q. Zhang, "Application of the Chaos, Fractal and Wavelet Theories to the Feature Extraction of Passive Acoustic Signal," *Acta Acustica*, vol. 24, pp. 197-203, 1999.
- [5] A. Petry, D. Augusto, and C. Barone, "Speaker Identification Using Nonlinear Dynamical Features," *Chaos, Solitons, and Fractals*, vol. 13, pp. 221-231, 2002.
- [6] M.E.H. Benbouzid, "Bibliography on Induction Motors Faults Detection and Diagnosis," *IEEE Trans. Energy Conversion*, vol. 14, pp. 1065-1074, 1999.
- [7] E. Manios, G. Fenelon, T. Malacky, A.L. Fo, and P. Brugada, "Life Threatening Ventricular Arrhythmias in Patients with Minimal or No Structural Heart Disease: Experience with the Implantable Cardioverter Defibrillator," available at <http://www.heartweb.org/heartweb/0197/icd0003.htm>, 1997, cited Dec. 2000.
- [8] F. Takens, "Detecting Strange Attractors in Turbulence," *Proc. Dynamical Systems and Turbulence*, pp. 366-381, 1980.
- [9] T. Sauer, J.A. Yorke, and M. Casdagli, "Embedology," *J. Statistical Physics*, vol. 65, pp. 579-616, 1991.
- [10] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*. Cambridge: Cambridge Univ. Press, 1997.
- [11] H.F.V. Boshoff and M. Grotelap, "The Fractal Dimension of Fricative Speech Sounds," *Proc. South African Symp. Comm. and Signal Processing*, pp. 12-61, 1991.
- [12] Y. Ashkenazy, P.C. Ivanov, S. Havlin, C.-K. Peng, A.L. Goldberger, and H.E. Stanley, "Magnitude and Sign Correlations in Heartbeat Fluctuations," *Physical Rev. Letters*, vol. 86, pp. 1900-1903, 2001.
- [13] V. Schulte-Frohlinde, Y. Ashkenazy, P.C. Ivanov, L. Glass, A.L. Goldberger, and H.E. Stanley, "Noise Effects on the Complex Patterns of Abnormal Heartbeats," *Physical Rev. Letters*, vol. 87, pp. 068104/1-4, 2001.
- [14] D. Sciamarella and G.B. Mindlin, "Topological Structure of Chaotic Flows from Human Speech Chaotic Data," *Physical Rev. Letters*, vol. 82, p. 1450, 1999.
- [15] C. Casdagli, "Nonlinear Prediction of Chaotic Time Series," *Physica D*, vol. 35, pp. 335-356, 1989.
- [16] H.D.I. Abarbanel, T.A. Carroll, L.M. Pecora, J.J. Sidorowich, and L.S. Tsimring, "Predicting Physical Variables in Time-Delay Embedding," *Physical Rev. E*, vol. 49, pp. 1840-1853, 1994.
- [17] J.D. Farmer and J.J. Sidorowich, "Exploiting Chaos to Predict the Future and Reduce Noise," *Evolution, Learning, and Cognition*, Y.C. Lee, ed., pp. 277-330, World Scientific, 1988.
- [18] J. Kadtke, "Classification of Highly Noisy Signals Using Global Dynamical Models," *Physics Letters A*, vol. 203, pp. 196-202, 1995.
- [19] T.K. Moon, "The Expectation-Maximization Algorithm," *IEEE Signal Processing Magazine*, pp. 47-59, 1996.
- [20] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren, and V. Zue, "TIMIT Acoustic-Phonetic Continuous Speech Corpus," *Linguistic Data Consortium*, 1993.
- [21] N.A. Demerdash and J.F. Bangura, "A Time-Stepping Coupled Finite Element-State Space Modeling for Analysis and Performance Quality Assessment of Induction Motors in Adjustable Speed Drives Applications," *Proc. Naval Symp. Electric Machines*, pp. 235-242, 1997.
- [22] C.T. Lin and C.S.G. Lee, *Neural Fuzzy Systems—a Neuro-Fuzzy Synergism to Intelligent Systems*. Upper Saddle River, N.J.: Prentice-Hall, 1996.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.