Automatic Selection of the Threshold Value *r* for Approximate Entropy

Sheng Lu*, *Member, IEEE*, Xinnian Chen, Jørgen K. Kanters, Irene C. Solomon, and Ki H. Chon, *Senior Member, IEEE*

***Abstract*—Calculation of approximate entropy (ApEn) requires *a priori* determination of two unknown parameters, *m* and *r*. While the recommended values of *r*, in the range of 0.1–0.2 times the standard deviation of the signal, have been shown to be applicable for a wide variety of signals, in certain cases, *r* values within this prescribed range can lead to an incorrect assessment of the complexity of a given signal. To circumvent this limitation, we recently advocated finding the maximum ApEn value by assessing all values of *r* from 0 to 1, and found that maximum ApEn does not always occur within the prescribed range of *r* values. Our results indicate that finding the maximum ApEn leads to the correct interpretation of a signal's complexity. One major limitation, however, is that the calculation of all choices of *r* values is often impractical due to the computational burden. Our new method, based on a heuristic stochastic model, overcomes this computational burden, and leads to the automatic selection of the maximum ApEn value for any given signal. Based on Monte Carlo simulations, we derive general equations that can be used to estimate the maximum ApEn with high accuracy for a given value of *m*. Application to both synthetic and experimental data confirmed the advantages claimed with the proposed approach.**

***Index Terms*—Approximate entropy, bounded random process, Brownian motion, heart rate variability, nonlinear determinism.**

I. Introduction

APPROXIMATE ENTROPY (ApEn) is a widely used method to provide a general understanding of the complexity of data [1]–[3]. Its popularity stems from the fact that it can be applied to both short- and long-term data recordings, and it is relatively easy to use. Consequently, it has found applications in many disciplines [3]–[29].

ApEn determines the conditional probability of similarity between a chosen data segment of a given duration and the next set of segments of the same duration; the higher the probability, the smaller the ApEn value, indicating less complexity of the data. Complexity refers to the difficulties we have in describing or predicting a signal.

Given a time series with N data points, the calculation of ApEn requires *a priori* determination of two unknown parameters, m and r. The parameter m determines the length of the sequences to be compared, and its selection can be estimated by calculating the false nearest neighbor [30]. The second parameter, r, is the tolerance threshold for accepting similar patterns between two segments, and has been recommended to be within 0.1–0.2 times the standard deviation of the data [1]–[3]. This recommendation was largely based on its application to relatively slow dynamic signals such as heart rate [4]–[17] and hormonal release data [23]. Our recent work suggests that these recommended values are not always appropriate for fast dynamic neural signals [36]. Furthermore, for a Brownian motion time series, with the selection of $r = 0.15$ times the standard deviation, ApEn value can be low as deterministic signals, which erroneously suggest low complexity of the signal.

Recently introduced variants of ApEn methods, sample entropy (SampEn) [31], [32] and multiscale entropy [33], [34], were developed to overcome the self-match problem associated with ApEn, and to provide a time-scale-dependent ApEn, respectively. However, these two methods also rely on the choices of both m and r. Therefore, these alternate methods are not immune to the sensitivity of the choice of r.

To this end, our recent research has provided the valuable insight that the most appropriate threshold value is the one that provides the maximum ApEn value [36]. In this study, computer simulation examples consisting of various signals with different complexity as well as neural respiratory data from cats, rats, and mice were compared. It was found that neither ApEn nor sample entropy methods were accurate in measuring signals' complexity when the recommended values (e.g., $m = 2$ and $r =$ 0.1–0.2 times the standard deviation of the signal) were strictly adhered to. However, when we selected the maximum ApEn value as determined by considering many different r values, we were able to correctly discern a signal's complexity for both synthetic and experimental data. However, this requires that many different choices of r values need to be considered. This is a very cumbersome and time consuming process. Thus, the primary goal of the present paper is to develop a method that can automatically select the appropriate threshold value r that corresponds to the maximum ApEn value, without resorting to the calculation of ApEn for each of the threshold values selected in the range of 0 and 1 times the standard deviation.

Manuscript received May 21, 2007; revised December 13, 2007. This work was supported in part by the National Institutes of Health under Grant HL069629. *Asterisk indicates corresponding author.*

*S. Lu is with the Department of Biomedical Engineering, State University of New York (SUNY) Stony Brook, NY 11794 USA (e-mail: shenglu123@yahoo.com).

X. Chen is with the Department of Biomedical Engineering, Shanghai Jiao Tong University, 200030 Shanghai, China (e-mail: xinnian@gmail.com).

J. K. Kanters is with the Coronary Care Unit, Department of Internal Medicine, Elsinore Hospital, DK-3000Helsingør, Denmark. He is also with the Laboratory of Experimental Cardiology, Copenhagen Heart Arrhythmia Research Center, 2100 Copenhagen, Denmark (e-mail: jkanters@mfi.ku.dk).

I. C. Solomon is with the Department of Physiology, State University of New York (SUNY) Stony Brook, NY 11794 USA (e-mail: irene.solomon@sunysb.edu).

K. H. Chon is with the Department of Biomedical Engineering, State University of New York (SUNY) Stony Brook, NY 11794 USA (e-mail: ki.chon@sunysb.edu).

Digital Object Identifier 10.1109/TBME.2008.919870

II. Methods

A. ApEn

In this section, we outline the step-by-step procedure for calculating ApEn. Given a signal $x(n) = x(1),\ x(2), \ldots,\ x(N)$, where *N* is the total number of data points, the ApEn algorithm can be summarized as follows [1]–[3], [29]

1) Form *m*-vectors, *X*(1) to *X*($N - m + 1$) defined by

$$X(i) = [x(i), x(i+1), \ldots, X(i+m-1)] \qquad i = 1, N-m+1.$$

2) Define the distance *d*[*X*(*i*), *X*(*j*)] between vectors *X*(*i*) and *X*(*j*) as the maximum absolute difference between their respective scalar components

$$d[X(i), X(j)] = \max_{k=0,m-1}[|x(i+k) - x(j+k)|].$$

3) Define for each *i*, for $i = 1,\ N - m + 1$

$$C_r^m(i) = V^m(i)/(N-m+1), \quad \text{where}$$

$$V^m(i) = \text{no. of } d[X(i), X(j)] \le r.$$

4) Calculate $\phi^m(r)$ by taking the natural logarithm of each $C_r^m(i)$, and average it over *i* as defined in step 3

$$\phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln(C_r^m(i)).$$

5) Increase the dimension to $m + 1$ and repeat steps 1–4.
6) Calculate ApEn values for a finite data length of *N*

$$\text{ApEn}(m, r, N) = \phi^m(r) - \phi^{m+1}(r).$$

As previously mentioned, calculation of ApEn requires *a priori* specification of two unknown parameters: *m*, the embedding dimension, and *r*, a tolerance value. The value of *m* can be estimated using the calculation of false nearest neighbor [30]. However, theoretical and varied clinical applications especially for slow dynamics (e.g., heart rate variability and growth hormone release) have shown that either $m = 1$ or 2 and *r* between 0.1 and 0.2 of the standard deviation of the data provide good statistical validity of ApEn. In general, *r* is essentially a filter where the type of filter depends on the choice of *r*. For example, a large *r* can be thought of as an all-pass filter since the number of self-matches will be large, whereas a small *r* can lead to a low-pass filter like behavior since it will lead to few self-matches and detailed information may be lost.

B. Automatic Selection of r That Corresponds to the Maximum ApEn Value

1) Justification for Choosing the Maximum ApEn Value: To illustrate our reasoning for choosing the maximum ApEn value rather than strictly heeding the recommendation that *r* be between 0.1–0.2 times the standard deviation of the signal, we show in Fig. 1 ApEn values as a function of *r* for three different time series with decreasing complexity: white noise, cross chirp, and sinusoidal signals. We note that for all three signals, increasing *r* at first results in a concomitant increase in ApEn

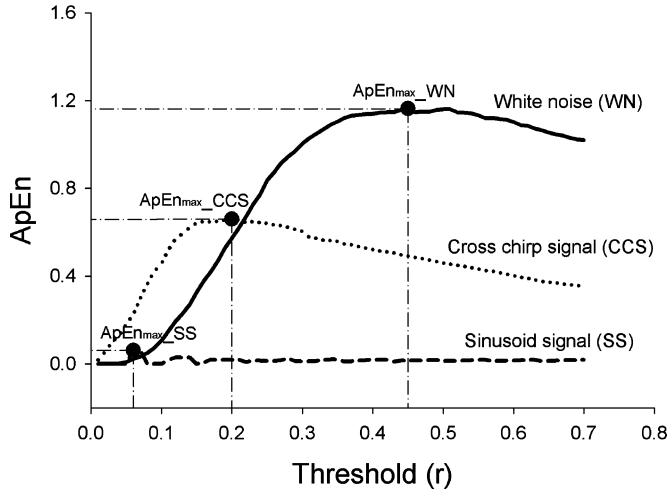


Fig. 1. ApEn values of white noise (WN), crosschirp signal (CCS), and sinusoid signal (SS) with various thresholds (*r*) are shown. Solid line indicates WN, dotted line indicates CCS, and dashed line indicates SS. The solid circle symbol indicates the maximum ApEn value.

with the maximum ApEn (denoted as $\text{ApEn}_{\max}$) found at different *r* values for all three signals. Thereafter, ApEn decreases with increasing *r*. If we choose $\text{ApEn}_{\max}$ for all three signals, there is no ambiguity as to which signal is more complex. However, if we were to choose $\text{ApEn}_{\max}$ based on the recommended *r* value being between 0.1–0.2 times the standard deviation of the signal, the cross-chirp signal has a higher ApEn value than the white noise signal. Certainly, this is a misleading result. Therefore, this simple example illustrates the pitfalls of the recommended *r* value selection process, and as an alternative, the most appropriate threshold value, *r*, can be simply determined by selecting the true $\text{ApEn}_{\max}$ value. Further examples demonstrating the appropriateness of selecting $\text{ApEn}_{\max}$ value instead of the recommended *r* value in the range of 0.1–0.2 times the standard deviation of the signal are provided in our recent study [36].

The significance of $\text{ApEn}_{\max}$ is that it denotes the largest information difference between data length *m* and $m + 1$ for any given *r*, thereby signifying the maximum complexity. We advocate the use of $\text{ApEn}_{\max}$ since it is less arbitrary than selecting the recommended *r* value within 0.1–0.2 times the standard deviation of the signal. Furthermore, as shown in Fig. 1, even within these suggested *r* values, there are wide variations in the ApEn values for all three signals, and the results can lead to the incorrect interpretation of complexities between these three signals. However, with the use of $\text{ApEn}_{\max}$, we obtain the correct information complexities for all three signals. Henceforth, we denote the *r* value to obtain the $\text{ApEn}_{\max}$ as the $r_{\max}$.

C. Automatic Selection of $r_{\max}$ Value

To automatically select $\text{ApEn}_{\max}$ without resorting to the calculation of every possible *r* value, our method is based on a theory about the behavior of the threshold value. We will show in the subsequent paragraph that the $r_{\max}$ value is dependent on the data record length and the square root of the ratio between short-term and long-term variability of the signal. To exploit

these relationships, the theory begins with a model of a bounded random process (BRP) [37]. Most biological signals exhibit both short- and long-term behaviors that can have wide ranges of complexity. The model of the BRP is defined by the following equation:

$$y(i) - y(i-1) = e(i)$$
$$\beta^- < y(i) < \beta^+ \quad (1)$$

where the time series, *y(i)* in (1), is the integrated white noise signal. Note that the time series *y*(*i*) differs from Brownian noise because of the boundaries β defined in the second line of (1). The top expression of (1) describes the short-term variability, in which differences between successive points are assumed to be random processes with the resultant time series, *e*(*i*), having zero mean. The standard deviation of *e*(*i*) is denoted as sd1. We use sd2 to denote the standard deviation of *y*(*i*) in (1), which can be thought of as long-term variability of the signal.

To simulate wide ranges of complexity, we used Monte Carlo simulations to generate 100 realizations of integrated independent and identically distributed Gaussian white noise signals ($e(i) \sim N(0,1)$), with each realization having different bound ($\beta^{\pm}$ is randomly selected from [± 2 to ± 20]). Thus, 100 realizations resulted in 100 different bounds. For each time series with different data lengths starting from 200 to 1000 at an increment of 100, ApEn values corresponding to threshold values ranging from 0.01 to 1, incrementing by 0.01, were computed for different embedding dimension values. We do not envision the embedding dimension to be higher than 4 for the relevant experimental data, so the embedding dimension values we used ranged from 2 to 4. Only the threshold value that provides the $\mathrm{ApEn}_{\max}$ for a particular embedding dimension was selected based on examining all ApEn values resulting from the set of threshold values.

A plot of the optimal *r* as a function of sd1/sd2 and data length, for embedding dimensions 2 and 3, is shown in Fig. 2. The right panels of Fig. 2 show 2-D illustration of the left panels for data lengths of 200, 600, and 1000 for the corresponding dimension. While the ApEn is based on choosing *r* between 0.1 and 0.2 of the standard deviation of the data, we emphasize the use of *r* as a function of the short-term variability, sd1, normalized by the long-term variability, sd2, for our method. We note a quasilinear relationship for both embedding dimensions. Thus, either a general linear or nonlinear equation can be derived for each embedding dimension. The general equations derived based on fitting multiple nonlinear least squares on the curves shown in Fig. 2 are provided next for each embedding dimension

$m = 2$:

$$\hat{r}_{\max} = (-0.02 + 0.23\sqrt{\mathrm{sd}_1/\mathrm{sd}_2})/\sqrt[4]{N/1000}. \quad (2)$$

$m = 3$:

$$\hat{r}_{\max} = (-0.06 + 0.43\sqrt{\mathrm{sd}_1/\mathrm{sd}_2})/\sqrt[4]{N/1000}. \quad (3)$$

$m = 4$:

$$\hat{r}_{\max} = (-0.11 + 0.65\sqrt{\mathrm{sd}_1/\mathrm{sd}_2})/\sqrt[4]{N/1000}. \quad (4)$$

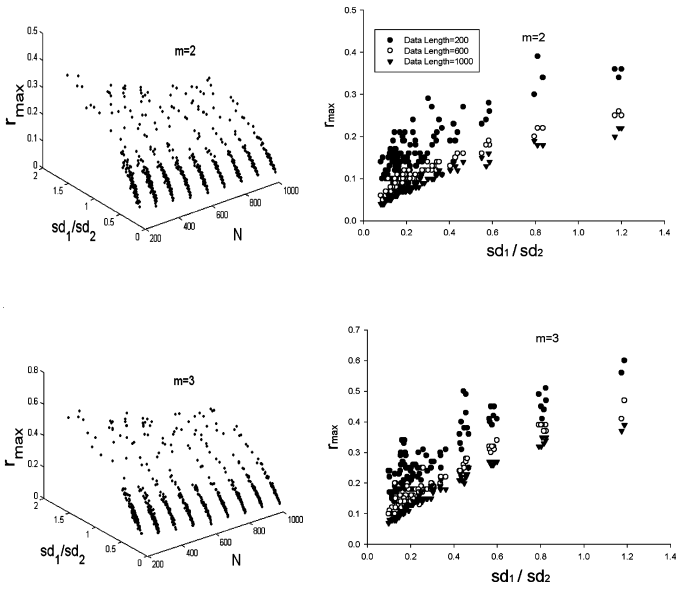


Fig. 2. Monte Carlo simulation plot of $r_{\max}$ as a function of sd1/sd2 and data length for two different embedding dimensions is shown. The right panels show 2-D illustrations of the left panels for data lengths of 200, 600, and 1000 for the corresponding dimension shown in the left panels. The time series ($N = 1000$) is generated by using the BRP model. $r_{\max}$ and sd1/sd2 are calculated for each subset of the time series ($N = 200, 300, \ldots 1000$).

We did not consider embedding dimensions higher than 4 because they are rarely used in practice as most ApEn users heed to the recommended ED value of 2. In addition, we will show in Section III that embedding dimensions of 2 and 3 are appropriate for human and cat data, respectively. Higher ED values (>4) can be estimated using the approach that we have outlined for interested readers.

Note that for the aforesaid equations, the estimated $r_{\max}$ value approaches a value of zero as *N* increases to infinity. This should not be a concern since ApEn is not usually calculated for data lengths larger than a few thousand points. An example result, the plot of the optimal *r* as a function of sd1/sd2 for embedding dimensions 2 and 3 for the data length of 1000 points, is shown as closed circles and closed inverted triangles in Fig. 3. Using (2) and (3), we obtain the estimated $r_{\max}$ values as a function of sd1/sd2, and they are shown as open circles and open inverted triangles in Fig. 3. In general, we observe excellent agreement between the actual and the estimated $r_{\max}$ values although the accuracy degrades a little as the ratio of sd1/sd2 increases. Note that the difference between the true and estimated $r_{\max}$ does not translate into a large discrepancy between the true and estimated $\mathrm{ApEn}_{\max}$ values; supporting evidence will be provided in the next section.

For experimental data, the embedding dimension is estimated first, followed by calculation of sd1 [standard deviation of *e*(*i*) in (1)] and sd2 [standard deviation of *y*(*i*)] from which the optimal threshold value is determined from the provided equations for a particular embedding dimension.

The solid circle points in Fig. 1 represent the actual $\mathrm{ApEn}_{\max}$ values. For all three signals, *m* was set to 3. The estimated $\mathrm{ApEn}_{\max}$ values were 1.157 for white noise (true $\mathrm{ApEn}_{\max} =$

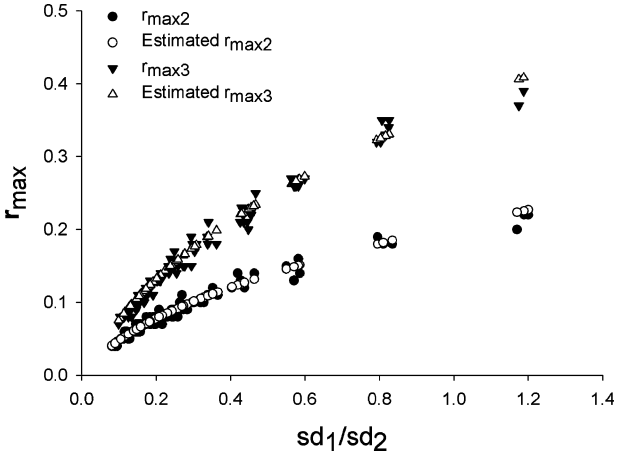


Fig. 3. Plot of the $r_{\max}$ as a function of sd1/sd2 for two different embedding dimensions for the data length of 1000 points is shown as closed circles ($m = 2$) and closed inverted triangles ($m = 3$). Open inverted triangles ($m = 3$) and open circles ($m = 2$) are values estimated using (2) and (3).

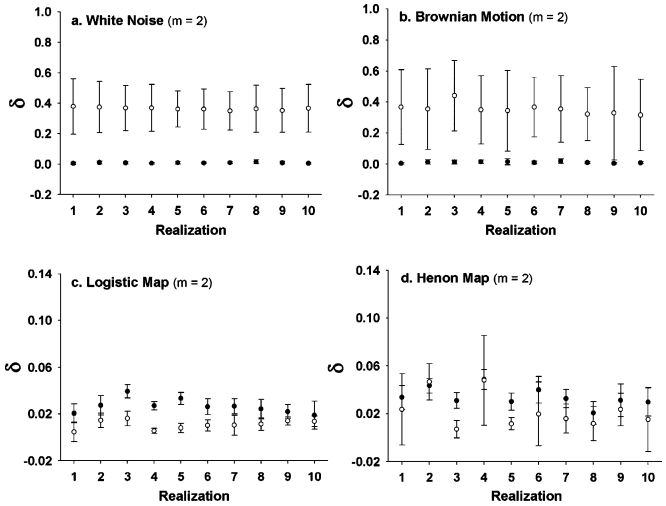


Fig. 4. $|\mathrm{ApEn}_{\max} - \mathrm{Ap\hat{E}n}_{\max}|$ and $|\mathrm{ApEn}_{\max} - \mathrm{Ap\hat{E}n}(2, 0.15)|$ (both are denoted as δ) for Gaussian white noise, Brownian motion, Logistic map, and Henon map series are shown (ten realizations for each type of signal). A solid circle indicates the estimated error by using (2) and an open circle indicates the error estimated by$\mathrm{Ap\hat{E}n}(2, 0.15)$.

1.163), 0.637 for cross-chirp (true $\mathrm{ApEn}_{\max} = 0.659$), and 0.015 for the sinusoidal signal (true $\mathrm{ApEn}_{\max} = 0.061$).

III. Results and Discussion

A. Synthetic Simulation Example Consisting of White Noise, Brownian Motion, and Henon and Logistic Map Series

To demonstrate the efficacy of our approach in automatically determining the $r_{\max}$ value, we generated ten independent realizations each for Gaussian white noise, Brownian noise, the logistic map, and Henon map time series. Every realization of the four types of signals contained 1000 data points. From each of these realizations, we generated nine new subrealizations of different data lengths by starting from 200 and incrementing by 100, up to the total data length of 1000 data points. The purpose of the last step was to examine the variability of the proposed and the original ApEn method on different data lengths. For every realization described earlier, the exact $\mathrm{ApEn}_{\max}$ values were determined for every curve as a function of *r* value, which was successively increased starting from 0.01 times the standard deviation up to 1 times the standard deviation, at an increment of 0.01. In addition, for every realization, we also determined estimates of $\mathrm{ApEn}_{\max}$ using (2) as well as the conventional estimates of ApEn values based on the arbitrary choice of $m = 2$ and with *r* set to 0.15 times the standard deviation of the signal. To examine how our proposed approach as well as the conventional approach compare to the true $\mathrm{ApEn}_{\max}$ values, we calculated the difference between $|\mathrm{ApEn}_{\max} - \mathrm{Ap\hat{E}n}_{\max}|$ and $|\mathrm{ApEn}_{\max} - \mathrm{Ap\hat{E}n}(2, 0.15)|$ for the nine subrealizations, to obtain mean and standard deviation values. This procedure was then repeated for each of the remaining nine realizations and their nine subrealizations. The outcome of this comparison is provided in Fig. 4. For stochastic signals, the proposed approach provides a more accurate estimate of $\mathrm{ApEn}_{\max}$ than the conventional method since the difference as denoted by δ is smaller. Even for deterministic signals, the proposed method provides a lower magnitude standard deviation around the true $\mathrm{ApEn}_{\max}$ value than the conventional approach. Further, the proposed approach is not affected by the varying data lengths since both error and standard deviation values are negligible. The conventional approach, however, has higher error and variability than the proposed approach especially for the stochastic signals.

B. Heart Rate Variability and Respiratory Neural Signals

Biological time series used to demonstrate the efficacy of the automatic selection of $\mathrm{ApEn}_{\max}$ include heart rate variability (HRV) series from ten healthy subjects (data length $N = 600$) and phrenic nerve discharge time-series data from ten □-chloralose anesthetized, vagotomized adult cats (data length $N = 2000$). HRV data used in this study consisted of the recordings of surface electrocardiogram (S-ECG). Measurements of S-ECG were sampled at 250 Hz to allow accurate detection and identification of QRS complexes in the ECG. The QRS complexes were used to identify beat locations. Once the timing of beats was determined, an instantaneous HR signal was created at a sampling rate of 4 Hz using the technique described in [35]. Raw phrenic nerve discharge data used for this study were recorded from the C_5 phrenic rootlet. The phrenic nerve discharge signals were amplified ($\times 10$ k), notch filtered at 60 Hz, and analog filtered to pass frequencies between 1 Hz and 500 Hz. Raw phrenic nerve activity was recorded on a computer at a sampling rate of 2 kHz (Chart 4.0, PowerLab, ADInstruments, Colorado Springs, CO). These data were then segmented to obtain data lengths corresponding to the inspiratory burst with little or no postinspiratory discharge, digitally bandpass-filtered between 20 and 250 Hz using a 4th-order Butterworth filter, and resampled with a sampling rate of 1 kHz [36].

For each of these experimental data sets, we performed a similar procedure of generating nine subrealizations starting from 200, with an increment of 100. We then calculated the

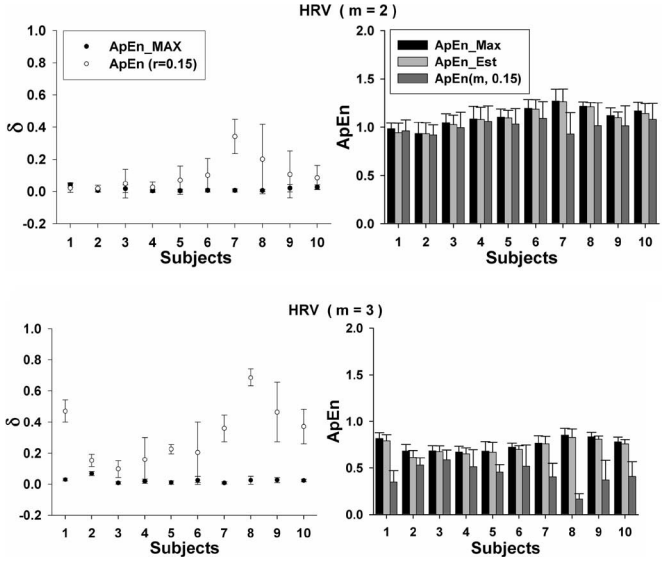


Fig. 5. $|\mathrm{ApEn}_{\max} - \mathrm{Ap\hat{E}n}_{\max}|$ and $|\mathrm{ApEn}_{\max} - \mathrm{Ap\hat{E}n}(m, 0.15)|$ (both are denoted as δ) for HRV from healthy subjects are shown in the left panels (ten subjects for each type of signal). A solid circle indicates the estimated error by using (2) and (3) and an open circle indicates the error estimated by$\mathrm{Ap\hat{E}n}(m, 0.15)$, where $m = 2$ or 3. The right panels show comparison of $\mathrm{Ap\hat{E}n}_{\max}$(black/1st column bar) and $\mathrm{Ap\hat{E}n}(m, 0.15)$(light gray/2nd column bar) to the true $\mathrm{ApEn}_{\max}$(gray/3rd column bar) for the same ten subjects.

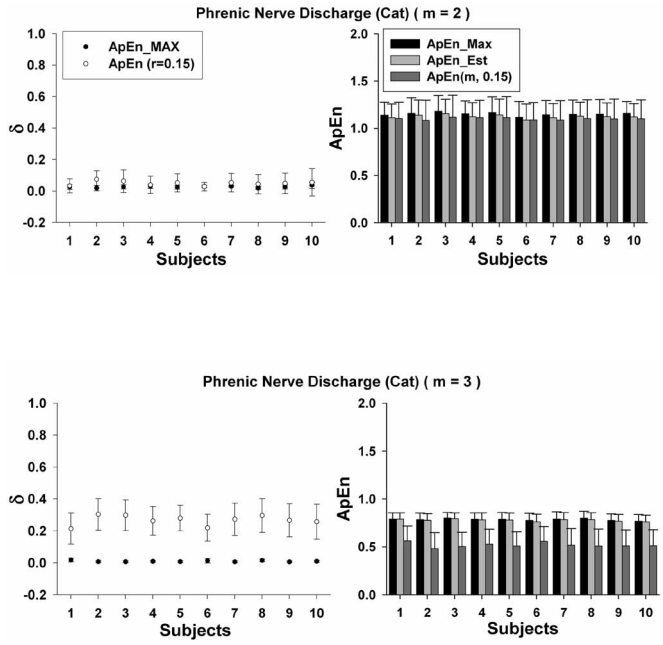


Fig. 6. $|\mathrm{ApEn}_{\max} - \mathrm{Ap\hat{E}n}_{\max}|$ and $|\mathrm{ApEn}_{\max} - \mathrm{Ap\hat{E}n}(m, 0.15)|$ (both are denoted as δ) for phrenic nerve discharge time series from adult cats are shown (ten subjects). A solid circle indicates the error estimated by using (2) and (3) and an open circle indicates the estimated error by$\mathrm{Ap\hat{E}n}(m, 0.15)$, where $m = 2$ or 3. The right panels show comparison of $\mathrm{Ap\hat{E}n}_{\max}$ (black/1st column bar) and $\mathrm{Ap\hat{E}n}(m, 0.15)$(light gray/2nd column bar) to the true $\mathrm{ApEn}_{\max}$ (gray/3rd column bar) for the same ten cats.

difference between $|\mathrm{ApEn}_{\max} - \mathrm{Ap\hat{E}n}_{\max}|$ and $|\mathrm{ApEn}_{\max} - \mathrm{Ap\hat{E}n}(m, 0.15)|$ for the nine subrealizations. The results from these comparisons are provided in the left panels of Fig. 5 for HRV data and Fig. 6 for phrenic nerve discharge data. In the right panels of Figs. 5 and 6, comparisons of $\mathrm{Ap\hat{E}n}_{\max}$ and $\mathrm{Ap\hat{E}n}(m, 0.15)$ to the true $\mathrm{ApEn}_{\max}$ are provided. For the conventional technique, we used *m* of either 2 or 3 depending on the data set and *r* of 0.15 times the standard deviation of the signal. For all data sets examined, whether HRV or phrenic nerve discharge, slow or fast dynamics, we noted a negligible difference between the true $\mathrm{ApEn}_{\max}$ and the estimated $\mathrm{ApEn}_{\max}$ using our proposed method. Similar to the results obtained with the simulation example given earlier, the conventional ApEn approach resulted in larger errors and greater variability than the proposed approach. Further, even when *r* was changed to either 0.1 or 0.2 times the standard deviation of the signal to estimate the difference between $|\mathrm{ApEn}_{\max} - \mathrm{Ap\hat{E}n}(m, r = 0.1 \text{ or } 0.2)|$, the conventional ApEn resulted in greater variability than the proposed approach. Also illustrated in the right panels of Figs. 5 and 6, the estimated $\mathrm{Ap\hat{E}n}_{\max}$values (2nd column bar of each subject) are much closer to the true $\mathrm{ApEn}_{\max}$ (1st column bar of each subject) than $\mathrm{Ap\hat{E}n}(m, 0.15)$ (3rd column bar of each subject) for both $m = 2$ and $m = 3$. The differences between the true $\mathrm{ApEn}_{\max}$ and $\mathrm{Ap\hat{E}n}(m, 0.15)$ are much more pronounced, especially with $m = 3$ for HRV and phrenic nerve discharge data, whereas they are negligible with $\mathrm{Ap\hat{E}n}_{\max}$. It should be noted that, for human data, the closest to the true $\mathrm{ApEn}_{\max}$ is obtained with $m = 2$ whereas for the phrenic nerve discharge data from cats, $m = 3$ provides the best approximation. These results imply that the selection of the embedding dimension needs to be higher with faster dynamic signals.

IV. Discussion

To date, determination of ApEn has been made using a recommended *r* value within the range of 0.1–0.2 times the standard deviation of the signal [1]–[3]. This recommendation was largely derived for slow dynamics signals, and the user was left with an arbitrary choice of *r* value within the range defined earlier [1]–[3]. However, as we have shown in this study, ApEn values vary significantly even within the defined range of *r* values. Furthermore, white noise signals have smaller ApEn than chirp signal for some of the recommended *r* values (see Fig. 1). A consequence of this lower ApEn value for white noise than Brownian noise is that it leads one to make an incorrect interpretation that the former is less complex than the latter. In an attempt to resolve this inherent deficiency and to investigate if the recommended *r* values could be improved upon for fast dynamic signals, our recent study suggested that the most appropriate solution is to look for the $\mathrm{ApEn}_{\max}$ value [36]. However, an intractable side effect of finding the $\mathrm{ApEn}_{\max}$ value is the computation of ApEn for every possible *r* value, which is computationally burdensome.

In this study, we developed a method to automatically determine the $\mathrm{ApEn}_{\max}$ without resorting to calculation of ApEn for every possible *r* value. The method is based on a heuristic model termed the bounded random process (BRP), which is a stochastic model that incorporates characteristics of both short- and long-term variability inherent in the signal. Using Monte Carlo simulations, we derived the general equations for determining the $\mathrm{ApEn}_{\max}$ for three different but commonly used *m*

values. The derived equations also took into account data length dependency of ApEn. We did not derive the general equations for *m* greater than 4 since it is rare that anybody uses *m* greater than 4 in practice, and because ApEn is less sensitive to the choice of *m* than *r*.

While the BRP model is a stochastic model, we have validated the accuracy of the proposed approach in finding the $\mathrm{ApEn}_{\max}$ values for many varieties of deterministic signals. Figs. 4–6 illustrate the efficacy of the proposed approach as the difference between the actual $\mathrm{ApEn}_{\max}$ and the estimated $\mathrm{ApEn}_{\max}$ values are negligible for a wide variety of signals (including deterministic signals), and the accuracy remains unaffected by different data lengths. Thus, the burden of automatically finding the $\mathrm{ApEn}_{\max}$ has been averted by the provided general equations (2)–((4)). Furthermore, these equations lead to a more accurate representation of the complexity than does the conventional method for most deterministic and stochastic signals. We believe obtaining accurate complexity of the signal is especially desired under circumstances in which a comparison is made between a baseline state and a pathological condition or pharmacological manipulation of the biological system. Thus, further studies are needed to examine how the calculation of the maximum entropy value can help in improving classification and quantitative data analysis between normal and diseased states.

REFERENCES

[1] S. M. Pincus and R. R. Viscarello, "Approximate entropy: A regularity measure for fetal heart rate analysis," *Obstet. Gynecol.*, vol. 79, pp. 249–255, 1992.

[2] S. M. Pincus, I. M. Gladstone, and R. A. Ehrenkranz, "A regularity statistic for medical data analysis," *J. Clin. Monit.*, vol. 7, pp. 335–345, 1991.

[3] S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 88, pp. 2297–2301, 1991.

[4] G. S. Dawes, M. Moulden, O. Sheil, and C. W. Redman, "Approximate entropy, a statistic of regularity, applied to fetal heart rate data before and during labor," *Obstet. Gynecol.*, vol. 80, pp. 763–768, 1992.

[5] D. T. Kaplan, M. I. Furman, S. M. Pincus, S. M. Ryan, L. A. Lipsitz, and A. L. Goldberger, "Aging and the complexity of cardiovascular dynamics," *Biophys. J.*, vol. 59, pp. 945–949, 1991.

[6] L. A. Fleisher, S. M. Pincus, and S. H. Rosenbaum, "Approximate entropy of heart rate as a correlate of postoperative ventricular dysfunction," *Anesthesiology*, vol. 78, pp. 683–692, 1993.

[7] A. L. Goldberger, J. E. Mietus, D. R. Rigney, M. L. Wood, and S. M. Fortney, "Effects of head-down bed rest on complex heart rate variability: Response to LBNP testing," *J. Appl. Physiol.*, vol. 77, pp. 2863–2869, 1994.

[8] K. K. Ho, G. B. Moody, C. K. Peng, J. E. Mietus, M. G. Larson, D. Levy, and A. L. Goldberger, "Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics," *Circulation*, vol. 96, pp. 842–848, 1997.

[9] C. W. Hogue, Jr., P. P. Domitrovich, P. K. Stein, G. D. Despotis, L. Re, R. B. Schuessler, R. E. Kleiger, and J. N. Rottman, "RR interval dynamics before atrial fibrillation in patients after coronary artery bypass graft surgery," *Circulation*, vol. 98, pp. 429–434, 1998.

[10] H. V. Huikuri, T. H. Makikallio, and J. Perkiomaki, "Measurement of heart rate variability by methods based on nonlinear dynamics," *J. Electrocardiol.*, vol. 36, Suppl., pp. 95–99, 2003.

[11] W. S. Kim, Y. Z. Yoon, J. H. Bae, and K. S. Soh, "Nonlinear characteristics of heart rate time series: Influence of three recumbent positions in patients with mild or severe coronary artery disease," *Physiol. Meas.*, vol. 26, pp. 517–529, 2005.

[12] J. T. Korpelainen, K. A. Sotaniemi, A. Makikallio, H. V. Huikuri, and V. V. Myllyla, "Dynamic behavior of heart rate in ischemic stroke," *Stroke*, vol. 30, pp. 1008–1013, 1999.

[13] T. A. Kuusela, T. T. Jartti, K. U. Tahvanainen, and T. J. Kaila, "Nonlinear methods of biosignal analysis in assessing terbutaline-induced heart rate and blood pressure changes," *Amer. J. Physiol. Heart Circ. Physiol.*, vol. 282, pp. H773–H783, 2002.

[14] T. H. Makikallio, T. Seppanen, M. Niemela, K. E. Airaksinen, M. Tulppo, and H. V. Huikuri, "Abnormalities in beat to beat complexity of heart rate dynamics in patients with a previous myocardial infarction," *J. Amer. Coll. Cardiol.*, vol. 28, pp. 1005–1011, 1996.

[15] J. A. Palazzolo, F. G. Estafanous, and P. A. Murray, "Entropy measures of heart rate variation in conscious dogs," *Amer. J. Physiol.*, vol. 274, pp. H1099–H1105, 1998.

[16] S. M. Pikkujamsa, T. H. Makikallio, K. E. Airaksinen, and H. V. Huikuri, "Determinants and interindividual variation of R–R interval dynamics in healthy middle-aged subjects," *Amer. J. Physiol. Heart Circ. Physiol.*, vol. 280, pp. H1400–H1406, 2001.

[17] P. Van Leeuwen, D. Cysarz, S. Lange, and D. Gronemeyer, "Increase in regularity of fetal heart rate variability with age," *Biomed. Tech. (Berl.)*, vol. 51, pp. 244–247, 2006.

[18] M. Akay, "Influence of the vagus nerve on respiratory patterns during early maturation," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 11, pp. 1863–1868, Dec. 2005.

[19] M. Akay and N. Sekine, "Investigating the complexity of respiratory patterns during recovery from severe hypoxia," *J. Neural Eng.*, vol. 1, pp. 16–20, 2004.

[20] N. Burioka, G. Cornelissen, F. Halberg, D. T. Kaplan, H. Suyama, T. Sako, and E. Shimizu, "Approximate entropy of human respiratory movement during eye-closed waking and different sleep stages," *Chest*, vol. 123, pp. 80–86, 2003.

[21] N. Burioka, G. Cornelissen, Y. Maegaki, F. Halberg, D. T. Kaplan, M. Miyata, Y. Fukuoka, M. Endo, H. Suyama, Y. Tomita, and E. Shimizu, "Approximate entropy of the electroencephalogram in healthy awake subjects and absence epilepsy patients," *Clin. EEG Neurosci.*, vol. 36, pp. 188–193, 2005.

[22] D. Caldirola, L. Bellodi, A. Caumo, G. Migliarese, and G. Perna, "Approximate entropy of respiratory patterns in panic disorder," *Amer. J. Psychiatry*, vol. 161, pp. 79–87, 2004.

[23] X. Chen, K. H. Chon, and I. C. Solomon, "Chemical activation of pre-Botzinger complex in vivo reduces respiratory network complexity," *Amer. J. Physiol. Regul. Integr. Comp. Physiol.*, vol. 288, pp. R1237–R1247, 2005.

[24] R. Ferenets, T. Lipping, A. Anier, V. Jantti, S. Melto, and S. Hovilehto, "Comparison of entropy and complexity measures for the assessment of depth of sedation," *IEEE Trans. Biomed. Eng.*, vol. 53, pp. 1067–1077, Jun. 2006.

[25] M. Koskinen, T. Seppanen, S. Tong, S. Mustola, and N. V. Thakor, "Monotonicity of approximate entropy during transition from awareness to unresponsiveness due to propofol anesthetic induction," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 669–675, Jun. 2006.

[26] K. Natarajan, U. R. Acharya, F. Alias, T. Tiboleng, and S. K. Puthusserypady, "Nonlinear analysis of EEG signals at different mental states," *Biomed. Eng. Online*, vol. 3, no. 7, pp. 1–11, 2004.

[27] C. Papadelis, C. Kourtidou-Papadeli, P. D. Bamidis, N. Maglaveras, and K. Pappas, "The effect of hypobaric hypoxia on multichannel EEG signal complexity," *Clin. Neurophysiol.*, vol. 118, no. 1, pp. 31–52, 2006.

[28] J. D. Veldhuis, W. S. Evans, and C. Y. Bowers, "Impact of estradiol supplementation on dual peptidyl drive of GH secretion in postmenopausal women," *J. Clin. Endocrinol. Metab.*, vol. 87, pp. 859–866, 2002.

[29] R. Hornero, M. Aboy, D. Abasolo, J. McNames, and B. Goldstein, "Interpretation of approximate entropy: Analysis of intracranial pressure approximate entropy during acute intracranial hypertension," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 10, pp. 1671–1680, Oct. 2005.

[30] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. A, Gen. Phys.*, vol. 45, pp. 3403–3411, 1992.

[31] D. E. Lake, J. S. Richman, M. P. Griffin, and J. R. Moorman, "Sample entropy analysis of neonatal heart rate variability," *Amer. J. Physiol. Regul. Integr. Comp. Physiol.*, vol. 283, pp. R789–R797, 2002.

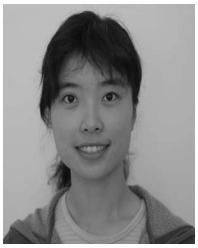
[32] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Amer. J. Physiol. Heart Circ. Physiol.*, vol. 278, pp. H2039–H2049, 2000.

[33] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Phys. Rev. Lett.*, vol. 89, pp. 068102-1–068102-4, 2002.

[34] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy to distinguish physiologic and synthetic RR time series," *Comput. Cardiol.*, vol. 29, pp. 137–140, 2002.
[35] S. Z. Abildstrom, B. T. Jensen, E. Agner, C. Torp-Pedersen, O. Nyvad, K. Wachtell, M. M. Ottesen, and J. K. Kanters, "Heart rate versus heart rate variability in risk prediction after myocardial infarction," *J. Cardiovasc. Electrophysiol.*, vol. 14, pp. 168–173, 2003.
[36] X. Chen, I. C. Solomon, and K. H. Chon, "Parameter selection criteria in approximate entropy and sample entropy with application to neural-respiratory signals," *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, to be published.
[37] S. Lu and J. K. Kanters, "Understanding approximate entropy in heart rate variability time series," *Biomed. Sig. Proc. Control*, to be published.

Sheng Lu received the B.S. degree in automatic control from the University of Science and Technology of China, Hefei, China, the M.S. and Ph.D. degrees in electrical engineering from City University of New York, New York.

Since 2002, he has been a Postdoctoral Researcher in the Department of Biomedical Engineering, State University of New York, Stony Brook. In 2005, he spent 1 year as a Postdoctoral Researcher at Weill Medical College, Cornell University, Itaca, NY, to conduct physiological experimental training. His current research interests include cardiovascular signal processing, and system identification and modeling of physiological systems.



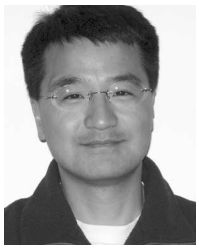
Xinnian Chen received the B.S. degree in mechano-electronic engineering from Xidian University, Xi'an, China, in 2000, and the M.S. degree in biomedical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2003.

Currently, she is working toward the Ph.D degree in the Department of Biomedical Engineering, State University of New York, Stony Brook. Her current research interests include biomedical signal processing and system identification in neuroscience.

Jørgen K. Kanters, photograph and biography not available at the time of publication.

Irene C. Solomon received the B.S. degree in physiology from the University of California, Davis, the M.S. degree in exercise physiology from California State University, Sacramento, and the Ph.D. degree in physiology from the University of California, Davis.

She spent 2 years as a Postdoctoral Fellow at the University of Medicine and Dentistry of New Jersey (UMDNJ)—Robert Wood Johnson Medical School, Newark, where she was engaged in research on central respiratory neural control. In 1996, she joined the State University of New York, Stony Brook, where she is currently an Associate Professor of Physiology and Biophysics. Her current research interests include brainstem mechanisms in respiratory rhythm generation and pattern formation, mechanisms of central hypoxic excitation, mechanisms of fast oscillatory rhythms in motor discharges, and computational modeling and simulation of respiratory network interactions.



Ki H. Chon received the B.S. degree in electrical engineering from the University of Connecticut, Storrs, the M.S. degree in biomedical engineering from the University of Iowa, Iowa City, and the M.S. degree in electrical engineering and the Ph.D. degree in biomedical engineering from the University of Southern California, Los Angeles.

He spent 3 years as a National Institutes of Health Postdoctoral Fellow in the Harvard—Massachusetts Institute of Technology Division of Health Science and Technology, 1 year as a Research Assistant Professor in the Department of Molecular Pharmacology, Physiology, and Biotechnology, Brown University, Providence, RI, and 4 years as an Assistant and Associate Professor in the Department of Electrical Engineering at the City College of the City University of New York, New York. Currently, he is a Professor in the Department of Biomedical Engineering, State University of New York, Stony Brook. His current research interests include medical instrumentation, biomedical signal processing, and identification and modeling of physiological systems.

Prof. Chon is an Associate Editor of the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING and *International Journal of Bioelectromagnetism*.