HEART PATHOLOGY DETERMINATION FROM ELECTROCARDIOGRAM SIGNALS BY APPLICATION OF DETERMINISTIC CHAOS MATHEMATICS

N. E. Clapp
L. M. Hively
Oak Ridge National Laboratory

R. E. Stickney
Physio-Control Corporation
11811 Willows Road Northeast
Post Office Box 97006
Redmond, WA 98073-9700

Prepared for
Laboratory Directed Research and Development Program
DOE Activity No. 473

Prepared by the
Engineering Technology Division
Oak Ridge National Laboratory
Oak Ridge, TN 37831-7294
Managed by
Lockheed Martin Energy Research
For the
U.S. Department of Energy
under contract DE-AC05-84OR21400

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED
DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.
HEART PATHOLOGY DETERMINATION FROM ELECTROCARDIOGRAM SIGNALS BY APPLICATION OF DETERMINISTIC CHAOS MATHEMATICS

N. E. Clapp
L. M. Hively
Oak Ridge National Laboratory

R. E. Stickney
Physio-Control Corporation
# Table of Contents

ABSTRACT ........................................................................... v

PURPOSE AND BACKGROUND .............................................. 1

SCOPE OF WORK ............................................................... 1
  Breakdown of Accomplishments by Tasks ............................ 1
  Task 1: EKG Data Selection, Handling, and Formatting .......... 1
  Task 2: CTSA Algorithm Modification and Data Analysis ....... 1
  Task 3: Interpretation and Application of Findings ................ 1
  Task 4: Reporting .......................................................... 2

NONLINEAR ANALYSIS DEFINITIONS ................................. 2
  Length Scale .................................................................... 2
  Mutual Information ......................................................... 2
  Principal Component Analysis .......................................... 3

EKG DATA ANALYSIS ......................................................... 4

CONCLUSIONS .................................................................... 6

REFERENCES ..................................................................... 7

APPENDIX A. ................................................................. A-1

APPENDIX B. ................................................................. B-1

APPENDIX C. ................................................................. C-1

APPENDIX D. ................................................................. D-1
ABSTRACT

It is well known that the electrical signals generated by the heart exhibit nonlinear, chaotic dynamics. A number of heart pathologies alter heartbeat dynamics and/or the electrical properties of the heart, which, in turn, alter electrocardiogram signals. Electrocardiogram techniques in common use for diagnosing pathologies have limited sensitivity and specificity. This leads to a relatively high misdiagnosis rate for ventricular fibrillation. It is also known that the linear analysis tools utilized (such as fast Fourier transforms and linear statistics) are limited in their ability to find subtle changes or characteristic signatures in nonlinear chaotic electrocardiogram signals. In contrast, our research indicates that chaotic time series analysis tools that we have developed allow quantification of the nonlinear nature of dynamic systems in the form of nonlinear statistics, and also enable characteristic signatures to be identified. The goal of this project is to modify these tools to increase and enhance the medically useful information obtained from electrocardiogram signals through the application of chaotic time series analysis tools. In the one year of the project, the tools have been extended to enhance the capabilities for detecting ventricular fibrillation. Chaotic time series analysis provides a means to increase sensitivity in detecting general heart dynamics. Oak Ridge National Laboratory specialists have worked with Physio-Control and their medical collaborators to extend the capabilities of state-of-the-art electrocardiogram systems and interpretation of results.
PURPOSE AND BACKGROUND

The purpose of this Cooperative Research and Development Agreement (CRADA) between Lockheed Martin Energy Research, Inc., also known as the Oak Ridge National Laboratory (ORNL), and Physio-Control Corporation, is to increase and enhance medically useful information from electrocardiogram (EKG) signals through the use of new signal analysis techniques. The developments in this study provide enhanced capabilities for detecting ventricular fibrillation and analysis of variations in EKG signals. The major tool, chaotic time series analysis (CTSA), appears well suited to significantly increase the sensitivity and specificity of EKG to certain coronary pathologies and heart dynamics in general.

Recent years have brought rapid development in the theory and application of deterministic chaos, aided by the wide availability of high-performance computing. It is now well known that many nonlinear dynamical systems treated as stochastic (dominated by random fluctuations), actually are governed by deterministic chaos. This study applies nonlinear techniques to selected EKG data.

SCOPE OF WORK

Research efforts have centered on data analysis using clinically documented EKG data available from a library of archived data provided by Physio-Control. The analysis tools were applied to selected EKG data sets to develop the capability to detect, quantify, and categorize heart fibrillation. The findings were applied to the development and design of algorithms suitable for new EKG-defibrillation products.

Breakdown of Accomplishments by Tasks

Task 1: EKG Data Selection, Handling, and Formatting

Research staff members from both ORNL and Physio-Control selected EKG data sets from a library. The selection criteria included data quality, contiguous length, and specific pathology. The selected EKG data sets were forwarded to ORNL.

Task 2: CTSA Algorithm Modification and Data Analysis

The major effort was to analyze the data sets supplied by the Physio-Control Corporation. The EKG data sets were drawn from patients with specific fibrillation related disorders. A code was written to access the data provided by Physio-Control and convert it to ASCII. Part of this task involved modifications to the CTSA algorithms to maximize their effectiveness with EKG data.

Task 3: Interpretation and Application of Findings

The output from Task 2 was reviewed by the Physio-Control Corporation and their medical consultants.
Task 4: Reporting

This document constitutes the final report. All algorithms developed are available for further evaluation and use. Both ORNL and Physio-Control Corporation will agree on what part of this report may be published in the open literature.

NONLINEAR ANALYSIS DEFINITIONS

Length Scale

A length scale is used to discriminate between large and small changes in magnitude of a signal varying with time, allowing noise to be removed while maintaining dynamic range.

The length scale used in this analysis is normalized using the absolute average deviation (AAD) for a given time window. AAD is calculated using the following:

$$ AAD = \frac{1}{NT} \sum_{i=1}^{NT} abs(d(i)),$$

where

- $d(i) =$ sampled data,
- $NT =$ number of points in the time window, and
- $AAD =$ absolute average value.

Mutual Information

Mutual information is a nonlinear version of the linear auto-correlation function, and was originally developed by Fraser and Swinney [1]. Mutual information measures the certainty with which a measurement can be predicted, given the outcome of another measurement. The mutual information indicates the average information that can be inferred from one measurement about a second measurement. As reported by Fraser and Swinney,

$$ MI(S,Q) = MI(Q,S) = H(Q) + H(S) - H(S,Q),$$

where

$$ H(S) = - \sum_i P_S(S_i) \log[ P_S(S_i) ],$$

$$ H(Q) = - \sum_j P_Q(Q_j) \log[ P_Q(Q_j) ],$$

and
\[ H(S, Q) = - \sum_{i,j} P_{SQ}(S_i, Q_j) \log[P_{SQ}(S_i, Q_j)]. \]

S denotes the whole system that consists of a set of states (binned values of S) with associated probabilities \(P_S(S_i)\) where \(i = 1\) to the total number of bins. Q denotes a set of states similar to S except at a different time. The function \(P_{SQ}(S_i, Q_j)\) is the joint probability of both states occurring simultaneously [2].

**Principal Component Analysis**

A principal component analysis is used to decompose a signal into a number of orthogonal time series. The resulting time series are then evaluated to find those that contain detail most useful in visualizing the specific problem.

The principal component analysis is used to determine the projection values to use in three-dimensional portraits. The principal component analysis uses a correlation matrix to determine eigenvalues and eigenvectors to use in the decomposition of the original signal. The elements in the correlation matrix are determined by the following equation:

\[
cm_{kj} = \frac{1}{NT - nd \cdot lag} \sum_{i=1}^{i=NT-nd*lag} d(i)^* d(i + abs(k - j)^* lag), \]

where

- \(d(i)\) – time series data,
- \(k\) – matrix row number,
- \(j\) – matrix column number,
- \(cm_{kj}\) – element in matrix for row \(k\) column \(j\),
- \(NT\) – number of points in the time window,
- \(nd\) – number of dimension used in correlation matrix, and
- \(lag\) – sample lag used in the phase plane projection.

Using the resulting matrix, eigenvalues are calculated and ordered by value with the largest being first. In this analysis second, third, and fourth eigenvalues were judged to contain the critical information to be evaluated and were used to obtain eigenvectors to generate the phase portraits from time series data.

\(EV_{mj}\) is the \(m^{th}\) eigenvector as determined by the \(m^{th}\) eigenvalue, and it has \(nd\) terms indexed by \(j\). The second, third, and fourth eigenvectors were used to generate the values for \(x\), \(y\), and \(z\) for the three-dimensional plot. In our case, \(nd\) equals ten, thus the eigenvector values have the effect of a finite impulse response digital filter on the time series data. The values for \(x\), \(y\), and \(z\) are determined by:
This is the principal component process that is used to generate the three-dimensional phase portraits of the EKG signals.

**EKG DATA ANALYSIS**

The set of figures in Appendix A shows the analysis for five Ventricular Tachyarrhythmia data sets received from Physio-Control Corporation. Each of these data sets shows the onset of ventricular fibrillation. The data was obtained by digitizing amplified signals from patient monitors and constitute a single channel. The signals as received by ORNL, were digitized using a 12 bit analog to digital converter, low-pass filtered by a two-pole filter with a 70 Hz break point and sampled at a rate of 250 Hz.

For each of the data sets, five types of analysis plots are presented. In Fig. A1, seven traces that we will call global signal features are shown. The first trace (MAX/MIN) indicates the maximum and minimum values of the EKG signal using a 40 s time window to generate one point. The time window is moved by 20 s to obtain the next 40 s window. This process provides a trace of maximum and minimum values of the EKG signal with time window for 40 s with a 20 s overlap. The second trace (SD/AAD) shows the standard deviation (SD) and the average absolute deviation (AAD, shown as dots) using the same windowing techniques. The third trace (SK/KT) indicates the behavior over the time period for skewedness (SK) and kurtosis (KT, shown by dots). The fourth trace (TS/CYC) shows the average time steps per beat using a window size of 40 s with an overlap of 20 s to generate the points in the trace. The fifth trace (M1) indicates the first minimum of the mutual information function using 20 bins with the bin size determined by (MAX-MIN)/(number of bins) for each of the 40 s intervals. The sixth trace [LG(K)] plots the log of entropy determined for each of the 40 s intervals using the bin size determined for the mutual information calculation. The last trace [correlation dimension (CD)] indicates the behavior of the correlation dimension using the same 40 s interval of data and a length scale of 2.0 AAD.

Figure A2 shows the details of the mutual information calculation for each of the 40 s intervals with 20 s of overlapped data. A constant vertical distance is used to separate each of the curves for a given 40-s interval.
Figure A3 shows how the entropy varies with time using the same 40 s window with a 20 s overlap. The top trace in this figure is the entropy when 1.0 AAD is used as the length scale. The length scale is increased by 1.0 AAD for each trace, going from the top to the bottom of the figure, starting with an AAD of 1.0 in the first trace. Thus, the bottom trace uses a length scale of 7.0 AAD to determine entropy. As the length scale is increased from 1.0 AAD to 7.0 AAD, the number of bins used to calculate entropy is decreased since the number of bins is determined by

\[ \text{number of bins} = \frac{\text{MAX} - \text{MIN}}{\text{length scale}}. \]

As the number of bins is decreased, the ability of the entropy to capture the structure of the time series data is lost. This is seen by the value of entropy decreasing as the number of bins is decreased. If the bin size is reduced so that all points in the time series fall into a single bin, the entropy would be zero. At this point, the entropy calculation would not show any structure for the data. The results of this variation indicates that a scale length of 1.0 AAD or 2.0 AAD is needed to show variation in data structure of the time series.

Figure A4 shows the CD vs length scale in AADs. The top left curve is for the first 40 s of the time series data. The next curve to the right shows 40 s of data shifted 20 s. This continues through the time series data moving left to right and top to bottom. This indicates that as the length scale is increased, the number of embedding dimensions required to project the data is decreased. For length scale between 1.0 AAD and 2.0 AAD, a correlation dimension of three will capture the dynamics. This type of analysis shows a definite change in behavior of the time series by the change in the CD vs length scale curve for a given time block. The correlation dimension is a good indicator of the length scale of noise in the time series data. The length scale of noise is determined by the point where the correlation dimension increases rapidly.

Figure A5 shows the three-dimensional plot for each of the time windows of 40 s. A principal component decomposition using a 10-dimension correlation matrix was used to determine the eigenvectors. Eigenvectors corresponding to the second, third, and fourth eigenvalues were selected to obtain the three-dimensional plots. The same format of left to right and top to bottom used to show the variation in the correlation dimension is used here to show the three-dimensional projections. As was seen in the correlation dimension curves, the three-dimensional plots reveal when there is a change in the structure of the time series data by a change in shape of the three-dimensional plot.

Figures A6–A10 provides the same plots for patient 2. Similarly, Figs. A11–A15, A16–A20, and A21–A25 provide plots for patients 3, 4, and 5, respectively.

Appendix B includes a set of figures (B1–B52) that shows the variation in the EKG signal of a single patient that has a ventricular fibrillation event. The time variation for a three-dimensional phase portrait, determined by a principal component decomposition of the EKG data set, is also provided. The data shown in Appendix B are from one of four additional data sets supplied by the Physio-Control Corporation and analyzed in this report. Each figure shows the three two-dimensional projections, a three-dimensional plot, the time series of the EKG signal, and a time
plot of the radius from the center of the three-dimensional plot to the points of the trajectory. The two-dimensional plot showing the variation in the x and y direction is shown in the top left portion of the figure. The plot showing the relationship between x and z is located just below the x and y plot. To the right of the x and y graph is the plot showing y and z directions. The three-dimensional plot show x (horizontal), y (vertical), and z (at an angle). The upper time series curve is the original EKG signal. The lower time series curve is the radius determined from the three-dimensional plot of the EKG signal.

Each of the phase portraits shown in Appendix B covers a 40 s time interval to indicate the change in behavior as the patient enters a ventricular fibrillation condition. Figure B49 shows the start of fibrillation for this patient.

Using the radius values generated by the three-dimensional plot, an entropy value was calculated for each of the 40 second data records. The value of this entropy versus time is shown for the four additional EKG data sets in Appendix C. Three of the four plots indicate a sharp rise in entropy as the patient goes into the ventricular fibrillation condition.

Appendix D shows the time series data for all the patients analyzed in this report.

CONCLUSIONS

The five data sets analyzed and presented in Appendix A indicated that entropy calculation using a length scale of 1.0 AAD provided a sensitive measure of ventricular fibrillation for the data from five patients. This global analysis also indicated that three dimensions could be used to show the dynamic behavior of the EKG signal. The values of the mutual information provided the nonlinear time constant of the data and allowed determination of the lag period to be used in the three-dimensional phase portraits. The windowed three-dimensional portraits provided a means of seeing transitional behavior of the EKG signal.

The phase portrait for a new patient, which is one of four new data sets analyzed, for the total data set windowed by 40 s is shown in Appendix B. Phase portraits are seen to be a useful way to emphasize EKG behavior and to show changes in EKG patterns.

In examining the time series value of the radius, a very slow amplitude modulated characteristic is seen. The character of this modulation points to a possible interaction between breathing and the EKG signal. The basic form of the trajectory, as shown in the three-dimensional plots, remains the same with slight changes in the size of the trajectories. The effect of a pre-ventricular contraction is seen in Fig. B35. The trajectory of the pre-ventricular contraction is very different from the normal EKG signal. This indicates a means of detecting abnormal EKG signals using the phase portraits.

The four additional data sets with ventricular fibrillation were used to generate the entropy plots shown in Appendix C. Figures C1, C2, and C4 show a sharp rise in entropy when ventricular fibrillation occurs. Figure C3 shows a gradual increase in entropy instead of a sharp rise.
REFERENCES


APPENDIX A

Appendix A provides chaos data treatment for five fibrillation related data sets.

<table>
<thead>
<tr>
<th>Figures</th>
<th>ASCII File</th>
<th>Patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1–A5</td>
<td>cu02</td>
<td>1</td>
</tr>
<tr>
<td>A6–A10</td>
<td>cu03</td>
<td>2</td>
</tr>
<tr>
<td>A11–A15</td>
<td>cu04</td>
<td>3</td>
</tr>
<tr>
<td>A16–A20</td>
<td>cu05</td>
<td>4</td>
</tr>
<tr>
<td>A21–A25</td>
<td>cu06</td>
<td>5</td>
</tr>
</tbody>
</table>
Fig. A1. Global signal features for patient 1.
Fig. A2. Mutual information windowed data for patient 1.
Fig. A3. Windowed entropy values for patient 1.
Fig. A4. Windowed correlation dimension versus length scale for patient 1
Fig. A5. Windowed principal component three-dimensional plot for patient 1.
Fig. A6. Global signal features for patient 2.
Fig. A7. Mutual information windowed data for patient 2.
Fig. A8. Windowed entropy values for patient 2.
Fig. A9. Windowed correlation dimension versus length scale for patient 2.
Fig. A10. Windowed principal component three-dimensional plot for patient 2.
Fig. A11. Global signal features for patient 3.
Fig. A12. Mutual information windowed data for patient 3.
Fig. A13. Windowed entropy values for patient 3.
Fig. A14. Windowed correlation dimension versus length scale for patient 3.
Fig. A15. Windowed principal component three-dimensional plot for patient 3.
Fig. A16. Global signal features for patient 4.
Fig. A17. Mutual information windowed data for patient 4.
Fig. A18. Windowed entropy values for patient 4.
Fig. A19. Windowed correlation dimension versus length scale for patient 4.
Fig. A20. Windowed principal component three dimensional plot for patient 4.
Fig. A21. Global signal features for patient 5.
Fig. A22. Mutual information windowed data for patient 5.
Fig. A23. Windowed entropy values for patient 5.
Fig. A24. Windowed correlation dimension versus length scale for patient 5.
Fig. A25. Windowed principal component three-dimensional plot for patient 5.
APPENDIX B

Appendix B provides nonlinear phase space treatment of one complete data set.

Each figure provides the three-dimensional phase space representation above the ASCII, time trace, and the radial component taken from the three-dimensional phase space plot. Each plot contains successive 40 s sequences until fibrillation is developed.

The top left plot is the standard horizontal x axis and vertical y axis. The top right shows the horizontal z axis and vertical y axis. The lower left has the horizontal x axis and the vertical z axis. The lower right plot is the full three-dimensional presentation with x horizontal, y vertical, and z axis coming out of the paper.

As one pages through the figures, the larger loops gradually fill in, indicating more variability followed by a general collapse with fibrillation beginning in Fig. B49.
Fig. B1. Three-dimension plot of 40 s of data starting at 4.0 s.
Fig. B2. Three-dimension plot of 40 s of data starting at 44.0 s.
Fig. B3. Three-dimension plot of 40 s of data starting at 84.0 s.
Fig. B4. Three-dimension plot of 40 s of data starting at 124.0 s.
Fig. B5. Three-dimension plot of 40 s of data starting at 164.0 s.
Fig. B6. Three-dimension plot of 40 s of data starting at 204.0 s.
Fig. B7. Three-dimension plot of 40 s of data starting at 244.0 s.
Fig. B8. Three-dimension plot of 40 s of data starting at 284.0 s.
Fig. B9. Three-dimension plot of 40 s of data starting at 324.0 s.
Fig. B10. Three-dimension plot of 40 s of data starting at 364.0 s.
Fig. B11. Three-dimension plot of 40 s of data starting at 404.0 s.
FIG. B12. Three-dimensional plot of 40 s of data starting at 444.0 s.
Fig. B13. Three-dimension plot of 40 s of data starting at 484.0 s.
Fig. B14. Three-dimension plot of 40 s of data starting at 524.0 s.
Fig. B15. Three-dimension plot of 40 s of data starting at 564.0 s.
Fig. B16. Three-dimension plot of 40 s of data starting at 604.0 s.
Fig. B17. Three-dimension plot of 40 s of data starting at 644.0 s.
FIG. B18. Three-dimensional plot of 40 s of data starting at 684.0 s.
Fig. B19. Three-dimension plot of 40 s of data starting at 724.0 s.
Fig. B20. Three-dimension plot of 40 s of data starting at 764.0 s.
Fig. B21. Three-dimension plot of 40 s of data starting at 804.0 s.
Fig. B22. Three-dimension plot of 40 s of data starting at 844.0 s.
Fig. B23. Three-dimension plot of 40 s of data starting at 884.0 s.
Fig. B24. Three-dimension plot of 40 s of data starting at 924.0 s.
Fig. B25. Three-dimension plot of 40 s of data starting at 964.0 s.
Fig. B26. Three-dimension plot of 40 s of data starting at 1004.0 s.
Fig. B27. Three-dimension plot of 40 s of data starting at 1044.0 s.
Fig. B28. Three-dimension plot of 40 s of data starting at 1084.0 s.
Fig. B29. Three-dimension plot of 40 s of data starting at 1124.0 s.
Fig. B30. Three-dimension plot of 40 s of data starting at 1164.0 s.
Fig. B31. Three-dimension plot of 40 s of data starting at 1204.0 s.
Fig. B32. Three-dimension plot of 40 s of data starting at 1244.0 s.
Fig. B33. Three-dimension plot of 40 s of data starting at 1284.0 s.
Fig. B34. Three-dimension plot of 40 s of data starting at 1324.0 s.
FIG. B35. Three-dimension plot of 40 s of data starting at 136.4 s.
Fig. B36. Three-dimension plot of 40 s of data starting at 1404.0 s.
Fig. B37. Three-dimension plot of 40 s of data starting at 1444.0 s.
Fig. B38. Three-dimension plot of 40 s of data starting at 1484.0 s.
Fig. B39. Three-dimension plot of 40 s of data starting at 1524.0 s.
Fig. B40. Three-dimension plot of 40 s of data starting at 1564.0 s.
Fig. B41. Three-dimension plot of 40 s of data starting at 1604.0 s.
Fig. B42. Three-dimension plot of 40 s of data starting at 16444.0 s.
Fig. B43. Three-dimension plot of 40 s of data starting at 1684.0 s.
Fig. B44. Three-dimension plot of 40 s of data starting at 1724.0 s.
Fig. B45. Three-dimension plot of 40 s of data starting at 1764.0 s.
Fig. B46. Three-dimension plot of 40 s of data starting at 1804.0 s.
Fig. B47. Three-dimension plot of 40 s of data starting at 1844.0 s.
Fig. B48. Three-dimension plot of 40 s of data starting at 1884.0 s.
Fig. B49. Three-dimension plot of 40 s of data starting at 1924.0 s.
Fig. B50. Three-dimension plot of 40 s of data starting at 1964.0 s.
Fig. B51. Three-dimension plot of 40 s of data starting at 2004.0 s.
Fig. B52. Three-dimension plot of 40 s of data starting at 2044.0 s.
APPENDIX C

Appendix C provides entropy plots for the four additional sets of data. Three of these show a clear onset of fibrillation.
Fig. C1. Entropy for data set 8201 using 10000 points to calculate entropy.
Fig. C2. Entropy for data set 8202 using 10000 points to calculate entropy.
Fig. C3. Entropy for data set 8203 using 10000 points to calculate entropy.
Fig. C4. Entropy for data set 8204 using 10000 points to calculate entropy.
APPENDIX D

Appendix D contains time series data printed directly from the ASCII files. Each line contains 60 s of data, each succeeding line beginning where the line before ends.

Figures D1–D5 provide the time series data for the analyses in Appendix A.

Figures D6–D9 provide the time series data for the analyses in Appendix C.

Observation of the time series data directly shows numerous interesting features and onset of fibrillation is clear. Note that an objective of this study is to provide a digital calculation that identifies the onset.
Fig. D1. Time series EKG signal for Appendix A patient 1.
Fig. D2. Time series EKG signal for Appendix A patient 2.
Fig. D3. Time series EKG signal for Appendix A patient 3.
Fig. D4. Time series EKG signal for Appendix A patient 4.
Fig. D5. Time series EKG signal for Appendix A patient 5.
Fig. D6. Time series EKG signal for patient used in Appendix B and the first patient analyzed in Appendix C.
Fig. D7. Time series EKG signal for the second patient analyzed in Appendix C.
Fig. D8. Time series EKG signal for the third patient analyzed in Appendix C.
Fig. D9. Time series EKG signal for the fourth patient analyzed in Appendix C.
DISTRIBUTION LIST

1. M. A. Akerman
2–6. N. E. Clapp (5)
7–11. L. H. Hively (5)
12. R. M. Schilling
13–17. R. E. Stickney (5)
18. Central Research Library
19. Laboratory Records-RC
20–21. Laboratory Records for submission to OSTI