# Analyzing Biological Time Series with Higher Order Cumulants

by

Beth Tamara Gerstein

B.S.E., Duke University (1995)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1997



Author .....
Department of Electrical Engineering and Computer Science
May 23, 1997

Certified by .....
Ary L. Goldberger, M.D.
Associate Professor, Harvard Medical School
Thesis Supervisor

Certified by .....
Roger G. Mark, M.D. Ph.D.
Professor, MIT
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chair, Department Committee on Graduate Students

# Analyzing Biological Time Series with Higher Order Cumulants

by

Beth Tamara Gerstein

Submitted to the Department of Electrical Engineering and Computer Science
on May 23, 1997, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering

## Abstract

Extensive linear and nonlinear analysis have been applied to heart rate variability (HRV) to gain insight into how the complex feedback system regulated by the autonomous nervous system controls the heart. Additionally, previous studies suggest a difference in the dynamics of healthy and pathologic interbeat intervals, possibly resulting from an altered control of the cardiovascular system. In my thesis, I analyze HRV in healthy and pathologic subjects to investigate possible differences in the dynamics of the interbeat interval. Because linear analysis suppresses important nonlinear interactions, and many previously applied nonlinear methods assume the heart rate is a specific type of deterministic chaos, these methods are limited and cannot be generously applied. Therefore, I apply higher order spectral analysis (HOSA), specifically third and fourth order cumulants, to HRV to investigate nonlinear properties of healthy and pathologic heart rate time series. HOSA has been previously employed to detect nonlinearities arising from a phase coupling of Fourier components in general time series. In addition, higher order spectral analysis has been used to suppress Gaussianities present in a signal. To this end, by applying third and fourth order cumulants, I demonstrate a distinctive nonlinear mode locking present in pathologic patients that is dramatically different than the complex nonlinearity encountered in healthy subjects.

Thesis Supervisor: Ary L. Goldberger, M.D.
Title: Associate Professor, Harvard Medical School

Thesis Supervisor: Roger G. Mark, M.D. Ph.D.
Title: Professor, MIT

# Acknowledgments

This work was partly supported by the grants to Dr. Ary Goldberger from the G.Harold and Leila Y. Mathers Charitable Foundation and the National Aeronautics and Space Administration. I would like to take this opportunity to thank a number of people without whom this thesis would not have been possible. First and foremost, I would like to thank my thesis advisor, Dr. Ary Goldberger, whose valuable discussions both motivated and enlightened me. It has truly been a pleasure to work with him. Next, I would like to thank C.-K. Peng, for his clear explanations, willingness to help, never-ending ideas, and last, but not least, for introducing me to the wonderful world of higher order spectral analysis. In addition, this thesis results from the academic guidance and support from several Beth Israel and BU Physics Department colleagues. Specifically, I would like to thank Joe Mietus, George Moody, Jeff Hausdorff, Plamen Ivanov, Eugene Stanley, Gandhi Viswanathan, and Roger Mark.

In addition, I would like to thank Andrew Allen, for revising my thesis at all hours of the night, and making me smile when no one else can. Working on my thesis with him has been a truly special experience.

Most of all, I would like to thank everyone in my family (my Mom, my Dad, Ellen, and Jennifer) for their never ending encouragement, love and support. My family makes me feel that there is nothing in this world that I cannot accomplish. Without them, I would not be where I am today. I dedicate my thesis to them.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

The normal heart rhythm has the deceptive appearance of regularity and consistency — after all, what could be more dependable than the pacemaker of life? However, measurements reveal that this seemingly uniform, metronomic signal is in fact highly variable, even with a constant level of physical activity. While the presence of these beat-to-beat variations has been recognized for quite some time, this variability is still often overlooked or treated as noise [1]. Furthermore, the mean heart rate, an indication of the level of cardiac exertion, is traditionally regarded as the principal measurement of interest.

In the past decade, considerable attention has been devoted to analyzing subtle heart rate dynamics. These fluctuations in the heart rate, known as heart rate variability (HRV), are studied primarily for two reasons [2]:

- They yield insight into the physiological mechanisms governing cardiovascular control.
- They provide clinical information regarding the overall condition of the cardiovascular system.

"Linear" methods such as power spectral analysis are used to assist clinical diagnosis of pathologies ranging from fetal distress syndrome to congestive heart failure (CHF) [3]. However, linear analysis is inadequate because it suppresses Fourier phase information. Because all nonlinear processes are characterized by Fourier phase inter-

actions, linear analysis is insufficient for describing such processes. Complex cardiac behavior suggests that the heart rate is the output of a highly nonlinear system.[1] Thus, linear analysis may not fully describe physiologic or pathologic heart rate dynamics. To more completely characterize HRV, nonlinear analysis is increasingly applied. Such analysis investigates temporal structure, or "phase information" within a time series using methods such as phase portraits and Lyapunov exponents [5]. However, these methods are also limited because they generally assume the heart rate is "chaotic," a specific type of nonlinear *deterministic* process.[2]

In my thesis, I introduce a method of nonlinear analysis known as higher order spectral analysis (HOSA), which extracts phase information from more general nonlinear processes (i.e., not necessarily chaotic). The specific tools for my analysis are the third and fourth order cumulants. By applying third order cumulants to heart rate time series, I investigate the presence of nonlinearities in both healthy and diseased patients to better understand the physiological processes in each system. Goldberger et al. [6] have hypothesized that the complex nonlinearity found in healthy HRV breaks down in diseased patients. Using HOSA, I test this hypothesis. Ultimately, the results from this analysis may be useful for both the mathematical modeling of the cardiovascular system and for clinical diagnosis of patients.

Before applying HOSA to "real world" heart rate time series, it is necessary to address some key questions. First, while HOSA theoretically can detect nonlinearities, how practically effective is HOSA for detecting nonlinearity in simulations? Specifically, it is important to consider finite-size effects (i.e., the number of data sets to be averaged and the length of the data needed to obtain an accurate estimate) and the unexpected numerical problems that can arise in computation. Additionally, can HOSA be used in practice to both distinguish and quantify different types of nonlinearities? While HOSA has often been applied in previous work to study frequency interactions, these interactions are usually quite simple, involving a phase coupling

[1] We must be careful in describing the heart rate as nonlinear, as this is a somewhat controversial conclusion that has not been concretely proven [4].

[2] The issues introduced in this paragraph will be explained in further detail in Section 2.1.

between only two or three frequencies (modes). However, in heart rate time series, the power-law scaling behavior of the power spectrum may indicate more complex nonlinear frequency interactions. Therefore, a third question to be addressed is whether HOSA is sensitive to power-law nonlinear frequency interactions. To better understand the use of HOSA for the study of nonlinear systems, we first apply HOSA to fully characterized mathematical models. This topic is explored in Chapter 3, which discusses the application of HOSA to two Gaussian random processes with a known autocorrelation $c_2(\tau)$ "fed" into a quadratic nonlinear system.

After determining the effect of nonlinearities in the mathematical simulations, I apply the knowledge obtained from these simulations to analyzing actual heart rate time series. Chapter 4 discusses the use of HOSA to both healthy subjects and CHF patients.

In addition to examining nonlinearities in the heart rate, Chapter 5 introduces the use of the fourth order cumulant to detect the presence of low amplitude, low frequency oscillations in the heart rate of patients with CHF. These important oscillations, associated with Cheyne-Stokes breathing, may not be readily apparent in the power spectrum. Because HOSA suppresses Gaussian noise, this method is extremely useful for detecting oscillations "hidden" in large amplitude noise. To this end, I develop a new method of analysis using HOSA to detect oscillations in heart failure patients.

The general goal in these series of investigations is to more clearly characterize the physiologic system governing the beat-to-beat interval, and compare the nonlinear HRV dynamics under selected healthy and pathologic conditions. The three specific inter-related aims of my thesis project are:

1. Determine the utility of third order cumulants for detecting and quantifying general nonlinearities in mathematical simulations (Chapter 3).

2. Apply third order cumulant analysis to compare healthy and pathologic heart rate time series (Chapter 4).

3. Use fourth order cumulants to detect low amplitude, low frequency oscillations

in pathologic heart rate time series (Chapter 5).

# Chapter 2

# Background

## 2.1 Physiological Motivation

The healthy, adult human heart spontaneously beats at a rate of approximately 50 to 80 times a minute at rest. Each of these beats is the end result of an elaborate sequence of electrical and mechanical events [7]. Electrically, the normal heartbeat originates in the sino-atrial (SA) node (located in the upper right atrium) which acts as the natural pacemaker, and hence is known as normal sinus rhythm. The electrical activity beginning in the SA node propagates radially through the atria, causing them to contract and pump blood into the ventricles. The excitation is next carried through the atrio-ventricular (AV) node to the specialized conduction system consisting of the bundle of His, bundle branches, and the network of Purkinje fibers. Through this conduction system, the electrical impulses activate the ventricular muscle, causing the ventricles to contract and the heart to pump blood to the lungs and system circulation.

To precisely calculate the interbeat interval time series used in HRV analysis, we utilize the electrocardiogram (ECG), a record of the net electrical activity of cardiac cells. The synchronized depolarization and repolarization of atrial and ventricular cells trace out well-defined waveforms illustrated in the ECG in Figure 2-1. The "P" wave reflects atrial depolarization, the "QRS" complex corresponds to the electrical activation of the ventricles, and the "T" wave reflects ventricular repolarization (re-

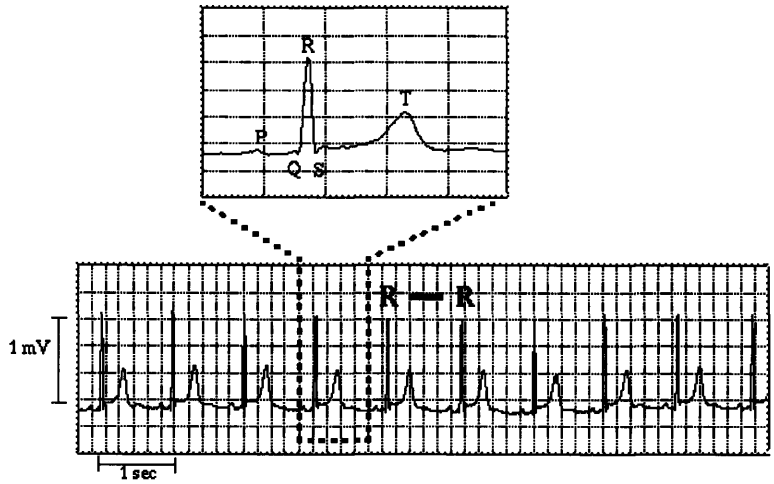



Figure 2-1: Electrocardiogram record of cardiac electrical activity.

covery). From the ECG, we determine the cardiac interbeat interval by calculating the time between two consecutive "R" waves (R-R interval). While the "R" wave does not precisely represent the "onset" of the mechanical heart beat, it is chosen to measure the period between beats because it is less affected by measurement and quantization noise and is more readily identifiable [2]. We obtain the overall heart rate time series by calculating successive cardiac interbeat intervals. The ECG is often recorded from an ambulatory device known as a Holter monitor, which stores and digitizes the data. Ectopic beats, which do not originate in the SA node, are usually removed before HRV analysis. These abnormal beats may occur even in healthy hearts, and often dominate the analysis if not excluded.

### Autonomic Control of the Heart Rate

The firing activity of the SA node is strongly regulated by autonomic influences from the brainstem, mechanics of respiration and circulation, and circulating neurohormones [8]. This complicated feedback control system is schematized in Figure 2-2. Although local factors such as tissue stretch and temperature changes affect the discharge frequency of the SA node, the two branches of the autonomic nervous system (ANS) are the most important factors in determining the heart rhythm. These

two branches, the parasympathetic (vagal) and sympathetic system, are regulated by higher centers located in the brain. Cardiac parasympathetic fibers originate in the dorsal motor nucleus (medulla oblongata), while sympathetic fibers originate in the vasomotor center [8]. Both the parasympathetic and the sympathetic system innervate the sino-atrial node, and thus the net effect of these systems impact the overall heart rate. The effects of these systems are mediated through neurotransmitters (norepinephrine in sympathetic, acetylcholine in vagal), which alter the electrical activity of pacemaker cells. Specifically, parasympathetic stimulation decreases the firing rate of the pacemaker cells, while sympathetic stimulation increases the firing rate.

Alterations in heart rate can be evoked by changes in blood pressure [3]. Transducers known as baroreceptors located in the aortic arch and carotid sinuses send feedback to the central nervous system. A change in blood pressure will elicit a change in heart rate through reciprocal changes in activity in the two autonomic divisions. The heart rate is also significantly affected by mechanical aspects of circulation, primarily respiration. Specifically, during inspiration, the heart rate increases, while in expiration, the heart rate decreases. This coupling of the heart rate to respiration is known as "respiratory sinus arrhythmia." It is important to note that this variability is indicative of a healthy condition, not a diseased one [2]. In addition, the cardiovascular system is connected to many relatively independent resistances within the circulation network. By varying its resistance, each tissue regulates its own blood flow and pressures. The cardiovascular system is regulated by these individual resistances [8].

The central nervous system acts by integrating information from various sensors within the cardiovascular system and controlling the rate of the heart beat through the peripheral ANS. Recently, there has been interest in studying HRV to gain insight into the nonlinear competition between sympathetic and parasympathetic mechanisms that govern the cardiovascular system [9]. A direct measurement of the parasympathetic and sympathetic activity is not clinically feasible; HRV, while indirect, provides a non-invasive indicator of integrated neuroautonomic control. The complex control

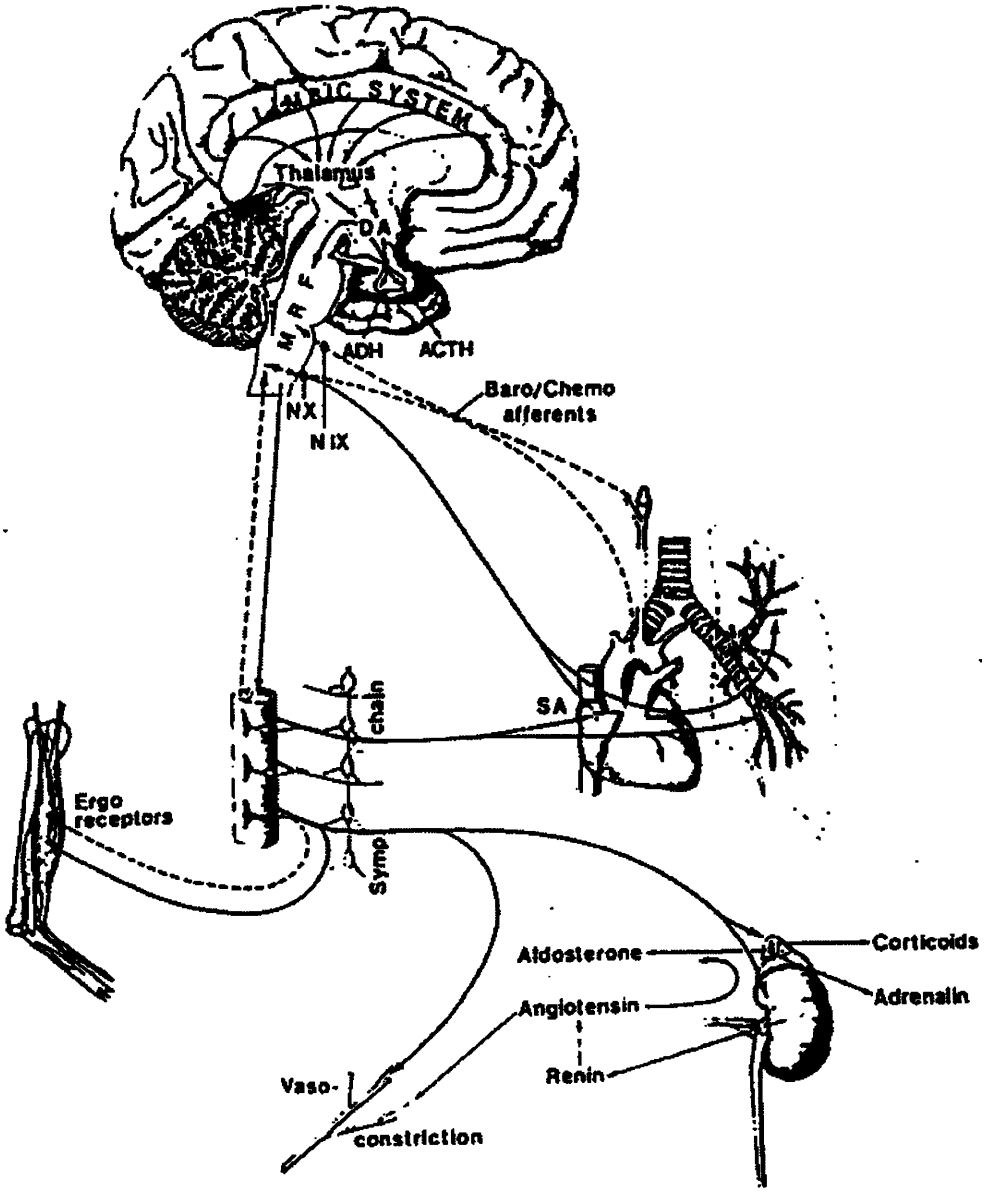



Figure 2-2: Complex feedback systems operating over wide range of temporal and spatial scales that regulate the dynamics of the heart rate. ADH — Antidiuretic hormone; ACTH — Adrenocorticotrophic hormone; SA — sino-atrial node.

feedback system of the cardiovascular system is not fully understood. However, both nonlinear and linear analysis of HRV have considerably advanced our current understanding of this system.

### 2.1.1 Linear Methods of Assessing Heart Rate Variability (HRV)

The beat-to-beat variability of the heart can be considered the integrated regulatory response of the cardiovascular system to intrinsic and extrinsic perturbations [1]. Examples of these perturbations range from changes in posture to adjustments in local vascular resistance in different tissue beds. A classical interpretation applied to a wide variety of physiologic systems is that the cardiovascular system acts in response to these disturbances to maintain *homeostasis* and achieve a physiological steady state [10]. According to this theory, developed by Walter B. Cannon of Harvard Medical School, we expect the heart rate to be relatively constant until perturbed and to remain at a constant value until perturbed again. This traditional homeostatic viewpoint assumes that the cardiovascular system under analysis is "well-behaved" and relatively "linear" — a change in the input elicits a proportional change in the output.[1] The assumption of the heart rate as linear is widespread, and many techniques have utilized this principle for both clinical application and physiologic understanding of the cardiovascular system.

Traditional statistics to quantify HRV include the mean and the standard deviation (first and second moments). These straightforward measures provide useful clinical application. For example, CHF is often associated with a decreased variability, as measured by standard deviation, compared to normal heart rate dynamics [11]. This characteristic is evident in the interbeat interval time series of a CHF patient shown in Figure 2-3 part (b), as contrasted to the heart rate time series of a healthy patient, shown in part (a). In addition, several studies have demonstrated that a lower standard deviation is associated with an increased mortality after my-

[1] A detailed description of linearity will be introduced in Section 2.2.

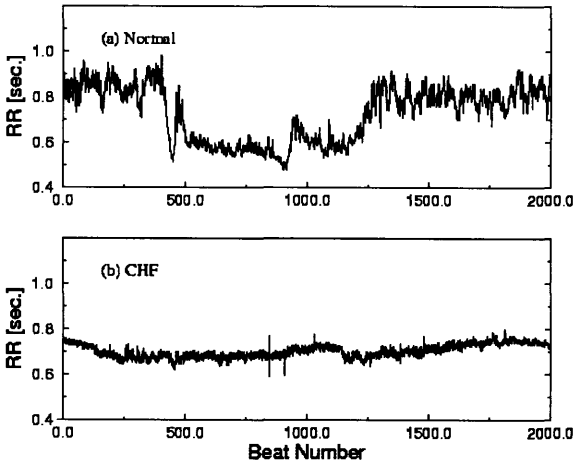



Figure 2-3: Heart rate time series for (a) healthy subject and (b) patient with congestive heart failure (CHF). Note the higher mean resting heart rate (lower R-R interval) and reduced variance in CHF.

ocardial infarcation. Thus, traditional HRV analysis plays an important prognostic role, especially because the patient can be monitored non-invasively.

Because the mean and standard deviation yield limited insight into the beat-to-beat *dynamics* of the heart rate, linear analysis is frequently applied. In a linear process, all frequency components are uncorrelated; thus, Fourier phases are randomized. Linear analysis, e.g. power spectral analysis, is traditionally used to characterize such processes. This technique uses Fourier methods to decompose a process into a superposition of sinusoids with different amplitudes and frequencies. The power spectrum presents the squared amplitude of these sinusoids as a function of frequencies. With this technique, oscillations present in the heart rate can be identified and correlated with physiological phenomena [2]. For example, the healthy heart rhythm, which is modulated by respiratory activity, oscillates with a frequency of approximately .25 Hz (corresponding to one respiratory cycle every four seconds). An oscillation may also be apparent at approximately .1 Hz. While the origin of this 10 second oscillation is not fully understood, it is known that it is importantly influenced by the baroreflex. Studies have shown that fluctuations in heart rate that occur at frequencies greater than approximately .15 Hz are mediated by parasympathetic activity, while lower

frequencies are jointly mediated by the sympathetic and parasympathetic nervous system.

The presence or absence of an oscillation (a peak in the frequency domain) can indicate altered autonomic nervous activity, thus providing clinical information. For example, diabetes, a disease accompanied by dysfunction of the ANS, is associated with a decrease in respiratory sinus arrhythmia, as well as changes in the 10 second oscillation induced by body posture changes. In addition, the heart rate of some CHF patients contains a one cycle per minute oscillation. This pathologic pattern, associated with Cheyne-Stokes breathing, will be discussed in more detail in Chapter 5.

It is important to recognize the clinical as well as physiologic importance of linear analysis — indeed, certain aspects of heart rate control can be deduced without dealing with the nonlinearities inherent in the cardiovascular system. However, the complex dynamic nature of the beat-to-beat interval cannot be fully explained with conventional linear methods. Furthermore, the constancy predicted by homeostasis cannot account for the dynamic, apparently far-from-equilibrium behavior observed in "regular" sinus rhythm: the intrinsic, highly irregular variability of the normal heart rate does not seem to correspond to a steady state [10]. In recent years, the assumption of HRV as a linear process has been challenged by increased evidence that many complex nonlinear interactions exist in physiologic systems, including the cardiovascular system. Consequently, researchers employ nonlinear analysis to more fully explain complex cardiac dynamics [5].

### 2.1.2 Nonlinear Methods of Assessing HRV

Nonlinear dynamics are used to study systems that respond disproportionately to stimuli [12]. Such a system cannot simply be dissected into its components, since the modes of the system interact. As a result, a nonlinear mechanism can produce a highly complex output, such as the heart rate. In addition, in nonlinear systems, small changes in a control parameter can lead to large changes in behavior. A wide class of different behaviors can typically result from even a simple nonlinear process. A typical feature of heart activity is the heterogeneous, intermittent patterns that

change with time.[2] When nonlinear dynamics are used to characterize HRV, these sudden dramatic changes in cardiac behavior do not need to be attributed to an equally dramatic change in cardiovascular control.

Chaos is a branch of deterministic nonlinear dynamics that has been applied recently to quantify and explain erratic cardiac behavior [5]. The term "deterministic" implies that chaotic systems are not strictly random, but rather the equations, parameters, and initial conditions are known and can be completely described mathematically. A chaotic process demonstrates aperiodic dynamics and demonstrates a sensitive dependence on the initial conditions: a minor change in the initial conditions may cause a marked difference in the outcome (known as the "butterfly effect"). Nonlinear chaos is not characterized by a complete state of disorganization as the vernacular suggests, but rather a constrained type of randomness which is associated with fractals — the geometric concept of scale-invariance or self-similarity which applies to the structure of chaotic (strange) attractors.

The properties of heart rate time series as well as the underlying physiology controlling the cardiovascular system demonstrate the possibility of chaotic or fractal influences in cardiac behavior [12]. Careful inspection of the heart rate reveals similar fluctuations on multiple different orders of temporal magnitude. Figure 2-4 demonstrates this scale-invariant characteristic, which is absent in any non-fractal process. Self-similar processes are found in a range of physical systems ranging from earthquakes to other biological systems. These processes are characterized by a power spectrum with a "$1/f$" power law decay that contains the same spectral power in any decade of frequency. Figure 2-5 displays this $1/f$ power spectrum of a healthy heart rate time series analyzed over several hours. This scale-invariant "fractal" manner of the heart rate is consistent with, but not diagnostic of chaos, since fractals may also arise from stochastic mechanisms.

The broad spectrum of the heart containing superimposed peaks in frequency (from the baroreceptor reflex and respiration) is also consistent with (but not necessarily indicative of) chaos [12]. In addition, phase space trajectories of the heart rate

[2]This non-stationarity increases the difficulty in analyzing the dynamics of the heart rate.

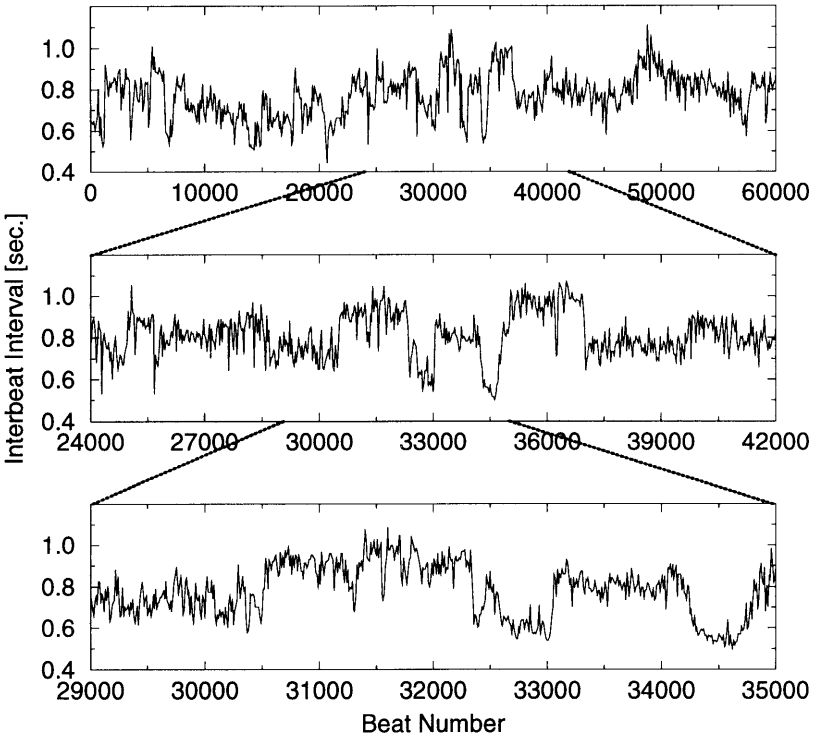



Figure 2-4: Self-similarity property of the fractal-like heart rate.

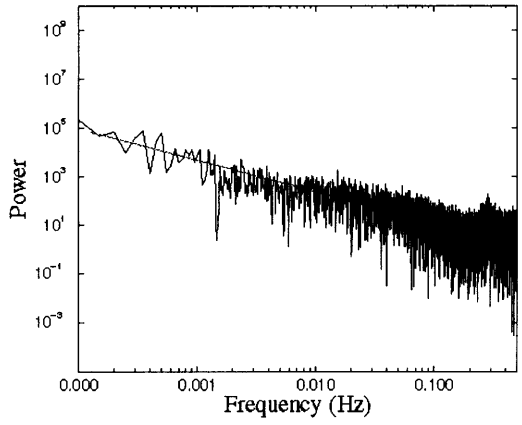



Figure 2-5: $1/f$ power spectrum of healthy heart rate.

reveal a complicated "attractor," rather than a periodic or fixed-point attractor seen in linear or regular processes.

What are the physiological mechanisms which induce this complex HRV? Heart rate control is influenced by the interaction of several central nervous system oscillators and control loops. The interaction of these physiologic control systems operating on different time scales may induce irregular time courses that create scale-invariance in the heart rate [9]. Additionally, the feedback loops present in the heart produce delays in response time. Glass and Mackey [12] have emphasized the importance of time delays in chaotic systems. Another proposed mechanism which causes complex behavior in HRV arises from competing neuroautonomic inputs of the parasympathetic and sympathetic branches of the nervous system.

The highly variable nature of the heart rate supports several functional purposes [10]. First, the erratic fluctuations in the healthy heart rate serve as an important mechanism for creating adaptability, or the ability to respond to unpredictable and changing perturbations. In addition, the absence of a characteristic time scale creates a broad frequency response that prevents the system from becoming mode-locked (locked to a single frequency). Finally, the long-range (fractal) correlations found in healthy patients indicate an overall principle for structuring highly complex nonlinear processes that generate fluctuations over a wide range of time scales [9].

Not surprisingly, a number of pathologies are characterized by a breakdown of the complex nonlinear dynamics observed in healthy heart function [6]. This breakdown may be manifest by a loss of complex variability, sometimes with the emergence of highly periodic behavior. For example, spectral analysis of ECG waveforms associated with sudden cardiac death reveals a narrow spectrum, rather than a broad spectrum apparent in chaotic systems. In addition, patients with severe left ventricular failure typically display two patterns: a strong, low frequency oscillation in heart rate (.01-.06 Hz) and reduced beat-to-beat variability. Similar pathologic patterns have also been observed in fetal distress syndrome. In general, emerging regularities indicate a decreased nonlinear complexity in pathologic patients,

There are a number of existing methods which quantify nonlinear dynamics demon-

strating deterministic chaos. Two such measures used to quantify phase-space plots are the correlation dimension (CD), a measure of the complexity of the process, and the Lyapunov exponent (LE), a measure of the predictability of the process [13]. These measures have been applied to cardiac behavior with conflicting results [14]. The reliability of these methods is questionable, since these methods have difficulty distinguishing deterministic chaos from linear correlated stochastic processes [6]. In general, characterizing a process as "chaotic" with complete certainty is a difficult task.

Previous application of chaos-related metrics demonstrates the potential usefulness of using dynamic nonlinear analysis to understand the complexities of the cardiovascular system. However, chaos is a sub-topic of the field of nonlinear dynamics, specifically applicable only to deterministic systems. It is unlikely that the cardiovascular system is purely deterministic since stochastic influences are likely to affect the overall state of the system. While the methods of CD and LE inherently assume that the system is "chaotic," other methods of statistical analysis have investigated more general nonlinearities in HRV. For example, Ivanov et al. [4] have used wavelet-based time series analysis to elucidate the nonlinear phase interactions between the different frequency components of the beat-to-beat interval.

In nonlinear systems (deterministic and stochastic), the Fourier phases of the signal interact in a non-random fashion. Thus, the harmonic components are not statistically uncorrelated. This Fourier phase dependence, a defining feature of the nonlinearity present, cannot be addressed with linear techniques. A technique that has previously been used to describe nonlinear interactions between Fourier phases is called higher order spectral analysis (HOSA). Although HOSA has been applied to a variety of physical systems, its application to HRV is relatively uncharted. By applying HOSA to the heart rate, the nonlinearly induced phase-couplings can be investigated, potentially improving our understanding of the complex processes of the ANS.

## 2.2 Higher Order Spectral Analysis (HOSA)

Since their first application in 1963, higher order statistics have been used to analyze a range of complex systems including telecommunications, sonar, radar, plasma physics, economic series, and biomedicine [15, 16, 17]. These statistics, known as cumulants, and their associated Fourier transforms, known as higher order spectra, extend beyond the second order characteristics of a signal to extract phase information as well as information due to deviations from Gaussianity. For general signal analysis purposes, there are two main motivations for using HOSA [17]:

1) To detect and quantify nonlinear interactions between Fourier components of a time series.

2) To suppress Gaussian noise processes for detection and classification problems.

Before we can explore these applications of HOSA, we must first introduce more rigorous definitions of linearity and nonlinearity, as well as the definitions of cumulants.

### 2.2.1 Gaussian Processes

By definition, a Gaussian, or "linear", process is fully described by its first order characteristics (i.e. the mean) and second order characteristics (i.e. the autocorrelation or power spectrum). The autocorrelation, $c_{2x}(\tau)$ (also known as the two-point cumulant), of a process $x(k)$ of finite length N, is defined as:

$$c_{2x}(\tau) = \langle x(k)x(k+\tau)\rangle - \langle x(k)\rangle^2 \approx \frac{1}{N-\tau}\sum_{k=1}^{N-\tau} x(k)x(k+\tau) - \left(\frac{1}{N}\sum_{k=1}^{N} x(k)\right)^2 \quad (2.1)$$

where $\langle\cdot\rangle$ denotes the expected value operation. For finite length stationary, ergodic processes[3], we approximate the ensemble expected value as a time average operation. This operation can be visualized as the process in Figure 2-6. For a set time lag, $\tau$,

[3]A process, $x(k)$ is ergodic if all its moments can be determined from a single observation. This implies that all time averages of all possible sample realizations equals the ensemble average [16, 17]. In addition, stationarity implies that the statistics of $y(k)$ do not change with time. In our analysis, we will assume all processes are ergodic.

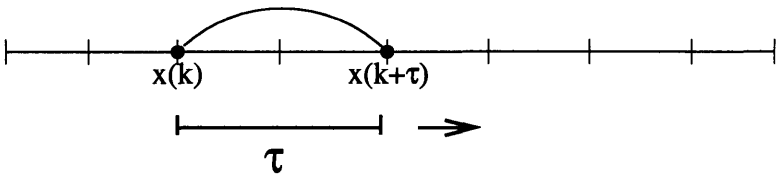



Figure 2-6: Autocorrelation for $x(k)$.

we multiply $x(k)$ with $x(k+\tau)$ (represented as the arc), and average this value for all $k$ (sliding the connected points along the axis). To gain a physical intuition of autocorrelation, consider flipping an unbiased two-sided coin. Each point on the x axis, $k$, corresponds to one flip of the coin. A flip resulting in a head corresponds to a value of 1, whereas a tail corresponds to a value of -1. Because there is an equal probability of producing a 1 and a -1 (a head or a tail), the average of all products $x(k)\cdot x(k+\tau)$ will be zero. Thus, for any time lag $\tau \neq 0$, there is no correlation between $x(k)$ and $x(k+\tau)$: knowledge of the present value of $x(k)$ provides no information for a future value at $x(k+\tau)$.

Now, consider the following recursively defined equation:

$$x(k+1) = \alpha x(k) + w(k) \tag{2.2}$$

where $w(k)$ is a stationary white Gaussian noise sequence. For a given $\tau$, the value of $x(k+\tau)$ depends on a previous value $x(k)$ by definition of recursion. Thus, this relationship between two points, called the autocorrelation, completely describes the process defined by Equation 2.2. More generally, the autocorrelation is sufficient for describing *any* zero-mean Gaussian process. Consequently, for such processes, all higher order moments provide no additional information because they can be expressed in terms of second order characteristics. For a set of zero mean (jointly) Gaussian random variables, $\{x_1, x_2, ...x_L\}$, with autocorrelation[4] $\gamma$, higher order mo-

[4]We represent $\gamma(\tau)$ as $\gamma$ for convenience.

ments can be decomposed as follows [18]:

$$\langle x_1 x_2 x_3 \cdots x_L \rangle = \begin{cases} 0 & \text{L odd} \\ \sum \gamma_{j_1 j_2} \gamma_{j_3 j_4} \gamma_{j_{L-1} j_L} & \text{L even} \end{cases} \tag{2.3}$$

where the summation is over all distinct pairings $\{j_1, j_2\}, ...\{j_{L-1} j_L\}$ of the set of integers $1, 2, ...L$.

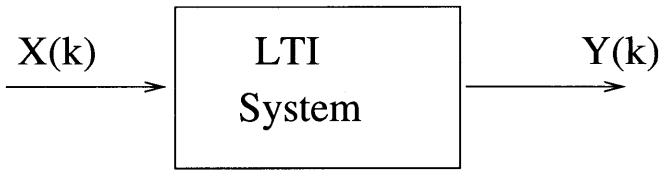
Alternatively, the power spectrum, which is the Fourier transform of the autocorrelation, is an equivalent second-order characteristic which also completely describes a Gaussian process. Formally, the power spectrum, $S_{2x}(f)$, is defined for a process $x(k)$ with corresponding Fourier transform $X(f)$ as[5]

$$S_{2x}(f) = X(f)X^*(f) = |X(f)|^2 \tag{2.4}$$

This classical method evaluates the power of each frequency component, suppressing all phase information [17]. An important characteristic of Gaussian processes is that all frequency components are uncorrelated, i.e., Fourier phases do not interact and are random. It therefore follows that phase-blind statistics completely describe Gaussian processes since they have randomized phases.

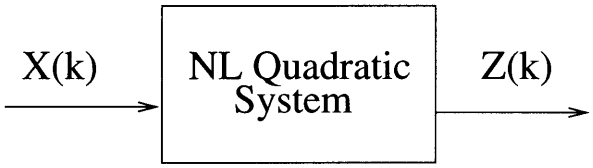
Because Gaussianity is preserved under linear operations, all linear systems with Gaussian inputs generate Gaussian (randomized phase) outputs. For example, Figure 2-7 shows several sinusoids with varying amplitudes, phases, and frequencies as the input, $X(k)$, to a linear time-invariant (LTI) system. Due to linearity, the output, $Y(k)$, consists of the superposition of these sinusoids with the same frequency components scaled and shifted. However, because Fourier components of $X(k)$ are uncorrelated, no coupling occurs between phases, $\phi_m$, of the sinusoids $Y(k)$ [17].

[5] $X^*$ is the complex conjugate of x




$$X(k) = \sum_m A_m \exp\{j(\lambda_m k + \varphi_m)\}$$
$$Y(k) = \sum_m B_m \exp\{j(\lambda_m k + \phi_m)\}$$

Figure 2-7: Linear time invariant (LTI) system with sinusoidal inputs, and uncorrelated harmonic components at the output.




$$X(k) = \sum_m A_m \exp\{j(\lambda_m k + \varphi_m)\}$$
$$Z(k) = \sum_m \sum_n C_m C_n \exp\{j(\lambda_m + \lambda_n)k + (\varphi_m + \varphi_n)\}$$

Figure 2-8: An example of a nonlinear (NL) quadratic system with sinusoidal inputs, and correlated harmonic components at the output.

### 2.2.2 Nonlinear Processes

In a nonlinear process, a (frequently complex) interaction exists between Fourier phases of a time series. Consider the quadratic nonlinear system shown in Figure 2-8. This system clearly induces phase coupling at the output $Z(k)$ [17]. To examine such a process (i.e., nonrandom phase), we cannot rely on classical linear methods. Rather, we turn to a more sophisticated method — higher order statistics. We begin our study of HOSA by introducing the third order cumulant, a higher order statistical technique for examining nonlinear processes.

For random variables $x_1$, $x_2$, and $x_3$, the third order cumulant is defined as:

$$\begin{aligned} cum(x_1, x_2, x_3) \equiv\ & \langle x_1 x_2 x_3 \rangle - \langle x_1 \rangle \langle x_2 x_3 \rangle - \langle x_2 \rangle \langle x_1 x_3 \rangle \\ & - \langle x_3 \rangle \langle x_1 x_2 \rangle + 2 \langle x_1 \rangle \langle x_2 \rangle \langle x_3 \rangle \end{aligned} \tag{2.5}$$

Given a strictly stationary, random process $x(k)$, the third order cumulant of $x(k)$, denoted as $c_{3x}(\tau_1, \tau_2)$, is given by [17]:

$$\begin{aligned} c_{3x}(\tau_1, \tau_2) &= cum\{x(k)x(k+\tau_1)x(k+\tau_2)\} \\ &= \langle x(k)x(k+\tau_1)x(k+\tau_2)\rangle - \langle x(k)\rangle\{\langle x(k)x(k+\tau_1)\rangle + \langle x(k)x(k+\tau_2)\rangle \\ &\quad +\langle x(k+\tau_1)x(k+\tau_2)\rangle\} + 2\langle x(k)\rangle^3 \end{aligned} \tag{2.6}$$

obtained by substituting $x_1 = x(k)$, $x_2 = x(k+\tau_1)$, and $x_3 = x(k+\tau_2)$ into Equation 2.5 and noting that $\langle x(k)\rangle = \langle x(k+\tau_1)\rangle = \langle x(k+\tau_2) = m_x$ by stationarity. Defining the moment function as

$$m_{nx}(\tau_1, \tau_2, ..\tau_{n-1}) \equiv \langle x(k)x(k+\tau_1)...x(k+\tau_{n-1})\rangle \tag{2.7}$$

we can rewrite Equation 2.6 as

$$c_{3x}(\tau_1, \tau_2) = m_{3x}(\tau_1, \tau_2) - \left(m_x(m_{2x}(\tau_1) + m_{2x}(\tau_1) + m_{2x}(\tau_2 - \tau_1)) - 2m_x^3\right) \tag{2.8}$$

Additionally, the third order cumulant can alternatively be written as:

$$c_{3x}(\tau_1, \tau_2) = m_{3x}(\tau_1, \tau_2) - m_{3x}^G(\tau_1, \tau_2) \tag{2.9}$$

where $m_{3x}(\tau_1, \tau_2)$ is the third order moment function of $x(k)$ and $m_{3x}^G(\tau_1, \tau_2)$ is the third order moment function of a Gaussian random process with the same first and second order characteristics as $x(k)$ (which we will denote as $x^G(k)$). We derive this result by first observing that all odd moments of a *zero mean* Gaussian random process are zero (from Equation 2.3). To obtain zero mean random processes, denoted by $\tilde{x}^G(k)$, $\tilde{x}^G(k+\tau_1)$, $\tilde{x}^G(k+\tau_2)$, we must subtract the mean from each process:

$$\tilde{x}^G(k) = x^G(k) - m_x \tag{2.10}$$

$$\tilde{x}^G(k+\tau_1) = x^G(k+\tau_1) - m_x \tag{2.11}$$

$$\tilde{x}^G(k+\tau_2) = x^G(k+\tau_2) - m_x \tag{2.12}$$

Hence,

$$\langle \tilde{x}^G(k)\tilde{x}^G(k+\tau_1)\tilde{x}^G(k+\tau_2)\rangle = 0 \tag{2.13}$$

Substituting Equations 2.10 through 2.12 into Equation 2.13, we obtain

$$\langle \Big(x^G(k) - m_x)(x^G(k+\tau_1) - m_x)(x^G(k+\tau_2) - m_x\Big)\rangle = 0 \tag{2.14}$$

Multiplying all terms, and after some algebra we obtain:

$$m_{3x}^G(\tau_1,\tau_2) = \langle x^G(k)x^G(k+\tau_1)x^G(k+\tau_2)\rangle = m_x(m_{2x}(\tau_1)+m_{2x}(\tau_2)+m_{2x}(\tau_2-\tau_1))-2m_x^3 \tag{2.15}$$

This decomposition process can be visualized with Figure 2-9 (with the arc representing the moment operation, and the blue circle representing the mean). An important result of Equation 2.9 is that for a (randomized phase) Gaussian process, $x(k)$, the third order cumulant is zero [17]:

$$\boxed{\begin{array}{c} m_{3x}(\tau_1,\tau_2) = m_{3x}^G(\tau_1,\tau_2) \\ \longrightarrow c_{3x}(\tau_1,\tau_2) = 0 \end{array}}$$

In addition, we note that for zero mean processes, the third order cumulant and the third order moment operation are equivalent, thus Equation 2.9 simplifies to

$$c_{3x}(\tau_1,\tau_2) = m_{3x}(\tau_1,\tau_2) \tag{2.16}$$

Whereas the autocorrelation examines the relationship between *two* points, the third-order cumulant examines the relationship between combinations of three points within a time series (Figure 2-9), removing Gaussianities. By only preserving non-randomized phase information, the third-order cumulant captures nonlinearities.[6]

[6]Likewise, fourth order cumulants to be discussed in Chapter 5 share this property.

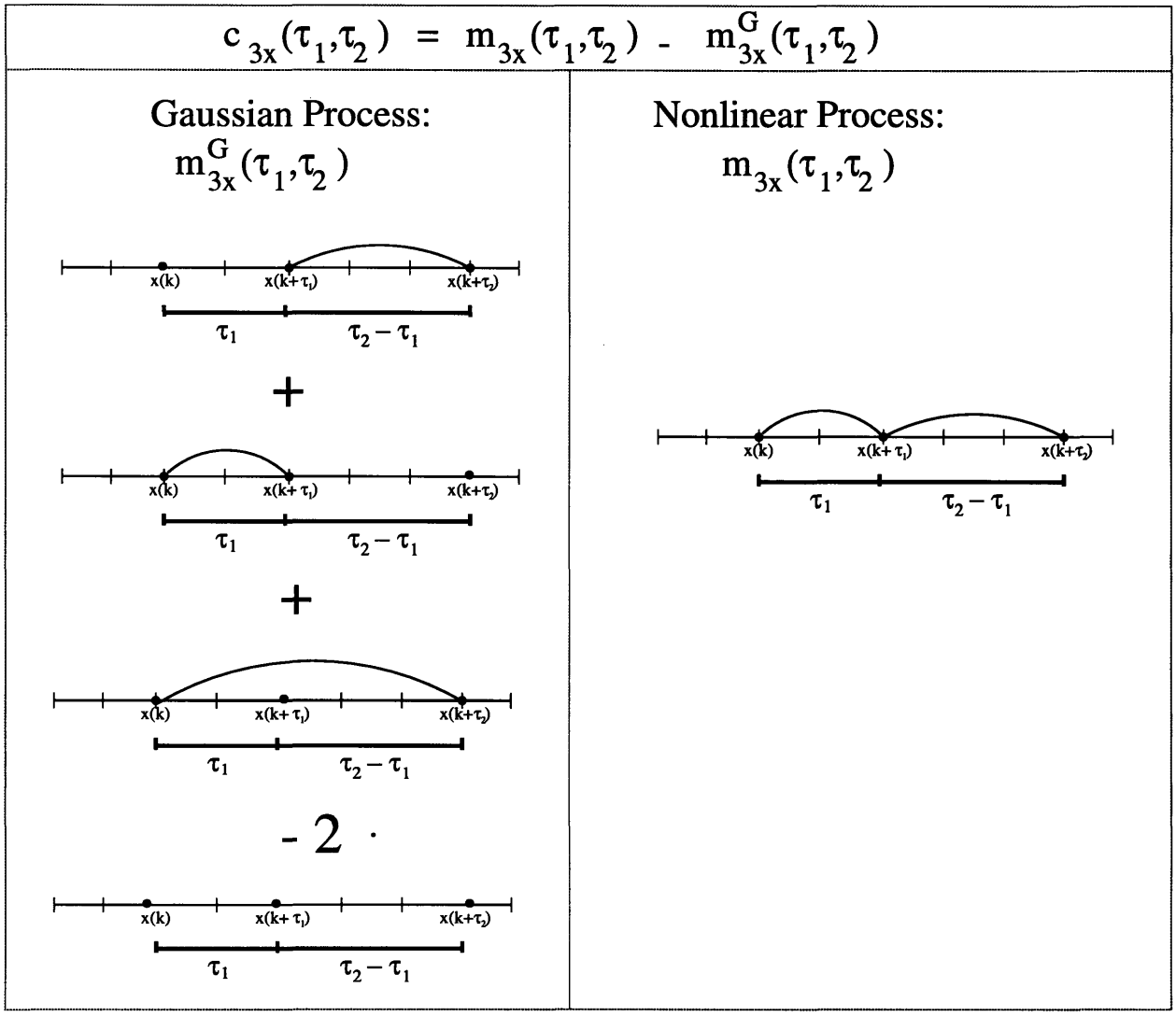



Figure 2-9: $c_{3x}(\tau_1, \tau_2)$ for a process $x(k)$.

### 2.2.3 Quadratic Phase Coupling Example

To further demonstrate why second order characteristics are insufficient for describing nonlinearly induced phase coupling, consider the time series given by

$$s(t) = cos(\omega_1 t + \theta_1) + cos(\omega_2 t + \theta_2) + cos(\omega_3 t + \theta_3) \tag{2.17}$$

where

$$\omega_3 = \omega_1 + \omega_2 \tag{2.18}$$

and $\theta_1, \theta_2$ are randomly distributed in $[0, 2\pi)$. Consider two cases:

Case 1: $\theta_3$ randomly distributed in $[0, 2\pi)$

Case 2: $\theta_3 = \theta_1 + \theta_2$.

If $\theta_3$ is independent of $\theta_1, \theta_2$ (case 1), then a time series consisting of many realizations of $s(t)$ with different sets of random phases will have Gaussian statistics. However, in case 2, the phases $\theta_1$ and $\theta_2$ of $s(t)$ are "completely" coupled. This phenomenon, known as *quadratic phase coupling (QPC)* [15], results from a quadratic nonlinear system which induces interaction between Fourier components (causing contributions to the power at sum and/or difference frequencies). The power spectrum of $s(t)$, shown in Figure 2-10, conceals phase relations of harmonic components, and therefore fails to discriminate case 1 from case 2. For cases 1 and 2, the third order cumulants can be easily obtained [17]. For case 1,

$$c_{3s}(\tau_1, \tau_2) \equiv 0 \tag{2.19}$$

However, for case 2:

$$\begin{aligned} c_{3s}(\tau_1, \tau_2) = \frac{1}{4}\{&cos(\omega_2\tau_1 + \omega_1\tau_2) + cos(\omega_3\tau_1 - \omega_1\tau_2) + \\ &cos(\omega_1\tau_1 + \omega_2\tau_2) + cos(\omega_3\tau_1 - \omega_2\tau_2) + \\ &cos(\omega_1\tau_1 - \omega_3\tau_2) + cos(\omega_2\tau_1 - \omega_3\tau_2)\} \end{aligned} \tag{2.20}$$

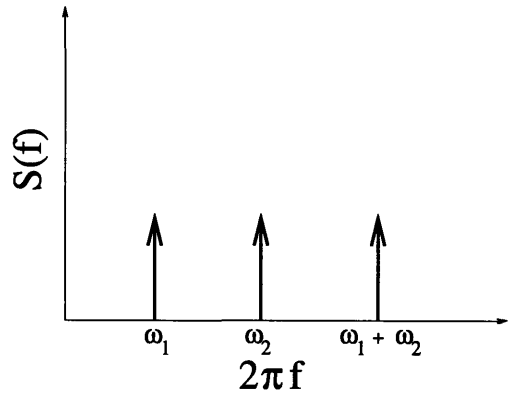



Figure 2-10: Power spectrum, $S_{2x}(f)$, for $s(t)$ with both randomized phases (case 1) and quadratic phase coupling (case 2).

To see this result more intuitively, we turn to the frequency domain and introduce the auto-bispectrum, which is formally defined as the Fourier transform of the third-order cumulant:

$$S_{3x}(f_1, f_2) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c_{3x}(\tau_1, \tau_2) \cdot e^{j2\pi f_1 \tau_1 - j2\pi f_2 \tau_2} d\tau_1 d\tau_2 \tag{2.21}$$

Alternatively, the bispectrum can be defined as

$$S_{3x}(f_1, f_2) = \langle X(f_1) X(f_2) X^*(f_1 + f_2) \rangle \tag{2.22}$$

From Equation 2.22, we can see that if the Fourier components are uncorrelated (as in case 1), the average triple product of the Fourier components is zero, producing a zero bispectrum. However, in case 2, the coupling of the Fourier components results in a nonzero bispectrum with a peak at $(f_1, f_2)$, indicating the oscillation at $f_1 + f_2$ results from a nonlinear interaction between $f_1$ and $f_2$. This result, shown in Figure 2-11, demonstrates the utility of HOSA for detecting quadratic phase coupling in signals.

### 2.2.4 Conclusion

HOSA has previously been applied to a variety of systems. Specifically, these systems contain "quadratic" nonlinearities, which induce phase coupling between triads of

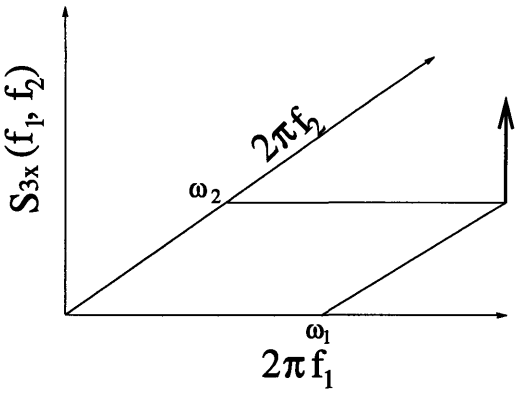



Figure 2-11: Bispectrum displaying peak at $(\omega_1, \omega_2)$ due to phase coupling between $\omega_1$ and $\omega_2$.

Fourier modes (such as discussed in the QPC example) [15]. By applying bispectral analysis, the cross-spectral transfer of energy resulting from these nonlinearities can be detected and quantified. While HOSA has been applied in previous works to study frequency interactions, these interactions are usually quite simple, involving a phase coupling between only two or three frequencies. However, in many nonlinear processes, such as those with $1/f$ power spectra, more complex nonlinear frequency interactions exist. Because of these complex frequency relationships, performing the higher order analysis in the time domain will be just as informative as the frequency domain[7]. Before applying HOSA to highly complex heart rate time series, we begin by applying HOSA to well described processes with calculable third-order cumulants to gain further understanding of the uses and potential limitations of this method.

[7]Generally, higher order spectra rather than cumulants are computed to analyze nonlinearities.

# Chapter 3

# Basic Investigation of HOSA: Mathematical Modeling

To better understand the use of cumulants for the study of nonlinear "real world" systems, we first apply higher order cumulants to fully characterized mathematical models. By applying higher order cumulants to well-known nonlinear processes, we can address several key questions. Specifically, we can examine finite-size effects, nonstationarity effects, and unexpected numerical problems that may arise in computation. Furthermore, we can assess the effectiveness of HOSA for both detecting and quantifying general nonlinearities, and determine the sensitivity of the third order cumulant to power-law nonlinear frequency interactions.

## 3.1 Computation of $c_3(\tau_1, \tau_2)$ for Gaussian Random Processes Applied to a Nonlinear System

To explore HOSA and its limitations, it is most informative to examine nonlinear processes, $y(k)$, arising from the most general nonlinear systems:

$$x(k) \longrightarrow \boxed{\text{Nonlinear System}} \longrightarrow y(k)$$

Brillinger [17] suggests that many nonlinear relationships may be approximated by the system:

$$y(k) = x(k) + \beta x^{n-1}(k) \tag{3.1}$$

for different orders n. For $n = 2$, the system is linear ($y(k) = (1 + \beta)x(k)$). However, $n = 3$ corresponds to the quadratic nonlinear system we will use in our analysis:

$$y(k) = x(k) + \beta x^2(k) \tag{3.2}$$

This quadratic nonlinear system, while simple, describes a wide variety of physical systems.

The input to this quadratic nonlinear system will be a zero-mean, stationary Gaussian random process $x(k)$ described by its autocorrelation, $c_{2x}(\tau)$. From Section 2.2, we recall that higher (third and fourth) order cumulants of a Gaussian random process are zero. However, when this process serves as the input to a nonlinear system, the output is a non-Gaussian random process with non-zero higher order cumulants. In this chapter, we compute the third order cumulant of $y(k)$ for two different inputs $x(k)$. One of these inputs is characterized by an exponentially decaying autocorrelation, the other by long-term correlations.

For *any* Gaussian input, $x(k)$, the third order cumulant of $y(k)$ (described by Equation 3.2), $c_{3y}(\tau_1, \tau_2)$, can be expressed as a combination of higher order moments of $x(k)$. In addition, because $x(k)$ is Gaussian, all higher order moments of $x(k)$ can be decomposed into a combination of first and second order moments as described in Equation 2.3. Because the first moment (the mean) of $x(k)$ is zero, $c_{3y}(\tau_1, \tau_2)$ can be written explicitly in terms of only the second order characteristics of $x(k)$, $m_{2x}(\tau)$. It is important to note that this relationship is *independent* of the choice of $m_{2x}(\tau)$. From Equation 2.1, and from the fact that $x(k)$ is zero mean, we note that the second order moment of $x(k)$, $m_{2x}(\tau)$, defined as:

$$m_{2x}(\tau) = \langle x(k)x(k+\tau) \rangle \tag{3.3}$$

is equivalent to $c_{2x}(\tau)$. Thus, we will use $m_{2x}(\tau)$ (rather than $c_{2x}(\tau)$) to denote the autocorrelation of a zero mean random process.

By definition, the third order cumulant of y(k) is:

$$\begin{aligned} c_{3y}(\tau_1,\tau_2) &= \langle y(k)y(k+\tau_1)y(k+\tau_2)\rangle - \langle y(k)\rangle\{\langle y(k)y(k+\tau_1)\rangle + \langle y(k)y(k+\tau_2)\rangle + \\ &\quad \langle y(k+\tau_1)y(k+\tau_2)\rangle\} + 2\langle y(k)\rangle^3 \qquad (3.4) \\ &= m_{3y}(\tau_1,\tau_2) - m_y\left(m_{2y}(\tau_1) + m_{2y}(\tau_2) + m_{2y}(\tau_2-\tau_1)\right) + 2m_y^3 \qquad (3.5) \end{aligned}$$

To determine $c_{3y}(\tau_1,\tau_2)$, we must first obtain $m_y$, $m_{2y}(\tau)$, and $m_{3y}(\tau_1,\tau_2)$. We begin by finding $m_y$ and $m_{2y}(\tau)$, the first and second moments of $y(k)$. Using Equation 3.2 and the fact that $\langle x(k)\rangle = 0$, we have

$$\begin{aligned} m_y &= \langle x(k) + \beta x^2(k)\rangle \\ &= \langle x(k)\rangle + \beta\langle x^2(k)\rangle \\ &= \beta\langle x(k)\cdot x(k+0)\rangle \\ &= \beta m_{2x}(0) \qquad (3.6) \end{aligned}$$

Similarly,

$$m_{2y}(\tau) = \langle\Big(x(k) + \beta x^2(k)\Big)\Big(x(k+\tau) + \beta x^2(k+\tau)\Big)\rangle \qquad (3.7)$$

Furthermore, we expand the terms in Equation 3.7. Because $x(k)$ is Gaussian, (from Equation 2.3) we eliminate all odd order moments. Thus, we have

$$\begin{aligned} m_{2y}(\tau) &= \langle x(k)x(k+\tau)\rangle + \beta^2\langle x^2(k)x^2(k+\tau)\rangle \\ &= m_{2x}(\tau) + \beta\langle x^2(k)x^2(k+\tau)\rangle \qquad (3.8) \end{aligned}$$

To decompose the fourth order moment $\langle x^2(k)x^2(k+\tau)\rangle$, we consider all possible combinations of second order moments. We graphically represent the second order moment $\langle x_1x_2\rangle$ as shown in Figure 3-1. For any given combination, we multiply all second order moments together. Finally, we sum the products from all combinations. It is helpful to visualize this decomposition process.


Figure 3-1: Depiction of second order moment, $\langle x_1 x_2 \rangle$.

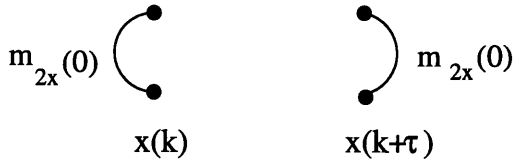



Figure 3-2: Combination 1 for fourth order moment $\langle x^2(k)x^2(k+\tau)\rangle$.

To facilitate this decomposition, we consider each combination separately. One possible combination of the points $x(k), x(k), x(k+\tau)$ and $x(k+\tau)$ is shown in Figure 3-2. The product of the second order moments within this combination is $m_{2x}^2(0)$. Next, we consider the combination shown in Figure 3-3. This combination yields the product $m_{2x}^2(\tau)$. Likewise, the same product results from the combination in Figure 3-4. Summing these combinations, we obtain

$$\langle x^2(k)x^2(k+\tau)\rangle = m_{2x}^2(0) + 2m_{2x}^2(\tau) \tag{3.9}$$

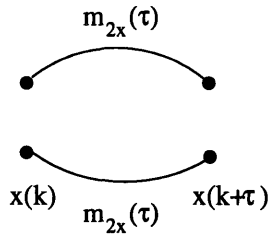



Figure 3-3: Combination 2 for fourth order moment $\langle x^2(k)x^2(k+\tau)\rangle$.

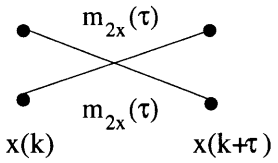



Figure 3-4: Combination 3 for fourth order moment $\langle x^2(k)x^2(k+\tau)\rangle$.

We can now determine the second order moment:

$$m_{2y}(\tau) = m_{2x}(\tau) + \beta^2 m_{2x}^2(0) + 2\beta^2 m_{2x}^2(\tau) \tag{3.10}$$

The next step in calculating $c_{3x}(\tau_1, \tau_2)$ is to determine $m_{3y}(\tau_1, \tau_2)$.

$$m_{3y}(\tau_1, \tau_2) = \langle \Big(x(k) + \beta x^2(k)\Big) \Big(x(k+\tau_1) + \beta x^2(k+\tau_1)\Big) \Big(x(k+\tau_2) + \beta x^2(k+\tau_2)\Big)\rangle \tag{3.11}$$

After expanding terms in Equation 3.11, all odd moments are again removed. All even moment terms are collected, and can further be broken down into a superposition of second order moments.

$$\begin{aligned} m_{3y}(\tau_1, \tau_2) &= \langle \beta x(k)x(k+\tau_1)x^2(k+\tau_2) + \beta x(k)x^2(k+\tau_1)x(k+\tau_2) + \\ &\quad \beta x^2(k)x(k+\tau_1)x(k+\tau_2) + \beta^3 x^2(k)x^2(k+\tau_1)x^2(k+\tau_2)\rangle \end{aligned} \tag{3.12}$$

We determine each term separately:

$$\beta\langle x(k)x(k+\tau_1)x^2(k+\tau_2)\rangle = \beta\left(m_{2x}(\tau_1)m_{2x}(0) + 2m_{2x}(\tau_2)m_{2x}(\tau_2-\tau_1)\right) \tag{3.13}$$

$$\beta\langle x(k)x^2(k+\tau_1)x(k+\tau_2)\rangle = \beta\left(m_{2x}(\tau_2)m_{2x}(0) + 2m_{2x}(\tau_1)m_{2x}(\tau_2-\tau_1)\right) \tag{3.14}$$

$$\beta\langle x^2(k)x(k+\tau_1)x(k+\tau_2)\rangle = \beta\left(m_{2x}(\tau_2-\tau_1)m_{2x}(0) + 2m_{2x}(\tau_1)m_{2x}(\tau_2)\right) \tag{3.15}$$

$$\begin{aligned} \beta^3\langle x^2(k)x^2(k+\tau_1)x^2(k+\tau_2)\rangle &= \beta^3(m_{2x}^3(0) + 2m_{2x}(0)m_{2x}^2(\tau_2-\tau_1) \\ &\quad +2m_{2x}(0)m_{2x}^2(\tau_2) + 2m_{2x}(0)m_{2x}^2(\tau_1) \\ &\quad +8m_{2x}(\tau_1)m_{2x}(\tau_2)m_{2x}(\tau_2-\tau_1)) \end{aligned} \tag{3.16}$$

Combining Equations 3.10 through 3.16, the final expression of $c_{3y}(\tau_1, \tau_2)$ yields:

$$\begin{aligned} c_{3y}(\tau_1, \tau_2) &= 2\beta(m_{2x}(\tau_1)m_{2x}(\tau_2 - \tau_1) + m_{2x}(\tau_2)m_{2x}(\tau_2 - \tau_1) \\ &\quad + m_{2x}(\tau_1)m_{2x}(\tau_2)) + 8\beta^3 m_{2x}(\tau_1)m_{2x}(\tau_2)m_{2x}(\tau_2 - \tau_1) \end{aligned} \quad (3.17)$$

Equation 3.17 demonstrates that both the autocorrelation of a zero mean process $x(k)$ and the parameters of the nonlinear system determine the theoretical third order cumulant $c_{3y}(\tau_1, \tau_2)$[1]. As expected, an absence of nonlinearity, $\beta$, (theoretically) causes $c_{3y}(\tau_1, \tau_2)$ to be equivalently zero.

By calculating the theoretical third-order cumulant for various Gaussian random inputs, we can determine the effect of changing the magnitude of the nonlinearity $\beta$. In addition, we can compare the theoretical results to experimental results obtained from simulations to determine the effects of numerical artifacts or other problems that may arise.

Specifically, we study two mathematical simulations. The input to these simulations consists of a zero mean Gaussian random process, $x(k)$, with a known autocorrelation $m_{2x}(\tau)$ "fed" into the quadratic nonlinear system described in Equation 3.2. The input to the first simulation is a Gaussian random process with a short-term, exponentially decaying autocorrelation. This simple model will allow the initial investigation of the third order cumulant. Second, we study a nonlinear process with a $1/f$ spectrum. This process contains long-term correlations found in many physiologic systems, such as the heart rate. This model both allows us to determine how the "non-white" power spectrum of the input affects the cumulant, and provides the ability to test if "whitening" the output spectrum is necessary.

[1]The general relationship between the (n-1)th order polyspectra of $y(k)$ and the power spectrum of $x(k)$ is given in Brillinger [17].

## 3.2 Simulation 1: Short-term, Exponentially Decaying Autocorrelation

### 3.2.1 Theoretical Calculation of $c_{3y}(\tau_1, \tau_2)$

A zero mean Gaussian random process with a short-term exponentially decaying autocorrelation can be generated with the following recursion:

$$x(k+1) = \alpha x(k) + w(k) \tag{3.18}$$

where $w(k)$ is a zero-mean, unit variance ($\langle w^2(k) \rangle = 1$) stationary white[2] Gaussian noise sequence and $x(0)$ is a Gaussian random variable with zero mean and variance 1 that is independent of (which implies uncorrelated to) $w(k)$. For $x(k)$ to be stationary, $\alpha$ must be strictly less than 1.

To demonstrate that the process generated in Equation 3.18 is zero mean, we take the expectation of both sides:

$$\langle x(k+1) \rangle = \alpha \langle x(k) \rangle \tag{3.19}$$

Since $\langle x(0) \rangle = 0$, Equation 3.19 implies that $x(k)$ is zero mean.

To verify that the autocorrelation of $x(k)$ indeed decays exponentially, we must first determine $m_{2x}(0)$, or $\langle x^2(k) \rangle$ [18]. Squaring both sides of Equation 3.18 and taking expectations, we have

$$\langle x^2(k+1) \rangle = \alpha^2 \langle x^2(k) \rangle + 1 \tag{3.20}$$

Because the process is stationary, we can assume that

$$\langle x^2(k+1) \rangle = \langle x^2(k) \rangle = m_{2x}(0) \tag{3.21}$$

[2] $w(k)$ is white implies that $c_{2w}(\tau) = 0$, for $\tau \neq 0$, or each value of $x(k)$ is independent of all other values.

and solve for $m_{2x}(0)$:

$$m_{2x}(0) = \frac{1}{1-\alpha^2} \tag{3.22}$$

In addition, we observe that $x(k)$ and $w(k)$ are independent, and 3.18 can be rewritten as:

$$x(k+\tau) = \alpha^\tau x(k) + \sum_{k=0}^{\tau-1} \alpha^k w(k) \tag{3.23}$$

Therefore, the autocorrelation can be computed by exploiting the independence of $x(k)$ and $w(k)$[3]:

$$\begin{aligned} m_{2x}(\tau) &= \langle x(k)x(k+\tau)\rangle \\ &= \langle x(k)\left(\alpha^\tau x(k) + \sum_{k=0}^{\tau-1}\alpha^k w(k)\right)\rangle \\ &= \alpha^\tau \langle x^2(k)\rangle = \alpha^\tau m_{2x}(0) \\ &= \alpha^\tau \frac{1}{1-\alpha^2} \end{aligned} \tag{3.24}$$

By substituting Equation 3.24 into Equation 3.17, we determine $c_{3y}(\tau_1, \tau_2)$ (for different values of $\beta$ as $\tau_1$ and $\tau_2$ are varied) with code created using the software package *MATLAB*[4]. To ensure that $x(k)$ is stationary, we let $\alpha = .5$. To determine the effect of changing $\beta$ on the third order cumulant, we find $c_{3y}(\tau_1, \tau_2)$ for both $\beta = .4$ and $\beta = .8$. For all three-dimensional plots, different colors represent different amplitudes. In addition, the three-dimensional image is projected onto the XY-plane and shown as a contour plot. Parts (a) and (c) of Figure 3-5 show the theoretical $c_{3y}(\tau_1, \tau_2)$ for $\beta = .4$ and $\beta = .8$ respectively as $\tau_1$ and $\tau_2$ are varied from -50 to 50. For both values of $\beta$, we observe that the third order cumulant, $c_{3y}(\tau_1, \tau_2)$, is nonzero for only a few time lags centered around the origin. In addition, by increasing the nonlinearity $\beta$, the qualitative structure of the third order cumulant *does not change.* The increased nonlinearity in the system merely increases the magnitude of the peak present in $c_{3y}(\tau_1, \tau_2)$. By doubling the nonlinearity present, the maximum value of $c_{3y}(\tau_1, \tau_2)$ changes disproportionately from 2.1 to 6.0.

[3]Independence implies that $\langle x(k)w(k)\rangle = 0$.

[4]MATLAB and FORTRAN code is attached in Appendix A.

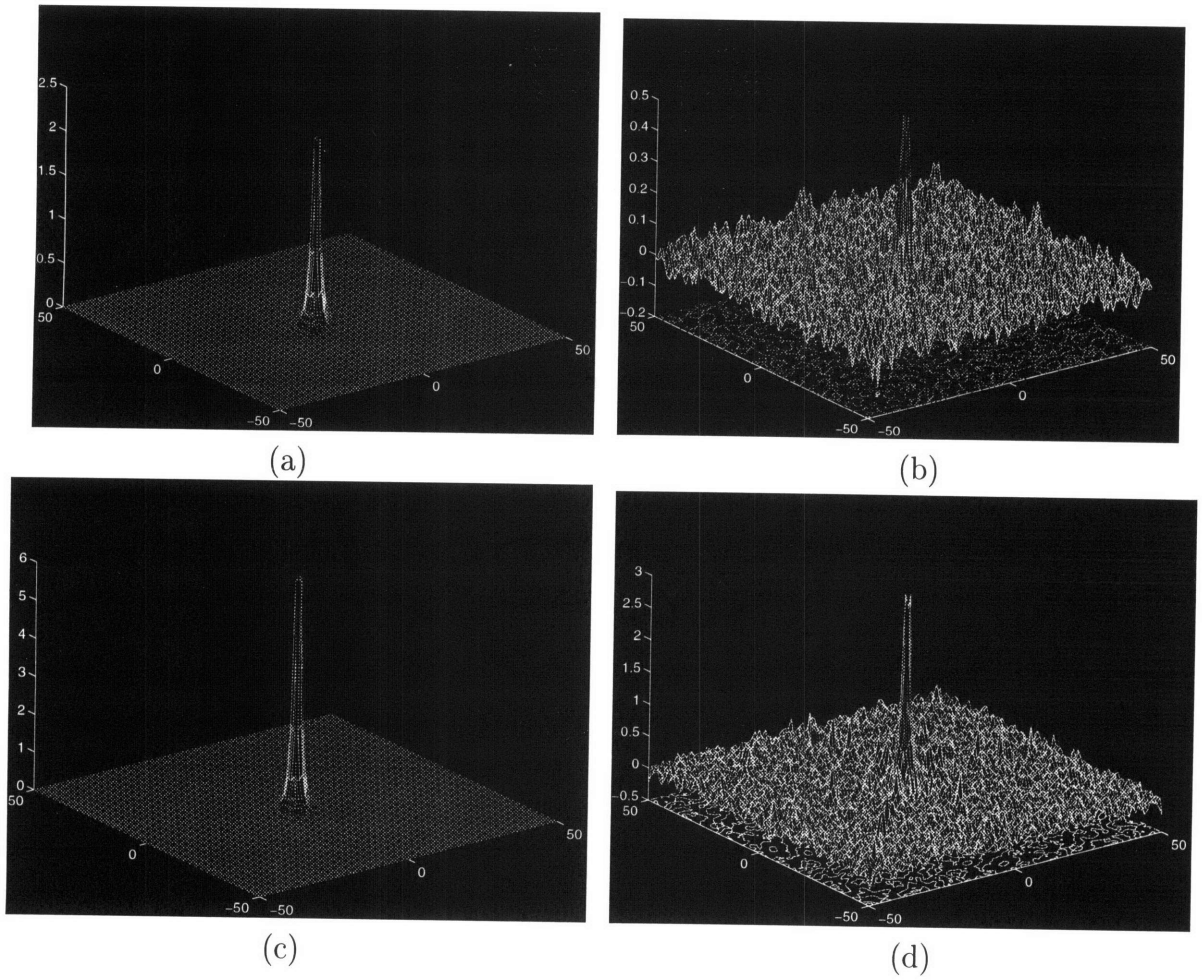



Figure 3-5: Third order cumulant for Gaussian random process with short-term, exponentially decaying autocorrelation filtered through a quadratic nonlinear system. Different colors represent different amplitudes. In addition, the three-dimensional image is projected onto the XY-plane and shown as a contour plot. (a) Theoretical $c_{3y}(\tau_1, \tau_2)$, $\beta = .4$, maximum value of 2.8; (b) Experimental $c_{3y_{nl}}(\tau_1, \tau_2)$, $\beta = .4$, maximum value of .50; (c) Theoretical $c_{3y}(\tau_1, \tau_2)$, $\beta = .8$, maximum value of 6.0; (d) Experimental $c_{3y_{nl}}(\tau_1, \tau_2)$, $\beta = .8$, maximum value of 2.9. The nonlinearity $\beta$ causes a significant nonzero $c_{3y}(\tau_1, \tau)$ for only a few lags centered around the origin. Additionally, the experimental results closely match the theoretical results.

### 3.2.2 Experimental Calculation of $c_{3y}(\tau_1, \tau_2)$

A stochastic realization of $y^{(i)}(k)$ is experimentally obtained (using *MATLAB*) by generating $x(k)$ recursively from Equation 3.18 and inputting it into the nonlinear system defined in Equation 3.2. From $y^{(i)}(k)$, $c_{3y}^{(i)}(\tau_1, \tau_2)$ must be estimated since only a finite length realization of $y^{(i)}(k)$ is numerically realizable. To obtain a consistent estimate, we use the following procedure suggested by [17]. First, to simplify calculations, the average value of $y^{(i)}(k)$ is subtracted from $y^{(i)}(k)$. For a zero mean process $y^{(i)}(k)$, the natural sample estimate of $c_{3y}^{(i)}(\tau_1, \tau_2)$ reduces to

$$c_{3y}^{(i)}(\tau_1, \tau_2) = \langle y^{(i)}(k) y^{(i)}(k+\tau_1) y^{(i)}(k+\tau_2) \rangle \approx \frac{1}{N} \sum_{k=k_1}^{k_2} y^{(i)}(k) y^{(i)}(k+\tau_1) y^{(i)}(k+\tau_2) \tag{3.25}$$

where $k_1 = 1 - min(0, \tau_1, \tau_2)$ and $k_2 = N - max(0, \tau_1, \tau_2)$. The final estimate, $c_{3y}(\tau_1, \tau_2)$, is obtained by averaging over several realizations $i = 1, 2, ...K$:

$$c_{3y}(\tau_1, \tau_2) = \frac{1}{K} \sum_{i=1}^{K} c_{3y}^{(i)}(\tau_1, \tau_2) \tag{3.26}$$

The calculations of all third order cumulants were performed using the program **c3div.f** created in Fortran code (attached in Appendix A). Under certain conditions [16, 17], this estimate will be asymptotically unbiased, with variance approaching zero as $K \to \infty$, or as $N \to \infty$ for $K = 1$. This asymptotic consistency implies that sample estimates of cumulants converge in probability to their true values.

By computing the third order cumulant experimentally, we can accomplish two goals. First, we can determine a suitable control to compare $c_{3y}(\tau_1, \tau_2)$ to confirm that the nonzero cumulant can only be attributed to nonlinear phase interaction rather than numerical artifact or other characteristics of the time series. Second, we can determine the length of the realization, $N$, and the number of realizations, $K$, to "sufficiently" detect the nonlinearity $\beta$.

### Determining a Suitable Control

A characteristic feature of quadratic nonlinear systems is mode coupling, which produces a non-random phase structure in the output, $y(k)$. By applying HOSA (specifically third order cumulants), we hope to detect and quantify this nonlinearly induced phase coupling in $y(k)$. However, we must be certain that the resulting third order cumulant, obtained experimentally, arises *only* from nonlinearity, rather than numerical artifact. Thus, our goal is to determine a suitable control on which to perform parallel HOSA analysis to serve as a comparison. If the surrogate data is *linear*, all higher order (third and fourth) cumulants will theoretically be zero. Thus, any nonzero experimentally obtained third order cumulant (of the surrogate data) *only* results from artifact.

By randomizing its Fourier phases, the nonlinearity present in $y(k)$ is essentially removed, while preserving the linear statistics contained in the power spectrum or autocorrelation. Thus, the surrogate data, $y_{rp}(k)$ (where rp denotes randomized phases), is obtained by Fourier transforming the original time series $y(k)$, preserving the magnitude of the Fourier transform while randomizing the phases[5], and performing an inverse Fourier Transform[6]:

| $y(k) \Longleftrightarrow \lvert Y(f)\rvert e^{j\angle Y(f)}$ |
|---|
| $y_{rp}(k) \Longleftrightarrow \lvert Y(f)\rvert e^{j\theta_{rand}}$ |

Comparing the zero mean realizations of $y(k)$ and $y_{rp}(k)$ (termed phase randomized surrogate), displayed in parts (a) and (b) (respectively) of Figure 3-6, we note a striking difference in their distributions. Examining the histogram of each realization, we observe that $y(k)$ has an asymmetric distribution (part (c) of Figure 3-6). Not surprisingly, the distribution of $y_{rp}(k)$ (part (d) of Figure 3-6) is more "Gaussian-like".

To ensure that any difference found in $c_{3y_{rp}}(\tau_1, \tau_2)$ and $c_{3y}(\tau_1, \tau_2)$ cannot be attributed to the different shapes of their histograms, we change the distribution of $y_{rp}(k)$ to "match" the distribution of $y(k)$. Thus, we retain the relative temporal structure of $y_{rp}(k)$, while changing its actual values to span the same range as $y(k)$.

[5] We randomly generate $\angle Y(f)$ from a uniform distribution from $[0..2\pi)$.

[6] This procedure has been previously applied to "linearize" nonlinear processes [4, 5].

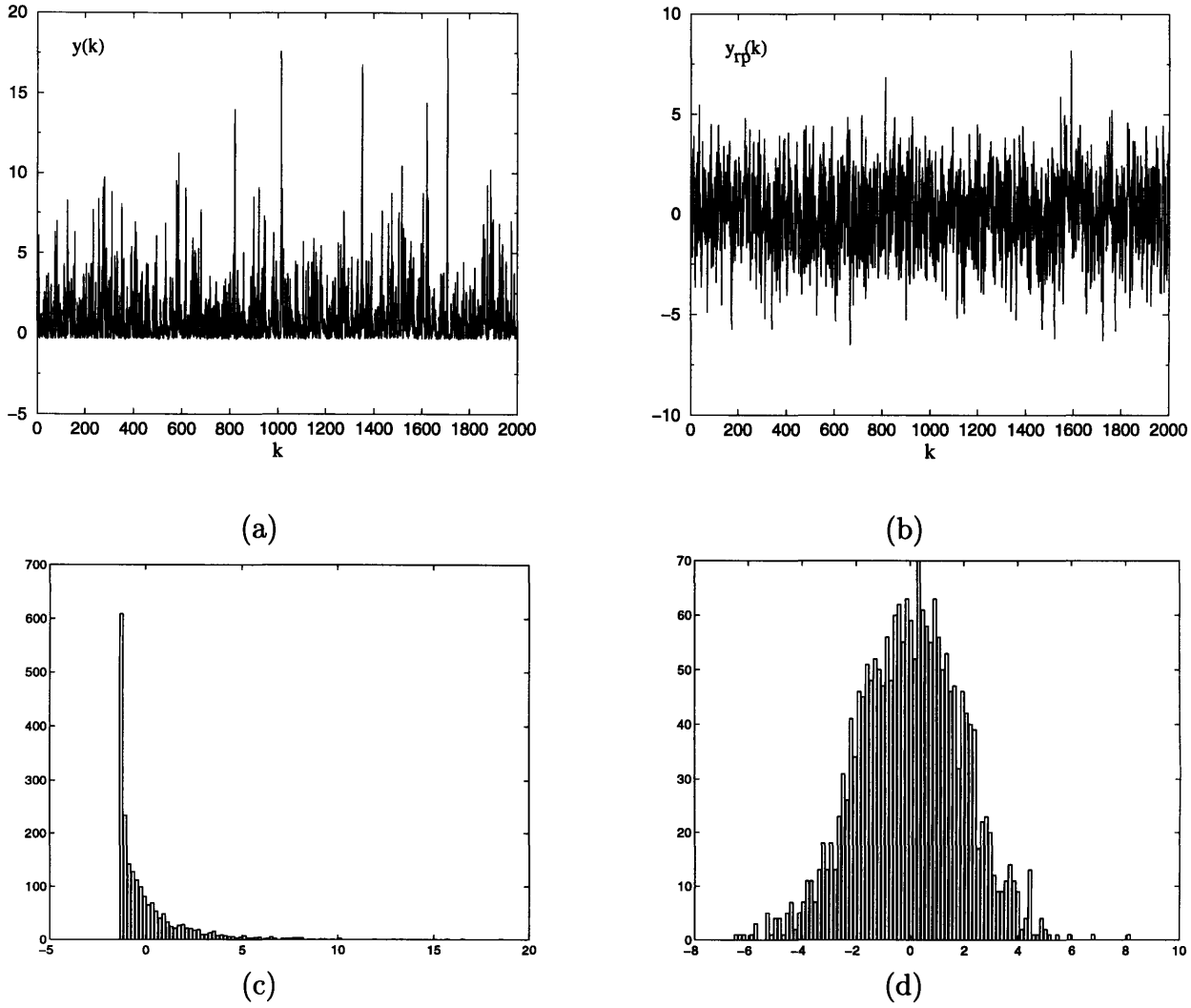
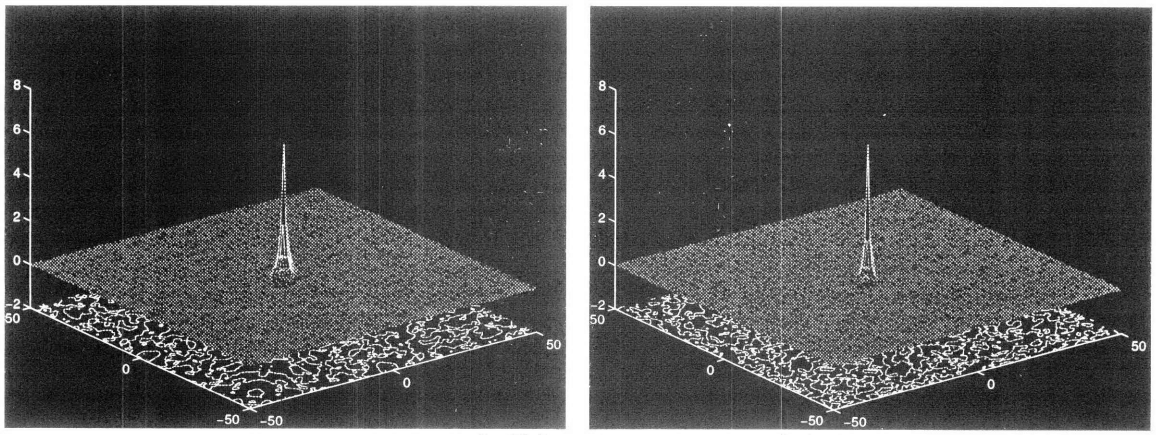



Figure 3-6: A different histogram results from randomizing the Fourier phases of $y(k)$. (a) $y(k) = x(k) + \beta x^2(k)$, $\beta = .4$ displaying an asymmetric distribution shown in (c); (b) randomized phase ("linearized") $y_{rp}(k)$ with Gaussian distribution shown in (d)




(a) (b)

Figure 3-7: Third order cumulant for Gaussian random process with short-term, exponentially decaying autocorrelation filtered through a quadratic nonlinear system, with $\beta = .4$. (a) Experimentally obtained $c_{3y}(\tau_1, \tau_2)$; (b) Experimentally obtained $c_{3y_{rpsd}}(\tau_1, \tau_2)$ calculated from $y_{rpsd}(k)$, where rpsd denotes phase randomized, same histogram of $y(k)$. There is no noticeable difference between parts (a) and (b).

We denote the resultant process as $y_{rpsd}(k)$ (where rpsd is an abbreviation for "randomized phases, same distribution").

To determine if $y_{rpsd}(k)$ is a suitable control, we apply third order cumulant analysis to both $y(k)$ and $y_{rpsd}(k)$. The length of each realization of $y(k)$ and $y_{rpsd}$ was empirically chosen to be $2^{11}$ (or 2048) data points. Using Equation 3.25, we compute the third order cumulant, $c_{3y}^{(i)}(\tau_1, \tau_2)$, for $y^{(i)}(k)$. We obtain the final estimate, $c_{3y}(\tau_1, \tau_2)$, by averaging over 10 realizations (i.e., $K = 10$ in Equation 3.26). The plot of $c_{3y}(\tau_1, \tau_2)$ is displayed in part (a) of Figure 3-7 as $\tau_1$ and $\tau_2$ are varied from -50 to 50 for $\beta = .4$. We perform parallel analysis on $y_{rpsd}(k)$ to compute $c_{3y_{rpsd}}(\tau_1, \tau_2)$, shown in part (b) of Figure 3-7.

Because $y_{rpsd}(k)$ is a linear process, its third order cumulant, $c_{3y_{rpsd}}(\tau_1, \tau_2)$, is theoretically zero. However, from part (b) of Figure 3-7, we observe that (experimentally obtained) $c_{3y_{rpsd}}(\tau_1, \tau_2)$ is *substantially* different from zero. We hypothesize that the resulting nonzero third order cumulant, $c_{3y_{rpsd}}(\tau_1, \tau_2)$, arises from the effect of the asymmetric shape of the histogram since there is no phase information present.

To ensure that the experimentally obtained nonzero cumulant results *only* from nonlinear phase interactions (and spurious numerical artifacts such as finite size ef-

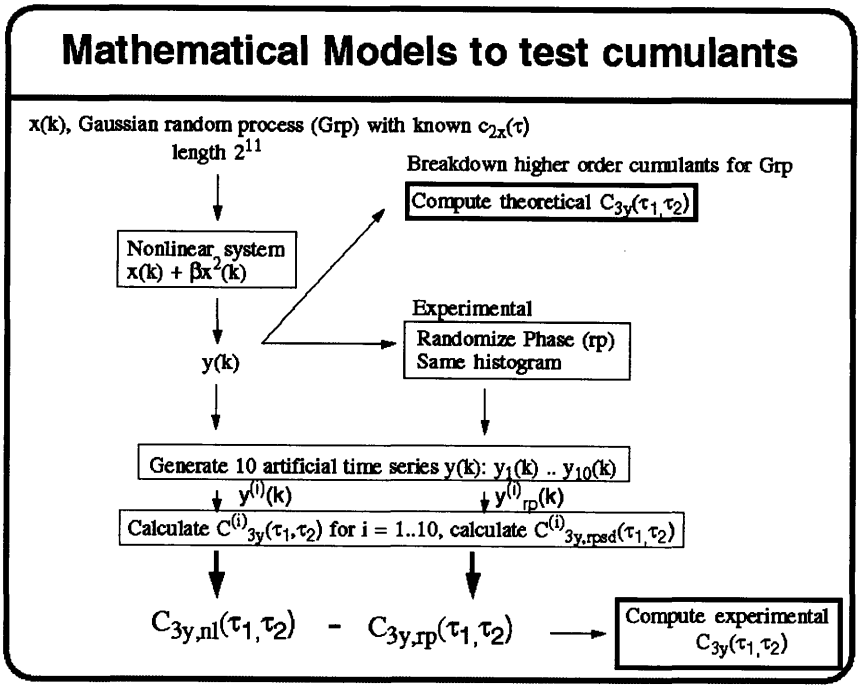



Figure 3-8: Summary of experimental procedure for determining $c_{3y}(\tau_1, \tau_2)$.

fects), we must remove the strong effect of the histogram on HOSA. To accomplish this, we calculate a new statistic:

$$c_{3y_{nl}}(\tau_1, \tau_2) = c_{3y}(\tau_1, \tau_2) - c_{3y_{rpsd}}(\tau_1, \tau_2) \tag{3.27}$$

$c_{3y_{nl}}(\tau_1, \tau_2)$ is therefore *only* influenced by nonlinear phase information present in $y(k)$. Figure 3-8 shows a diagram summarizing the experimental procedure for determining $c_{3y_{nl}}(\tau_1, \tau_2)$ in the context of the simulations studied.

The experimentally obtained $c_{3y_{nl}}(\tau_1, \tau_2)$ for both $\beta = .4$ and $\beta = .8$ are shown in parts (b) and (d) of Figure 3-5 respectively as $\tau_1$ and $\tau_2$ are varied from -50 to 50. Because the experimentally obtained third order cumulant is only an *estimate* computed from finite length data, the experimental results are not exactly equivalent to the theoretical results. However, we do observe strong qualitative similarities between the experimental calculations and the theoretical calculations of $c_{3y}(\tau_1, \tau_2)$. In addition, an increase in the nonlinearity, $\beta$, from .4 to .8 produces a corresponding

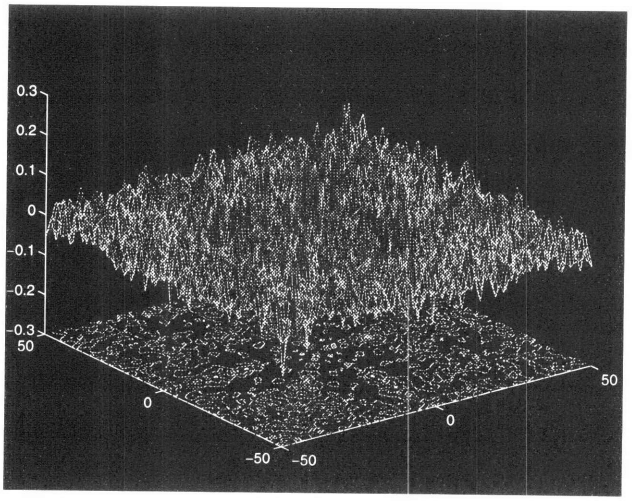



Figure 3-9: The "control", $c_{3y_{control}}(\tau_1, \tau_2)$, demonstrates the third order cumulant when no linearity is present.

change in the magnitude of $c_{3y_{nl}}(\tau_1, \tau_2)$ from .50 to 2.9.

To test whether the observed nonzero $c_{3y_{nl}}(\tau_1, \tau_2)$ arises from the nonlinearity present in $y(k)$, we introduce a new control. Specifically, from $y^{(i)}(k)$, we generate two randomized phase, same distribution processes $y^{(i)}_{rpsd,1}(k)$ and $y^{(i)}_{rpsd,2}(k)$ from each of the ten realizations of $y^{(i)}(k)$. After computing $c_{3y_{rpsd,1}}(\tau_1, \tau_2)$ and $c_{3y_{rpsd,2}}(\tau_1, \tau_2)$, we calculate our control:

$$c_{3y_{control}}(\tau_1, \tau_2) = c_{3y_{rpsd,1}}(\tau_1, \tau_2) - c_{3y_{rpsd,2}}(\tau_1, \tau_2) \tag{3.28}$$

The peak present in $c_{3y_{nl}}(\tau_1, \tau_2)$ (and in the theoretically derived third order cumulant) is absent in $c_{3y_{control}}(\tau_1, \tau_2)$, shown in Figure 3-9. However, because $y^{(i)}(k)$ is finite length, and we are averaging over a finite number of realizations, $c_{3y_{control}}(\tau_1, \tau_2)$ contains numerical artifact, causing it to deviate somewhat from zero. Ultimately, by comparing $c_{3y_{control}}(\tau_1, \tau_2)$ and $c_{3y_{nl}}(\tau_1, \tau_2)$, we can determine whether the nonzero cumulant arises from finite size effects or from actual phase interactions induced from nonlinearities.

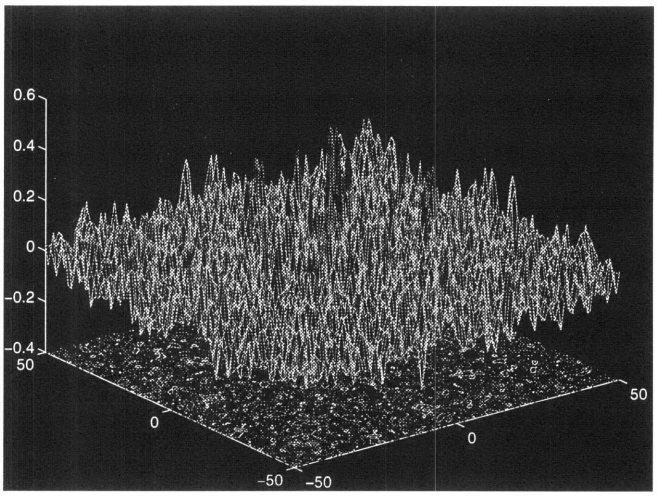



Figure 3-10: $c_{3y_{nl}}(\tau_1, \tau_2)$ obtained by averaging over only one realization ($K = 1$), for $N = 2^{11}$, $\beta = .4$. The noise level obscures the nonlinearity present.

**Other Considerations: Averaging Over Several Realizations; Stationarity Effects**

Averaging over more realizations increases the accuracy of the estimate of $c_{3y_{nl}}(\tau_1, \tau_2)$. With only one realization, a poor estimate is obtained as a result of the high noise level, and the nonlinearity present in $y(k)$ is unobservable (Figure 3-10). An increased number of realizations decreases the noise level present, thus detecting the peak in $c_{3y_{nl}}(\tau_1, \tau_2)$ becomes substantially easier. Empirically, we determine that for a process $y^{(i)}(k)$ of length $N = 2^{11}$, averaging $c^{(i)}_{3y_{nl}}(\tau_1, \tau_2)$ over $K = 10$ realizations is sufficient for detecting nonlinearity for both $\beta = .4$ and $\beta = .8$.

Additionally, from our analysis, we determine the importance of stationarity at the input $x(k)$ (resulting in a stationary process $y(k)$ at the output) in order to produce a consistent estimate, $c_{3y_{nl}}(\tau_1, \tau_2)$. Initially, $\alpha$ was chosen to be .9. Theoretically, this generates a stationary process $x(k)$. However, experimentally, $x(k)$ was found to be nonstationary (as a result of the finite length). This generated many different estimates of $c_{3y_{nl}}(\tau_1, \tau_2)$, none of which resembled the theoretical $c_{3y}(\tau_1, \tau_2)$ computed.

## 3.3 Simulation 2: Long-term Correlations - $1/f$ Nonlinear Process

### 3.3.1 Theoretical Calculation of $c_{3y}(\tau_1, \tau_2)$

$1/f$ processes, or self-similar processes demonstrating long term correlations, are inherent in many physical systems, including biological systems [19]. Many of these systems are quite complex and are frequently nonlinear. However, previous models used to simulate these complex processes only capture the $1/f$ spectrum, while ignoring the nonlinear component present in the Fourier phases [19]. In this section, we determine a method for generating *nonlinear* $1/f$ processes. In addition, we determine a robust signal processing method for detecting nonlinearities in $1/f$ processes for real world applications. Additionally, we apply the third order cumulant to more fully understand nonlinear $1/f$ processes.

Stationarity is a necessary condition for higher order cumulant analysis [17]. For a $1/f^{\delta}$ process to be stationary, the condition $0 < \delta < 1$ is necessary. To generate a stationary nonlinear $1/f^{\delta}$ process, a zero mean Gaussian random process with power spectrum, $S_{2x}(f) = 1/f^{\lambda}$ $(1/2 < \lambda < 1)$, will serve as the input to the quadratic nonlinear system defined in Equation 3.2:

$$S_{2x}(f) = 1/f^{\lambda}(1/2 < \lambda < 1) \quad \longrightarrow \quad \boxed{x(k) + \beta x^2(k)} \quad \longrightarrow \quad S_{2y}(f) \propto 1/f^{\delta}(0 < \delta < 1)$$

To verify that the output spectrum, $S_{2y}(f)$, of the quadratic nonlinear system is proportional to $1/f^{\delta}$, and that $1/2 < \lambda < 1$ is necessary at the input for $y(k)$ to be stationary, we:

1) Determine $c_{2x}(\tau)$.

2) Obtain $c_{2y}(\tau)$ expressed in terms of $c_{2x}(\tau)$.

3) Find $S_{2y}(f)$, the corresponding Fourier Transform of $c_{2y}(\tau)$.

[20] derives that a process $x(k)$ with power spectrum $S_{2x}(f) = 1/f^{\lambda}$ has a corresponding autocorrelation $c_{2x}(\tau)$ (defined for $\tau > 0$):

$$\begin{aligned} c_{2x}(\tau) &= \frac{1}{2}[(\tau+1)^{2H} - 2\tau^{2H} + (\tau-1)^{2H}] \\ \text{where } H &= \frac{1}{2}(\lambda+1) \end{aligned} \tag{3.29}$$

Because the autocorrelation is symmetric about zero, for $\tau < 0$ we have

$$c_{2x}(-\tau) = c_{2x}(\tau) \tag{3.30}$$

To simplify analysis, let $x(k)$ have unit variance:

$$c_{2x}(0) = 1 \tag{3.31}$$

For larger $\tau$, Equation 3.29 asymptotically approaches:

$$\begin{aligned} c_{2x}(\tau) &\approx \alpha\tau^{-(1-\lambda)} \\ \text{where } \alpha &= H(2H-1) \end{aligned} \tag{3.32}$$

Therefore, we can approximate the following Fourier transform pair:

$$\begin{aligned} S_{2x}(f) &\Longleftrightarrow c_{2x}(\tau) \\ 1/f^{\lambda} &\Longleftrightarrow \alpha\tau^{-(1-\lambda)} \end{aligned} \tag{3.33}$$

Next, using Equations 2.1 and 3.10, we determine $c_{2y}(\tau)$, the autocorrelation of $y(k)$:

$$\begin{aligned} c_{2y}(\tau) &= m_{2y}(\tau) - m_y^2 \\ &= m_{2x}(\tau) + 2\beta^2 m_{2x}^2(\tau) \end{aligned} \tag{3.34}$$

Recognizing that $m_{2x}(\tau) = c_{2x}(\tau)$ since $x(k)$ is zero mean, we substitute Equation 3.32

into Equation 3.34 to obtain:

$$c_{2y}(\tau) \approx \alpha\tau^{-(1-\lambda)} + 2\beta^2\alpha^2\tau^{-2(1-\lambda)} \tag{3.35}$$

Using Equation 3.33, we obtain the corresponding power spectrum $S_{2y}(f)$:

$$S_{2y}(f) \approx 1/f^{\lambda} + (2\beta^2\alpha)1/f^{1-2(1-\lambda)} \tag{3.36}$$

For $y(k)$ to be stationary,

$$\begin{aligned} 0 < 1 - 2(1-\lambda) < 1 \\ \Longrightarrow 1/2 < \lambda < 1 \end{aligned}$$

For our analysis, we arbitrarily choose $\lambda$ to be 5/8, (which corresponds to a value of $H = 13/16$).

To find the theoretical third order cumulant, (because $c_{2x}(\tau) = m_{2x}(\tau)$) we substitute Equation 3.29 into Equation 3.17. $c_{3y}(\tau_1, \tau_2)$ is again calculated for $\beta = .4$ and $\beta = .8$ as $\tau_1$ and $\tau_2$ are varied from -50 to 50. Similar to the example in Section 3.2.1, we observe that increasing $\beta$ does not change the qualitative structure of $c_{3y}(\tau_1, \tau_2)$, but rather increases the magnitude proportionately. This effect is demonstrated in parts (a) and (c) of Figure 3-11. The maximum amplitude of $c_{3y}(\tau_1, \tau_2)$ changes from 5.0 to 15.9 as $\beta$ is increased from .4 to .8.

In addition, we observe that the three-dimensional graph of $c_{3y}(\tau_1, \tau_2)$ is more complex, with a nonzero value near the origin as well as along one-dimensional slices. The more complex nonlinear frequency interactions present in $y(k)$ result in a more complex third order cumulant structure compared to the third order cumulant obtained in the example in Section 3.2.1.

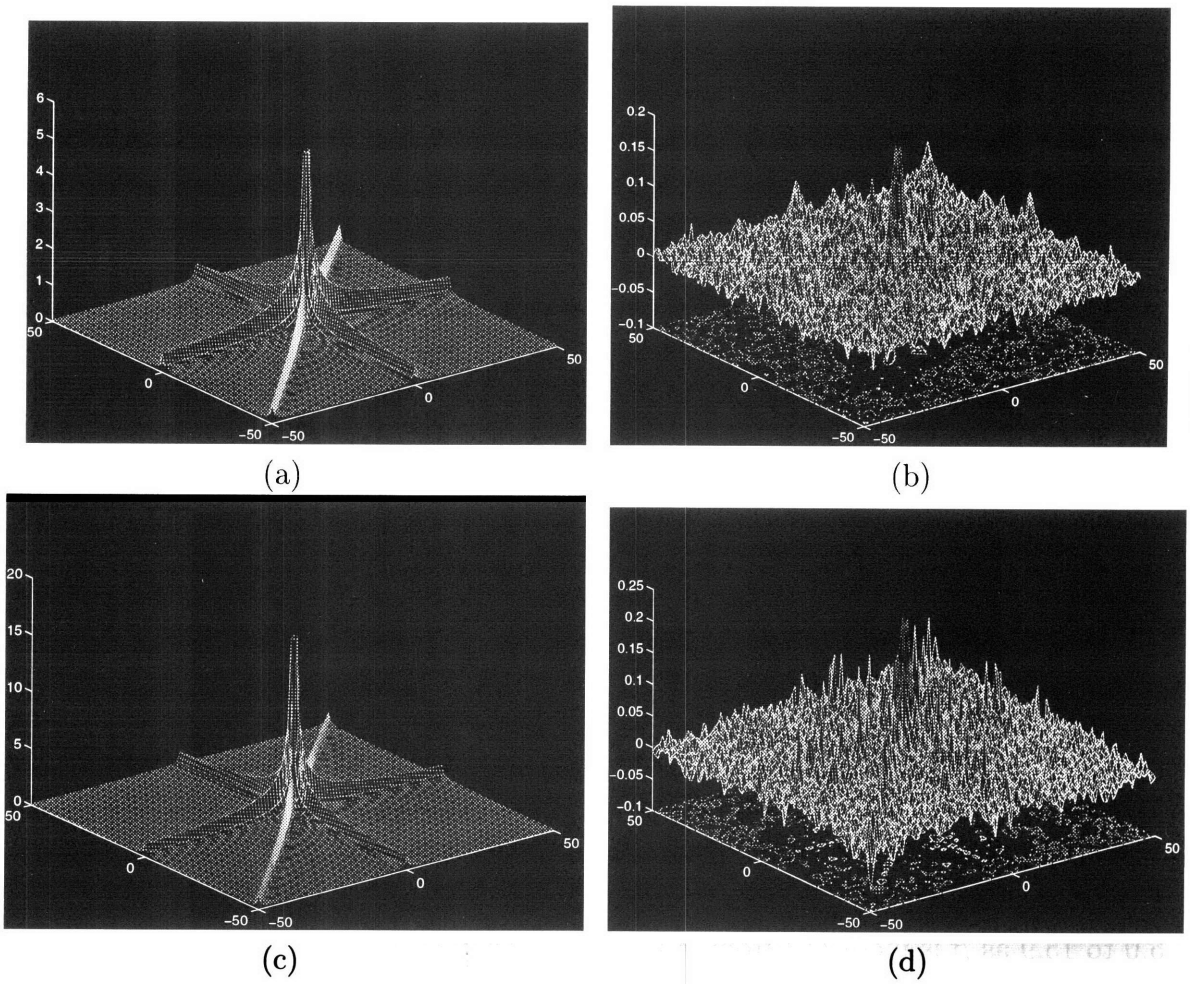



Figure 3-11: Third order cumulant for $1/f$ nonlinear process. (a) Theoretical $c_{3y}(\tau_1, \tau_2)$, $\beta = .4$, maximum value of 5.0 (b) Experimentally obtained $c_{3y_{nl}}(\tau_1, \tau_2)$, $\beta = .4$, maximum value of .17; (c) Theoretical $c_{3y}(\tau_1, \tau_2)$, $\beta = .8$, maximum value of 15.9; (d) Experimentally obtained $c_{3y_{nl}}(\tau_1, \tau_2)$, $\beta = .8$, maximum value of .22. The complex nonlinear frequency interactions present in $y(k)$ result in a complex third order cumulant structure which is somewhat obscured experimentally.

### 3.3.2 Experimental Calculation of $c_{3y}(\tau_1, \tau_2)$

We wish to generate a Gaussian random process, $x(k)$, with autocorrelation $c_{2x}(\tau)$ given in Equation 3.29, and corresponding power spectrum

$$S_{2x}(f) = |X(f)|^2 = 1/f^{\frac{5}{8}} \tag{3.37}$$

This is most easily accomplished by constructing $x(k)$ in the frequency domain.

$$x(k) \Longleftrightarrow |X(f)|e^{j\angle X(f)} \tag{3.38}$$

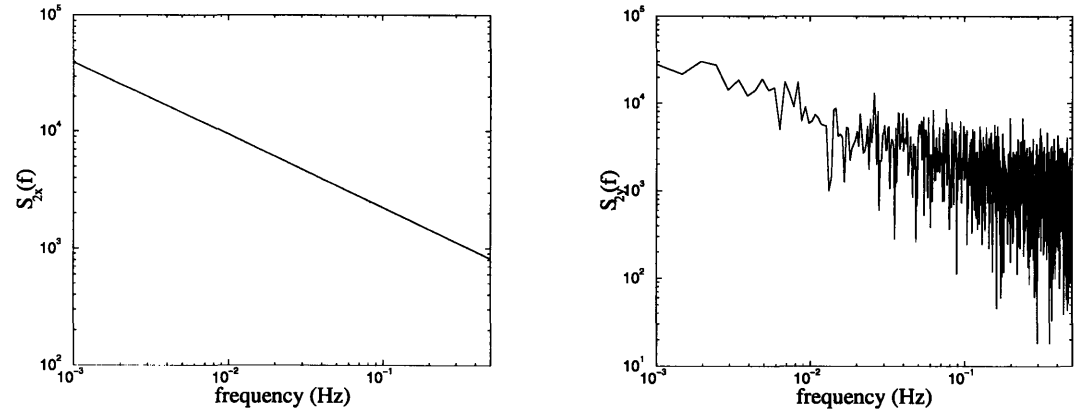
where

$$\angle X(f) = \theta \quad \text{randomly distributed in} \quad [0, 2\pi) \tag{3.39}$$

$$|X(f)| = 1/f^{\frac{5}{16}} \tag{3.40}$$

Because $x(k)$ is a Gaussian random process, it contains no information in its Fourier phases. Thus, we randomly generate $\angle X(f)$ from a uniform distribution from $[0..2\pi)$ and multiply $\angle X(f)$ with $|X(f)|$ given in Equation 3.40 to obtain $X(f)$. To determine $x(k)$, we take the inverse Fourier transform. In addition, we normalize the variance of $x(k)$ to be 1 to be consistent with the theoretical derivation (Equation 3.31). We generate $y(k)$, a stationary nonlinear $1/f$ process with phase interactions, by letting $x(k)$ serve as the input to the quadratic nonlinear system. The power spectrum of both $x(k)$ and $y(k)$ are displayed in parts (a) and (b) of Figure 3-12 respectively.

The same procedure described in Figure 3-8 (also applied in Section 3.2.2) is applied to experimentally obtain the third order cumulant, $c_{3y_{nl}}(\tau_1, \tau_2)$ (i.e., each realization contains $2^{11}$ data points and 10 realizations of $c_{3y}^{(i)}(\tau_1, \tau_2)$ are averaged). Parts (b) and (d) of Figure 3-11 display the experimentally obtained $c_{3y_{nl}}(\tau_1, \tau_2)$ for $\beta = .4$ and $\beta = .8$ respectively. We observe that the nonzero information along the one-dimensional slices present in the theoretical third order cumulant is somewhat




(a) (b)

Figure 3-12: Power spectrum of (a) $1/f^{\frac{5}{8}}$ process, $S_{2x}(f)$; (b) $1/f^{\frac{5}{8}}$ process through nonlinear system (with $\beta = .4$), $S_{2y}(f)$.

obscured in $c_{3y_{nl}}(\tau_1, \tau_2)$ due to the low signal to noise ratio. In addition, the maximum amplitude of $c_{3y}(\tau_1, \tau_2)$ is attenuated.

Experimentally, we find more complex nonlinear frequency interactions are more difficult to detect with the third order cumulant (compared to the example in Section 3.2.1). However, by averaging over more realizations, we can decrease the noise level, and thus accentuate the nonzero cumulant due to nonlinearity. Despite this numerical artifact, the experimental results are comparable to the theoretical results, thus verifying the validity of our procedure to calculate $c_{3y_{nl}}(\tau_1, \tau_2)$.

### Whitening the Spectrum

Higher order cumulants are theoretically affected only by nonlinear phase information and not by linear information contained in the power spectrum. However, experimentally, it is quite possible that the cumulant *is* affected by linear characteristics. Thus, it is possible that whitening the spectrum could remove artifacts due to the power spectrum and enhance the experimental $c_{3y_{nl}}(\tau_1, \tau_2)$. This could be an important consideration in preprocessing a signal before analysis with HOSA. To determine if higher order cumulants are affected by the shape of the histogram, we whiten the spectrum, $S_{2y}(f)$, before applying third order cumulant analysis to $y(k)$. This is achieved by

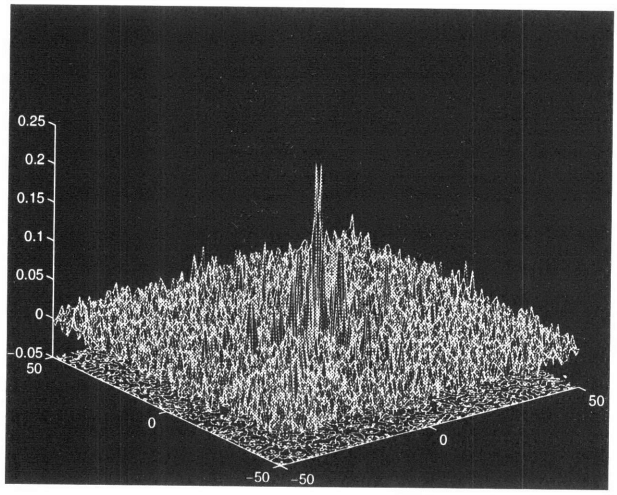



Figure 3-13: Third order cumulant $c_{3y,whitened}(\tau_1, \tau_2)$ for $1/f$ nonlinear process with whitened spectrum.

taking the Fourier Transform, preserving the Fourier phase information, normalizing the Fourier magnitude (to maintain a unit variance for $y(k)$), and taking the inverse Fourier Transform.

$$\boxed{y(k) \Leftrightarrow |Y(f)|e^{j\angle Y(f)}} \longrightarrow \boxed{\begin{array}{c} y_{whitened}(k) \Leftrightarrow ce^{j\angle Y(f)} \\ \text{where } < y^2_{whitened}(k) >= 1 \end{array}}$$

The third order cumulant, $c_{3y_{whitened}}(\tau_1, \tau_2)$, shown in Figure 3-13 for $\beta = .4$, is obtained experimentally by averaging ten realizations of $c^{(i)}_{3y_{whitened}}(\tau_1, \tau_2)$. Comparing this result to Figure 3-11, we observe that the "whitening" procedure does not enhance the cumulant (i.e., $c_{3y_{whitened}}(\tau_1, \tau_2)$ does not more closely resemble the theoretical $c_{3y}(\tau_1, \tau_2)$). Thus, we determine that whitening the spectrum is unnecessary before applying higher order cumulant analysis.

## 3.4 Summary

After applying two Gaussian random processes (characterized by different autocorrelations) to a quadratic nonlinear system, we calculated the third order cumulant, $c_{3y}(\tau_1, \tau_2)$ both theoretically and experimentally. Through theoretical calculations, we discover the following:

- The nonlinearity, $\beta$, present in the quadratic nonlinear system causes a nonzero third order cumulant, $c_{3y}(\tau_1, \tau_2)$. Increasing $\beta$ increases the magnitude of $c_{3y}(\tau_1, \tau_2)$, without changing the overall qualitative structure. *Thus, the nonlinearity present in the system may be quantified by the maximum amplitude of* $c_{3y}(\tau_1, \tau_2)$.

- Different third order cumulants result from different Gaussian inputs to the quadratic nonlinear system. Specifically, the third order cumulant is more complex for processes with more complex nonlinear $(1/f)$ frequency interactions.

By comparing theoretical results to experimental simulations, we arrived at several other conclusions:

- The shape of the histogram of a process $y(k)$ substantially affects the third order cumulant. To remove this undesired effect, we determined a new statistic, $c_{3y_{nl}}(\tau_1, \tau_2)$.

- The control, $c_{3y_{control}}(\tau_1, \tau_2)$, is necessary for ensuring that any nonzero cumulant results from nonlinearity, rather than numerical artifact.

- To produce a consistent estimate of $c_{3y}(\tau_1, \tau_2)$, $y(k)$ must be stationary.

- Averaging over more realizations to obtain $c_{3y_{nl}}(\tau_1, \tau_2)$ accentuates the nonzero structure of the third order cumulant resulting from nonlinear frequency interactions, while attenuating the noise level.

- "Whitening" the process, $y(k)$, provides no advantage for using third order cumulants to investigate nonlinearities.

In summary, this chapter provides guidelines for applying third order cumulants. In addition, it demonstrates the effectiveness of the the third order cumulant in investigating general nonlinearities, as well as $1/f$ nonlinear processes relevant to the analysis of healthy heart rate time series.

# Chapter 4

# Applying HOSA To Detect Nonlinearities In Heart Rate Time Series

In this chapter, we apply HOSA to heart rate time series to investigate nonlinearly induced phase-coupling in both healthy and diseased subjects. While nonlinear analysis has been previously applied to irregular heart rate dynamics, the validity of such approaches is often questioned. Therefore, it is not universally accepted that heart rate variability is a nonlinear process. One of the goals of this study is to use higher order statistics to unambiguously confirm the presence of nonlinearity in healthy heart rate time series.

Additionally, a number of pathologies are postulated to be characterized by a breakdown of the complex nonlinear dynamics observed in healthy heart function (this breakdown results in a loss of complex variability sometimes with the emergence of periodicities) [6]. Thus, we will also apply HOSA to identify underlying differences in nonlinear dynamics between healthy heart rate time series and subjects with congestive heart failure (CHF). Specifically, we apply the third order cumulant, $c_3(\tau_1, \tau_2)$, to probe these nonlinearities in healthy and pathologic heart rate time series. Ultimately, this nonlinear analysis may yield insight into the physiological process of the autonomic nervous system (ANS), in comparison to conventional linear analysis.

## 4.1 Methods

### 4.1.1 Subjects

We analyzed cardiac interbeat data from two different groups of subjects: 8 healthy adults without clinical evidence of heart disease (age range: 29–64 years, mean 44) and 8 adults with severe heart failure[1] (age range: 22–71 years; mean 56). Data from each subject consist of approximately 24 hours of ECG recording. Data from patients with heart failure due to severe left ventricular dysfunction are likely to be particularly informative in analyzing nonlinear dynamics under pathologic conditions since these individuals have abnormalities in both the sympathetic and parasympathetic control mechanisms [23] that regulate beat-to-beat variability.

### 4.1.2 Physiological Time Series Preprocessing

The time series was obtained by plotting the sequential intervals between beat $k$ and beat $k+1$. Parts (a) and (b) of Figure 4-1 display time series for both a healthy subject and a subject with CHF. The healthy beat-to-beat interval, which is modulated to respiratory activity, has been shown to include oscillations with a frequency of approximately .25 Hz (corresponding to one respiratory cycle every four seconds). This feature, known as "respiratory sinus arrhythmia" corresponds to heart rate increasing during inspiration and decreasing during expiration [2]. To remove the strong respiratory signal embedded in the heart rate time series, the heart rate was first averaged over a non-overlapping window of six beats. This procedure is important because we wish to focus our analysis primarily on nonlinear cardiac behavior over relatively long time scales (less than .1 Hz).

An immediate problem facing researchers applying time series analysis to interbeat interval data is that the heartbeat time series is often highly non-stationary. To reduce

[1]ECG recordings of Holter monitor tapes were processed both manually and in a fully automated manner using our computerized beat recognition algorithm (Aristotle). Abnormal beats were deleted from each data set. The deletion has practically no effect on the cumulant analysis since less than 1% of total beats were removed. Patients in the heart failure group were receiving conventional medical therapy prior to receiving an investigational cardiotonic drug; see Ref. [21].

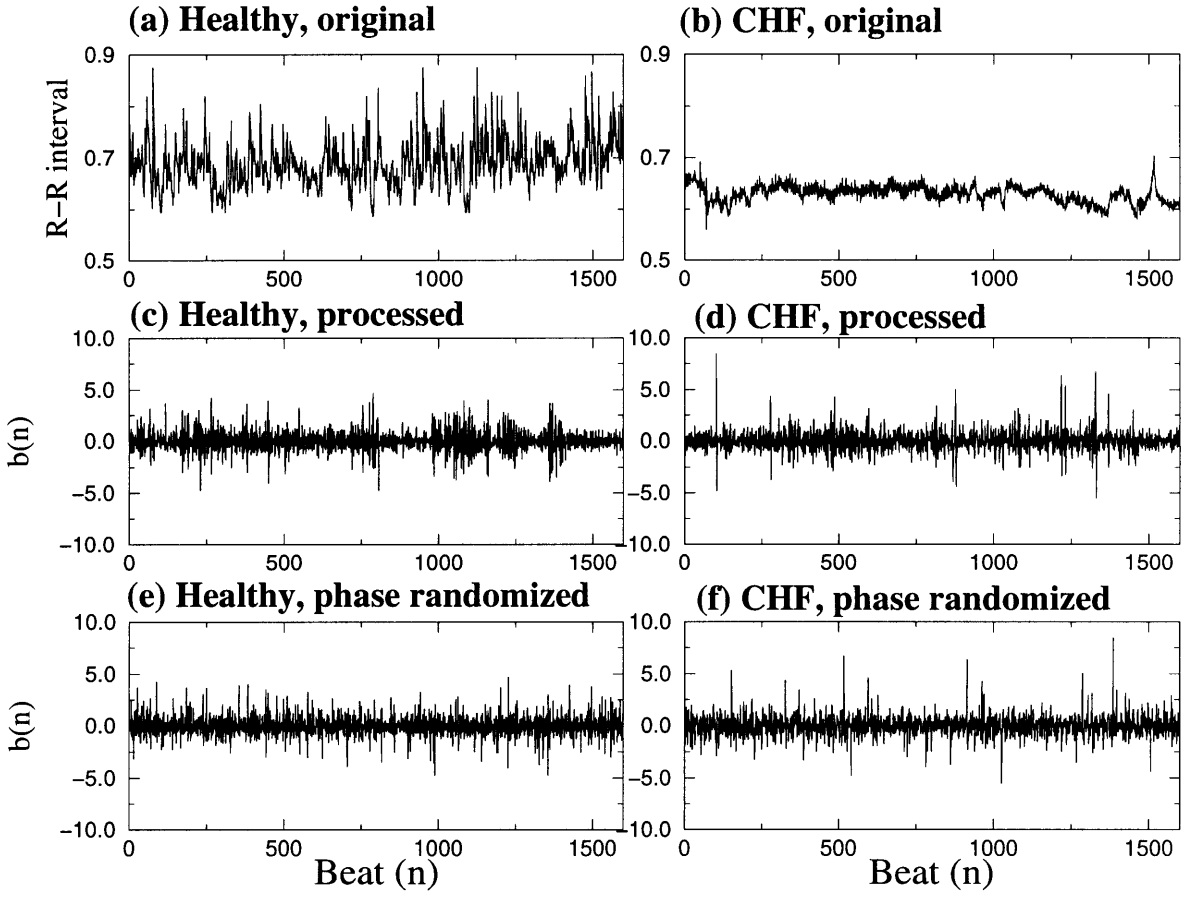



Figure 4-1: Heart rate time series for healthy subject n9723 (left), patient with congestive heart failure (CHF) m9723 (right) before processing (a),(b); after processing (c),(d); phase randomized (e),(f).

the effect of nonstationarity present after averaging (specifically due to a slow drift), we compute the first difference of the signal [24]. Finally, we truncate all time series to have 16,600 consecutive data points to eliminate spurious effects arising from length differences.

### 4.1.3 Third Order Cumulant Analysis

To obtain the third order cumulant, $c_{3b_{nl}}(\tau_1, \tau_2)$, for the 16 heart rate time series, we followed the procedure described in Section 3.2.2. Specifically, for the 16 heart rate time series, we first generate 10 realizations of each time series by segmenting the processed 16,600 length signal into 10 data sets[2]. Each realization is made zero mean by subtracting the average value from each segment. In addition, to maintain uniformity among all data sets, we normalize the standard deviation of each realization to 1. The resulting time series, denoted by $b^{(i)}(k)$, is shown in parts (c) and (d) of Figure 4-1 for a healthy patient and a pathologic patient respectively. For each zero mean realization, Equation 3.25 (with $N = 1660$) is used to calculate $c_{3b}^{(i)}(\tau_1, \tau_2)$. The overall estimate $c_{3b}(\tau_1, \tau_2)$ is obtained using Equation 3.26 by averaging over $K = 10$ realizations.

Section 3.2.2 demonstrates that the third-order cumulant is sensitive to the shape of the histogram of the time series. Thus, it is first necessary to remove the contributions of the shape of the histogram to the third-order cumulant. To remove this effect, each nonlinear realization, $b^{(i)}(k)$, is linearized (by randomizing the Fourier phases)[3], while preserving the original histogram of the nonlinear time series. We denote this surrogate data as $b_{rpsd}^{(i)}$, where "rpsd" is an abbreviation for randomized phase, same distribution. Parts (c) and (d) of Figure 4-1 show $b_{rpsd}^{(i)}(k)$ for both a healthy patient and a pathologic patient, respectively. After computing the third-order cumulant of

[2]In Section 3.2.2 each realization contains 2048 data points. However, because we only have 16,600 points available in the total time series for each subject, the data length of each "realization" is only 1660.

[3]This entails Fourier transforming $b^{(i)}(k)$, preserving the magnitude of the Fourier transform while randomly generating the Fourier phases from a uniform distribution from $[0..2\pi)$. and performing an inverse Fourier Transform.

both the nonlinear time series (denoted $c_{3b}(\tau_1, \tau_2)$) and the randomized phase time series (denoted by $c_{3b_{rpsd}}(\tau_1, \tau_2)$), $c_{3b_{rpsd}}(\tau_1, \tau_2)$ is subtracted from $c_{3b}(\tau_1, \tau_2)$ to remove all effects due only to the shape of the histogram. The resulting statistic, $c_{3b_{nl}}(\tau_1, \tau_2)$, retains the nonlinear phase information present in the time series.

To test whether the observed nonzero $c_{3b_{nl}}(\tau_1, \tau_2)$ arises from the nonlinearity present in the beat-to-beat interval, we determine a control (from the procedure in Section 3.2.2). Specifically, we generate two randomized phase, same histogram processes $b^{(i)}_{rpsd,1}(k)$ and $b^{(i)}_{rpsd,2}(k)$ from each of the ten realizations of $b^{(i)}(k)$. After computing $c_{3b_{rpsd,1}}(\tau_1, \tau_2)$ and $c_{3b_{rpsd,2}}(\tau_1, \tau_2)$, we calculate:

$$c_{3b_{control}}(\tau_1, \tau_2) = c_{3b_{rpsd,1}}(\tau_1, \tau_2) - c_{3b_{rpsd,2}}(\tau_1, \tau_2) \tag{4.1}$$

Ultimately, by comparing $c_{3b_{control}}(\tau_1, \tau_2)$ and $c_{3b_{nl}}(\tau_1, \tau_2)$, we can determine whether the nonzero cumulant arises from finite size effects or from actual phase interactions induced from nonlinearities.

## 4.2 Results

Parts (a) and (b) of Figure 4-2 show $c_{3b_{nl}}(\tau_1, \tau_2)$ versus $\tau_1$ and $\tau_2$ for the healthy and CHF time series respectively. In addition, part (c) displays $c_{3b_{control}}(\tau_1, \tau_2)$ versus $\tau_1$ and $\tau_2$ for the surrogate time series. Any deviation from zero of the third order cumulant, $c_{3b_{control}}(\tau_1, \tau_2)$, (calculated from the surrogate data), results purely from numerical artifact (i.e., finite size effects). This effect produces a (somewhat) uniform "noisy" three dimensional $c_{3b_{control}}(\tau_1, \tau_2)$. In contrast, for both healthy and pathologic time series, we observe that $c_{3b_{nl}}(\tau_1, \tau_2)$ is *substantially* nonzero for a small number of time lags around the origin. Because all nonlinearity is removed in the surrogate data, this difference in third order cumulants indicates the presence of nonlinearity in both healthy and pathologic time series.

Qualitatively, the graphs of $c_{3b_{nl}}(\tau_1, \tau_2)$ for healthy and pathologic patients are very similar. However, we note a substantial difference in the magnitude of the maximum

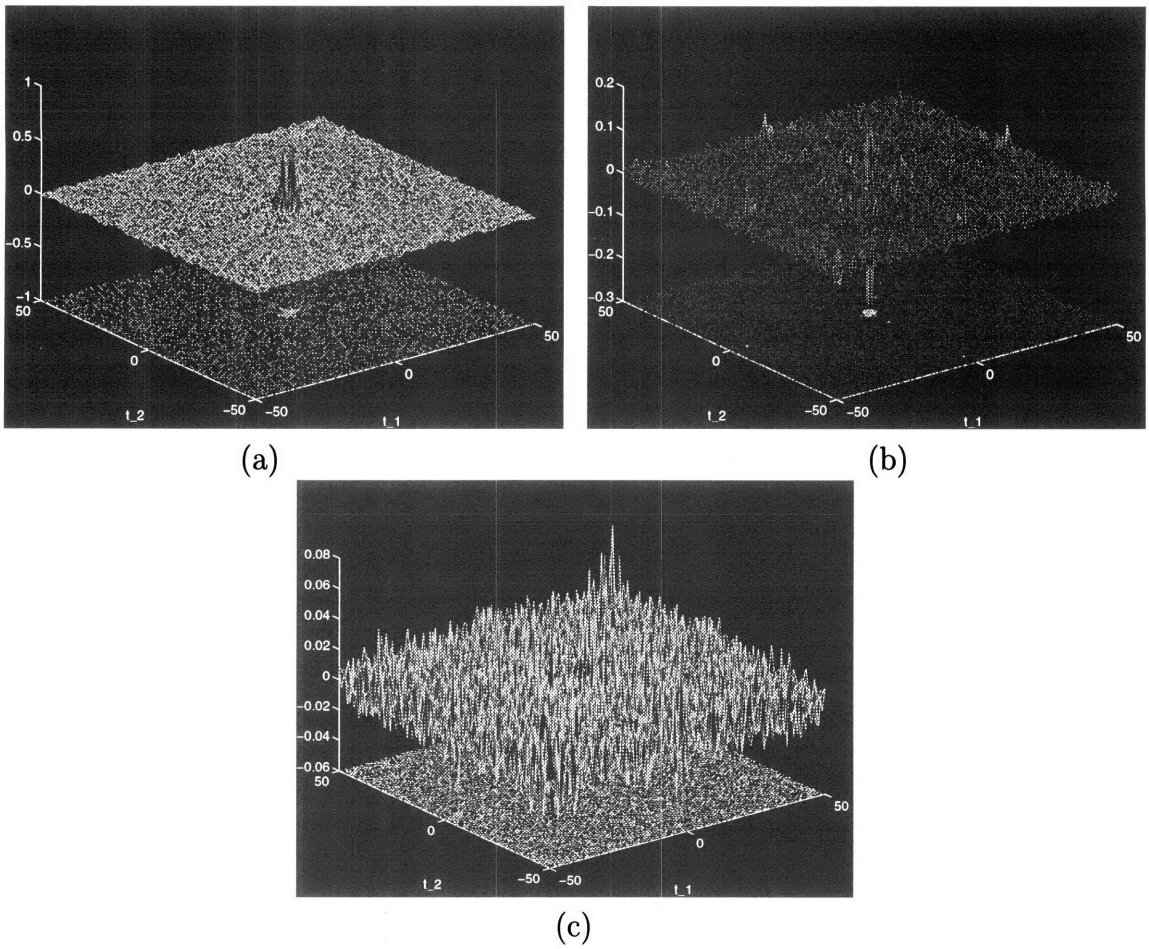



Figure 4-2: $c_{3b_{nl}}(\tau_1, \tau_2)$ for (a) a patient with congestive heart failure (CHF), $\beta_{max} = .62$; and (b) a healthy patient, $\beta_{max} = .16$. Part (c) shows "zero" amplitude $c_{3b_{control}}(\tau_1, \tau_2)$ (obtained from surrogate data), indicating an absence of nonlinearity.

amplitude ("peak") of $c_{3b_{nl}}(\tau_1, \tau_2)$. Chapter 3 suggests that the maximum amplitude of $c_{3b_{nl}}(\tau_1, \tau_2)$, which we will denote as $\beta_{max}$, is a measure of nonlinearity present in the system. For the healthy time series shown in part (a) of Figure 4-1, $\beta_{max} = .16$, while for the pathologic time series, shown in part (b) of Figure 4-1, $\beta_{max} = .62$.

For the group of eight healthy subjects, $\beta_{max} = .22 \pm .10$ (mean $\pm$ standard deviation), while for the subjects with CHF, $\beta_{max} = .72 \pm .33$. The maximum amplitude, $\beta_{max}$, of $c_{3b_{nl}}(\tau_1, \tau_2)$ (for $0 < \tau_1, \tau_2 < 50$) for all 16 data sets is shown in Table 4.1 and displayed (graphically) in Figure 4-3. With the exception of one healthy (n6483) and one pathologic (m9674) subject, $\beta_{max}$ for the two groups are distinctly separated. For all other healthy patients, $\beta_{max} < .3$, while for all other CHF patients, $\beta_{max} > .5$.

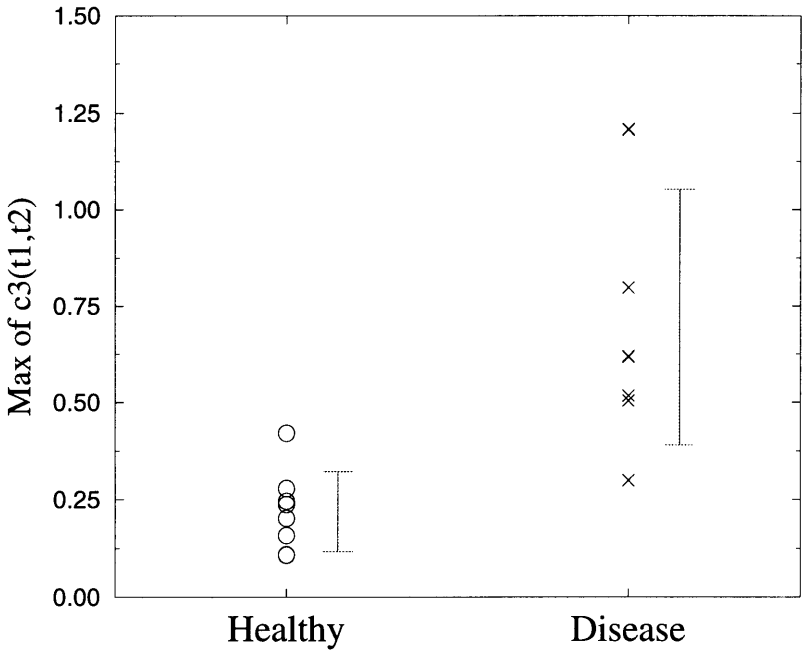



Figure 4-3: $\beta_{max}$, maximum amplitude of $c_{3b_{nl}}(\tau_1, \tau_2)$ (for $0 < \tau_1, \tau_2 < 50$) for 8 healthy heart rate time series and 8 heart rate time series with congestive heart failure, p-value $< .001$ for t-test.

Table 4.1: $\beta_{max}$, maximum amplitude of $c_{3b_{nl}}(\tau_1, \tau_2)$ (for $0 < \tau_1, \tau_2 < 50$) for: 8 healthy heart rate time series (prefixed by "n", e.g. n7453); 8 heart rate time series with congestive heart failure, (prefixed by "m", e.g. m8679); p-value $< .001$.

| Healthy HR Time Series | $\beta_{max}$ | Pathologic HR Time Series | $\beta_{max}$ |
|---|---|---|---|
| n6265 | .16 | m8679 | .62 |
| n6420 | .11 | m8988 | .62 |
| n6483 | .42 | m9435 | .80 |
| n6539 | .24 | m9643 | 1.20 |
| n6786 | .11 | m9674 | .30 |
| n7453 | .20 | m9706 | .52 |
| n8177 | .28 | m9723 | .50 |
| n8184 | .25 | m9837 | 1.21 |
| **Mean $\pm$ SD** | **.22 $\pm$ .10** | **Mean $\pm$ SD** | **.72 $\pm$ .33** |

## 4.3 Discussion

Two important conclusions can be drawn from these results. First, the noticeable difference between $c_{3b_{nl}}(\tau_1, \tau_2)$ (calculated from heart rate time series) and $c_{3b_{control}}(\tau_1, \tau_2)$ (calculated from the surrogate data) indicates the presence of nonlinearity in *all* (healthy and pathologic) heart rate time series studied. By concluding that heart rate time series are nonlinear processes, we are not denying the usefulness of linear analysis. Rather, we are emphasizing that other, more sophisticated techniques could be beneficial in analyzing and interpreting complex cardiac behavior. Additionally, any attempt to model heart rate dynamics should incorporate these phase interactions. Consequently, third order cumulant analysis could be used as a test for the ability of a model to reproduce these results.

Second, $\beta_{max}$ is an indirect measure of the nonlinearity of a process. As demonstrated in Chapter 3, when the nonlinearity, $\beta$, in the nonlinear quadratic system increases, $\beta_{max}$ increases. The pronounced difference in $\beta_{max}$ values for the healthy and CHF subjects suggests an inherent difference in the nonlinear dynamics of the healthy heart rate and the pathologic heart rate. This conclusion reaffirms that an altered physiologic control results from pathology. The more coherent peak, $\beta_{max}$, found in the CHF data demonstrates a distinctive, more intense pattern of mode interactions (mode locking). This result has potential clinical implications in using third order cumulant analysis to detect heart failure physiology. It will be necessary to validate this finding in a larger group of subjects to determine whether this tool can be used for diagnosis of CHF.

# Chapter 5

# Using HOSA To Detect Extremely Low Amplitude Heart Rate Oscillations

## 5.1 Introduction

The healthy heart rate time series, representing the output of a complex cardiovascular control system, fluctuates in an irregular, highly variable fashion. With congestive heart failure (CHF), this complex heart rate variability has been shown to be substantially reduced [9]. Additionally, previous studies have reported emerging intermittent, low frequency oscillations (about every minute) in the beat-to-beat interval of selected CHF patients. These oscillations are associated with periodic (Cheyne-Stokes) breathing, an abnormal breathing phenomenon associated with low cardiac output and circulatory delay[1]. However, it is not clear whether the heart rate oscillations accompanying Cheyne-Stokes physiology are due to entrainment of cardiac activity by respiration or to entrainment of both respiratory and cardiac dynamics to some central oscillatory mechanism [23]. These previously reported oscillations were detected either visually or with classical power spectral methods. In this chapter, we

[1]This respiratory pattern is characterized by a systematic decline in the depth of respiration until breathing stops momentarily, followed by a recommencement of deep breathing [2].

demonstrate that this emerging periodicity may be more ubiquitous than previously believed in CHF subjects. Previously, these oscillations were obscured by a high noise level, thus they could not be detected with conventional linear methods.

To detect these low amplitude, low frequency oscillations "buried" in noise, we apply a HOSA technique, namely the fourth order cumulant. The plot of the (projected) fourth-order cumulant presents a visual representation of low frequency oscillations that are not readily apparent in the power spectrum of the original time series. Because HOSA accentuates these subtle oscillations, it offers an attractive advantage over classical power spectral methods.

## 5.2 Methods

### 5.2.1 Subjects

We analyzed cardiac interbeat data from two different groups of subjects: 12 healthy adults without clinical evidence of heart disease (age range: 29–64 years, mean 44) and 12 adults with severe heart failure (age range: 22–71 years; mean 56). Data from each subject consists of approximately 24 hours of ECG recording.

### 5.2.2 Physiological Time Series Preprocessing

The time series was obtained by plotting the sequential intervals between beat $k$ and beat $k+1$. Because the interbeat interval at night time is typically more stationary, only the nocturnal portion of the time series is utilized (approximately six hours, or twenty thousand data points). In addition, to remove the strong respiratory signal (corresponding to one oscillation every four seconds) occurring from respiratory sinus arrhythmia, the heart rate is averaged over a non-overlapping window of ten beats. Averaging over ten beats allows us to more closely examine lower frequencies, which may be obscured otherwise. For uniformity among data sets, all time series (heart rate and artificial) were made zero mean and unit variance, and were truncated to have 2000 data points. To compare the sinusoidal nature of the time series, we performed

identical analysis on an artificially created sinusoid with additive white (uncorrelated) Gaussian noise (AWGN), $y(k)$, such that

$$y(k) = s(n) + .5w(n) \tag{5.1}$$

where

$$s(n) = cos(2\pi n/11) \tag{5.2}$$

and $w(n)$ is an uncorrelated Gaussian random process with zero mean, and variance 1.

### 5.2.3 Fourth Order Cumulant Analysis

In Chapters 3 and 4, HOSA was applied to both simulated and HRV data to extract phase information from nonlinear processes. In this chapter, we introduce another key motivation for using polyspectral analysis — to remove additive Gaussian noise processes. For Gaussian processes, all higher order cumulants (greater than second order) are theoretically zero by definition [17]. Thus, a time series corrupted with additive Gaussian noise can be analyzed in the cumulant domain to remove the noise present, thus accentuating the intrinsic properties of the signal.

Specifically, we utilize the fourth order cumulant, $c_4(\tau_1, \tau_2, \tau_3)$, which is defined for a zero mean signal $x(k)$ as [16, 17]:

$$\begin{aligned} c_{4x}(\tau_1, \tau_2, \tau_3) &= m_{4x}(\tau_1, \tau_2, \tau_3) - m_{2x}(\tau_1)m_{2x}(\tau_3 - \tau_2) - \\ & \quad m_{2x}(\tau_2)m_{2x}(\tau_3 - \tau_1) - m_{2x}(\tau_3)m_{2x}(\tau_2 - \tau_1) \end{aligned} \tag{5.3}$$

where $m_{nx}(\tau)$ (more commonly known as the moment function) is defined in Equation 2.7. Assuming $x(k)$ is ergodic and finite length N, we approximate all ensemble expected values as time average operations [17]. Thus, we can estimate $m_{4x}(\tau_1, \tau_2, \tau_3)$ as:

$$m_{4x}(\tau_1, \tau_2, \tau_3) = \langle x(k)x(k + \tau_1)x(k + \tau_2)x(k + \tau_3)\rangle$$

$$\approx \frac{1}{N}\sum_{k=k_1}^{k_2} x(k)x(k+\tau_1)x(k+\tau_2)x(k+\tau_3) \tag{5.4}$$

where $k_1 = 1 - min(0, \tau_1, \tau_2, \tau_3)$ and $k_2 = N - max(0, \tau_1, \tau_2, \tau_3)$. Also, we can approximate $m_{2x}(\tau)$ as:

$$m_{2x}(\tau) = \langle x(k)x(k+\tau)\rangle \approx \frac{1}{N-\tau}\sum_{k=1}^{N-\tau} x(k)x(k+\tau) \tag{5.5}$$

where $\langle\cdot\rangle$ denotes the expected value operation. Because the length of each data set is 2000, we do not segment the data set, as we did with third order cumulant analysis in Chapter 4. By substituting Equations 5.4 and 5.5 into Equation 5.3, we obtain our final estimate of $c_{4x}(\tau_1, \tau_2, \tau_3)$[2]. The calculations of all fourth order cumulants were performed using the program **c4.f** created in Fortran code (attached in Appendix A). We compute $c_4(\tau_1, \tau_2, 8)$ for all data sets, setting "lag" $\tau_3$ to eight arbitrarily, and varying the other two lags $\tau_1$ and $\tau_2$ from -50 to 50. To obtain a visual representation of $c_4(\tau_1, \tau_2, 8)$, the three dimensional cumulants were plotted, and their contours were projected onto the XY-plane.

### 5.2.4 Classical Power Spectral Method

Fourier analysis is classically employed to detect strong frequency components in signals. Therefore, for each processed data set, we compute the power spectrum. In order to compare the power spectral method to the cumulant technique to be described, the 2000 length signal is divided into two overlapping signals of length 1024 each. The power spectrum of each segment is computed and averaged. Assuming an averaged eighty beats per minute heart rate, the frequency axis is converted from cycles per every 10 beats to cycles per second (by dividing the frequency axis by 7.5).

[2]Note we only use one realization compared to averaging over several realizations in Chapters 3 and 4.

### 5.2.5 Averaged Fourier Transform of $c_4(\tau_1, \tau_2, \tau_3)$

To compare HOSA to classical power spectral methods, we used the following technique to obtain a new measure, which we term the averaged Fourier transform of the cumulant, $c_4(\tau_1, \tau_2, \tau_3)$. First, we determine $c_4(\tau_1, \tau_2, k_2)$ as we vary $\tau_1$ and $\tau_2$ from -512 to 512, while setting $\tau_3$ to a constant $k_2$. This created a three-dimensional signal of size 1024 x 1024 x 1024. By holding $\tau_2$ constant, we then partitioned the entire signal into individual two-dimensional signals, $c_4(\tau_1, k_1, k_2)$, of length 1024 ($\tau_1 = -512.. + 512$). First, each two-dimensional signal was linearly detrended (to make the signal stationary for FFT processing) by removing the best straight-line fit from the data. Next, the Fourier transform of each two-dimensional signal was computed. This procedure was repeated for all 1024 values of $\tau_2$, and the Fourier transform of the cumulant for all two-dimensional signals was averaged for a specific value $\tau_3$. Finally, this procedure was repeated for eleven different values of $\tau_3$ (which was incremented from -512 to 512 in steps of 100). A single two-dimensional signal, which is the cumulative Fourier transform of $c_4(\tau_1, \tau_2, \tau_3)$ was obtained by averaging the Fourier tranforms resulting from varying $\tau_1$ and $\tau_2$ from -512 to 512, for $\tau_3 = -512, -412, ...412, 512$. The calculation for the averaged Fourier transform of $c_4(\tau_1, \tau_2, \tau_3)$ is performed using the program **c4psav.f** created in Fortran code (attached in Appendix A). This technique was applied to all healthy and diseased data sets, as well as to the simulated sinusoid with additive white Gaussian noise.

The analysis performed on the heart rate is summarized in Figure 5-1.

## 5.3 Sinusoid with Additive White Gaussian Noise (AWGN) Example

An example that will be useful to compare to pathologic heart rate time series is the sinusoid with AWGN mathematically described by Equations 5.1 and 5.2. To compute the fourth order cumulant of $y(k)$, we note that an important property of HOSA is that the cumulant of the sum of statistically independent random processes equals

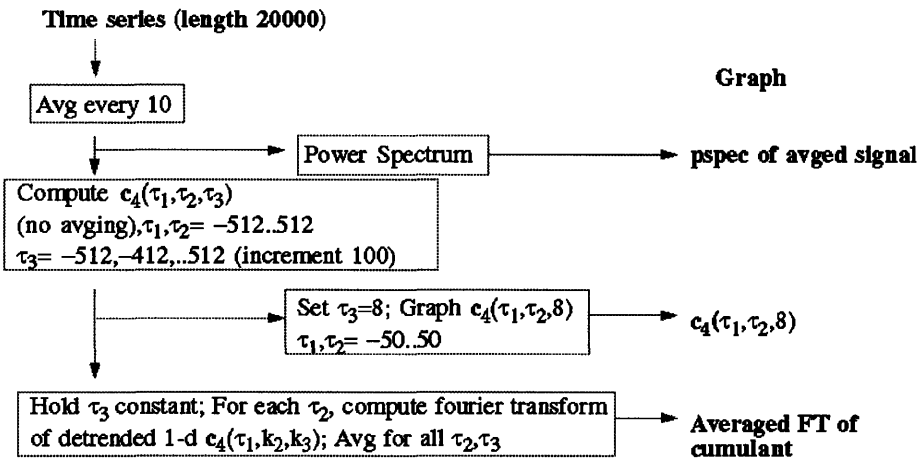



Figure 5-1: Summary of methods performed on the heart rate

the sum of the cumulants of the individual random processes (whereas the same is not true for higher order moments) [16, 17]. Therefore, because $s(k)$ (with corresponding fourth order cumulant $c_{4s}(\tau_1, \tau_2, \tau_3)$) and $w(k)$ (with corresponding $c_{4w}(\tau_1, \tau_2, \tau_3)$) are independent:

$$c_{4y}(\tau_1, \tau_2, \tau_3) = c_{4s}(\tau_1, \tau_2, \tau_3) + c_{4w}(\tau_1, \tau_2, \tau_3) \tag{5.6}$$

From Equation 5.6 and noting that higher order cumulants of Gaussian random processes $(w(k))$ are theoretically zero, we observe that

$$c_{4y}(\tau_1, \tau_2, \tau_3) = c_{4s}(\tau_1, \tau_2, \tau_3) \tag{5.7}$$

Generally, for a sinusoid,

$$s(n) = \alpha cos(\omega n + \phi) \tag{5.8}$$

we can compute the fourth order cumulant, $c_{4y}(\tau_1, \tau_2, \tau_3)$, (given in [16]):

$$c_{4y}(\tau_1, \tau_2, \tau_3) = -\frac{1}{8}\alpha^4[cos(\omega(\tau_1 - \tau_2 - \tau_3)) + cos(\omega(\tau_2 - \tau_3 - \tau_1)) + cos(\omega(\tau_3 - \tau_1 - \tau_2))] \tag{5.9}$$

Thus, we can theoretically compute $c_{4y}(\tau_1, \tau_2, \tau_3)$ (with $s(k)$ given in Equation 5.2) by setting $\alpha = 1$ and $\omega = 2\pi/11$. To compare the sinusoidal simulation to HRV analysis, we generate $y(k)$ (in *MATLAB*) and compute the estimate of the fourth

order cumulant experimentally using Equation 5.3 (by using the approximations given in Equations 5.4 and 5.5). By setting $\tau_3$ to a constant (8), and varying $\tau_1$ and $\tau_2$, we essentially obtain a projection of $c_{4y}(\tau_1, \tau_2, \tau_3)$ onto three dimensions. From Equation 5.9, we see that this projection contains sinusoidal components, resulting in an oscillatory structure in the fourth order cumulant projection (part (c) of Figure 5-4). Therefore, a signal with strong oscillations will have a sinusoidal fourth order cumulant projection, regardless of the noise present. Because the third order cumulant of a sinusoid is theoretically zero, it is necessary to compute the fourth order cumulant to retain the oscillations in our original sinusoid [16].

## 5.4 Results

A representative pathologic heart rate time series (before processing) is shown in part (a) of Figure 5-2 next to that of a healthy patient (part (b) of Figure 5-2), and an artificial sinusoid with AWGN (part (c) of Figure 5-2). Comparing the diseased heart rate time series to the artificial sinusoid, we do not observe strong similarities. Although further HOSA analysis suggests the presence of an oscillation repeating about every 100 beats, this oscillation is buried in noise, and therefore is not unambiguously visible. This oscillation is apparently absent in the healthy heart rate time series.

Parts (a),(b), and (c) of Figure 5-3 show the power spectra $S(f)$ (obtained in Equation 2.4), the square of the Fourier Transform amplitudes of the averaged heart rate time series shown in Figure 5-2. As expected, the sinusoid with additive white Gaussian noise (AWGN) displays a prominent peak at its frequency of 1/11 (which is normalized in Figure 5-3 for comparison). In addition, the healthy heart rate time series has the expected $1/f$ power spectrum. While the spectrum from the pathologic time series has an increased amplitude between the ranges of .075 to .015 Hz, there does not appear to be a single dominant frequency.

Parts (a), (b), and (c) of Figure 5-4 show $c_4(\tau_1, \tau_2, 8)$ of the healthy heart rate, diseased heart rate, and sinusoid with AWGN (respectively), with lag $\tau_3$ set to 8, and lags $\tau_1$ and $\tau_2$ varying from -50 to 50. The fourth order cumulant obtained from the

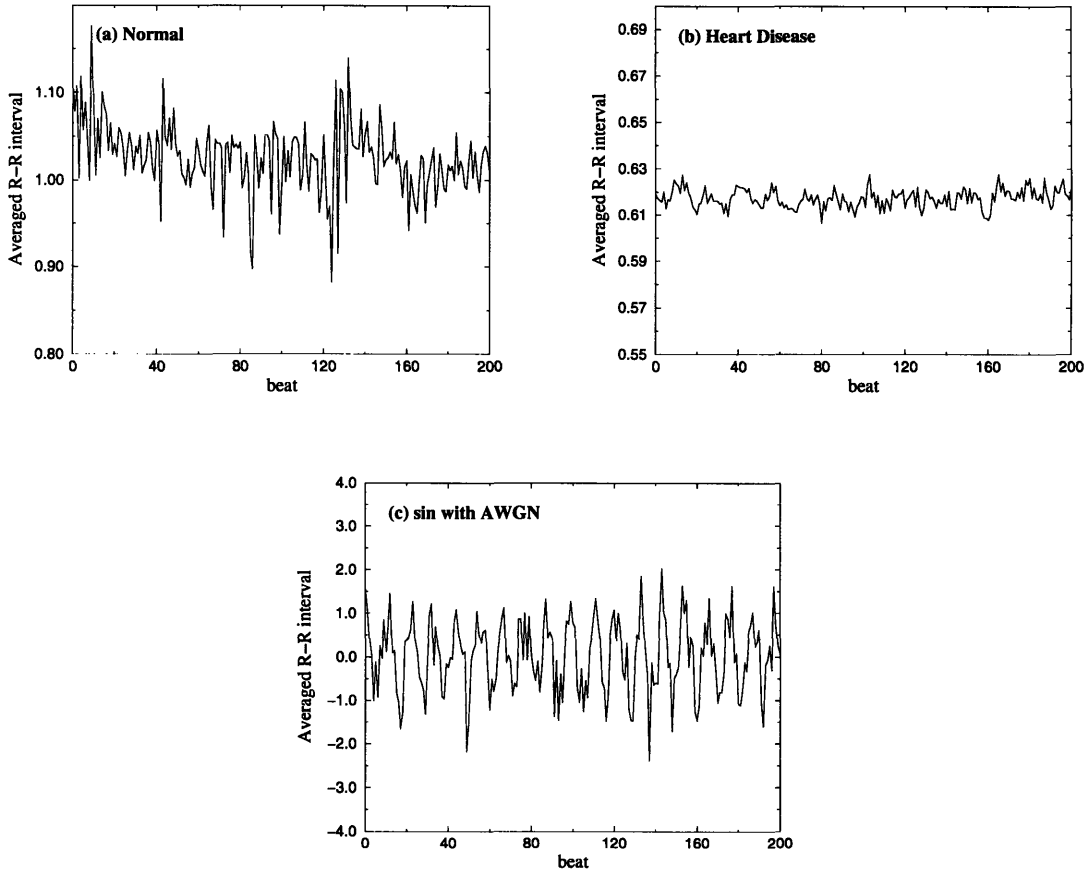



Figure 5-2: The interbeat interval after averaging for (a) a healthy subject and (b) a patient with cardiac disease. Figure (c) shows an artificial sinusoid with additive white Gaussian noise (AWGN) at half the level of the sinusoid. There are no obvious similarities between the data set from the patient with heart disease and the artificial sequence, although further analysis will show that they contain the same characteristic frequency. Oscillations (on the order of every 110 beats) are not visually present in the diseased patient. Also, note the decreased variability of the pathologic heart patient.

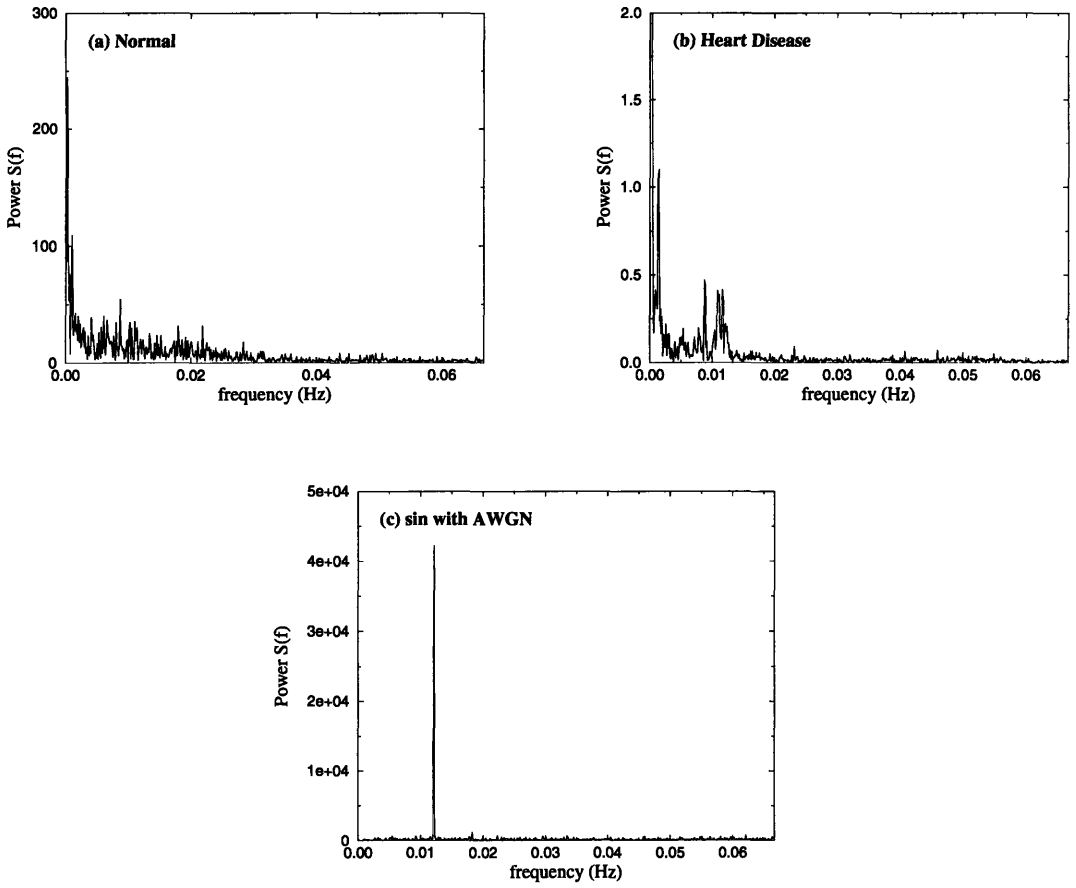



Figure 5-3: The power spectrum S(f) of the interbeat interval after averaging for (a) a healthy subject and (b) a patient with cardiac disease. Figure (c) shows the power spectrum of an artificial sinusoid with additive white Gaussian noise (AWGN) at half the level of the sinusoid. While a few peaks occur in the power spectrum of the pathologic patient, none appears to be the dominant feature of the signal (as compared to the artificial sinusoid).

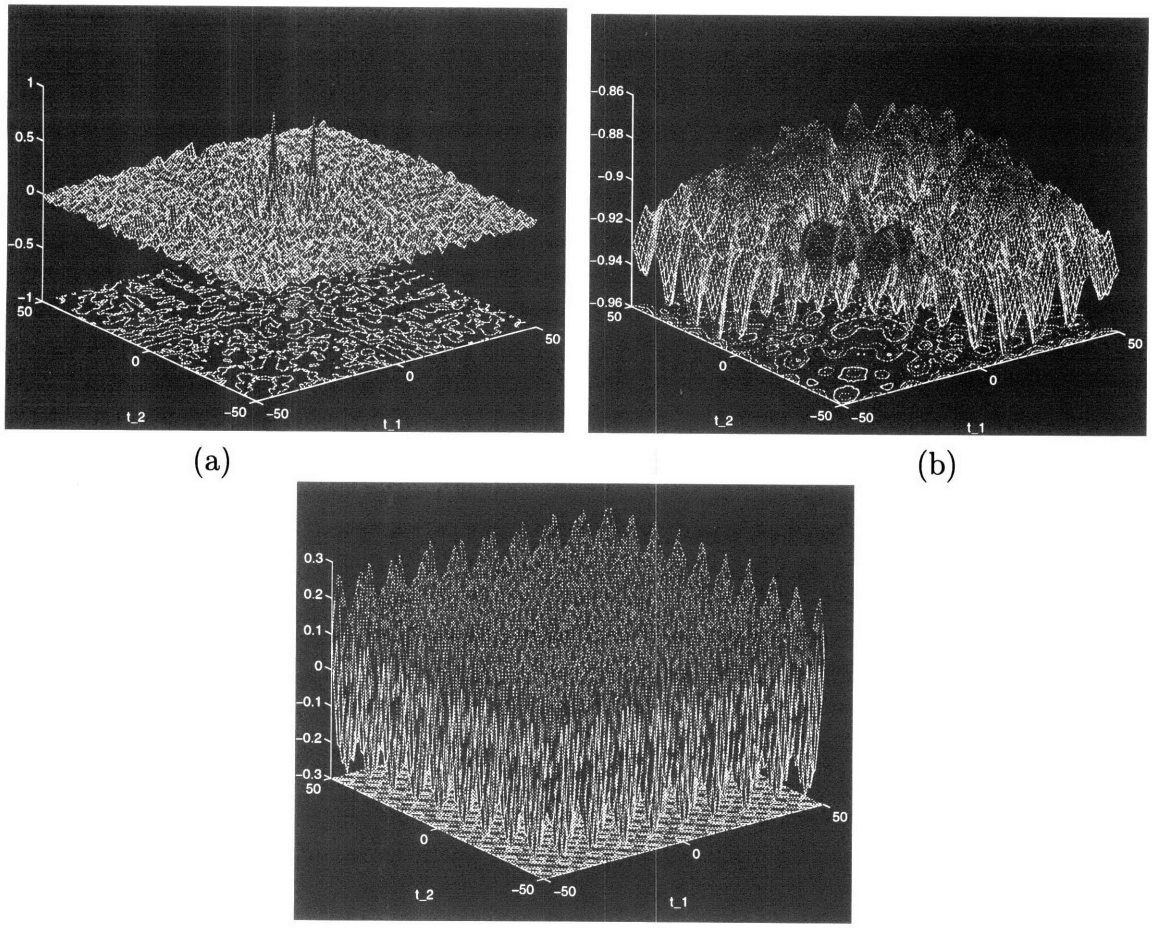



Figure 5-4: The higher order cumulant $c_4(\tau_1, \tau_2, \tau_3)$ (and its projected contour) of the interbeat interval after averaging for (a) a healthy subject, (b) patient with heart disease, and (c) an artificial sinusoid with AWGN. Notice the complexity of the healthy cumulant, compared to the structured oscillation observed in both the artificial signal and the pathologic signal.

healthy heart rate time series is erratic, possessing no easily identifiable structure or regularity. In contrast, $c_4(\tau_1, \tau_2, 8)$ for the diseased patient displays a structured oscillatory "lattice" which appears extrememly sinusoidal particularly when compared to the cumulant of the artficial sinusoid with AWGN. This visual display of a dominant frequency apparent in $c_4(\tau_1, \tau_2, 8)$ prompts further attention to the increased magnitude previously observed in the power spectrum between the frequencies of .005 and .01 Hz.

Finally, the averaged Fourier transform of the detrended cumulant is displayed in parts (c) and (d) of Figure 5-5 (for the healthy subject and diseased patient,

respectively). This method appears to accentuate the true oscillations of the signal, while attenuating the noise present. Comparing this method to the classical power spectral techniques (Figure 5-5 (a) and (b)), we see a great improvement in detecting the oscillation present. Specifically, the peak occurs at a frequency of .012 Hz, or every 82.5 seconds (equivalent to every 110 beats). This low frequency corresponds to the well-described coupling of heart rate to respiration in heart failure, termed Cheyne-Stokes breathing.

The same analysis performed on the other data sets yielded similar results. Ten of the twelve pathologic signals showed evidence of a single dominant frequency. Although these low frequency peaks were present in the power spectrum of the signal, they were not usually prominent. However, the Fourier transform of the cumulant accentuated these peaks. In addition, the two-dimensional graph of $c_4(\tau_1, \tau_2, 8)$ also served as a visual display for oscillations in the ten patients with heart disease. None of the healthy data sets displayed any evidence of this oscillatory behavior.

Analysis was also performed on a time series obtained by transforming the original sequence of R-R intervals into a series of time (in intervals of 5 seconds) versus number of beats. Although the analysis produced similar results, in many cases, the frequency of the oscillation was too low to be visible. This is because the averaged Fourier transform of the fourth order cumulant experimentally results in a low frequency $1/f^2$ behavior, which overpowers very low frequency oscillations. Thus, when detecting oscillations, it was more convenient to look at the beat-to-beat interval.

## 5.5 Discussion

In this study, we apply fourth order cumulants to pathologic beat-to-beat intervals to demonstrate that a dominant frequency is *indeed* present in a majority of the heart failure patients studied, thus suggesting that this oscillation is a prominent characteristic of the signal. The actual mechanisms underlying the periodic pattern reported here are not known. However, the sustained (non-damped) character of the oscillation is important in identifying the underlying dynamics as being nonlinear in character

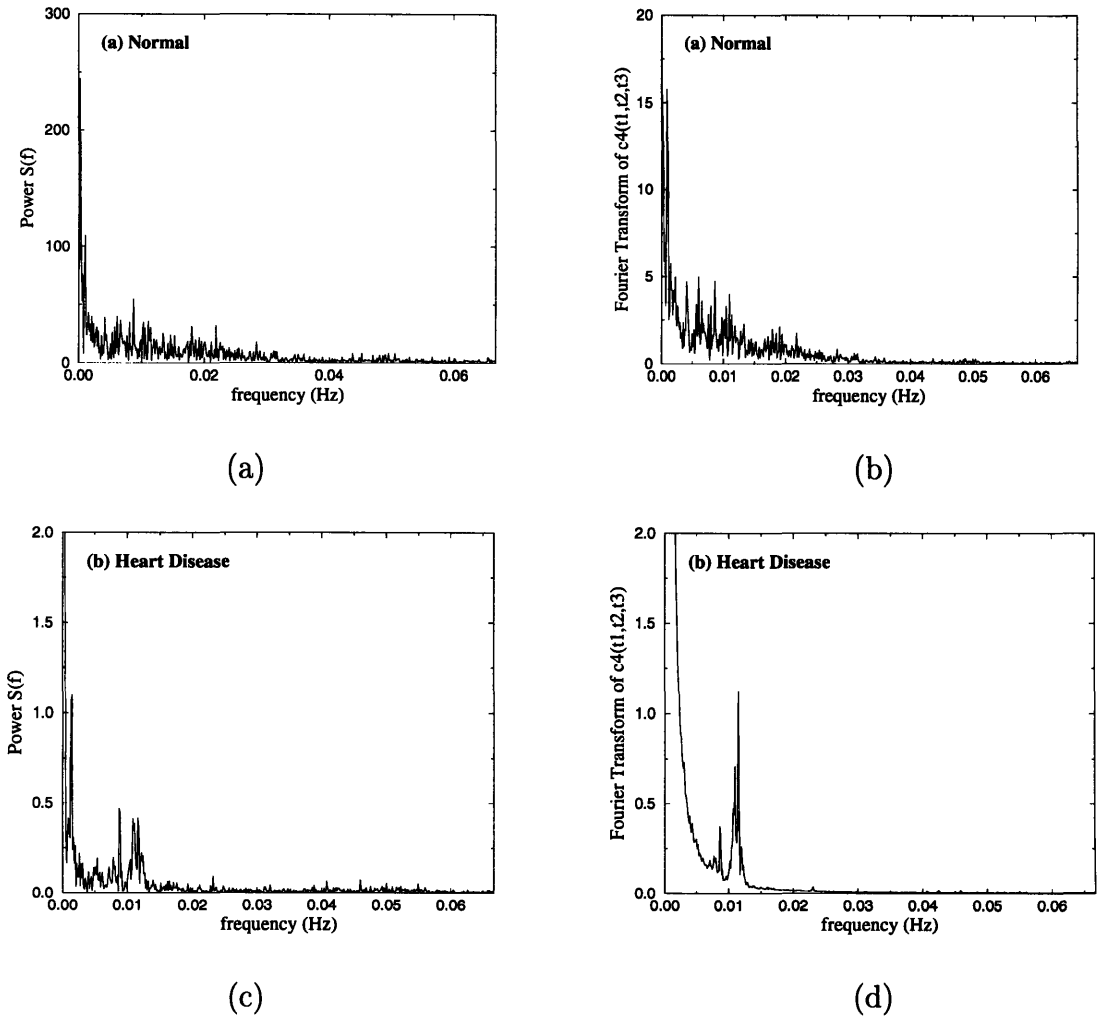



Figure 5-5: The averaged Fourier transform of $c_4(\tau_1, \tau_2, \tau_3)$ of the interbeat interval after averaging for (b) a healthy subject, (d) patient with heart disease. This technique removes the excess noise present, uncovering the dominant features of the signal. We compare this technique to the corresponding power spectrum displayed in parts (a) and (c). The peak from the pathologic patient is unambiguously present in the averaged Fourier transform of $c_4(\tau_1, \tau_2, \tau_3)$ compared to the power spectrum. In the healthy subject, neither the averaged Fourier transform of $c_4(\tau_1, \tau_2, \tau_3)$ nor the power spectrum displays any dominant peaks.

[23]. These results, as well as the results in Chapter 4, suggest that universal nonlinear dynamics in CHF subjects are inherently different from healthy subjects. These dynamics present in pathology may result from the cardiovascular system becoming "mode-locked" (i.e., being entrained) to one specific frequency. Thus, despite many dynamic inputs from the autonomic nervous system, the cardiovascular system can not adaptably respond. The complexity found in healthy beat-to-beat intervals is highly reduced, resulting in an emerging regularity.

The observed periodic heartrate dynamics have potential diagnostic and prognostic importance. To determine the full scope of these clinical applications, further analysis must be applied.

# Chapter 6

# Conclusions

Linear and nonlinear analysis have been widely applied to describe quantitatively the properties of normal and pathologic interbeat fluctuations. In addition, previous studies have suggested a difference in the dynamics of healthy and pathologic interbeat intervals, resulting from an altered autonomic control of the cardiovascular system. My investigations employed HOSA to explore the dynamics of both healthy and pathologic heart rate time series for two major applications: to investigate nonlinearity; and to detect the presence of low frequency, low amplitude oscillations in pathologic time series.

Conventional linear analysis, including power spectral methods, assumes the heart rate is a superposition of statistically uncorrelated harmonic components; therefore, it ignores all potential phase interactions between frequencies. However, in nonlinear deterministic and stochastic systems, the Fourier phases of the signal interact in a non-random fashion. This Fourier phase dependence, an important consequence of the nonlinearity present, cannot be addressed with traditional linear techniques. The importance of phase information has been utilized for several signal processing applications, including image processing. Oppenheim et al. [25] describes how in many contexts, the phase structure contains much of the essential "information" in a signal. Furthermore, one interpretation of the importance of phase is that it preserves the location of "events," or the dynamics of a signal. Chapter 4 uses third order cumulants to extract this important phase information from healthy and pathologic

time series, demonstrating that:

- Both healthy and pathologic beat-to-beat intervals possess nonlinearly induced (non-random) phase information.
- This nonlinearity is inherently different in healthy and pathologic time series.

Chapter 5 utilizes the fourth order cumulant to detect sinusoidal components within noisy pathologic heart rate time series resulting from an emergent periodicity associated with Cheyne-Stokes breathing. These results, as well as the results in Chapter 4, suggest universal nonlinear dynamics in CHF subjects which are intrinsically distinct from healthy subjects.

Specifically, the results from HOSA analysis in Chapters 4 and 5 demonstrate a distinctive type of mode interaction consistent with "mode-locking" present in pathologic interbeat intervals. This-mode locking may result from a breakdown in the complex nonlinear dynamics present in healthy heart rate time series.

These preliminary results emphasize the necessity to construct nonlinear mathematical models to provide testable explanations for physiologic heart rate fluctuations [23]. Finally, a mechanistic explanation for this different nonlinearity (as well as the periodicities) encountered in pathologic and healthy subjects should be provided by a unified model of the cardiovascular system and its autonomic control. To confirm our preliminary results, additional healthy and pathologic data sets should be analyzed with higher order cumulant analysis. Further results will determine the full clinical application of HOSA analysis for diagnostic and prognostic purposes.

# Appendix A

# Computer Code

```
************************************************
/* c3theor.m */
% To determine the theoretical third order
%cumulant c3(t1,t2) % for a zero mean Gaussian
%random process, x(k), with autocorrelation c(t)
%as the input to a quadratic nonlinear system,
%x(k) + b x^2(k)

t2 = -100:1:100;
b = .4;

for t1 = -100:100

  c3theor(:,t1+101) = ...
   (2*b.*(c(t2).*c(t2-t1) + c(t1).*c(t2-t1) + ...
   c(t1).*c(t2)) + 8*(b.^3)*c(t1).*c(t2).*c(t2-t1))';
end

************************************************
/* c.m */
% Short-term, Exponentially Decaying
% Autocorrelation c(t) for zero mean
% Gaussian random process x(k)

function c2t1 = c(t1);
c20 = 1 / (1-.5^2);
c2t1 = c20.*(0.5).^(abs(t1));

************************************************
/* c.m */
% Long-term autocorrelation c(t) for
%zero mean Gaussian random process x(k)

function c2t1 = c(t1);
  t1 = abs(t1);
  gamma = .3750;
  H = 1 - gamma/2;
  b = .4;

  c2x = .5*((t1+1).^(2*H) - 2*((t1).^(2*H))
            + (t1-1).^(2*H));

  if (find(t1==0))
     c2x(find(t1==0))= 1;
  else
  end

c2t1 = c2x + 2*b^2*(c2x.^2);
************************************************
/* create_nlecgrp.m*/
% Generate nonlinear process by inputting
% exponentially correlated gaussian random process
% into a quadratic nonlinear system

b=.4;
a = .5;
k = 11;
corrgrp(1) = randn(1,1);

for l = 1:2^k-1;
      corrgrp(l+1) = a.*corrgrp(l) + randn(1,1);
end

nlcorrgrp = corrgrp + b*(corrgrp.^2);
```

```
rpnlcorrgrp = rphase(nlcorrgrp');
  %control -- randomized phase of above signal
  %randomized phase nonlinear gaussian

rpsdnl = samehist(nlcorrgrp,rpnlcorrgrp);
  %randomized phase, same distribution
  %of nonlinear process

************************************************
\*rphase.m*\
%Randomizes the Fourier Phases of a signal,
%while preserving the magnitude of its FT

function ranphase = rphase(signal)

L = length(signal);
L2 = L/2;
L2a = L2 - 1;
mag = abs(fft(signal,L));

rph = 2*pi*rand(length(mag)/2-1,1);
phase = cos(rph) + j.*sin(rph);
phsconj = conj(phase);

spect1 = [mag(1); mag(2:L2).*phase; mag(L2+1);
            mag(L2:-1:2).*phsconj(L2a:-1:1)];
ranphase = real(ifft(spect1));

************************************************
/* samehist.m */
% Matches ``rphase'' histogram to ``sigtomatch''
function shist = samehist(sigtomatch,rphase)
[y,indexlist] = sort(rphase);
[sortedlist,j] = sort(sigtomatch);

for i = 1:length(indexlist)
        shist(indexlist(i)) = sortedlist(i);
end;

************************************************
/* create_oof.m*/
% Generate nonlinear process by inputting
% 1/f^(5/8) Power Spectrum of gaussian random process
% into a quadratic nonlinear system

b = .4;
a = .5;
k = 11;

x = randn(2^k,1);
oofsig = oof(2^k,x);

oofsig = oofsig / std(oofsig);
oofnl = oofsig + b*(oofsig.^2);
oofnl = oofnl / std(oofnltemp);
%   Normalize standard deviation at output

rpoofnl = rph(oofnltemp);
%   control -- randomized phase of above signal

rphistnloof = samehist(oofnltemp,rpoofnltemp);
%  randomized phase, same distribution
%  of nonlinear process

************************************************
/*oof.m*/
%  Creates a length N 1/f^(5/16) Fourier Transform
%  magnitude with angle ``origsig'', and returns
%  the inverse fourier transform, ``sig''

function sig = oof(N,origsig)

mag(1) = 0;

for k = 1:(N/2)
        mag(k+1) = 1/(k/N).^(5/16);
end;

mag(N/2+2:N) = mag(N/2:-1:2);
totangle = (angle(fft(origsig)))';
phs = (cos(totangle) + j.*sin(totangle));
spect1 = mag.*phs;
sig = real(ifft(spect1))';

************************************************
/* c3div.f */
      program main
      parameter(nmax=100000,lseg = 2048,nn=50)
      real  u(lseg), y(nmax)
      real  c3ave(-nn:nn,-nn:nn), c3(-nn:nn,-nn:nn)
```

```
      do i = 1, nmax
         read(5,*,end=101) y(i)
         ntot=i
      end do
 101  nseg = int(ntot/lseg)

      do i = 1, nseg
         do j = 1, lseg
            u(j) = y((i-1)*lseg+j)
         end do
         call Cthree(u,c3,lseg,nn)
         do m = -nn, nn
            do n = -nn, nn
               c3ave(m,n)=c3ave(m,n)+c3(m,n)
            end do
         end do
      end do

      do m = -nn, nn
         write(6,1001)
          (c3ave(m,n)/real(nseg),n=-nn,nn)
 1001    format(101(F14.8,1X))
      end do
      stop
      end



      SUBROUTINE Cthree(u,c3,lseg,nn)
      real  u(*), c3(-nn:nn,-nn:nn)
      real  ave, sd, s2
      integer t1, t2
*** calculate mean & s.d.
      ave = 0.
      s2 = 0.
      do i = 1, lseg
         s2 = s2 + u(i)**2
         ave=ave+u(i)
      end do
      ave = ave/real(lseg)
      sd = sqrt((s2/real(lseg))-ave**2)
*** normalize data to zero mean and unit s.d.
      do i = 1, lseg
c         u(i) = (u(i)-ave)/sd
c           depending on experiment
          u(i) = (u(i)-ave)
      end do

*** calculate C3           X....t1....X...t2-t1...X
      do t1 = -nn, nn
         do t2 = -nn, nn
            i1 = min(0,t1,t2)
            i2 = max(0,t1,t2)

            temp = 0.
            do i = 1-i1, lseg-i2
               temp = temp + u(i)*u(i+t1)*u(i+t2)
            end do
            tot = real(lseg-i2+i1)
            c3(t1,t2)=temp/tot
         end do
      end do
      return
      end

*****************************************************
/* c4.f */
      program main
      parameter(nmax=100000,nn=50)
      real  u(nmax)
      real  c4(-nn:nn,-nn:nn,-nn:nn)
      do i = 1, nmax
         read(5,*,end=101) u(i)
         lseg=i
      end do
 101  continue
      call Cfour(u,c4,lseg,nn)
      do m = -nn, nn
         write(6,1001) (c4(8,m,n),n=-nn,nn)
 1001    format(101(F14.8,1x))
      end do

      stop
      end



      SUBROUTINE Cfour(u,c4,lseg,nn)
      real  u(*), c4(-nn:nn,-nn:nn,-nn:nn)
      real  C2(0:1000)
      real  dc(0:101,0:101,0:101), temp(0:101,0:101,0:101)
      real  ave, sd, s2
```

```
      integer t1, t2, t3
      do t1=0,2*nn
         do t2=0,2*nn
            do t3=0,2*nn
               temp(t1,t2,t3)=0.
            end do
         end do
      end do
*** calculate mean & s.d.
      ave = 0.
      s2 = 0.
      do i = 1, lseg
         s2 = s2 + u(i)**2
         ave=ave+u(i)
      end do
      ave = ave/real(lseg)
      sd = sqrt((s2/real(lseg))-ave**2)
*** normalize data to zero mean and unit s.d.
      do i = 1, lseg
         u(i) = (u(i)-ave)/sd
      end do
*** calculate 2-point correlations before average.
      c2(0) = 1.*real(lseg)
      do i = 1, 2*nn
         c2(i) = 0.
         do j= 1, lseg-i
            c2(i)= c2(i)+u(j)*u(j+i)
         end do
      end do
*** finite-size corrections to 2-point correlations.
      nmax = 2*nn
      do n=0,nmax
         d1 = 0.
         do l = 0, nmax-n
            if(l .ne. 0) d1 = d1 + u(l)*u(l+n)
            d2 = 0.
            do m = l+n, nmax
               m1 = lseg-m+l+1
               if(m .ne. l+n)  d2 = d2 + u(m1)*u(m1+n)
               dc(l,m,n) = d1 + d2
            end do
         end do
      end do

*** calculate C4
      t1 = 8
      do t2 = -nn, nn
         do t3 = -nn, nn
            i1 = min(0,t1,t2,t3)
            j0 = 0 - i1
            j1 = t1 - i1
            j2 = t2 - i1
            j3 = t3 - i1
            call sort4(j0,j1,j2,j3)
            if(temp(j1,j2,j3) .ne. 0.) then
               c4(t1,t2,t3) = temp(j1,j2,j3)
            else
               temp(j1,j2,j3) = 0.
               do i = 1, lseg-j3
                  temp(j1,j2,j3) = temp(j1,j2,j3) +
     &                 u(i)*u(i+j1)*u(i+j2)*u(i+j3)
               end do
               tot = real(lseg-j3)
               temp(j1,j2,j3) = temp(j1,j2,j3)/tot
               x1=(c2(j1)-dc(0,j3,j1))/tot
               x2=(c2(j3-j2)-dc(j2,j3,j3-j2))/tot
               x3=(c2(j2)-dc(0,j3,j2))/tot
               x4=(c2(j3-j1)-dc(j1,j3,j3-j1))/tot
               x5=(c2(j3)-dc(0,j3,j3))/tot
               x6=(c2(j2-j1)-dc(j1,j3,j2-j1))/tot
               temp(j1,j2,j3)=temp(j1,j2,j3)-x1*x2-x3*x4-x
               c4(t1,t2,t3) = temp(j1,j2,j3)
            end if
         end do
      end do
      return
      end


      SUBROUTINE sort4(k1,k2,k3,k4)
      integer kk(4), kt
      kk(1)=k1
      kk(2)=k2
      kk(3)=k3
      kk(4)=k4
      do i = 3, 1, -1
         do j = 1, i
            if(kk(j) .gt. kk(j+1)) then
               kt = kk(j)
               kk(j) = kk(j+1)
```

```
            kk(j+1) = kt
          end if
        end do
      end do
      k1=kk(1)
      k2=kk(2)
      k3=kk(3)
      k4=kk(4)
      return
      end

************************************************
/* c4psav.f */
      program main
      parameter(nn=512,nmax=128000)
      real c4(-nn:nn,-nn:nn)
      real u(nmax)
      real x(1:2*nn),y(1:2*nn),sig(1:2*nn)
      real sum(0:nn)
      real psd(0:nn)
      integer t1, op

      do i = 1, nmax
         read(5,*,end=101) u(i)
         lseg=i
      end do
 101  continue



      op=100
      do t1 = -nn,nn,100

      call Cfour(u,c4,t1,lseg,nn)

      do i = 0,nn
         sum(i) = 0.
      end do

      count = 0
      do m = -nn, nn
            count = count + 1


            do n = -nn, nn-1
               y(n+nn+1) = c4(m,n)
            end do
            call detrend(x,y,sig,2*nn)
            call power(y,nn,psd)

            do i = 0,nn
               sum(i) = sum(i) + psd(i)
            end do

      end do
      op=op+1
      do m=1,nn
         write(op, *) (sum(m)/count)
      end do



      end do

      stop
      end



      SUBROUTINE Cfour(u,c4,t1,lseg,nn)
      real  u(*), c4(-nn:nn,-nn:nn)
      real  ave, sd, s2
      integer t1, t2, t3
*** calculate mean & s.d. of time series u(lseg)
      ave = 0.
      s2 = 0.
      do i = 1, lseg
         s2 = s2 + u(i)**2
         ave=ave+u(i)
      end do
      ave = ave/real(lseg)
      sd = sqrt((s2/real(lseg))-ave**2)
*** normalize data to zero mean and unit s.d.
      do i = 1, lseg
         u(i) = (u(i)-ave)/sd
      end do
*** calculate C4 by the slow (straightforward) way
      do t2 = -nn, nn
         do t3 = t2, nn
            i1 = min(0,t1,t2,t3)
            i2 = max(0,t1,t2,t3)
            x1=0.
            x2=0.
```

```
            x3=0.
            x4=0.
            x5=0.
            x6=0.
            do i = -i1+1, lseg-i2
               c4(t2,t3)=c4(t2,t3)+ u(i)*u(i+t1)*u(i+t2)*u(i+t3)
               x1 = x1 + u(i)*u(i+t1)
               x2 = x2 + u(i+t2)*u(i+t3)
               x3 = x3 + u(i)*u(i+t2)
               x4 = x4 + u(i+t1)*u(i+t3)
               x5 = x5 + u(i)*u(i+t3)
               x6 = x6 + u(i+t1)*u(i+t2)
            end do
            tot = real(lseg+i1-i2)
            c4(t2,t3)=c4(t2,t3)/tot
     &        -(x1*x2+x3*x4+x5*x6)/(tot*tot)
         end do
      end do
      do t2 = -nn, nn
         do t3 = -nn, t2
            c4(t2,t3)=c4(t3,t2)
         end do
      end do

      return
      end

*** Power spectrum subroutine
*** data(2*n): the time series (input data) of length 2*N
*** psd(0:n): power spectrum density (square of FT amplitude)
***** where psd(i) corrspond to frequency f=i/(2*N), i from 0 to N
      SUBROUTINE POWER(DATA,N,PSD)
      DIMENSION DATA(2*N), PSD(0:N)
** call FFT
      call REALFT(DATA,N,1)
** Calculate the PSD
      psd(0) = data(1)**2
      psd(n) = data(2)**2
      do i = 1, n-1
         psd(i) = data(2*i+1)**2+data(2*i+2)**2
      end do
      return
      end
```

```
**** FFT subroutine from Numerical Recipes
      SUBROUTINE REALFT(DATA,N,ISIGN)
      REAL*8 WR,WI,WPR,WPI,WTEMP,THETA
      DIMENSION DATA(*)
      THETA=6.28318530717959D0/2.0D0/DBLE(N)
      C1=0.5
      IF (ISIGN.EQ.1) THEN
        C2=-0.5
        CALL FOUR1(DATA,N,+1)
      ELSE
        C2=0.5
        THETA=-THETA
      ENDIF
      WPR=-2.0D0*DSIN(0.5D0*THETA)**2
      WPI=DSIN(THETA)
      WR=1.0D0+WPR
      WI=WPI
      N2P3=2*N+3
      DO 11 I=2,N/2+1
        I1=2*I-1
        I2=I1+1
        I3=N2P3-I2
        I4=I3+1
        WRS=SNGL(WR)
        WIS=SNGL(WI)
        H1R=C1*(DATA(I1)+DATA(I3))
        H1I=C1*(DATA(I2)-DATA(I4))
        H2R=-C2*(DATA(I2)+DATA(I4))
        H2I=C2*(DATA(I1)-DATA(I3))
        DATA(I1)=H1R+WRS*H2R-WIS*H2I
        DATA(I2)=H1I+WRS*H2I+WIS*H2R
        DATA(I3)=H1R-WRS*H2R+WIS*H2I
        DATA(I4)=-H1I+WRS*H2I+WIS*H2R
        WTEMP=WR
        WR=WR*WPR-WI*WPI+WR
        WI=WI*WPR+WTEMP*WPI+WI
11    CONTINUE
      IF (ISIGN.EQ.1) THEN
        H1R=DATA(1)
        DATA(1)=H1R+DATA(2)
        DATA(2)=H1R-DATA(2)
      ELSE
        H1R=DATA(1)
        DATA(1)=C1*(H1R+DATA(2))
        DATA(2)=C1*(H1R-DATA(2))
```

```
      CALL FOUR1(DATA,N,-1)
    ENDIF
    RETURN
    END


    SUBROUTINE FOUR1(DATA,NN,ISIGN)
    REAL*8 WR,WI,WPR,WPI,WTEMP,THETA
    DIMENSION DATA(*)
    N=2*NN
    J=1
    DO 11 I=1,N,2
      IF(J.GT.I)THEN
        TEMPR=DATA(J)
        TEMPI=DATA(J+1)
        DATA(J)=DATA(I)
        DATA(J+1)=DATA(I+1)
        DATA(I)=TEMPR
        DATA(I+1)=TEMPI
      ENDIF
      M=N/2
1     IF ((M.GE.2).AND.(J.GT.M)) THEN
        J=J-M
        M=M/2
      GO TO 1
      ENDIF
      J=J+M
11  CONTINUE
    MMAX=2
2   IF (N.GT.MMAX) THEN
      ISTEP=2*MMAX
      THETA=6.28318530717959D0/(ISIGN*MMAX)
      WPR=-2.D0*DSIN(0.5D0*THETA)**2
      WPI=DSIN(THETA)
      WR=1.D0
      WI=0.D0
      DO 13 M=1,MMAX,2
        DO 12 I=M,N,ISTEP
          J=I+MMAX
          TEMPR=SNGL(WR)*DATA(J)-SNGL(WI)*DATA(J+1)
          TEMPI=SNGL(WR)*DATA(J+1)+SNGL(WI)*DATA(J)
          DATA(J)=DATA(I)-TEMPR
          DATA(J+1)=DATA(I+1)-TEMPI
          DATA(I)=DATA(I)+TEMPR
          DATA(I+1)=DATA(I+1)+TEMPI
12        CONTINUE
          WTEMP=WR
          WR=WR*WPR-WI*WPI+WR
          WI=WI*WPR+WTEMP*WPI+WI
13      CONTINUE
        MMAX=ISTEP
      GO TO 2
      ENDIF
      RETURN
      END


      subroutine detrend(x,y,sig,ntot)
      real  y(*), x(*), sig(*)
      do i=1,ntot
         x(i) = i
      end do
** linear regression fit
      mwt = 0
      call FIT(x,y,ntot,SIG,mwt,A,B,SIGA,SIGB,CHI2,Q)
      do i=1,ntot
         y(i) = y(i) - (A + B*x(i))
         ave = ave + y(i)
      end do
** set mean to 0
      ave = ave/real(ntot)
      do i=1,ntot
         y(i) = y(i) - ave
      end do
      return
      end

      SUBROUTINE FIT(X,Y,NDATA,SIG,MWT,A,B,SIGA,SIGB,CHI2,Q)
      DIMENSION X(*),Y(*),SIG(*)
      SX=0.
      SY=0.
      ST2=0.
      B=0.
      IF(MWT.NE.0) THEN
        SS=0.
        DO 11 I=1,NDATA
          WT=1./(SIG(I)**2)
          SS=SS+WT
          SX=SX+X(I)*WT
          SY=SY+Y(I)*WT
11      CONTINUE
```

```
      ELSE
        DO 12 I=1,NDATA
          SX=SX+X(I)
          SY=SY+Y(I)
12      CONTINUE
        SS=FLOAT(NDATA)
      ENDIF
      SXOSS=SX/SS
      IF(MWT.NE.0) THEN
        DO 13 I=1,NDATA
          T=(X(I)-SXOSS)/SIG(I)
          ST2=ST2+T*T
          B=B+T*Y(I)/SIG(I)
13      CONTINUE
      ELSE
        DO 14 I=1,NDATA
          T=X(I)-SXOSS
          ST2=ST2+T*T
          B=B+T*Y(I)
14      CONTINUE
      ENDIF
      B=B/ST2
      A=(SY-SX*B)/SS
      SIGA=SQRT((1.+SX*SX/(SS*ST2))/SS)
      SIGB=SQRT(1./ST2)
      CHI2=0.
      IF(MWT.EQ.0) THEN
        DO 15 I=1,NDATA
          CHI2=CHI2+(Y(I)-A-B*X(I))**2
15      CONTINUE
        Q=1.
        SIGDAT=SQRT(CHI2/(NDATA-2))
        SIGA=SIGA*SIGDAT
        SIGB=SIGB*SIGDAT
      ELSE
        DO 16 I=1,NDATA
          CHI2=CHI2+((Y(I)-A-B*X(I))/SIG(I))**2
16      CONTINUE
CC        Q=GAMMQ(0.5*(NDATA-2),0.5*CHI2)
      ENDIF
      RETURN
      END

      FUNCTION GAMMQ(A,X)
      IF(X.LT.0..OR.A.LE.0.)PAUSE
      IF(X.LT.A+1.)THEN
        CALL GSER(GAMSER,A,X,GLN)
        GAMMQ=1.-GAMSER
      ELSE
        CALL GCF(GAMMCF,A,X,GLN)
        GAMMQ=GAMMCF
      ENDIF
      RETURN
      END

      SUBROUTINE GCF(GAMMCF,A,X,GLN)
      PARAMETER (ITMAX=100,EPS=3.E-7)
      GLN=GAMMLN(A)
      GOLD=0.
      A0=1.
      A1=X
      B0=0.
      B1=1.
      FAC=1.
      DO 11 N=1,ITMAX
        AN=FLOAT(N)
        ANA=AN-A
        A0=(A1+A0*ANA)*FAC
        B0=(B1+B0*ANA)*FAC
        ANF=AN*FAC
        A1=X*A0+ANF*A1
        B1=X*B0+ANF*B1
        IF(A1.NE.0.)THEN
          FAC=1./A1
          G=B1*FAC
          IF(ABS((G-GOLD)/G).LT.EPS)GO TO 1
          GOLD=G
        ENDIF
11    CONTINUE
      PAUSE 'A too large, ITMAX too small'
1     GAMMCF=EXP(-X+A*ALOG(X)-GLN)*G
      RETURN
      END

      SUBROUTINE GSER(GAMSER,A,X,GLN)
      PARAMETER (ITMAX=100,EPS=3.E-7)
      GLN=GAMMLN(A)
      IF(X.LE.0.)THEN
        IF(X.LT.0.)PAUSE
        GAMSER=0.
```

```
        RETURN
      ENDIF
      AP=A
      SUM=1./A
      DEL=SUM
      DO 11 N=1,ITMAX
        AP=AP+1.
        DEL=DEL*X/AP
        SUM=SUM+DEL
        IF(ABS(DEL).LT.ABS(SUM)*EPS)GO TO 1
11    CONTINUE
      PAUSE 'A too large, ITMAX too small'
1     GAMSER=SUM*EXP(-X+A*LOG(X)-GLN)
      RETURN
      END

      FUNCTION GAMMLN(XX)
      REAL*8 COF(6),STP,HALF,ONE,FPF,X,TMP,SER
      DATA COF,STP/76.18009173D0,-86.50532033D0,
     *     24.01409822D0,
     *    -1.231739516D0,.120858003D-2,
     *    -.536382D-5,2.50662827465D0/
      DATA HALF,ONE,FPF/0.5D0,1.0D0,5.5D0/
      X=XX-ONE
      TMP=X+FPF
      TMP=(X+HALF)*LOG(TMP)-TMP
      SER=ONE
      DO 11 J=1,6
        X=X+ONE
        SER=SER+COF(J)/X
11    CONTINUE
      GAMMLN=TMP+LOG(STP*SER)
      RETURN
      END
*************************************************
```

# Bibliography

[1] M. Appel, R. Berger, J. Saul, J. Smith, and R. Cohen, "Beat to Beat Variability in Cardiovascular Variables: Noise or Music?," *Journal of American College of Cardiology*, Vol. 14, No. 5, November 1989, pp. 1139–1148.

[2] D. Kaplan and M. Talajic, "Dynamics of Heart Rate," *Chaos*, Vol. 1, No. 3, 1991, pp. 251–256.

[3] O. Rompelman, A.J.R.M. Coenen, and R.I. Kitney, "Measurement of heart-rate variability: Part 1 - Comparative Study of Heart-Rate Variability Analysis Methods," *Medical and Biological Engineering and Computing*, pp. 233–239, May 1977.

[4] P.C. Ivanov, M.G. Rosenblum, C.-K. Peng, J.Mietus, S. Havlin, H.E. Stanley, and A.L. Goldberger, "Scaling behavior of heartbeat intervals obtained by wavelet-based time-series analysis," *Nature*, Vol. 383, September 26, 1997, pp. 323–327.

[5] D. Kaplan and A.L. Goldberger, "Chaos in Cardiology," *Journal of Cardiovascular Electrophysiology*, Vol. 2, August 1991, pp. 342–354.

[6] A.L. Goldberger, *Nonlinear dynamics for clinicians: Chaos theory, fractals and complexity at the bedside*, Lancet 1996, Vol. 347, pp. 1312–1314.

[7] R. Barr, *Bioelectricity — A Quantitative Approach*, Plenum Press, NY, 1993.

[8] R.M. Berne and M.N. Levy, *Cardiovascular Physiology*, CV Mosby Co., St. Louis, 5th Edition, 1986.

[9] C.K. Peng, J. Mietus, J. Hausdorff, S. Havlin, H.E. Stanley, and A.L. Goldberger, "Long-Range Anticorrelations and Non-Gaussian Behavior of the Heartbeat," *Physical Review Letters*, Vol. 70, No. 9, March 1993, pp. 1343–1346.

[10] A.L. Goldberger, "Is the normal heartbeat chaotic or homeostatic?," *News in Physiologic Science (Int Union Physiol Sci/Am Physiol Soc)*, Vol. 6,1991, pp. 87–91.

[11] M.Malik, and A.J. Camm, *Heart Rate Variability*, Futura Publishing Co. Inc., New York 1995.

[12] A.L. Goldberger, D.R. Rigney, and B.J. West, "Chaos and Fractals in Human Physiology," *Scientific American*, Vol. 262, No. 2, February 1990, pp. 42–49.

[13] D. Hoyer, K. Schmidt, R. Bauer, U. Zwiener, M. Kohler, B. Luthke, and M. Eiselt, "Nonlinear Analysis of Heart Rate and Respiratory Dynamics," *IEEE Engineering in Medicine and Biology*, January/ February 1997, pp. 31–39.

[14] P.E. Rapp, "Chaos in the neurosciences: Cautionary tales from the frontier," *Biologist*, Vol. 40, 1993, pp. 89–94.

[15] S. Elgar and V. Chandran, "Higher-Order Spectral Analysis to Detect Nonlinear Interactions in Measured Time Series and an Application to Chua's Circuit," *International Journal of Bifurcation and Chaos*, Vol. 3, No. 1,1993, pp. 19–34.

[16] J.M. Mendel, "Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications," *Proceedings of the IEEE*, Vol. 79, No. 3,1991, pp. 278–305.

[17] C.L. Nikias and A.P. Petropulu, *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework*, Prentice Hall Inc., New Jersey: Englewood Cliffs, 1993.

[18] A.S. Willsky, G.W. Wornell, and J.H. Shapiro, *Stochastic Processes, Detection and Estimation — 6.432 Supplementary Course Notes*, Department of Electri-

cal Engineering and Computer Science, Massachusetts Institute of Technology, Massachusetts: Cambridge, 1996.

[19] J.M. Hausdorff and C.-K. Peng, "Multiscaled randomness: A possible source of $1/f$ noise in biology," *Physical Review E*, Vol. 54, No. 1, July 1996, pp. 54–57.

[20] J. Beran, *Statistics for Long-Memory Processes*, Chapman & Hall, New York, 1994.

[21] D. S. Baim et al., "Survival of patients with severe congestive heart failure treated with oral milrinone," *Journal of American College of Cardiology*, Vol. 7, 1986, pp. 661–670.

[22] C.-K. Peng et al., "Fractal Mechanisms and Heart Rate Dynamics: Long-Range Correlations and Their Breakdown with Disease," *Journal of Electrocardiology*, Vol. 28, 1996, pp.59–65.

[23] A. L. Goldberger, D. R. Rigney, J. Mietus, E. M. Antman & S. Greenwald, "Nonlinear dynamics in sudden cardiac death syndrome: Heart rate oscillations and bifurcations," *Experientia*, Vol. 44, 1988, pp.983–987.

[24] L. Glass and D. Kaplan, "Time Series Analysis of Complex Dynamics in Physiology and Medicine," *Medical Progress through Technology*, Vol. 19, May 1993, pp. 115–128.

[25] A.V. Oppenheim and J.S. Lim, "The Importance of Phase in Signals," *Proceedings of the IEEE*, Vol. 69, No. 5, May 1981, pp. 529-541.

[26] A. Swami, J. Mendel and C.L. Nikias, *Higher-Order Spectral Analysis Toolbox User's Guide*, The MathWorks, Inc., Massachusetts: Natick, 1993.