# Nonlinear adaptive wavelet analysis of electrocardiogram signals

H. Yang,[1] S. T. Bukkapatnam,[1] and R. Komanduri[2,*]
[1]*Industrial Engineering and Management, 322 Engineering North, Oklahoma State University, Stillwater, Oklahoma 74078, USA*
[2]*Mechanical & Aerospace Engineering, 218 Engineering North, Oklahoma State University, Stillwater, Oklahoma 74078, USA*
(Received 20 March 2007; published 22 August 2007)

Wavelet representation can provide an effective time-frequency analysis for nonstationary signals, such as the electrocardiogram (EKG) signals, which contain both steady and transient parts. In recent years, wavelet representation has been emerging as a powerful time-frequency tool for the analysis and measurement of EKG signals. The EKG signals contain recurring, near-periodic patterns of *P*, *QRS*, *T*, and *U* waveforms, each of which can have multiple manifestations. Identification and extraction of a compact set of features from these patterns is critical for effective detection and diagnosis of various disorders. This paper presents an approach to extract a fiducial pattern of EKG based on the consideration of the underlying nonlinear dynamics. The pattern, in a nutshell, is a combination of eigenfunctions of the ensembles created from a Poincare section of EKG dynamics. The adaptation of wavelet functions to the fiducial pattern thus extracted yields two orders of magnitude (some 95%) more compact representation (measured in terms of Shannon signal entropy). Such a compact representation can facilitate in the extraction of features that are less sensitive to extraneous noise and other variations. The adaptive wavelet can also lead to more efficient algorithms for beat detection and QRS cancellation as well as for the extraction of multiple classical EKG signal events, such as widths of *QRS* complexes and *QT* intervals.

DOI: 10.1103/PhysRevE.76.026214 PACS number(s): 05.45.Tp, 02.50.−r, 07.05.Kf, 42.30.Sy

## I. INTRODUCTION

An electrocardiogram (EKG) is a time-varying signal that captures the ionic current flow responsible for the contraction and subsequent relaxation of the cardiac fibers. It is a typical near periodic but nonstationary signal (see Fig. 1). The surface EKG is obtained by recording the potential difference between two electrodes placed on the surface of the skin. A single normal cycle of the EKG represents the successive atrial and ventricular depolarization-repolarization events which occur with every heart beat and includes *P*, *Q*, *R*, *S*, *T*, and *U* waves (see Fig. 1):

(i) *P* wave: the sequential activation of the right and left atria.

(ii) *QRS* complex: right and left ventricular depolarization.

(iii) *T* wave: ventricular repolarization.

(iv) *U* wave: origin though not clear, is probably "postdepolarization" in the ventricles.

Patterns from the surface electrocardiogram (EKG) signals are widely used in the diagnosis of various cardiac disorders [1–3]. Similar to the pattern recognition techniques used successfully in speech, fingerprinting, etc. [4,5], EKG signal pattern matching may also provide the opportunity to recognize the same person (biometric identification) and diagnose heart diseases by means of the similarity comparison with the evaluated patterns of a typical patients' EKG in the collected databases. This paper presents a technique for customizing the wavelet functions adapted to the EKG signal pattern through the use of principles of nonlinear dynamics. In specific, we show that using tools such as Poincare section, one can achieve extremely compact signal representation (two orders of magnitude reduction in signal entropy) that can be more sensitive to different EKG signal patterns. The remainder of this paper is organized as follows: Sec. II gives a brief background of the applications of wavelet transform in surface EKG analysis; Sec. III is the research methodology used; Sec. IV presents the implementation of designing a customized EKG wavelet function and its comparison with standard wavelet library, and Sec. V presents the conclusions from the reported research and perspectives on future investigations.

## II. WAVELET SIGNAL REPRESENTATION OF EKG SIGNALS

EKG signal traces are gathered and stored in analytical instruments (e.g., EKG machines) in the form of time series.

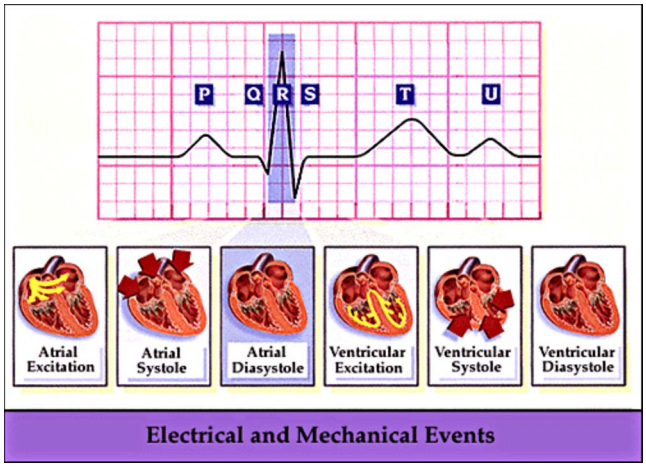


FIG. 1. (Color online) Electrical and mechanical events diagram during one heart beat [28].

*Corresponding author. ranga@ceat.okstate.edu

 

They can be transformed from the time domain into frequency, time-frequency (e.g., wavelets), or other domains depending on the nature of the information required. These signals exhibit recurrent patterns spread over multiple scales, and significant nonstationarity. Wavelet methods are most effective in capturing spatio-temporal content of such signals over multiple scales of resolution.

Wavelet functions $\psi(t)$ are building blocks that can be used to simultaneously decompose signal characteristics in both time and frequency domains. Wavelet representations are particularly useful for the analysis of transients, aperiodicity, and other nonstationary signal features. Subtle changes in signal morphology can be highlighted over the scales of interest through the interrogation of the transform [6]. Consequently, the *QRS* complex can be distinguished from high *P* and *T* waves, noise, baseline drift, and other artifacts in different scales. The relation between the characteristic points of EKG signals and those of modulus maximum pairs of its wavelet transforms are almost straightforward to establish [7].

Over the past few years, wavelets have been used in the analysis of physiological signals, such as EKG, Electroencephalogram (EEG), Electromyogram (EMG), blood pressure, and respiration signals [6]. Researchers have been exploring the applicability of wavelets for capturing complex nonlinear dynamics of EKG [8–10]. However, not much attention was given for the detection of certain complex patterns inside the EKG signals. Conceivably, much of the information necessary for early diagnosis of various ailments are buried in these complex patterns. Optimal matched wavelet for EKG signals can be extremely useful at finding occurrences of certain complex recurring patterns. Through effective use of recurrence analysis principles from nonlinear dynamic system theory and customized wavelets, it is possible to compact a multiscale representation by over two orders of magnitude (measured in terms of signal entropies) compared to a multiscale representation using any standard wavelet basis.

## III. RESEARCH METHODOLOGY

There are several existing families of standard wavelets, such as Haar, Daubechies, Coiflet, and so on [11]. The choice of wavelet basis functions plays a significant role in determining the compactness of the resulting wavelet representation. There is no consistent answer to the question: Which is the best wavelet? Some wavelets are better suited for detecting some particular problems, such as discontinuities and breakdown (e.g., Haar wavelet to detect discontinuities), while others are superior for long term estimation or compression [e.g., a sufficiently regular wavelet with $k$ ($k \geqslant 3$) vanishing moments will be better to compactly represent a smooth signal].

Nonetheless, it is generally understood that the closer the basis functions match the signal patterns, the more compact the representation will be. As shown in Fig. 2, this investigation presents a customized wavelet function design using least square fitting for the fiducial signal pattern extracted from the nonlinear system dynamic characterization of signal

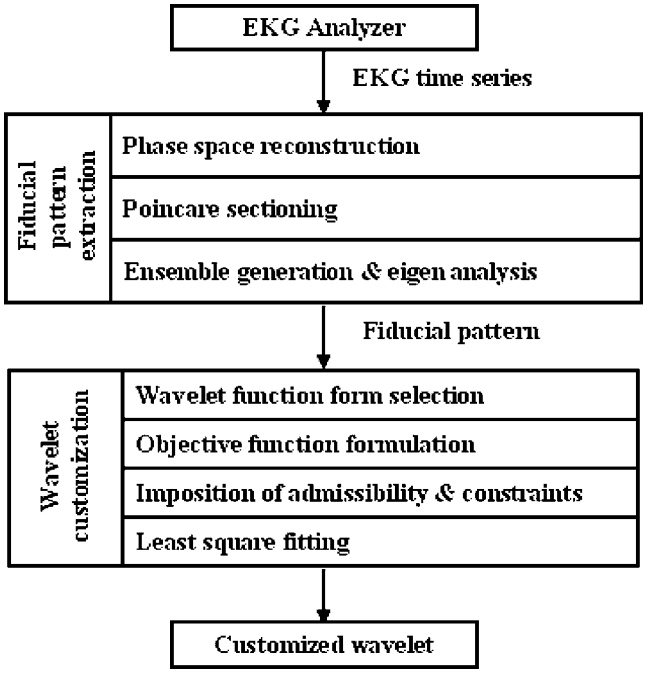


FIG. 2. Procedure for customized wavelet design from system nonlinear dynamics.

state space. In the first step, system phase space trajectory will be reconstructed from the observed time series. False nearest-neighbor test and mutual information test provide the necessary parameters, namely, embedding dimension $d_E$ and time delay $d_\tau$ [12]. After the reconstruction of phase space, those ensembles can be extracted from Poincare sectioning of the trajectory [13]. The Karhunen-Loeve (KL) transformation of those extracted signal ensembles provides the fiducial signal pattern for the least square customized wavelet design stage. We use polynomial wavelet structures and a constrained least square fit procedure [14]. To fit the structure to match as closely as possible, subject to admissibility and regularity constraints. The details of the two major steps of fiducial pattern extraction and wavelet fitting are presented in the following two subsections. Finally, representation entropy is calculated to compare the performance of various wavelets. Signal entropy $h$ is a measure of parsimony of representation [15,16]. Here, the normalized entropy with EKG signal energy is used as follows:

$$h = -\int\int_{\tau,s} p(\tau,s)\log_{10} p(\tau,s)d\tau ds, \tag{1a}$$

where

$$p(\tau,s) = \frac{|\Psi_x^\psi(\tau,s)|^2}{\|x(t)\|^2}, \tag{1b}$$

$\Psi_x^\psi(\tau,s)$ is a continuous wavelet transform coefficient at scale $s$ and translation $\tau$, and $\|x(t)\|^2$ is the signal energy. The smaller the value of signal entropy, the greater is the parsimony of the representation [17,18]. Intuitively, "high entropy" means the representation coefficients $\Psi_x^\psi(\tau,s)$ is from

a uniform (boring) distribution, i.e., the histogram of distribution of values of coefficients would be flat, and "low entropy" means that coefficients $\Psi_x^{\psi}(\tau,s)$ is from varied distribution (consisting of multiple distinct peaks and large valleys). The varied distribution of representation provides the accurate detection of signal pattern.

### A. Phase space fiducial pattern extraction

Natural systems generally show nonstationary and complex behaviors. The Karhunen-Loeve (KL) transform provides an optimal representation for second-order stochastic processes. The KL representation of a stochastic process $x(t)$ is given by

$$x(t) = \sum_j b_j \alpha_j(t). \tag{2}$$

Here, $b_j$'s are KL representation coefficients, and the basis function $\alpha j(t)$ are the linearly independent solutions of

$$\int_{\mathbf{R}} K(t,\tau)\alpha_j(\tau)d\tau = \vartheta_j \alpha_j(t), \tag{3}$$

where $K(t,\tau)$ is the autocovariance function calculated from a set of signal ensembles. It is evident that $\vartheta_j$ are eigenvalues, and $\alpha_j(t)$ are eigenfunctions of $K(t,\tau)$, and are therefore orthogonal. The order of eigenvalues, highest to lowest, indicates the components in order of significance. The KL basis can be approximated using a set of ensembles of the process $x(t)$ such that the mean square error (MSE) between a given set of ensembles and their projection to the subspaces spanned by each basis function of the KL representation is minimized.

However, in the absence of multiple realizations of $x(t)$, underlying nonlinear dynamics should be taken into account in order to extract ensembles needed to develop optimal representations (i.e., basis set) of signals from these systems. The nonlinear and stochastic dynamic characterization of a system is usually helpful in providing information on the dimensionality and the functional form of models that can capture the observed behaviors [12].

The time evolution of the phase space trajectories emanated from the system explains the underlying nonlinear dynamics. Usually, the measured observations of a process cannot include all possible state variables. Couplings among the system's components imply that each single component contains necessary information about the dynamics of the larger system. The embedding theorem of Takens [19] guarantees that the reconstructed trajectory portrays the dynamics in the higher dimensional state space. It states that a diffeomorphism exists between the reconstructed phase space with the state vector given by

$$\underline{x}(t_n) = [x(t_n), x(t_{n+d_\tau}), x(t_{n+2d_\tau}), \ldots, x(t_{n+(d_E-1)d_\tau})], \tag{4}$$

and original phase space, if $d_E \geqslant 2D+1$, where $d_E$ is the embedding dimension, $d_\tau$ is the time delay, and $D$ is the dimension of the compact manifold containing the attractor. This implies that the dimension and entropy spectra of the reconstructed attractor are the same as those of the original one.

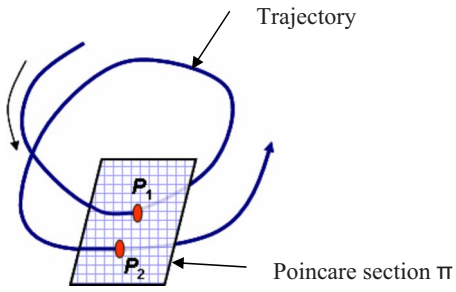


FIG. 3. (Color online) An illustration of trajectories of an attractor intersecting a (planar) Poincare section.

Poincare section is a $d_E$-1 dimensional hyperplane intersecting with the phase space trajectories (see Fig. 3). The recurrence property of a chaotic attractor **A** shows that for every $\varepsilon < 0$ and almost every $\underline{x}(0) \in \mathbf{A}, \exists t > 0$ such that $\|\underline{x}(0) - \underline{x}(t)\| < \varepsilon$, in effect, the trajectories with an attractor remain bounded. Those points $P_i$, $i=1,2,\ldots$ at which the trajectory intersects the Poincare section follow a return map. For strictly periodic trajectory, the points $P_i$, $i=1,2,\ldots$ will overlap (i.e., $\varepsilon \equiv 0$) such that the duration between $P_i$ to $P_{i+1}$ along the trajectory constitutes the period. For chaotic systems no two $P_i$'s overlap. For near-periodic signals, such as EKG, each strand emanating from a Poincare section intersection point $P_i$ and lasting approximately until the next intersection $P_{i+1}$ along the trajectory may be treated as a realization of a stochastic process from an invariant probability space [18]. Due to heart rate variability, some ensembles move faster, i.e., the two successive intersections occur over shorter intervals, compared to the others [20,21]. In our investigations the length (time duration) of the ensembles is taken as the time interval between the intersections of the fastest ensemble. Moreover, those collected near-periodic ensembles provide an effective way for the Karhunen-Loeve (KL) representation of this signal. In this EKG investigation, the largest eigenvalue $\vartheta_{\max}$ is 4.3026, which implies that 98.34% of the total variation occurs along the leading (the first one out of 41) eigenfunctions. Intuitively, each eigenfunction $\alpha(\cdot)$ captures the shape of a specific mode of variation (roughly, a degree of freedom) of $\underline{x}(\cdot)$. The fiducial pattern $\zeta(t)$ of a signal emerging from a process with $d$ degrees of freedom (or topological dimension $D$, where $d=\lceil D \rceil$) is obtained as the optimal projection of the ensembles onto the space spanned by $\alpha_1(\cdot), \alpha_2(\cdot), \cdots, \alpha_d(\cdot)$. For computational convenience (during matching wavelet design), the support of $\zeta(t)$ is rescaled so that $t \in [0,1]$.

### B. Least squares matching wavelet design

The continuous wavelet transform (CWT) $\Psi_x^{\psi}$ of the signal $x(t)$ using the analysis wavelet $\psi(t)$ is

$$\Psi_x^{\psi}(\tau,s) = \frac{1}{\sqrt{|s|}} \int_t x(t) \tilde{\psi}\left(\frac{t-\tau}{s}\right) dt, \tag{5}$$

where $\tilde{\psi}$ is the dual of $\psi(t)$ [22]. If $x(t)$ is similar to the wavelet basis functions, then the coefficients $\Psi_x^{\psi}$ will likely

be large only for a few basis functions. Thus, the customized wavelet function adapted to the EKG pattern may achieve the best performance. We design the wavelet function using a least squares approach [11,23,24]. A wavelet basis function $\psi(t)$ is approximated as a polynomial regression of degree $M$,

$$\psi(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_M t^M = \sum_{m=0}^{M} a_m t^M = U(t)\theta^T,$$

$$U(t) = [1 \; t \; t^2 \; \cdots \; t^M],$$

$$\theta = [a_0 \; a_1 \; a_2 \; \cdots \; a_M]. \tag{6}$$

Let us assume that $N$-sample time series of fiducial EKG pattern $\zeta(t)$ be available such that

$$Z_N = \{\zeta(t_n),\ 0 \leqslant t_n \leqslant 1\}. \tag{7}$$

Similarly, we can reduce operator $U(t)$ to a Vandermonde matrix $A$, whose elements are powers of $t$, will be as following:

$$A = \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^M \\ 1 & t_2 & t_2^2 & \cdots & t_2^M \\ \vdots & & \vdots & & \vdots \\ 1 & t_N & t_N^2 & \cdots & t_N^M \end{bmatrix} \quad 0 \leqslant t_1, \ldots, t_N \leqslant 1. \tag{8}$$

It is important that the function $\psi(t)$ which cannot only best fit the fiducial signal pattern $\zeta(t)$ but also satisfy the wavelet admissibility and regularity conditions [11]. The admissibility requirements for any valid real or complex-value continuous-time function $\psi(t)$ to be a wavelet basis function can be summarized as reconstruction, zero mean, finite energy, and regularity constraints. The first three defines the *wave*, and the last condition determines the rate of decay or the *let*. A function satisfied with all the four conditions can be a valid *wavelet* for CWT [11,23,24].

Since the fiducial pattern $\zeta(t)$ is a finite length signal, the finite energy requirement is automatically met. The zero mean condition $\int_{-\infty}^{\infty}\psi(t)dt=0$ implies that the Fourier transform of $\psi(t)$ vanishes at the zero frequency $|F_\psi(\omega)|^2|_{\omega=0}=0$, where $F_\psi(\omega)$ stands for the Fourier transform of $\psi(t)$. So, it will also make sure that $C_\psi=\int_{-\infty}^{\infty}\frac{|F_\psi(\omega)|^2}{|\omega|}d\omega$ is finite for the success of inverse continuous wavelet transform,

$$x(t) = \frac{1}{C_\psi}\int_s\int_\tau \Psi_x^\psi(\tau,s)\frac{1}{s^2}\psi\left(\frac{t-\tau}{s}\right)d\tau ds,$$

$$C_\psi = \int_{-\infty}^{\infty}\frac{|F_\psi(\omega)|^2}{|\omega|}d\omega. \tag{9}$$

Summarizing, the least square $M$th degree polynomial fitting with regularity $K$ for the wavelet design will add zero mean and regularity conditions, and the deduction of revised matrix $\tilde{A}$ and $Z_N$ is as following.

(1) Zero mean:

$$\int_0^1 \psi(t)dt = \left(a_0 t + \frac{a_1 t^2}{2} + \cdots + \frac{a_M t^{M+1}}{M+1}\right)\Bigg|_0^1 = 0. \tag{10}$$

(2) Regularity: The $i$th moment of the function $\psi(t)$ is defined as $\int_{-\infty}^{+\infty}t^i\psi(t)dt$. If the function's first $i$ moments are zero $\int_{-\infty}^{+\infty}t^i\psi(t)dt=0$ for $0\leqslant j\leqslant k$, the number of vanishing moment of the function $\psi(t)$ is $k+1$,

$$\int_0^t t^i\psi(t)dt = \left(\frac{a_0 t^{i+1}}{i+1} + \frac{a_1 t^{i+2}}{i+2} + \cdots + \frac{a_M t^{i+M+1}}{i+M+1}\right)\Bigg|_0^1 = 0,$$

$$1 \leqslant i \leqslant k. \tag{11}$$

New Vandermonde matrix $\tilde{A}$ will contain polynomial terms of $t$ and the wavelet constraints,

$$\tilde{A} = \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^M \\ \vdots & & \vdots & & \vdots \\ 1 & t_N & t_N^2 & \cdots & t_N^M \\ 1 & 1/2 & 1/3 & \cdots & 1/(M+1) \\ 1/2 & 1/3 & 1/4 & \cdots & 1/(M+2) \\ \vdots & & \vdots & & \vdots \\ 1/(k+1) & 1/(k+2) & 1/(k+3) & \cdots & 1/(k+M+1) \end{bmatrix} \quad 0 \leqslant t_1, \ldots, t_N \leqslant 1, \tag{12}$$

$$\theta = [a_0 \; a_1 \; a_2 \; \cdots \; a_M] \quad \text{and} \quad Z_N = [\zeta(t_1), \cdots, \zeta(t_N), 0, 0, \cdots, 0]^T. \tag{13}$$

The $N$-sample estimate of the coefficients vector $\hat{\theta}_N$ of the matching polynomial wavelet function can be determined to minimize the objective function $V_N(\theta, Z_N)$,

$$V_N(\theta, Z_N) = \frac{1}{N}\sum_{t=1}^{N}(Z_N - \tilde{A}^T\theta)^2, \tag{14}$$

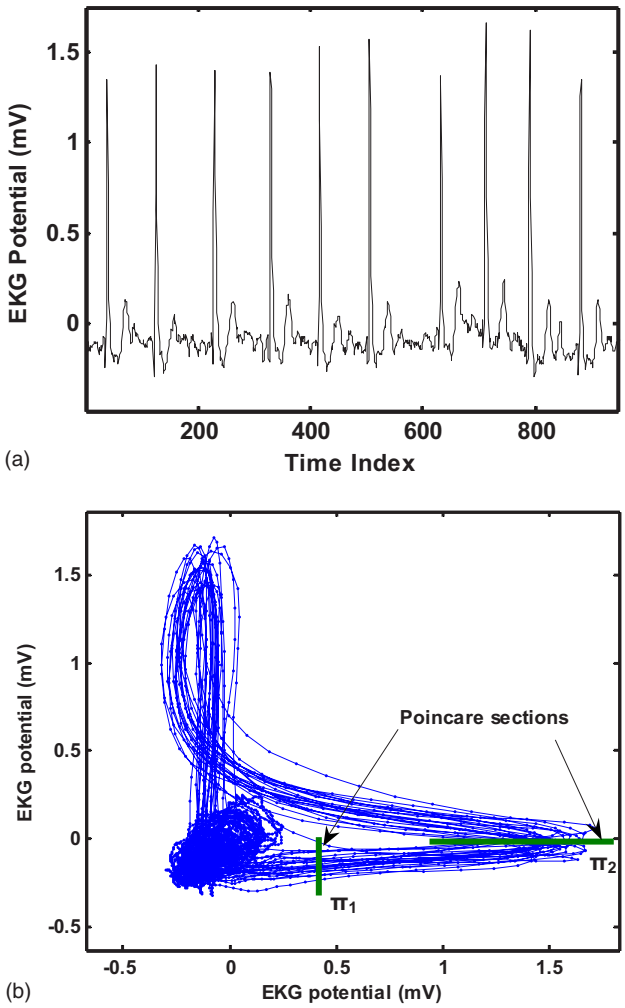


FIG. 4. (Color online) (a) Time domain plot of a representative EKG signal trace (a01). (b) State space portrait reconstructed from time delay coordinates.

$$\hat{\theta}_N = \arg\min_{\theta}\{V_N(\theta, Z_N)\}, \tag{15}$$

and the estimate can be obtained using the pseudoinverse

$$\theta^T = Z_N(\tilde{A}^T\tilde{A})^{-1}\tilde{A}^T. \tag{16}$$

## IV. IMPLEMENTATION OF WAVELET DESIGN ADAPTED TO EKG PATTERN

As discussed in the research methodology section, the coefficients of a compact wavelet representation need to be more sensitive to variation in the underlying processes (physiological) and less sensitive to noise variation. Thus, feature sets extracted from a compact representation tend to be lighter (i.e., fewer and more sensitive) and more effective in estimating various anomalies [here, different atrial fibrillation (AF) states]. For implementation and validation of the present approach, we utilized the EKG data from the 2004 PhysioNet challenge named "Spontaneous Termination of Atrial Fibrillation (AF)" , posted on the PhysioNet website [25,26]. Atrial fibrillation (AF) is one of the serious cardiac disorders that affect millions of human beings, and its early detection can have a significant bearing on the quality of healthcare. In this contest, classification needs to be made among the following three categories of AF patients test signals:

(1) *Group N*. Nonterminating AF (defined as AF that was not observed to have terminated for the duration of the long-term recording, at least an hour following the segment).

(2) *Group S*. Soon to be terminating (AF that terminates one minute after the end of the record).

(3) *Group T*. Terminating immediately (AF terminating within one second after the end of the record).

In all, 80 recordings of AF from 60 different subjects were made available in the database. Each record is a one-minute long segment containing two channel EKG signals (lead I and II), acquired at 128 sampling rate. Figure 4(a) contains EKG signal trace taken from a subject with AF for the duration of approximately 10 heartbeats. The trace shows a *QRS* complex with a significant *R* peak followed by a *T* wave. The signal is superposed with higher frequency (>6 Hz) atrial fibrillation *F* waves in lieu of *P* waves.

After the false nearest neighbor and mutual information test [12,27], the optimal embedding dimension and time delay were determined, and the state space from the EKG traces using time delay coordinates in $d_E$ embedding dimensions as $\underline{x}(t_n)=[x(t_n),x(t_{n+d_\tau}),x(t_{n+2d_\tau}),\ldots,x(t_{n+(d_E-1)d_\tau})]$ [see Fig. 4(b)]. The reconstructed state space shows a large loop of *QRS* complex extending top down and left to right, with

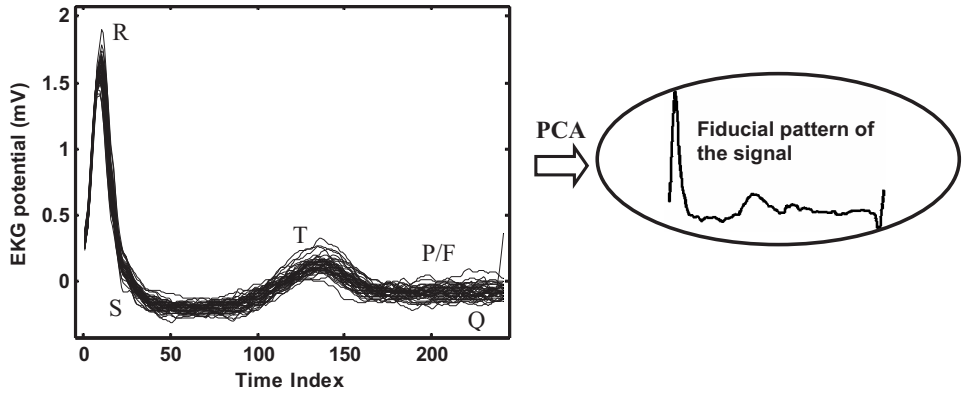


FIG. 5. EKG pattern ensembles extracted from the Poincare section.

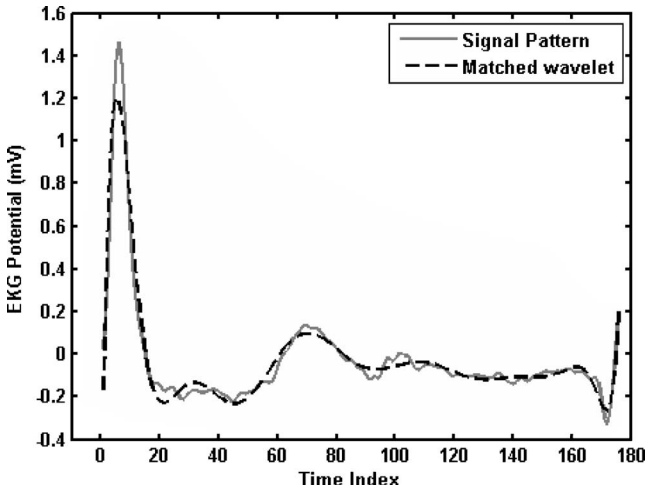


FIG. 6. Matching wavelet extracted from fiducial signal pattern.

the $R$ peaks occurring at the top and right extrema. A $T$ loop as well as an irregular ball formed by the $F$ waves are observed near the origin (bottom left corner). The Poincare sections $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ of the state portrait, as shown in Fig. 4(b), can be used to gather alternative sets of pattern ensembles.

The choice of the Poincare section $\boldsymbol{\pi}$ affects the shape of the pattern as well as the performances (entropy) of the representation. For example, one needs to choose the Poincare section such that the flow lines are directed about one or two flat planes, in which case $\boldsymbol{\pi}$ takes the form of a simple planar object. Also, it is desirable that the local Lyapunov exponent $\lambda_{max}$ of the state is close to or below zero about $\boldsymbol{\pi}$. This will help in making sure the resulting wavelet easily satisfies the constraints imposed by Eqs. (10) and (11). In this context, wavelet customized through Poincare section $\boldsymbol{\pi}_1$ can yield better performances than that through Poincare section $\boldsymbol{\pi}_2$ as shown in Fig. 4(b). Figure 5 shows the ensembles gathered from Poincare section $\boldsymbol{\pi}_1$. The fiducial EKG signal pattern used for the wavelet design stage can be extracted from

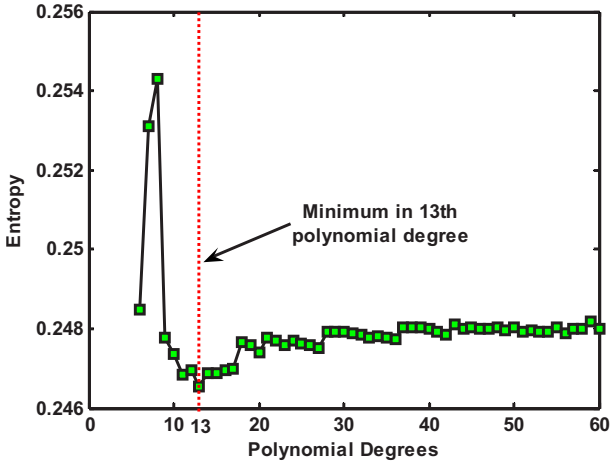


FIG. 7. (Color online) Variation of entropy of adaptive wavelet representation with polynomial degree.

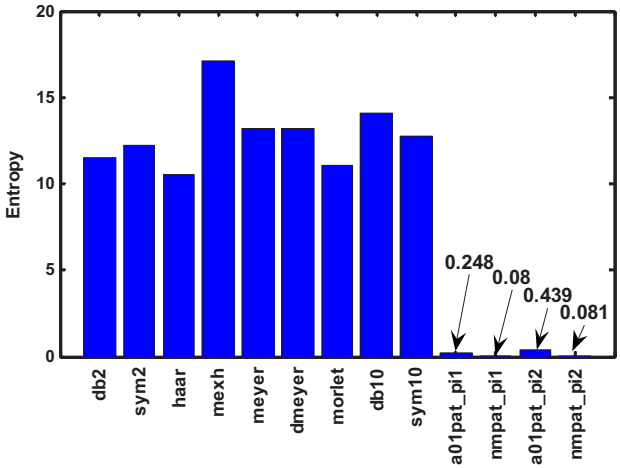


FIG. 8. (Color online) Entropy comparison between standard wavelets and matching wavelets.

dominant eigenfunctions estimated using the KL representation. Figure 6 shows the least square matching wavelet design result. The resulting wavelet $\psi(t)$ holds significant similarities to the fiducial pattern $\zeta(t)$ and capture a majority of the variations among the ensembles.

Figure 7 shows the variation of entropy [Eq. (1a)] with the polynomial degree of the wavelet function $\psi(t)$ that matches the fiducial pattern. The minimal entropy for matching wavelet (shown in Fig. 6) is found to have a polynomial degree of 13.

As summarized in Fig. 8, the customized wavelets of an EKG from a nonterminating AF case, obtained through Poincare sections $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ of the reconstructed state space, denoted a01pat_pi1 and a01pat_pi2, respectively, and customized wavelets from normal EKG pattern (nmpat_pi1 and nmpat_pi2), shown in the last four bars on the right side, are about 95% (approximately two orders of magnitude) more compact than all standard available wavelet bases investigated (shown on the left side), namely, db2 (Daubechies-2), sym2 (Symlet-2), haar (Haar), mexh (Mexican hat), meyr (Meyer), dmeyr (Discrete Meyer), morl (Morlet), db10

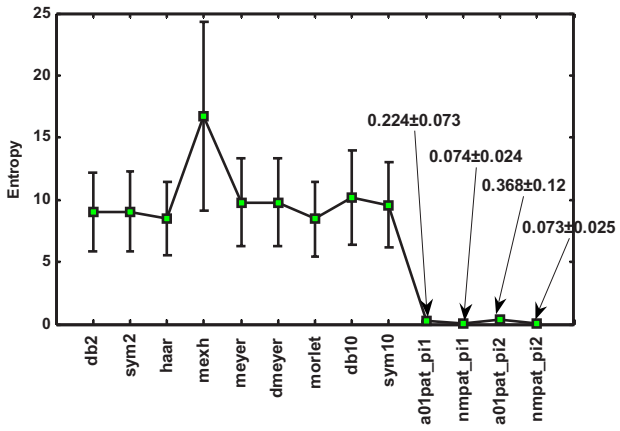


FIG. 9. (Color online) Entropy distribution for different wavelet representations, taken over all EKG signals in the database considered.

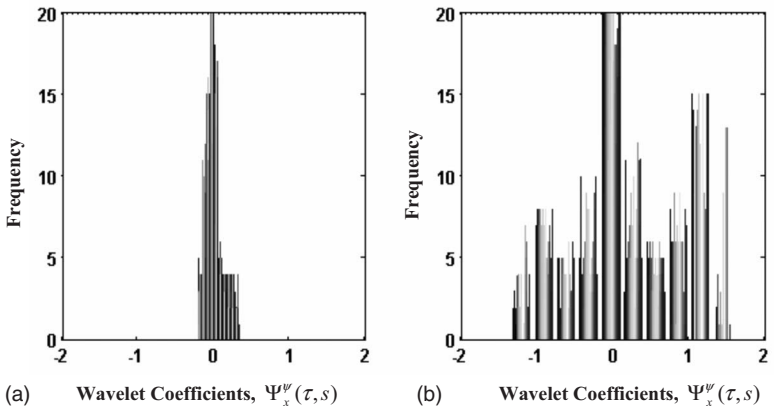


FIG. 10. Histogram of matching wavelet transform coefficients: (a) customized wavelet, (b) Morlet wavelet.

(Daubechies-10), and sym10 (Symlet-10). Also, the wavelets customized for a normal EKG signal seems to yield more compact representations for a signal from a nonterminating AF case (i.e., a01 signal in the PhysioNet [25,26].) The two orders of magnitude increase in the compactness of EKG signal representation with customized wavelet (measured in terms of entropy reduction) is further evidenced from the examination of the distribution of entropies for eleven signals (a01, a02, a03, b01, b02, n01, n02, s01, s02, t01, t02) in the PhysioNet database for the AF challenge with the eleven alternative wavelet bases including the standard and the customized wavelets (see Fig. 9). Interestingly, the wavelets customized for an EKG signal from a normal case yield about five times lower entropy compared to that from a nonterminating AF case. Also, it may be noted that, although both customization yield extremely low entropy compared to standard wavelet basis, the entropy of wavelet adapted from Poincare section $\boldsymbol{\pi}_2$ is twice as large as that from Poincare section $\boldsymbol{\pi}_1$. It is evidently due to the fact that the customized wavelet from $\boldsymbol{\pi}_1$ naturally closes to satisfying the wavelet constraint Eqs. (10) and (11).

Reasons for the compactness of the representations are further evident from an examination of the histograms of wavelet coefficients of the representations (see Fig. 10). The figure shows the distribution of wavelet coefficients within various bins of the histogram. The coefficients from the customized wavelet representation [see Fig. 10(a)] are concentrated around zero with few large coefficients. Such a low entropy distribution, with few large coefficients and several near-zero coefficients can lead to clearer demonstration and identification of various salient events in EKG signals. In contrast, coefficients of a representation from a standard library wavelet [see Fig. 10(b)] are spread uniformly, providing no clear distinction between significant and nonsignificant coefficients. Therefore, we have used the customized wavelets for *QRST* cancellation and feature extraction.

## V. CONCLUDING REMARKS

The choice of various basis functions is known to determine the compactness of a wavelet representation. In general, the closer the basis function captures the signal characteristics, the more compact is the representation, and more likely are the features sensitive to relevant EKG states and insensitive to variations in extraneous noise. In this investigation, we have customized the basis functions of a continuous wavelet representation by choosing polynomial wavelet basis functions that match the characteristics of a fiducial 1-beat long EKG signal pattern extracted from the Poincare sectioning of EKG state space. The customized representations were found to be roughly two orders of magnitude more compact (measured in term of signal entropy) than the wavelet basis functions available in the standard wavelet library. The unraveling of scale-time distribution of signal content in wavelet representation can facilitate the identification of various EKG events including the onsets, peaks, and offsets of various EKG waveforms for different beats. Further, it provides a means for QRST subtraction to suppress EKG signal components that emerge from ventricular sources.

## ACKNOWLEDGMENTS

The authors wish to thank the National Science Foundation (Grant No. DMI-0428356) for the generous support of the research reported in this paper. One of the authors (R.K.) also thanks A.H. Nelson, Jr., Endowed Chair in Engineering for additional financial support. This work is part of a larger project on biomedical Signal Analysis and Integrated Diagnostics (SAID) at the Okalhoma State University, Stillwater and OSU Center for Health Sciences, Tulsa.

[1] Z. D. Yum, J. Q. Xu, and G. P. Li,Proceedings of the 2003 IEEE International Symposium on Intelligent Control(IEEE, Houston, 2003), p. 642.
[2] Z. Huang, M.S. thesis, Michigan State University, 2002.
[3] R. Bousseljot and D. Kreiseler, Comput. Cardiol. **25**, 349 (1998).
[4] A. K. Jain, R. P. W. Duin, and J. Mao, IEEE Trans. Pattern Anal. Mach. Intell. **22**, 4 (2000).
[5] T. Irino and H. Kawahara, IEEE Trans. Signal Process. **41**, 3549 (1993).
[6] P. S. Addison, Physiol. Meas **26**, 155 (2005).
[7] C. Li, C. Zheng, and C. Tai, IEEE Trans. Biomed. Eng. **42**, 21 (1995).
[8] A. Stefanovska and M. Bracic, Contemp. Phys. **40**, 31 (1999).
[9] F. Ravelli and R. Antolini, Biol. Cybern. **67**, 57 (1992).
[10] H. E. Stanley, Physica A **270**, 309 (1999).
[11] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis* (Cambridge University Press, New York, 2000).
[12] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, New York, 1997).
[13] S. T. S. Bukkapatnam (unpublished).
[14] M. Misiti, Y. Misiti, G. Oppenheim, and J.-M. Poggi, *Wavelets and Teir Applications* (ISTE, London, 2007).
[15] M. V. Wickerhauser, Department of Mathematics Technical Report, Washington University, 1991 (unpublished).
[16] R. R. Coifman and M. V. Wickerhauser, IEEE Trans. Inf. Theory **38**, 713 (1992).
[17] S. T. S. Bukkapatnam, in Proceedings of ASME Design Engineering and Technical Conference, DETC-VIB 8068 (ASME, NY, 1999).
[18] S. T. S. Bukkapatnam, S. R. T. Kumara, and A. Lakhtakia, IMA J. Appl. Math. **63**, 149 (1999).
[19] F. Takens, *Detecting Strange Attractors in Turbulence* (Springer, Berlin, 1981), Vol. 898.
[20] F. Marciano, M. L. Migaux, D. Acanfora, G. Furgi, and F. Rengo, Comput. Cardiol. **21**, 577 (1994).
[21] G. D'Addio, G. D. Pinna, R. Maestri, G. Corbi, N. Ferrara, and F. Rengo, Comput. Cardiol. **31**, 457 (2004).
[22] I. Daubechies and B. Han, *Pairs of Dual Wavelet Frames from Any Two Refinable Functions* (Springer, New York, 2004).
[23] C. Valens, *A Really Friendly Guide to Wavelets*, http://perso.orange.fr/polyvalens/clemens/wavelets/wavelets.html, 1999–2007.
[24] G. Strang and T. Nguyen, *Wavelets and Filter Banks* (Wellesley-Cambridge Press, Cambridge, MA, 1996).
[25] G. Moody, Comput. Cardiol. **31**, 101 (2004).
[26] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng, and H. Stanley, Circulation **23**, e215 (2000).
[27] S. T. S. Bukkapatnam, A. Lakhtakia, and S. R. T. Kumara, Phys. Rev. E **52**, 2375 (1995).
[28] Marquette-KH, Marquette Electronics, http://library.med.utah.edu/kw/ecg/mml/ecg_em_events.html, 1996.