A Study of Heart Period Variability for the Statistical Detection of Congestive Heart Failure

by

Stephen Dueck

A thesis
submitted to the Faculty of Graduate Studies
in partial fulfillment of the degree of

MASTER OF SCIENCE

Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, Manitoba

Thesis Advisor: W. Kinsner, Ph.D., P.Eng.

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

A Study of Heart Period Variability for the Statistical Detection of Congestive Heart Failure

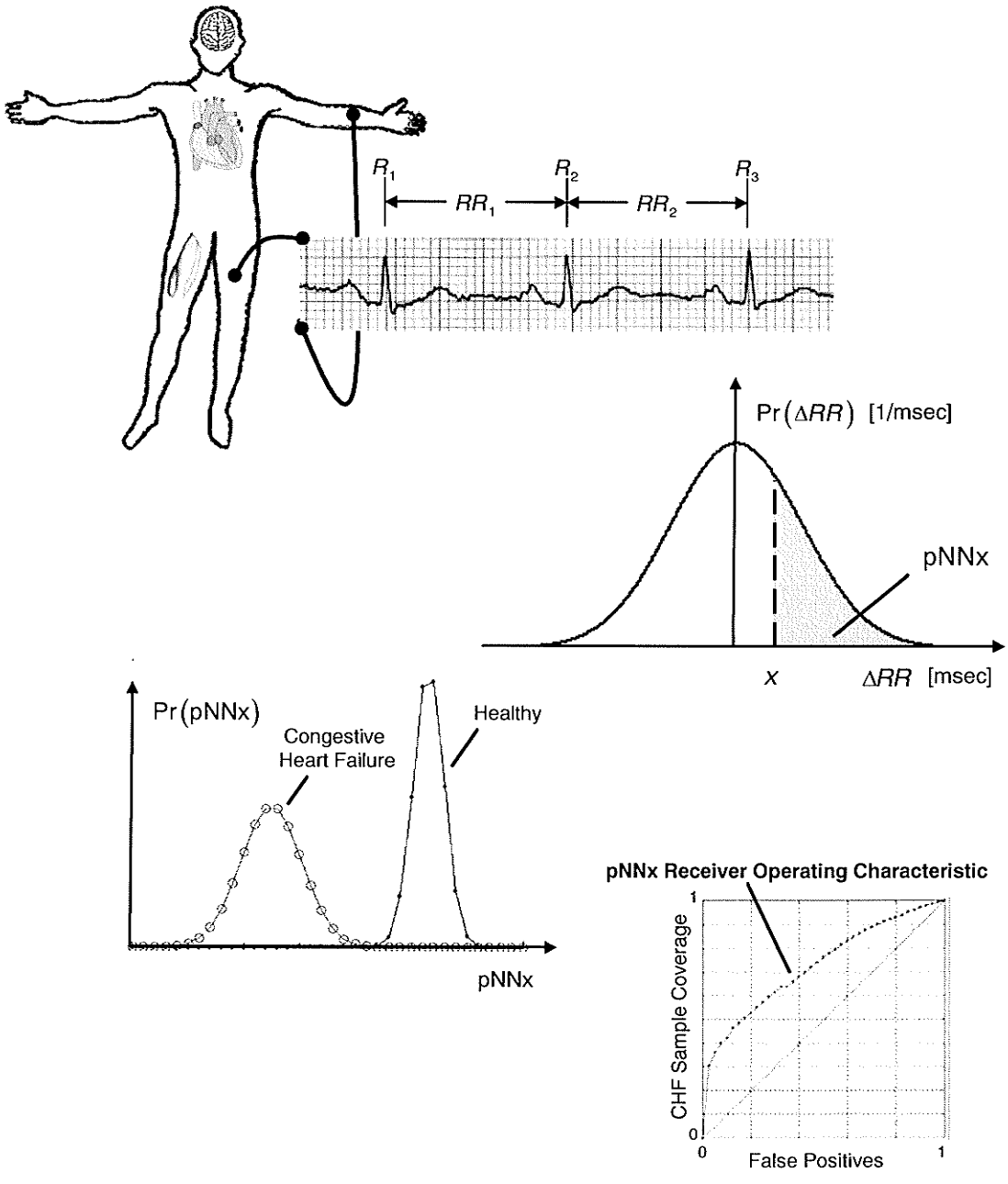
BY

Stephen Dueck

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement of the degree

MASTER OF SCIENCE

Visual Abstract



Abstract

A decrease in the variability in the time interval spanning cardiac contractions (as measured from R-complex to R-complex of the *electrocardiogram* (ECG)) has clinical significance for cardiovascular disease. Untransformed *heart period variability* (HPV) metrics, however, are often treated as minor variations on a similar theme and, as a result, both (i) the requirements of common HPV metrics from data, and (ii) the requirements of common data sets from HPV metrics remain misunderstood. This problem manifests itself in the prevalence of HPV papers in both medical and engineering peer-reviewed literature which use multiple metrics whose selection is unjustified and persists, in part, due to the absence of adequate quantitative assessments of the discrimination power of HPV metrics.

The results of this work support the thesis that (a) a metric with the power to statistically discriminate individuals with *congestive heart failure* (CHF) from healthy individuals can be derived from RR interval data, (b) its discrimination power can be characterized consistently with results existing in literature, (c) the results in literature can be expanded with more powerful discriminant assessments, and (d) these results will generalize given a different sample of comparable data. Results include, first, the statistical formulation and derivation of a HPV metric. This measure can be taken as a measure of cardiac acceleration and possesses the well-known HPV metric termed pNNx as a subset. Second, p values in agreement with literature were obtained. Third, the physiological variation manifests itself most powerfully for pNNx parameter for x=16 msec. Third, at the cost of a high proportion of false positives, pNNx raised the visibility of the CHF sample (2% prevalence) by 11.4% and 34.1% while finding 95% and 85% of CHF subjects, respectively. A low rate of false positives (approximately 1 in 4) can be achieved if consideration is restricted to a CHF subset representing 15% of the total sample. Fourth, cross validation suggests these results are repeatable given a similar CHF sample.

These results indicate, first, that low amplitude cardiac accelerations are important. The results indicate, second, that the treatment of CHF can benefit from the use pNNx but that either it must be combined with a second, complementary metric or placed at the output of a subset-isolating screening stage to improve efficiency.

Acknowledgements

I would like to acknowledge several people who have made contributions to this work in supporting roles as well as three people who have been key to both the work itself and the larger effort of which this work is part. I would like begin by thanking my sister, Christine, who is my twin in work ethic and humor, for her support and Jessica Carvalho for her support early on in this work. In addition, I can not imagine this work without the articles procured by the fine Document Delivery department of the University of Manitoba Libraries or the many discussions with members of the Delta Group including Michael Potter, Aram Faghfouri, Sharjeel Siddiqui, Neil Gadhok, and Hakim El-Boustani. I must also emphatically thank my employer, IDers, for allowing me flexibility in the terminal stages of this project. Lastly, I would like thank the independent reviewers of this work: Andrew Bugera, Dr. Christine Dueck, Justin Olivier, Dr. Jan Roberts, Kory Zelickson, and, particularly, Michael Potter.

The first of three key people I would like to recognize is Dr. Kinsner, who I have worked with since I was an undergraduate. Specifically, I would like to acknowledge his ability to select problems of fundamental importance for his research group, for allowing me to develop intellectual tools for their comprehension, and for holding me to the same intellectual standards he holds himself. It has been of immeasurable benefit to my ability to independently sift complex ideas and critically appraise arguments.

I would like to acknowledge, secondly, the mind and temperament of Michael Potter, who has improved the clarity of this work throughout its evolution and with whom I possess a working relationship by which all others are measured.

Lastly, I would like to thank my Mom, whose intellectual, moral, and financial support have given me the luxury of questioning the nature of heart period variability. Her strength of character, determination, and balance of work/life accomplishments are my gold standard. If I am written by multiple authors, my Mom is the first.

Dedicated to Jan Roberts, M.D., Ph.D., M.o.M.

Table of Contents

List of Abbreviations

ASK	Amplitude shift keying.
ANS	Autonomic nervous system.
AV	Atrioventricular.
BSMS	Blind signal mixture separation.
cdf	Cumulative distribution function.
CHF	Congestive heart failure.
cmdf	Cumulative mass distribution function.
CNS	Central nervous system.
CO_2	Carbon dioxide.
CVS	Cardiovascular system.
ECG	Electrocardiogram.
FNP	False negative proportion.
FPP	False positive proportion.
HPV	Heart period variability.
LOO	Leave one out cross validation.
NN50	Number of RR interbeat interval changes exceeding 50 milliseconds.
NYHA	New York Heart Association
pdf	Probability density function.
pmf	Probability mass function..
pNN50	Percentage of RR interbeat interval changes exceeding 50 milliseconds.
pNNx	Percentage of RR interbeat interval changes exceeding x milliseconds.
PNS	Peripheral nervous system.
QPSK	Quadrature phase shift keying.
rMSSD	Root mean square of successive differences.

RR	Time between two successive R complexes of the electrocardiogram.
ROC	Receiver operating characteristic.
SA	Sinoatrial.
SE	Standard error.
SEM	Standard error on the mean.
TNP	True negative proportion.
TPP	True positive proportion.

List of Symbols

α	Classifier amplification.
β	Rate of TPP classifications.
$B(\cdot)$	Euler's Beta function.
C	Classification class.
δ	Degrees of freedom.
δ_p	Pooled degrees of freedom.
D	M-dimensional subspace.
η	Classification efficiency.
$E[\cdot]$	Statistical expectation function.
$f(\cdot)$	Probability mass function.
$\hat{f}(\cdot)$	Probability mass function estimate.
$F(\cdot)$	Cumulative mass distribution function.
$\hat{F}(\cdot)$	Cumulative mass distribution function estimate.
γ	Classification accuracy.
H_0	Null hypothesis.
H_1	Alternative hypothesis.
I	Set of all integers.
κ	Classifier coverage.
ℓ_{ab}	Likelihood ratio of random variable a to random variable b.
μ	Heart period variability measure function.
$M_{C[j]}$	Value of μ for CHF subject j.
$M_{H[i]}$	Value of μ for healthy subject i.
$\Psi(\cdot)$	Continuous window function for kernel density estimation.
$\varphi(x)$	Discrete window function for kernel density estimation.

$p(\cdot)$	Probability density function.
p	"p" value.
$\hat{p}(\cdot)$	Probability density function estimate.
$p(x\|C)$	*A priori* probability density function for the random variable x given the class C.
$p(C\|x)$	*Posteriori* probability measure function for the class C given the random variable x.
P	The probability of an event.
Q	Digital amplitude quantization step.
ρ	Disease prevalence.
$\mathbb{R}$	Set of real numbers.
R	Set of all RR interbeat intervals.
$\overline{R}$	Set of the successive differences of all RR interbeat intervals.
R_C	Set of all CHF RR interbeat intervals.
$\overline{R}_C$	Set of the successive differences of all CHF RR interbeat intervals.
R_H	Set of all healthy RR interbeat intervals.
$\overline{R}_H$	Set of the successive differences of all healthy RR interbeat intervals.
$R_{C[j]}$	Set of all RR interbeat intervals for CHF subject j.
$\overline{R}_{C[j]}$	Set of the successive differences of all RR interbeat intervals for CHF subject j.
$R_{H[i]}$	Set of all RR interbeat intervals for healthy subject i.
$\overline{R}_{H[i]}$	Set of the successive differences of all RR interbeat intervals for healthy subject i.
$\hat{\sigma}$	Standard deviation estimate.
$\hat{\sigma}_N$	Standard error estimate calculated from N samples.
$\hat{\sigma}_{P[N]}$	Pooled standard error estimate calculated from N samples.
S_C	Class of all CHF subjects.

S_H	Class of all healthy subjects.
t	Statistic associated with the t-test.
T	Decision threshold; a scalar

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1.1 Motivation

Making evidence-based decisions in the face of uncertainty is a problem of universal importance. Decisions are impeded by evidence which does not clearly support a single alternative. This lack of clear support is due to either the natural variability of the evidence or the presence of confounding variables. Confounding variables may be present in both the raw data chosen for evidence or introduced when it is measured.

Evidence, in the context of computer engineering, is often derived from a signal. A *signal* is a measurement of electrical potential or an electrical value (such as voltage or current) which is assigned in proportion to a physical quantity of interest (such as water pressure).

Confounding variables, within computer engineering, are often referred to as "noise." The field of signal processing concerns itself with reducing the impact of unintentional noise on the decision. In the presence of cheap and reliable storage media or bandwidth (eliminating the need for source/channel encoding), signal processing for decision making consists chiefly of *blind signal mixture separation* (BSMS), feature extraction, and classification. The decision, in this context, is based on the state of the physical system, which is statistically inferred by classifying features of the physical system. Features, in turn, reflect the state of the physical system. As an example, the BSMS problem may involve distinguishing the voice of a friend from other voices with which it is entwined at a party. The feature extraction problem in this case may be detecting

characteristic sounds (such as sharp consonants) from the friend's voice and, then, classifying groups of these characteristic sounds into words in order to decide on a response. This process is summarized in Fig. 1.1 and two real-life examples involving an unmanned aerial vehicle and space satellite controller are described in papers by Dueck, *et al.* [DDFH05] and Dueck, *et al.* [DGKK03], respectively. Lastly, it should be noted that the decision regarding the control of the physical system may in general be made by a person, a computer, or both.

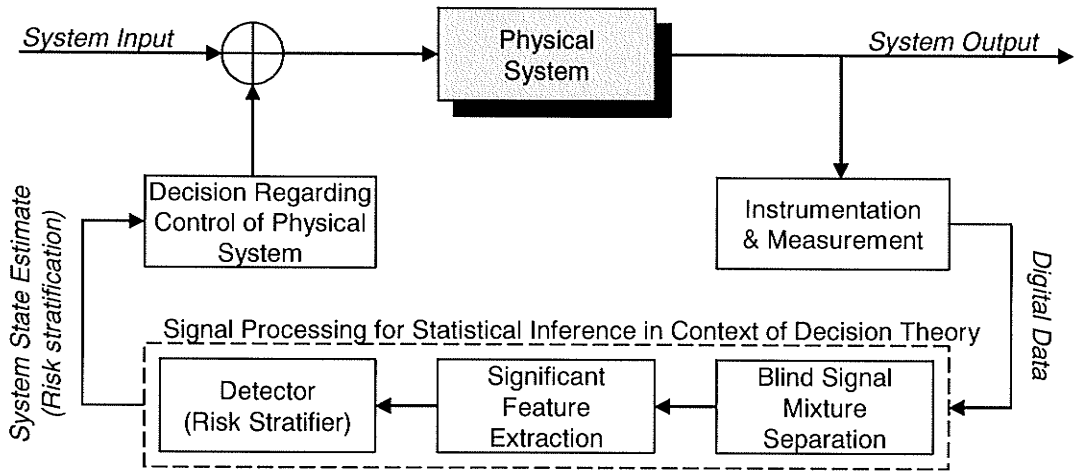


Fig. 1.1 Closed-loop computer engineering system for the control of a physical system.

Risk of decision error is proportional to noise amplitude and inversely proportional to signal amplitude. Human-engineered systems typically enjoy a very low decision error rate – typically on the order of one wrong decision in a million. This is because the language used to communicate between the, say, digital transmitter and receiver is clearly defined and agreed upon in advance. Decisions made on signals "transmitted" from natural systems, such as the human body, must be content with much lower success rates because the language is often unknown and the physical origins are difficult to model.

When a classification system possesses a low risk of error a definitive choice between decisions can be made by simply choosing the alternative with the highest likelihood. When decisions are valuable even though they are likely wrong (due to the high worth of the occasional right decisions), however, they are dealt with through the use of risk stratification, also known as decision tree analysis. In other words, in systems with higher error rates a more generalized multi-stage decision structure is used.

Within medicine, risk stratification is defined as the categorization of illness severity for treatment and diagnosis. The aim of risk stratification, in this context, is to ascribe certainty of having a disease and incur the associated (and proportionate) cost incrementally. The approach assumes that if there is a cost per classification – say, money, time, or the invasiveness of a medical procedure – that, for a given cost, more classifications can be performed without loss of accuracy if an expensive low error rate test is replaced by a multi-tier collection of low cost, high error rate tests which acquire both cost and knowledge incrementally. Risk stratification can be regarded as a hierarchical tree of statistical inferences to be traversed at minimum cost, as will be seen shortly.

Medical screening tests are one such example of risk stratification and evidence, particularly in the realm of clinical lab testing [CaSG78], shows that their correct usage can enhance decisions while incorrect usage inhibits them. As an illustration of cost savings, consider, for example, a screening test in which a positive result means a patient has the disease 40% of the time (and is false-positive 60% of the time) for a disease which has a general prevalence in the population of 0.1% (such as AIDS in Cuba [RKDL06]). Assume the test costs $50 to perform, has a false-negative rate of 0%, and must be followed by a so-called gold standard test (which is right 100% of the time and costs, say, $1,000) if a positive screening result is obtained. The cost of administering the screening – gold standard test combination to a population of 10,000 people will be

$500,000. For the same cost, the gold standard test on its own could be applied to only 65% of the entire population. The risk stratification structure is shown in Fig. 1.2.

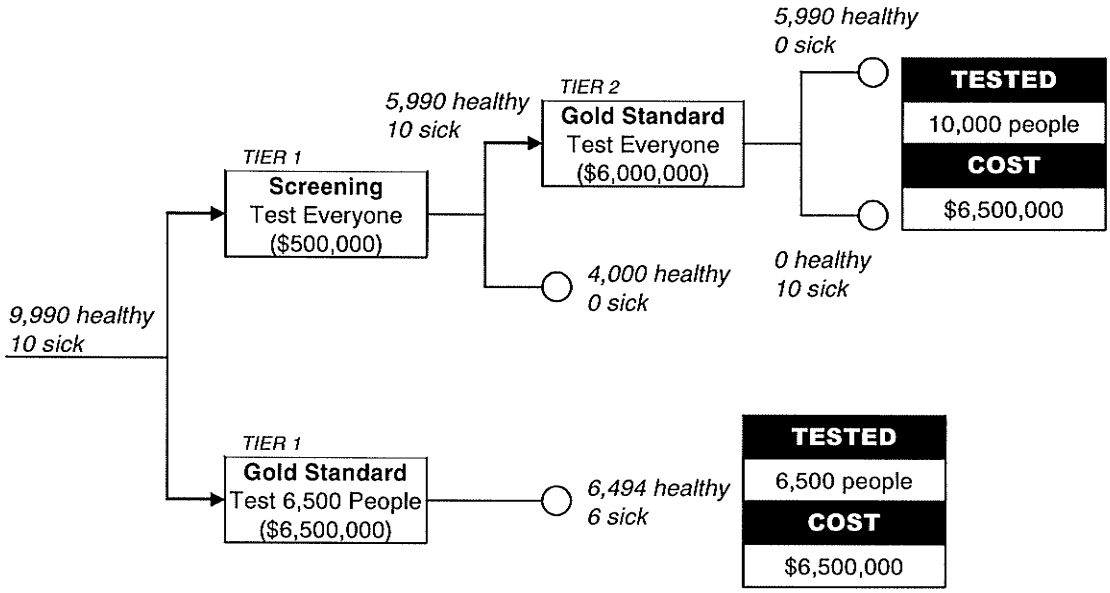


Fig. 1.2 Risk stratification as a multi-tiered decision process.

Signal processing requires a background in instrumentation and measurement, a sub-field of electrical and computer engineering. The IEEE Instrumentation and Measurement Society defines instrumentation and measurement as being concerned with electrical and electronic instruments for the purpose of measuring, monitoring or recording various physical phenomena [IEEE03]. Signal processing also requires a background in computing because the BSMS, feature extraction, classification and control steps of Fig. 1.1 are often performed in the digital domain. The Association of Computing Machinery has stated that computing can be taken to mean any goal-oriented activity requiring, benefiting from, or creating computers including designing and building hardware and software systems, processing, structuring, and managing various kinds of information, making computer systems behave intelligently, and creating and

using communications media, among others [ACM05]. The methodology of signal processing, therefore, involves reducing the impact of noise introduced during instrumentation, measurement, and computing on the final decision in a manner which does not alter the fundamental nature of the signal.

The resolution of many problems within the field of medicine are closely related to quality of human life. Therefore, their resolution is of importance. Specifically, there is a need for diagnosing proven relationships between observation and health in a clinical setting and identifying unrecognized or emerging relationships. Note, also, that the block diagram of Fig. 1.1, which omitted signal processing for communications, could be extended along the lines of Fig. 1.3 with a communications component to achieve medical diagnosis and control over distance (i.e. telehealth) as shown in Fig. 1.4. Note the presence of a low error rate statistical bit decision block within the communications extension of Fig. 1.3.

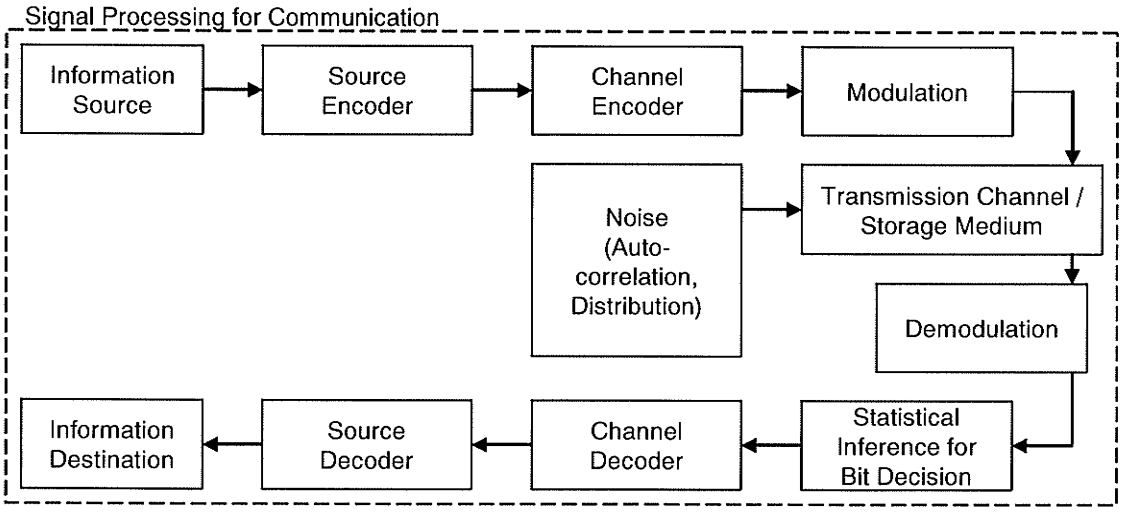


Fig. 1.3 Block diagram showing major conceptual components of signal processing for digital communications.

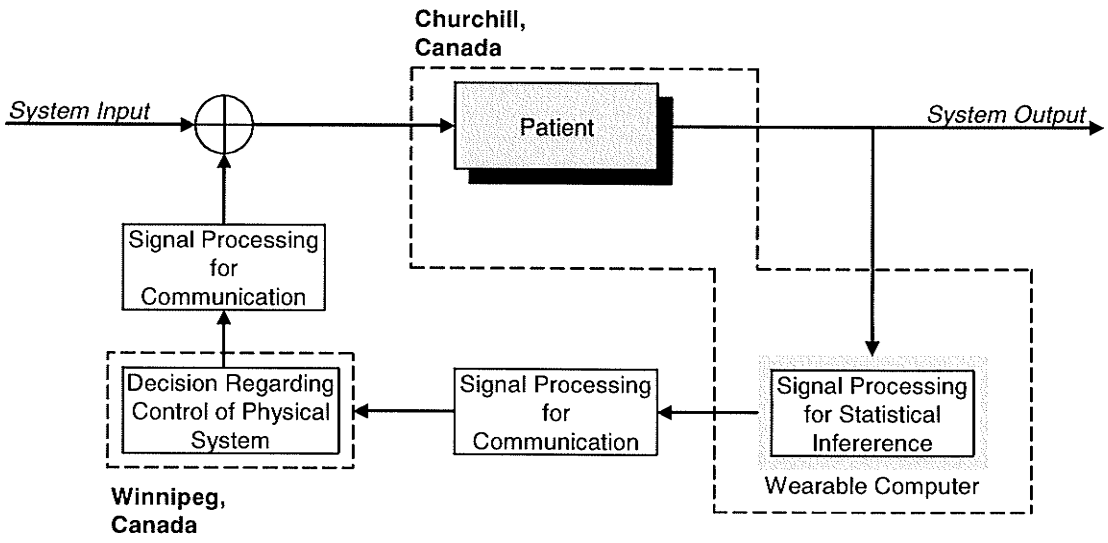


Fig. 1.4 Extension of signal processing for statistical inference by signal processing for communications to implement telehealth.

Both engineering and medical problems within medicine often involve signals produced by human organ systems which, as a result, are challenging to classify. This complexity is hypothesized to be a physiologic homeostasis which, contrary to the traditional model [Cann32], is irregular in health and becomes more regular – or "decomplexified" – in disease due to nonlinear, multiplicative feedback mechanisms [Gold92] [IAGH99] [BaLW994]. Consequently, signals of this type are useful for exposing the strengths and limitations of a given signal processing technique.

A significant relationship has been established between the behaviour of the *autonomic nervous system* (ANS) and cardiovascular health. In particular, it has been shown that decreased parasympathetic activation (nervous system relaxation) within the nervous system increases the risk of life-threatening alterations in heart beat and increased sympathetic activity (nervous system excitation) decreases this risk [MaES83] [LoVe76]. Sympathetic and parasympathetic activity influence the force and time between contractions of the heart muscle and, consequently, the nervous system is

manifest in changes in heart beat period [HSSL03]. As expected, a decrease in *heart period variability* (HPV) – also commonly known as heart rate variability [BAGC86] – has been associated with an increased risk of sudden death [SMMW88]. As a consequence, it has been suggested that treatments for cardiovascular health may be better aimed at neurophysiologic triggers rather than simply the heart [LoVe76]. The American College of Cardiology Cardiovascular Technology Assessment Committee regards HPV as a useful metric for risk stratification [ACC93] For a recent review of clinical implications of HPV see Ref. [KlSB05].

A desirable candidate organ for observing the ANS will have an aspect of its operation under direct autonomic nervous system control. Potential candidate organs include the eyes, salivary glands, lungs, heart, digestive tract (including bladder), and reproductive organs. The operation of the heart is very involuntary (more so than, say, the lungs). In addition, its nature is highly electrical, making it particularly accessible to non-invasive study.

The most frequent cardiac measurement based on electrical potential is the *electrocardiogram* (ECG). The ECG is the gold standard for the diagnosis of gross irregularities in heart beat period, also called arrhythmias due to the scope of information it carries regarding contraction effectiveness [Braun97]. The ECG is preferred over other cardiac imaging techniques because it is a one-dimensional time series. Increasing the number of data dimensions exponentially increases the data requirements for classification [DuHa73]. It is already clear that long-term (on the order of 24 hours) ECG records are required to capture infrequent but important behaviour of the cardiovascular system [EwNT84].

Measuring HPV requires identifying a feature within the ECG signal which is symbolic of heart muscle contraction. The most striking feature of the prototypical ECG signal is a sharp upward pulse, denoted 'R' in Fig. 1.5, which marks the largest and most widespread electrically-induced contraction of the heart. HPV data based on the so-called

R-complex is referred to as RR interbeat interval data and the clear variability of the R-complex in the ECG is visible in Fig. 1.6.

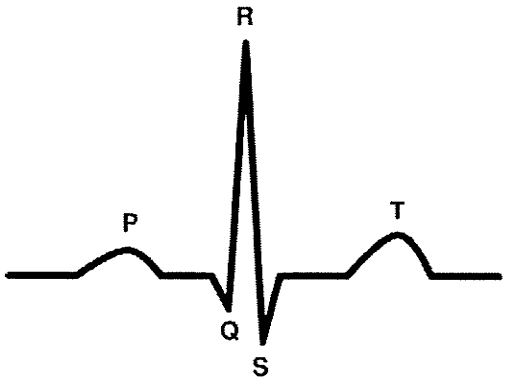


Fig. 1.5 Typical representation if a single heart contraction as measured by ECG. (After [Atki07]).

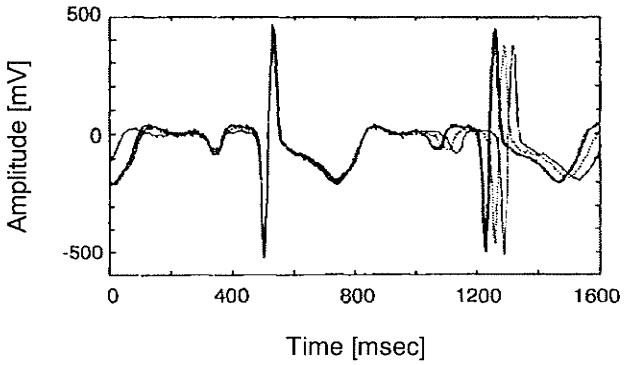


Fig. 1.6 Illustration of heart rate variability. (After [KaSc98]).

Nonlinear dynamics are techniques designed to illuminate the behaviour of nonlinear systems such as those hypothesized to underpin the behaviour of HPV. While results from nonlinear dynamics are promising they continue to lack a complete physiological basis, a state of knowledge represented by two important works in the Proceedings of the Royal Society [GGCS87] and Nature [IAGH99]. With regard to ECG-based HPV

analysis, it is generally believed that the processes underlying HPV are too complex to be accessible through RR interval data alone [KaHA94]. As a consequence, following the work of noted researchers Kantz and Schreiber, the more conservative hypothesis is adopted that the process which governs the contraction onset is effectively stochastic, superimposed by the regulations of the autonomous nervous system which control the average heart rate [KaSc98]. This conservative approach serves as a baseline for the performance of more complex feature extraction techniques such as nonlinear dynamics.

Given data modeled by a stochastic process, it may be beneficial to perform a transformation on the data. The effectiveness of a transform is related to the clarity with which the generating mechanism can be discerned. For example, if an organism is exhibiting its stochastic and time varying state, $s(t)$, in the phase of some observable, $x(t)$, it is useful to have a transform which will reveal $s(t)$ from, say, $x(t) = \sin\left(wt + s(t)\right)$, such as the Fourier transform. The form, however, of $x(t)$ for HPV is unknown, again due to the complexity of the heart contraction initiation timing mechanism. Therefore, the HPV data in this work will be examined without a transformation into another domain, such as the Fourier domain. This is because, even using a transform as familiar as Fourier, large portions of the Fourier spectrum remain poorly understood [KlSB05] [ChKa01].

The complexity of the data also motivates the use of a classifier which is data-driven as opposed to an analytical, parameterized approach based on a model. Moreover, it is useful for the data-driven classifier to possess a linear decision boundary both to focus on the need for effective features over complex classifiers and so the classifiers serve as a basis for approaches which utilize nonlinear decision boundary approaches such as neural networks. The dangers of using complex, nonlinear classifiers in the domain of medical imaging, for example, are well documented in Ref. [ScVS00].

Congestive heart failure (CHF) is a significant cardiovascular condition and causes a significant compensatory response from the human nervous system. Consequently, it is reasonable to expect a detectable change in HPV in people with CHF.

As noted in the Canadian Cardiovascular Atlas, CHF is associated with significant morbidity, mortality and health resource use [LJGH06] [HCGR99] [CGPK99] [TuZh99]. CHF is a growing public health concern because the incidence of heart failure increases with age, and the elderly are the most rapidly expanding demographic group [KaBe91]. The cost (including in-patient, out-patient, and transplantations) of CHF in the United States in 1991 is estimated at $38.1 billion – which is 5.4% of the estimated $700 billion total healthcare expenditure for that year [OCBr93].

Thus, there is motivation, under the baseline assumption of stochasticity, to quantify the success with which a data-driven, derived metric of HPV can be used to discriminate, for the purposes of the medical risk stratification, health from CHF, a condition of wide influence both within the cardiovascular system and modern society.

1.2 Problem Formulation

From the motivation, it is clear that problems may be considered under four main headings: medical decision-making; classification; feature extraction (which elicits the metric used for classification); and all of the above as it is applied to the problem of CHF.

The requirements of an engineering solution for a medical problem will be assessed in Sec. 1.2.1. Classification requirements which emerge from this analysis will be assessed and extended in Sec. 1.2.2. Requirements of feature extraction necessary to support the classification process in the field of HPV analysis will be examined in Sec. 1.2.3. Lastly, present problems in the use of HPV for detecting CHF will be reviewed in Sec. 1.2.4.

1.2.1 Medical Decision Making

In the context of medical decision-making, the medical community is expected to evaluate new and effective methods of pathology classification discovered by engineering within the pertinent risk stratification framework. The unknown in this evaluation include the location of the method within the framework and the resulting cost/benefit. Consider, for example, the possibility of upgrading a widely-deployed ECG monitor with the ability to perform HPV-based CHF detection. Is there sufficient motivation to pay for engineering this upgrade? Are there factors which would, in fact, dissuade a hospital from purchasing a device such as this for its physicians or dissuade a physician from using it in clinical practice? Ultimately, this integration task demands that the pathology classifier be both assessed realistically and in a manner which is credible within the medical community. The issue of credibility has been experienced first-hand by the author's work within mathematical modeling for epidemiology [RKDL06] and will be discussed first.

Credibility is impeded, first, by differences in both the aims and languages of the applied sciences of medicine and engineering. A key weakness, in this regard, of much research within signal processing engineering is the use of mathematics available in software "toolboxes" without appreciation for their underlying assumptions. This manifests itself in the form of studies which, without justification, select several feature extraction techniques for evaluation and present numerical results without a substantial discussion of discrepancies. On the other hand, there is evidence to suggest that medicine suffers from a very low rate of "statistical literacy" [CaSG78] [Cohe94]. Both are weaknesses contributing to the same language barrier problem.

In addition, the credibility of work done in HPV will be scrutinized carefully because the area is new and complex. For example, there are four main compensatory responses to decreased cardiac output: increased contraction force, increased heart rate,

vasoconstriction, and adjustments to total circulating volume [HSSL03]. The relationships between volume, pressure, and the multi-stage contraction of the heart have been studied extensively in the form of pressure-volume diagrams. The pressure-volume diagram has a strong theoretical foundation within thermodynamics. The nuances of heart period, however, have not been as extensively studied nor can their basis be as clearly visualized. By this it is meant that the initiation of a non-arrhythmic contraction of the heart is described simply as a single number in units of beats per minute – with 60 to 100 beats per minute being normal and all other rates being pathological [Lily03].

1.2.2 Classification

The integration task discussed in Sec. 1.2.1 requires a realistic assessment of the pathology classifier. This assessment is statistical and, within the social and life sciences, nearly always in terms of the so-called p value [Cohe94]. The p value is meant, essentially, to estimate the proportion of a sample which will be misclassified (or misdiagnosed) and is calculated via hypothesis testing. This implies a classifier assessment performed in this work should also incorporate the p value to increase credibility. There is, however, abundant criticism of the p value and the somewhat arbitrarily chosen value of $p = 0.05$ (which is often used to indicate a "significant" difference exists between classes and that they, therefore, may be classified). Prominent criticisms of the p value include the following:

- the relationship of a small p value to the ability to replicate the results exists but is not widely-known [TvKa71];
- it is possible to achieve an arbitrarily small p in the limit of sample size [Rose93]; and

- p values characterize the effectiveness of the test but do not characterize the effectiveness of the test in the context of the environment in which it will be performed [Cohe94].

For a full discussion of the limitations of the p value see Ref. [Cohe94] [StSm01].

These weaknesses are addressed by complementing the use of the p value with metrics based on the related statistical assessment technique of Bayesian analysis. Bayesian analysis directly addresses the need for a decision and the steps involved in its calculation include those necessary which achieve the intent of the p value. The components of the Bayesian calculation allow the calculation of misdiagnosis proportions directly from data (without a hypothesis test). Bayesian analysis itself estimates the rate at which these proportions occur while making decisions in the field, essentially providing an estimate of classifier efficiency. If a disease has a particular prevalence then the correct rate of decisions in the field relative to this prevalence represents the increase in visibility the disease has attained as a result of the classifier (or screening test).

The challenge of generalizability and replication faces all studies. This challenge asks the question: how well does the sample under consideration reflect the population at large? This problem requires the use of cross validation, which essentially uses a portion of the data to derive a decision rule and reserves the rest of the data for testing the decision rule.

1.2.3 Feature Extraction

A problem common to both methods of classification is that of choosing an adequate feature. This choice determines how clearly the underlying process is revealed and, as mentioned in Sec. 1.1, feature extraction can be thought of as consisting of BSMS, data reduction, and feature selection.

BSMS, which incorporates both denoising [Dono95] and demixing [Pott06], is not necessary for this work because the HPV data set will be constructed from QRS complexes (as shown in Fig. 1.5), a feature which is extremely large in comparison to the largest source of noise in the analog signal – weakly correlated electrical signals from the muscles, eyes, and brain.

Noise cannot be discounted, however, and additive noise is introduced during HPV data set construction to two ways. First, the relatively low sampling rate of most ECG recorders serves to round sharp edges in the ECG signal. This inhibits the reliability of QRS complex detection software which, in turn, has its own error rate. Empirically determining the magnitude of this noise is a problem which must be solved.

The feature selection stage will be data-driven given the complexity of the physiological generating mechanism. The challenge, therefore, in arriving at a feature selection metric is developing an awareness of the characteristics of the data.

1.2.4 Congestive Heart Failure and HPV

A literature review using the phrases "heart rate variability" and "heart failure" for the time period January 2001 to June 2007 in the following journals was performed:

- British Medical Journal;
- Canadian Medical Association Journal;
- All IEEE journals;
- Journal of the American Medical Association; and
- New England Journal of Medicine.

All HPV metrics used in the articles found were reviewed. In addition, if a study used the phrases "heart rate variability" and "heart failure" but referenced another study instead of doing original work then the HPV metrics used in the referenced studies were reviewed as well.

In total, 55 papers were reviewed. On average, 2.8 metrics were used in each paper though, in IEEE literature, the average was 3.6 and the maximum 11. The detectability metric used in most cases was the p value and this was followed by simple visual inspection in frequency. A Bayesian analysis was performed (on a non-linear feature) only once. The relative frequency of detectability metrics are shown in Fig. 1.7.

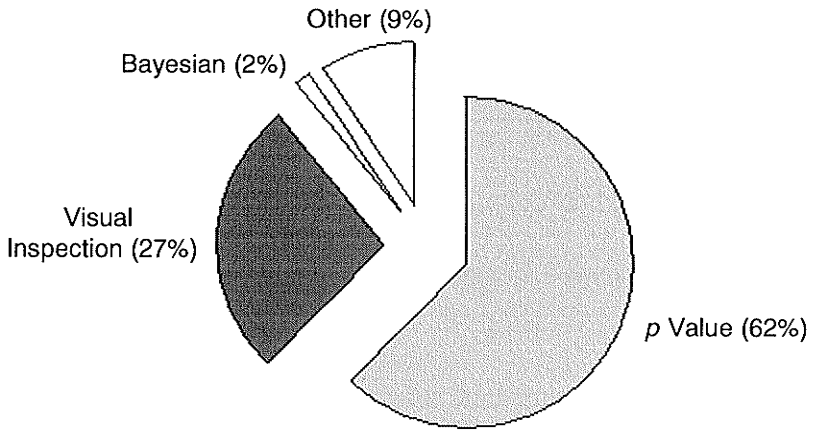


Fig. 1.7 Relative frequency of detectability metrics used in literature.

Standard deviation may automatically assumed to be an inadequate metric unless, for example, the nervous system is assumed to be linear and the central limit theorem applicable. A Bayesian analysis was performed only once in the 55 papers reviewed and 5 instances were found where the importance of HPV was without a reference or without defining how HPV was calculated.

The clinical significance of HPV has existed for about 20 years but the significance of many existing measures of HPV are not appreciated and the conclusions of many studies are on weak or absent evidential basis, as observed in an important report by Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology[Task96].

The ongoing prevalence of multi-metric papers, therefore, suggests the need, firstly, for a forethought and careful metric selection. Secondly, this findings illuminate the need for Bayesian analysis. Thirdly, these findings illuminate the need for Bayesian analysis to suggest limitations on clinical utility. Lastly, these findings highlight the need for cross validation to suggest whether significant deviations from study to study should be expected.

1.3 Thesis Statement

The thesis of this work is that (a) a metric with the power to statistically discriminate individuals with CHF from healthy individuals can be derived from RR interval data, (b) its discrimination power and its limitations can be characterized consistently with results existing in literature (namely via the *t*-test), (c) the results in literature can be expanded on with more powerful discriminant assessments such as error rates per class (termed classifier *coverage*), per decision (termed classifier *efficiency*), and per decision relative to CHF prevalence (termed classifier *amplification*), and (d) these results will generalize given a different sample of comparable data.

The most important outcome of part (a) of the thesis statement is a conclusion as to whether the data of this work is naturally suited to a single metric or whether multiple metrics must be used, as is commonly done in literature. A secondary outcome of part (a) is that physiological implications of the observed statistical character of HPV may be commented on. Parts (b) and (c) will provide insight into

- the proportion of variability which is physiological versus variability which is due to instrumentation,
- the cost at which coverage may be purchased by the HPV-based detector in terms of population-based correct diagnosis rates (efficiency) for an unscreened CHF population (prevalent at the 2% level, as discussed in the background)

- how this cost scales for a population which is screened prior to the HPV test (thereby increasing the prevalence beyond 2%),
- how much more visible unscreened populations (2% prevalence) are after HPV-based detection relative to their prevalence prior to detection, as well as
- how much more visible these screened populations (prevalence $\geq 2\%$) are after HPV-based detection relative to their prevalence prior to detection, and
- what opportunities exist for improving the HPV-based detector.

Part (d) addresses the reliability of conclusions drawn regarding the above outcomes and the repeatability of these conclusions by other researchers.

The thesis statement consists of the following objectives:

(i) to use ECG data with a history of publication in peer-reviewed journals and which is publicly available – this will contribute to the credibility (as defined in Sec. 1.2) of the research outcomes;

(ii) to perform a statistical characterization of the relevant system physiology through a characterization of RR probability distributions – this will provide the rationale to be used in deriving an appropriate metric of HPV;

(iii) to develop a metric for the detection of CHF under an assumption of stochasticity through the statistical characterization of objective (ii) – this is to serve as the basis for a data-driven classifier;

(iv) to obtain an estimate of the noise introduced during instrumentation via an inventory of the steps required to produce the RR data set to be used for analysis – this will be useful in distinguishing what variability is physiological and what variability is due to instrumentation;

(v) to use the *p* value, a metric in wide use within the life and social sciences, as a measure of generalizability and predictor of discrimination effectiveness through the use of a nil hypothesis test – this will address issues of credibility discussed in the problem;

(vi) to obtain the proportion of the CHF sample and proportion of the healthy sample correctly diagnosed as a function of classification metric threshold through the use of *receiver operating characteristic* (ROC) curves – this will expand objective (v) and give insight into test *accuracy* (which represents the quality of the information given by the test);

(vii) to obtain the proportion of positive test results which are true positives and false positives when the prevalence of CHF is considered through the use of Bayesian statistics – this will expand objective (vi) and allow insight into the *efficiency* and *amplification* of the test, which is a stepping stone to assessing test *efficacy* via cost/benefit assignment;

(viii) to draw on objective (vi) and objective (vii) to consider the cost (in terms of efficiency) at which the accuracy of the test is obtained through the study of both specific and general cases – this will provide insight into refinements which should be made to the test and/or the ECG data upon which the test is performed;

(ix) to define the relationship between Bayesian statistics and both the p value and ROC curve during the theoretical development within this work through a literature review and mathematical analysis – this will unite the three test evaluations of objective (v), (vi), and (vii) and establish techniques used in objective (vii) as an expansion of objective (v);

(x) to reconcile results obtained here in literature through a literature review – this will give a sense of the applicability of the results obtained here; and

(xi) to assess the repeatability of this work through the use of cross validation – the outcome of which will indicate whether the results of this work can be expected to generalize given a sample with similar characteristics.

1.4 Thesis Scope

An important strength of this thesis is its use of publicly available data which possesses a publication history. This implies the results of this work may be compared to existing results and duplicated by others. It also, however, implies that the data is not problem-specific. On one hand, this implies results obtained in this work have wide applicability to a general population but, on the other, it increases the number of uncontrolled and unknown variables which, in a sense, become noise. Furthermore, control is lost over reporting of the conditions under which data was gathered.

In addition, how to incorporate the results of this work into medical diagnosis or a risk stratification framework will not be addressed.

1.5 Thesis Organization

The physiological basis for the HPV signal to be analyzed in this work – including the physiology of CHF – is established in the first sub-section of the background, Sec. 2.1. From there, instrumentation, HPV data set creation, feature extraction (including common HPV metrics), and statistical classification (including the relationship of the p value to Bayesian statistics) are described in Sec. 2.2, Sec. 2.3, Sec. 2.4, and Sec. 2.5, respectively. The HPV data set is statistically characterized and an appropriate HPV metric which suits this characterization is chosen in Sec. 3, which fulfills part (a) of the thesis statement. Experiments which evaluate the ability of this metric to discriminate between health and CHF are evaluated according the p value in Ch. 4, fulfilling part (b) of the thesis statement. This discrimination assessment is expanded in Ch. 5, fulfilling part (c) of the thesis statement. Lastly, an experiment which tests the generizability of the results of Ch. 3, Ch. 4, and Ch. 5 is performed in Ch.6, fulfilling part (c) of the thesis statement. Results are synthesized and cohesive conclusions drawn in Ch. 7.

1.6 Summary

This chapter has motivated the use of HPV for medical decision making, referencing numerous examples of the demonstrated clinical relevance of HPV. This chapter has also motivated the use of the electrocardiogram as a general marker of nervous system and cardiac performance and the R-complex of the electrocardiogram as a specific basis for HPV in order to mitigate noise. Congestive heart failure was subsequently identified as a significant cardiovascular condition of modern relevance which may be detectable using HPV. Lastly, this chapter established a central theme for the work – the thesis statement – based on eleven objectives which, in turn, were drawn on problems important to signal processing for medical decision making. At its core, involves the derivation of an HPV metric from data and testing its ability to discriminate health from CHF. Background for these activities is presented in the next chapter, as shown in Fig. 1.8.

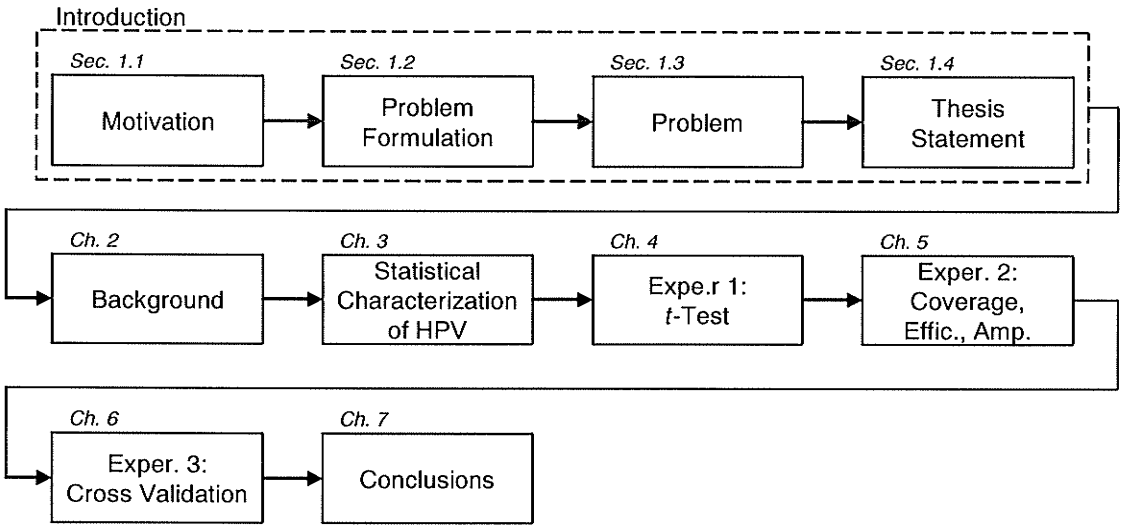


Fig. 1.8 The introduction in the context of other chapters.

Chapter 2

BACKGROUND

2.1 Introduction

The physiological basis for the HPV signal to be analyzed in this work – including the physiology of CHF – is established in the first sub-section of the background, Sec. 2.1. From there, the instrumentation for acquiring the analog ECG signal and converting into digital data is discussed in Sec. 2.2. The intent of feature extraction and the process whereby a data-driven feature will be derived in Sec. 3 are discussed in Sec. 2.3. Lastly, statistical classification (including the relationship of the p value to Bayesian statistics) are described in Sec. 2.5. The relationship between the background in the context of signal processing task it is meant to support is shown in Fig. 2.1.

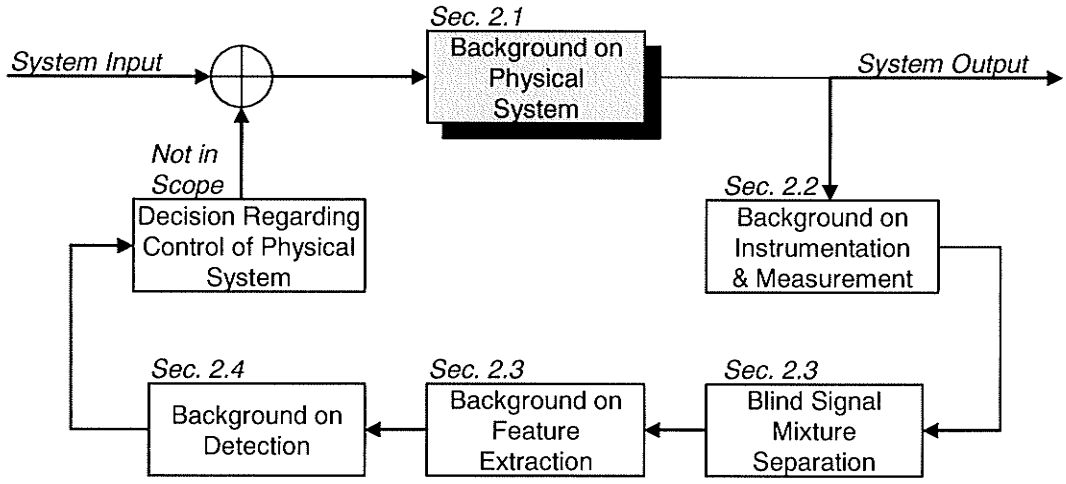


Fig. 2.1 Overview of information covered in the background.

2.2 Background on the Physical System

The physical system – the cardiovascular system – will be examined in terms of its major organs, how these organs are regulated by the nervous system, and how the cardiovascular system behaves when experiencing CHF.

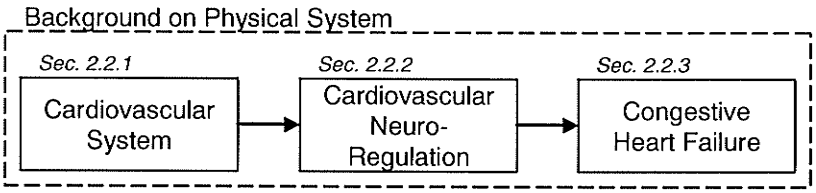


Fig. 2.2 Overview of the physical system sub-section.

2.2.1 Cardiovascular System

The *cardiovascular system* (CVS), also known as the circulatory system, exists to perform thermoregulation and transport substances to and from cells. The most important of these substances are oxygen and *carbon dioxide* (CO_2) and, therefore, the heart's role in oxygen-CO_2 exchange is its most important one. This is recognized by the dominance of the heart and lungs within the cardiovascular system, shown in Fig. 2.3. These organs are complemented by the blood, which is the substance exchange medium, and the vasculature, which exchanges blood between the heart and lungs and serves as the blood conduit to the rest of the body. Together, these entities form the four major building blocks of the CVS.

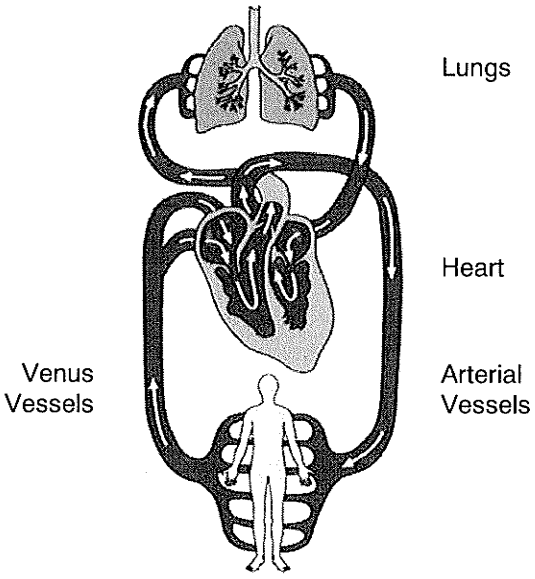


Fig. 2.3 Major components of the CVS. (From [ATR05]).

The CVS operates as a closed, variable volume circulatory system with blood circulating at mean positive pressure. It can be visualized as consisting of two vascular beds. The pulmonary vascular bed receives deoxygenated blood from the right side of the heart and performs oxygen-CO_2 exchange via the lungs. The systemic bed, consisting of the rest of the body, receives and consumes oxygenated blood from the left side of the heart. The heart draws blood passively from the beds into holding chambers, called atria, and then distributes it using muscular blood pumps, called ventricles. These structures and activities are summarized in Fig. 2.4.

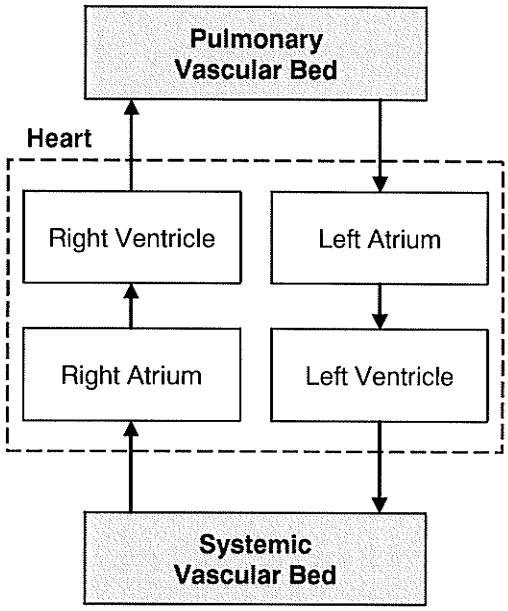


Fig. 2.4 The two vascular beds of the CVS mediated by the heart.

The atrial/ventricular muscles are fixed to a structure of dense connective tissue called the cardiac skeleton, shown in Fig. 2.5. The atria, ventricles, and cardiac skeleton are, in turn, suspended in a low friction sac named the pericardium, shown in a diagram of the heart in relation to the rest of the body in Fig. 2.6. The heart, minus the major vessels, is approximately cone shaped and, in the rear, posses a network of large vessels carrying blood to the heart, called arteries, and away from the heart, called veins. The aorta and superior/inferior vena cava are examples of an artery and vein, respectively. The flow of blood between chambers in the heart is controlled using valves. These structures are shown below in Fig. 2.7. It is interesting to note that the circulatory system is arranged so that circulation in health depends heavily on peripheral vascular factors. For example, for 90-95% aortic constriction, it is estimated that only 25% of the ventricular potential energy is transmitted to circulation [PBGH72].

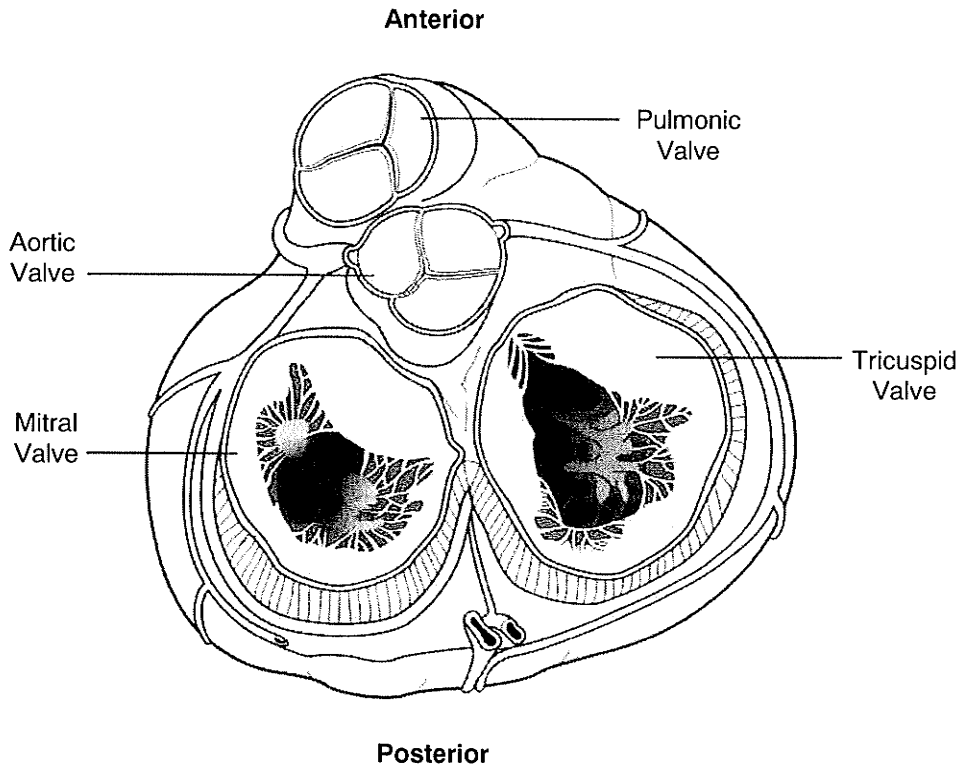


Fig. 2.5 The cardiac skeleton as viewed form above. (From [MaEL03]).

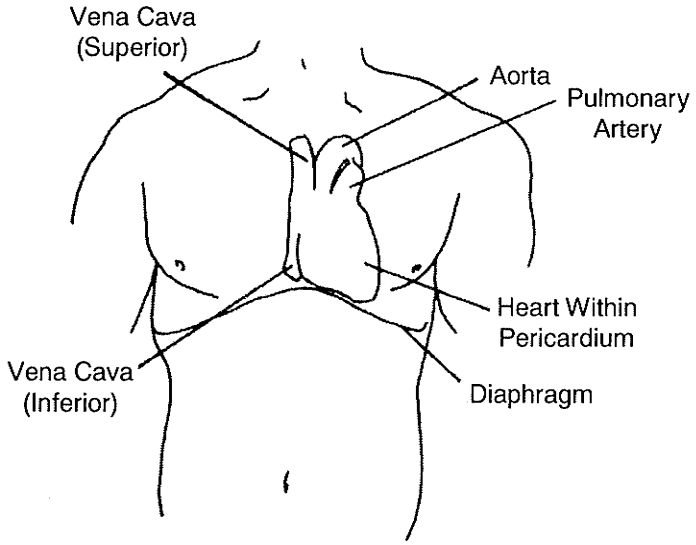


Fig. 2.6 The position of the heart within the chest. (From [MaEL03]).

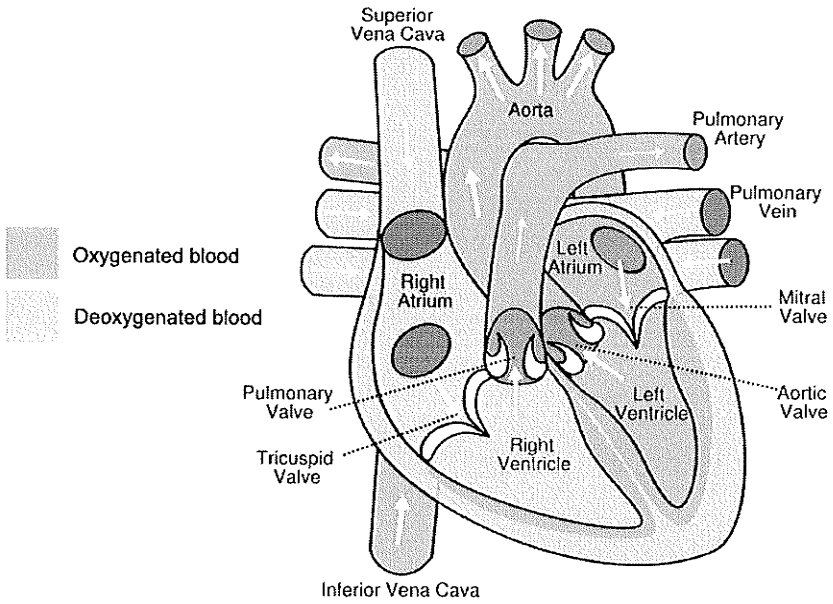


Fig. 2.7 The anatomy and flow of blood within the heart. (After [Pier06]).

Contraction of the heart muscles to effect pumping is caused by the delivery of an electrical trigger. Muscle cells are a member of a special class of cells called "excitable cells" which have a negative electric potential relative to their environment like other cells but, in addition, are capable of both influencing and being influenced by the electrical potential of adjacent cells. Steady state electrical potential results from the diffusion of K^+ ions to the outside of the cell balanced against a resulting net negative charge within the cell which, in turn, restricts K^+ diffusion.

Non-steady state (dynamic) cell potentials occur when cells influence or are influenced by other cells. The act of "influencing" is very binary – the potential is either sufficient to cause an action or it is not. The receipt of a so-called "action potential" in excess of a certain threshold sets off a chain of events to create (essentially re-transmit) a similar action potential. Essentially, the permeability of the cell membrane to ions is voltage dependent and the reception of an action potential initiates a flow of

hypopolarizing ions, raising the potential of the cell from its steady state negative potential. This is referred to as phase 0 by Lily [Lily03] and others. Next, a flow of ions begins which prevents run-away hypopolarization and which begins returning the cell to its resting potential. The simultaneous activation of hypopolarizing and hyperpolarizing ionic flow is phase 1 and 2. During the last dynamic polarization phase, phase 3, the cell quickly repolarizes to its resting potential. The resting potential steady state is known as phase 4 and it is only possible for the next depolarization to occur once stage 4 has been reached. This process is summarized in terms of cell membrane conductance in Fig. 2. 8 and cell potential in Fig. 2.9. On a larger scale, this process proceeds as illustrated in Fig. 2.10.

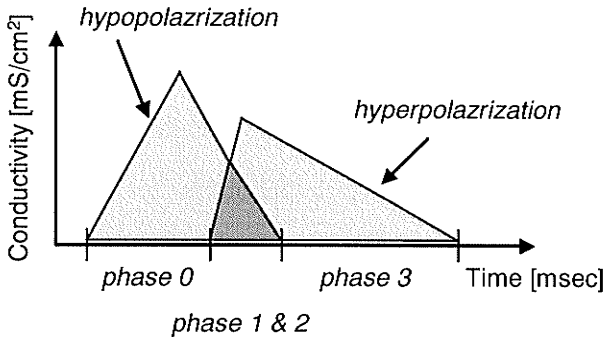


Fig. 2.8 Changes in ionic conductance of the cell membrane during depolarization. (After [Clar97] [HoHu52]).

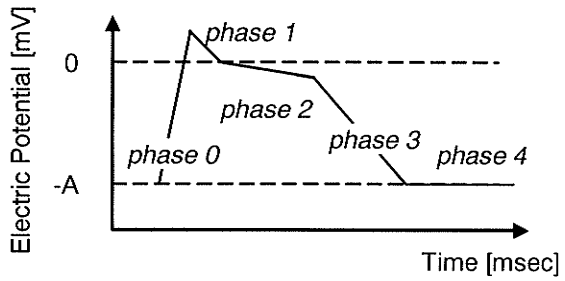


Fig. 2.9 Change in cell potential during depolarization from a resting potential of -*A* mV. (After [Clar97]).

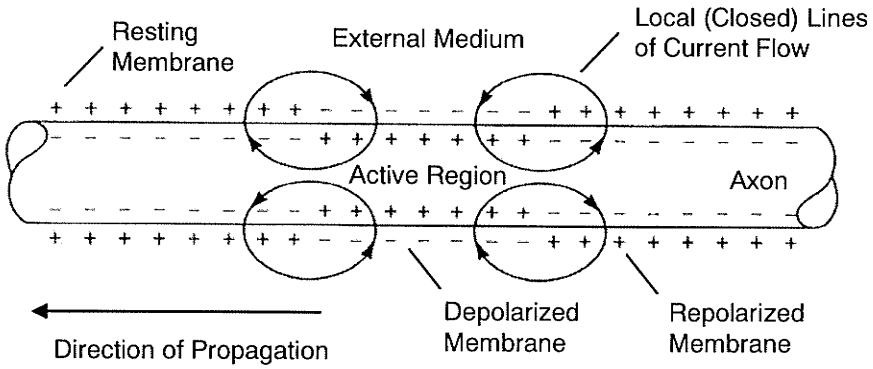


Fig. 2.10 Act of signaling via action potential cascade. (From [Clar97]).

Within the heart, this wave of depolarization is initiated by special excitable cells named pacemaking cells which possess the ability to self-trigger. Pacemaking cells gradually depolarize during phase 4 of Fig. 2.9 until a threshold is reached, at which point an action potential is generated. Within the electrical conduction system of the heart, shown in Fig. 2. 11, the ability to self-trigger (or *automaticity*), is possessed by the *sinoatrial* (SA) node, *atrioventricular* (AV) node, bundle of His, and Purkinje fibres. Note that, within the heart, the atria are electrically isolated from the ventricles and that conduction occurs via the AV node.

Conduction and, therefore, contraction of the associated structures begins at the SA node, proceeds to the AV node and Bundle of His, and then very rapidly to the Purkinje fibres. The last conduction event results in a fast, synchronous, and large-area depolarization of the ventricles. The SA node initiates contraction despite the possession of automaticity by other structures because it discharges more quickly, preventing spontaneous depolarization elsewhere. Arrhythmias result when the sequence of action potential transmission is interrupted, such as when signals arise from other, ectopic sites. It should be noted, despite this high-level understanding of heart excitation, that

conduction geometry is very complex and many clinically observed arrhythmias have not been analyzed or synthesized [Glas99].

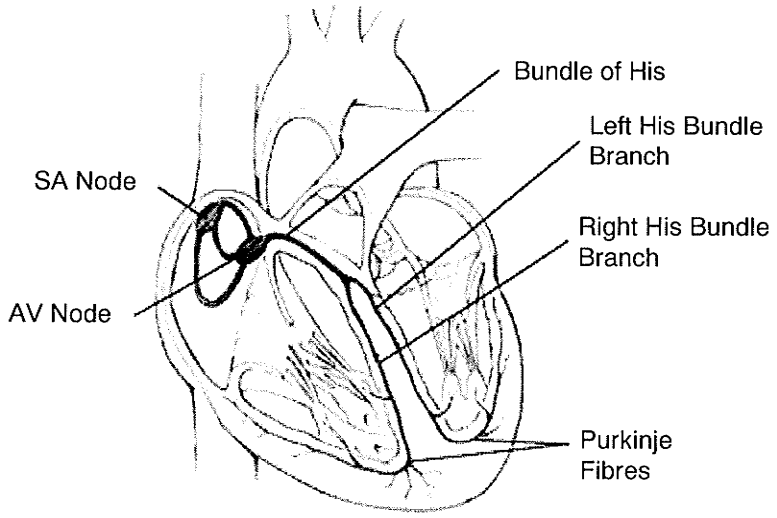


Fig. 2.11 Electrical conduction system of the heart. (From [StJu07]).

The order and effectiveness of contraction can also be visualized through the pressure-volume diagram, which is familiar to students of thermodynamics. A standard pressure-volume diagram for the heart is shown in Fig. 2.12. Aspects of Fig. 2.12 and their relationship to contraction are as follows:

- Segment ab: filling of the left ventricle;
- Point b: mitral valve closes;
- Segment bc: left ventricle contracts;
- Point c: aortic valve opens;
- Segment cd: left ventricle ejection; and
- Segment da: left ventricle relaxation.

The period of contraction is also referred to as *systole* and the period of rest after contraction as *diastole.*

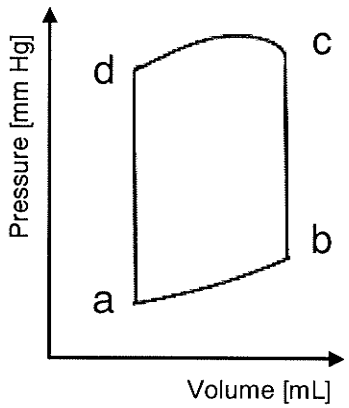


Fig. 2.12 A standard pressure-volume curve for heart contraction

2.2.2 Cardiovascular Regulation

The purpose of the human nervous system, in general, is to coordinate the muscles, monitor the organs, process and inhibit input from the senses, and initiate actions. With respect to the cardiovascular system, the nervous system exists to maintain blood pressure to perfuse the vital organs [HSSL03].

The nervous system, at the highest level, is composed of the *central nervous system* (CNS), which consists of the brain and spinal cord, and the *peripheral nervous system* (PNS), which consists of nerves that reside outside of the CNS to service limbs and organs. Within the PNS exists the *autonomic nervous system* (ANS), which, together with a structure called the medulla oblongata (shown in Fig. 2.13), is responsible for regulating the body's internal state without conscious awareness or intervention. The medulla oblongata serves as both the ANS control centre and as a message relay between the spinal cord and the brain.

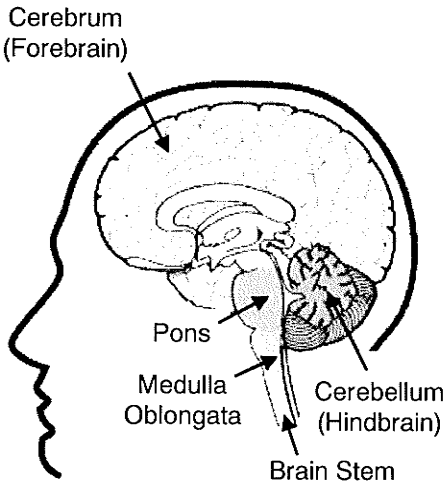


Fig. 2.13 The medulla oblongata in relation to other major brain structures. (After [LeMo07]).

The function of the nervous system with respect to the cardiovascular system, as stated earlier, is to maintain blood pressure to perfuse the vital organs. The ANS is the nervous sub-system which accomplishes this. As shown in Fig. 2.14, regulation is accomplished via four main mechanisms: contractility, heart rate, vasoconstriction, and circulating volume. These influence the pressure-volume diagram and the rate at which it is traversed.

The behaviour of the ANS is accomplished by two complementary behavioural systems called the sympathetic nervous system and the parasympathetic nervous system. The sympathetic system is typically performance-enhancing and the parasympathetic system performance relaxing. Sympathetic communication is accomplished through innervation of the adrenal gland, which releases adrenaline, and the nerve fibres routed through the spinal cord, which control the familiar fight or flight response. Parasympathetic innervation of nearly all thoracic viscera occurs via the tenth cranial nerve, also known as the vagus nerve, which is located within the medulla oblongata.

Feedback is accomplished through the baroreflex mechanism [ZWBS95]. The high level relationships of the regulation systems discussed thus far are summarized in Fig. 2.15.

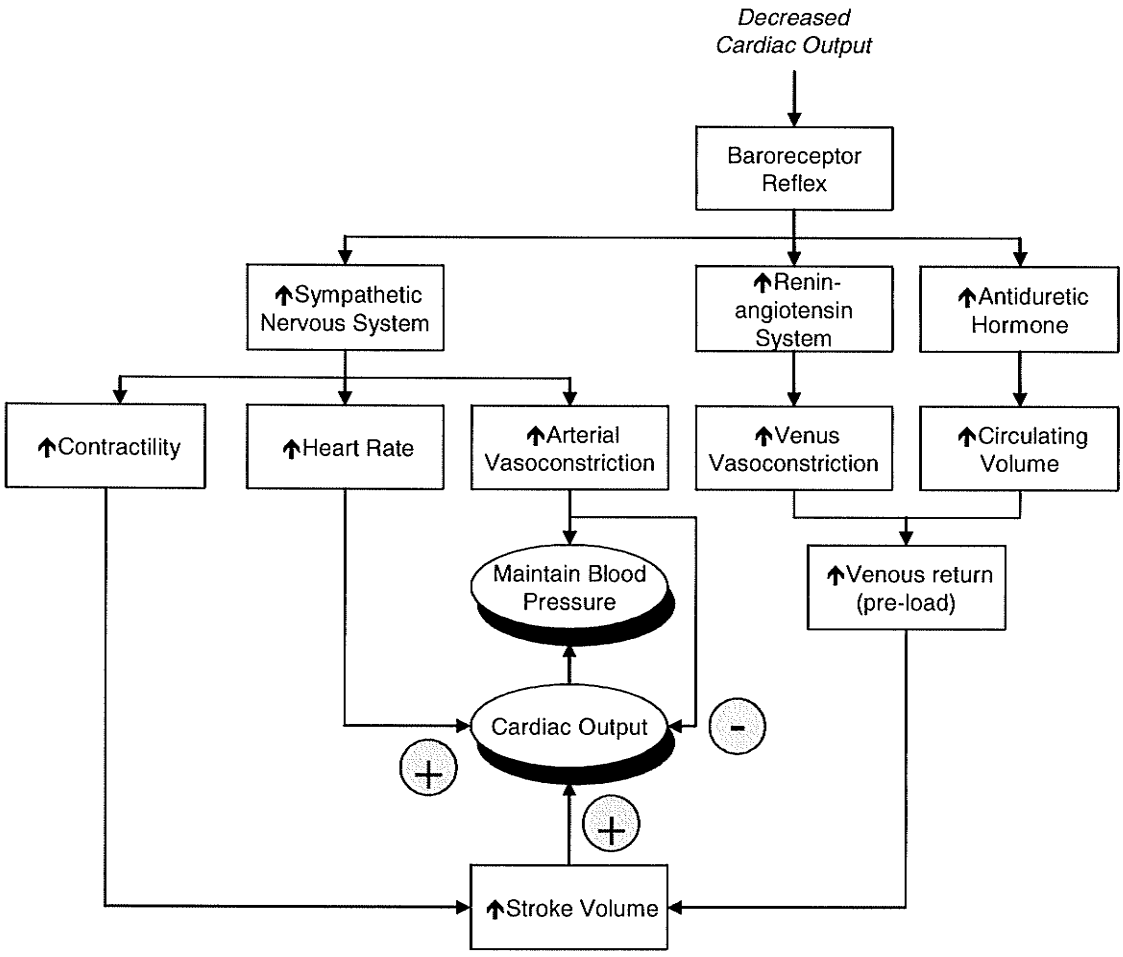


Fig. 2.14 Compensatory neurohormonal stimulation in response to reduced forward cardiac output and blood pressure of heart failure. (After [HSSL03]).

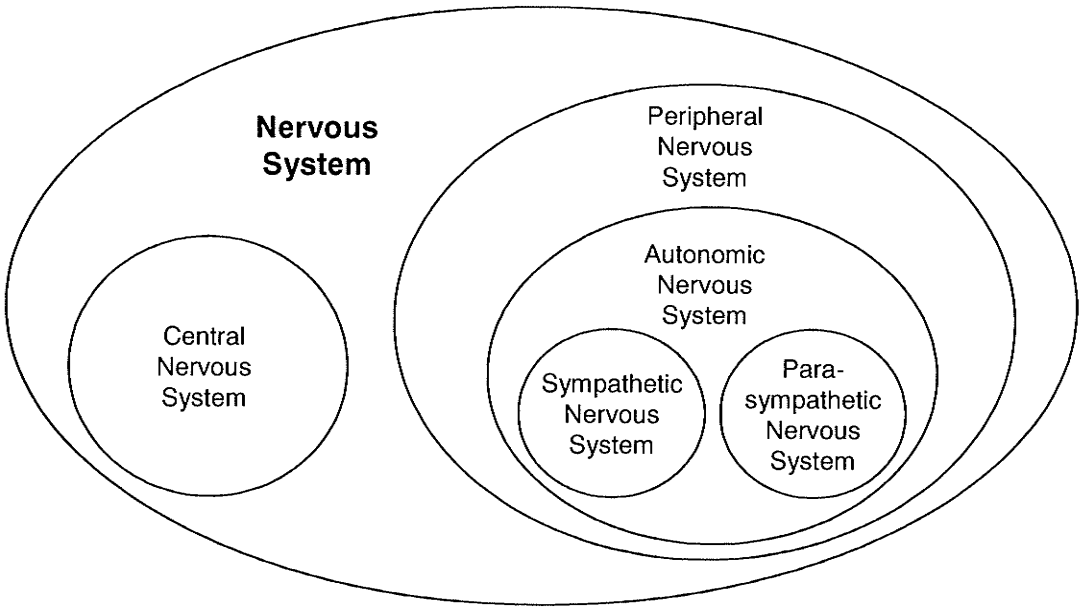


Fig. 2.15 Relationship of major nervous sub-systems to each other.

Communication between the medulla oblongata and the organs it controls is either hormonal or neuronal. Neurons are cells within the nervous system that transmit information via chemically-generated electricity. The structure of a typical neuron is shown in Fig. 2.16 and consists, for the purposes of this discussion, of the following:

- a dendrite which receives the electrical communication from other cells ;
- a series of axons which conduct the electrical communication to the axon terminal; and
- an axon terminal which relays the communication to other neurons and to non-neuronal cells such as muscles or glands.

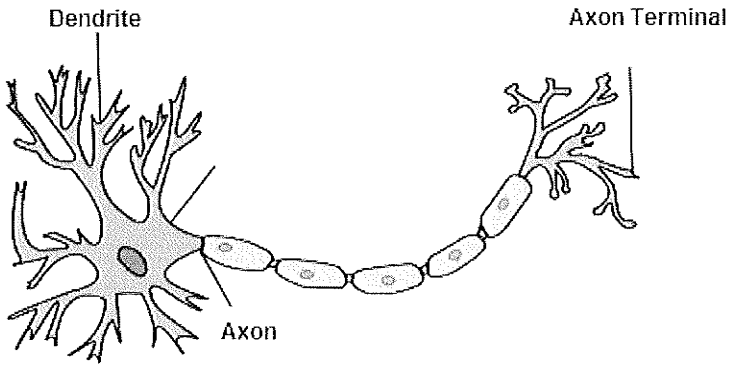


Fig. 2.16 Structure of a typical neuron. (After [Selk07]).

In the case of muscles, each muscle fibre receives inputs from (or in other words is *innervated* by) a single neuron which implies each muscle fibre is in contact with one axon terminal, as shown in Fig. 2.17. This process is often referred to as excitation-contractive coupling.

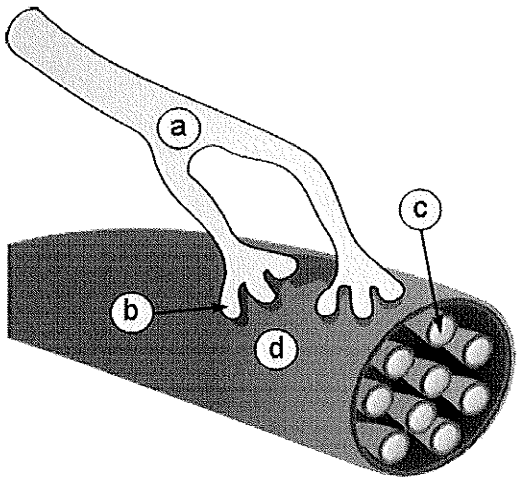


Fig. 2.17 Picture of innervation structure of a muscle showing (a) an axon (b) spreading its axon terminals (c) to activate the myofibrils (special contractive structures within muscle cells) within a (d) muscle fibre. (After [Dake05]).

The heart has been described as "metronomic" [SaBe95] – very regular in mechanical action – following transplant, a finding supported by the work of noted Harvard cardiologist Lily [SALS89] and others. In health, however, the nervous system exerts considerable influence over the cardiac pacemaking cells and clinically-significant variations in heartbeat period are visible in health. Causes of this variation include ventilation, baroreflexes, thermoregulation, circadian changes, exercise, and emotion [KlSB05] [ABSS89] [BKSR89] [BiDu90] [CBDJ84] [Eckb80] [FGCD90] [GrBW90]. This variability in period has been shown to decrease in pathology; for a substantial list of pathological influences consult Ref. [KlSB05].

These variations in cardiac pacemaking period are visible as variations in the pressure-volume of Fig. 2.12, as shown in Fig. 2.18. Figure 2.18, however, does not capture information about the time in which each segment of the pressure-volume diagram is traversed. In fact, the non-arrhythmic contraction of the heart is typically described simply as a single number in units of beats per minute – with 60 to 100 beats per minute being normal and all other rates being pathological [Lily03, p. 86] – despite being one of only four general methods the nervous system responds to decreased cardiac output (refer back to Fig. 2.14). Consequently, a major component of the theory of cardiovascular health is missing.

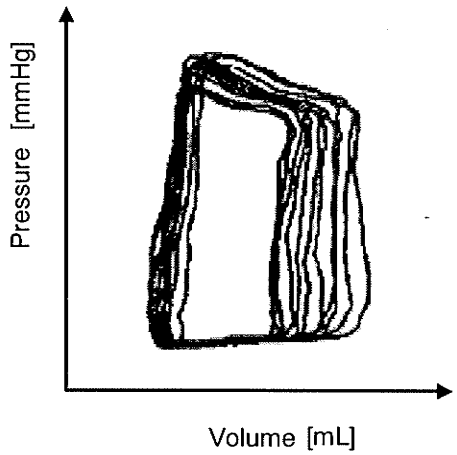


Fig. 2.18 The cardiac pressure-volume diagram exhibiting natural fluctuations. (From [HLOM99]).

2.2.3 Congestive Heart Failure

CHF, also referred to simply as heart failure, is defined as the inability of the heart to pump blood forward at a sufficient rate to meet the metabolic demands of the body ("forward failure") or the ability to do so only if the cardiac filling pressures are abnormally high ("backward failure") [DyFi03]. CHF derives its name from the tendency of heart dysfunction to cause an increase in blood pressure in order to maintain cardiac output. An example is shown in Fig. 2.19, in which an increase in left ventricle diastolic pressure compensates for a decrease in cardiac function. This can, alternately, be viewed as a shift in the volume-pressure diagram of Fig. 2.12.

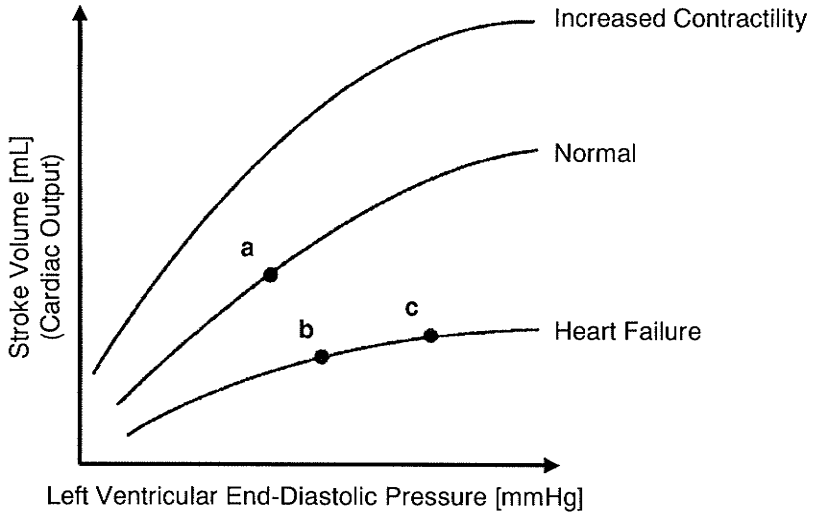


Fig. 2.19 Left ventricular performance for (a) a healthy individual at rest, (b) the same individual after developing systolic dysfunction and heart failure (after, say, a heart attack), and (c) compensatory response via blood pressure. (From [DyFi03]).

Classifying the conditions which can underlie CHF is important because of their large number. To illustrate, the causes of CHF for the dominant form of CHF – left ventricular CHF – are given below in Table 2.1. Causes may be classified in several ways including the following [DyFi03]:

- the side of the heart involved, (left heart failure versus right heart failure);
- whether the abnormality is due to contraction or relaxation of the heart (systolic heart failure vs. diastolic heart failure);
- whether the abnormality is due to low cardiac output or high systemic vascular resistance (low-output heart failure vs. high-output heart failure); or
- the degree of functional impairment conferred by the abnormality (as in the *New York Heart Association Functional Classification* (NYHA) Classes I - IV).

Table 2.1 Causes of CHF for the dominant form of CHF [DyFi03].

SYSTOLIC DYSFUNCTION		DIASTOLIC DYSFUNCTION	
Impaired Contractility	**Increased Systolic Pressure**	**Impaired Ventricular Relaxation**	**Obstruction of Left Ventricular Filling**
▪ Myocardial infarction ▪ Transient myocardial ischemia ▪ Chronic volume overload (mitral/aortic regurgitation) ▪ Dilated cardiomyopathy	▪ Aortic stenosis ▪ Uncontrolled hypertension	▪ Left ventricular hypertrophy ▪ Hypertrophic cardiomyopathy ▪ Restrictive cardomyopathy ▪ Transient myocardial ischemia	▪ Mitral senosis ▪ Pericardial constriction

The NYHA classification has gained widespread acceptance and was originally developed by NYHA, now part of the American Heart Association. Summarized in Table 2.2, the NYHA classification is used by ECG databases such as Physionet to classify CHF severity in a simple but meaningful way.

Table 2.2 NYHA CHF classification [Braun97].

Class	Description
I	Patients with documented heart disease of any type who are completely symptom free
II	Slight limitation of physical activity because symptoms (shortness of breath, chest pain) occur only with more than ordinary physical activity
III	Marked limitation of physical activity because symptoms occur even with ordinary physical activity (e.g., eating meals)
IV	Severe limitation of physical activity because symptoms occur even at rest (e.g., in a sitting or lying position)

CHF is a highly significant condition for health policy makers as it may be the final and most severe manifestation of nearly every form of cardiac disease including coronary atherosclerosis, myocardial infarction, valvular disease, hypertension, congenital heart disease, and the cardiomyopathies [DyFi03]. As noted in the Canadian Cardiovascular Atlas, CHF is associated with significant morbidity, mortality and health resource use [LJGH06] [HCGR99] [CGPK99] [TuZh99]. Within the United States, it is estimated to have a prevalence rate of 2% [AHA07] with 5 million people living with the condition, 400,000 new cases each year, and approximately 12 million medical office visits

[DyFi03]. The 1991 single-year cost (including in-patient, out-patient, and transplants) of CHF in the United States in 1991 is estimated in the Journal of Heart and Lung Transplantation at $38.1 billion – 5.4% of the estimated $700 billion total healthcare expenditure for that year [OCBr93]. Moreover, it is a growing public health concern because the incidence of heart failure increases with age, and the elderly are the most rapidly expanding demographic group [KaBe91].

2.3 Background on Instrumentation and Measurement

This sub-section charts the evolution of the ECG measurement from a continuous time, continuous amplitude analog signal (Sec. 2.3.1) to a discrete time, discrete amplitude digital signal via filtering and sampling (Sec. 2.3.2). This progression is shown in Fig. 2.20.

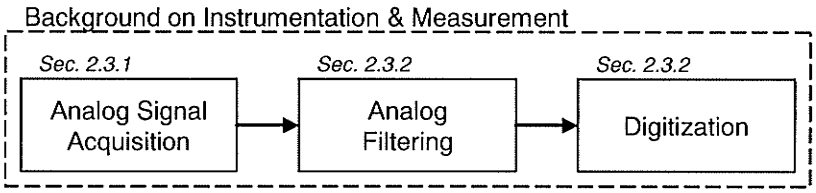


Fig. 2.20 Overview of instrumentation and measurement sub-section.

2.3.1 Analog Signal Acquisition

The acquisition of analog signals requires an exposure to the ECG acquisition signal chain, conduction vectors within the heart, lead placement strategies which are aligned with these conduction vectors, and the electrical waveforms which result from this placement.

The existence of elements within biopotential recording system will be acknowledged but not discussed in depth. Resistive and capacitive effects noted below are parameters

to be minimized and, in general, do not impede diagnosis. For an in-depth review consult Ref. [Neum98].

A block diagram of the analog signal acquisition system for the heart is shown below in Fig. 2.21. The heart muscle is the origin of the ionic current flow responsible for inducing current flow in the volume conductor that is the human body. The human body is generally considered resistive in nature and, therefore, mixing is additive and propagation speeds to the skin surface short (on the order of nanoseconds per metre if the speed of light in water is taken as an upper bound). The dermis is also considered linear but the epidermis and associated sweat glands exhibit a small capacitive effect, as does the gel/electrode interface. The gel and lead wire may be taken as purely resistive in the frequency range of interest for biological signals.

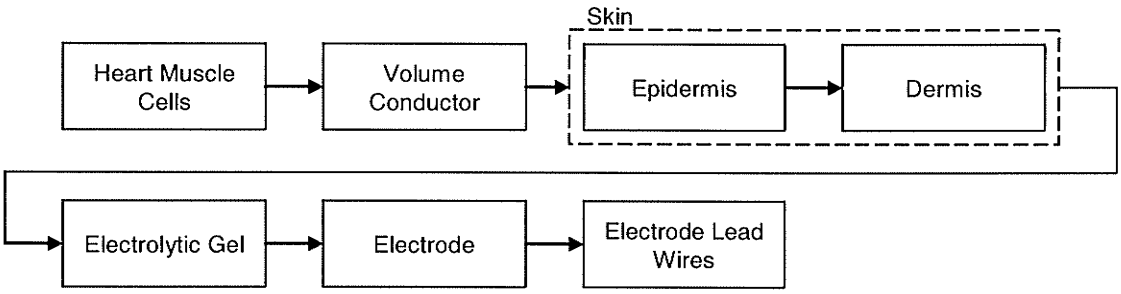


Fig. 2.21 Block diagram of the analog ECG signal acquisition chain.

Electrode lead wires must be placed properly to obtain a three-dimensional measure of depolarization direction and magnitude. In a single direction, this measurement occurs as shown in Fig. 2.22. If the single direction measurement of Fig. 2.22 is performed on a planar electrical wave then the outcome of the measurement will be the dot product of the one-dimensional measurement direction with the two-dimensional wave. Logically, a complete characterization of the plane could be obtained through multiple single-dimensional measurements at different orientations.

A full characterization of the heart is accomplished via measurements taken in the frontal (parallel to the length of the body) and transverse (perpendicular to the length of the body) anatomical planes. Different measurement directions will be more important than others at particular points in time depending on the direction of depolarization.

Leads are placed to each coordinate axes in six radial directions in both the frontal and transverse planes to achieve the standard 12-lead ECG. The lead placements and sample depolarization waveforms along each of these leads are shown in Appendix A. The clinically acquired set of waveforms is shown in Fig. 2.23 illustrate that different measurement directions will be more important than others at particular points in time depending on the direction of depolarization

The consequence of performing this measurement is the well-known prototypical form of the ECG shown in Fig. 1.5. Waveform features and the corresponding cardiac events are as follows:

- Segment PQ: delay of conduction at the AV node;
- Segment QRS: depolarization of the Purkinje network; and
- Point T: ventricular repolarization.
- Point P: depolarization of the atria;

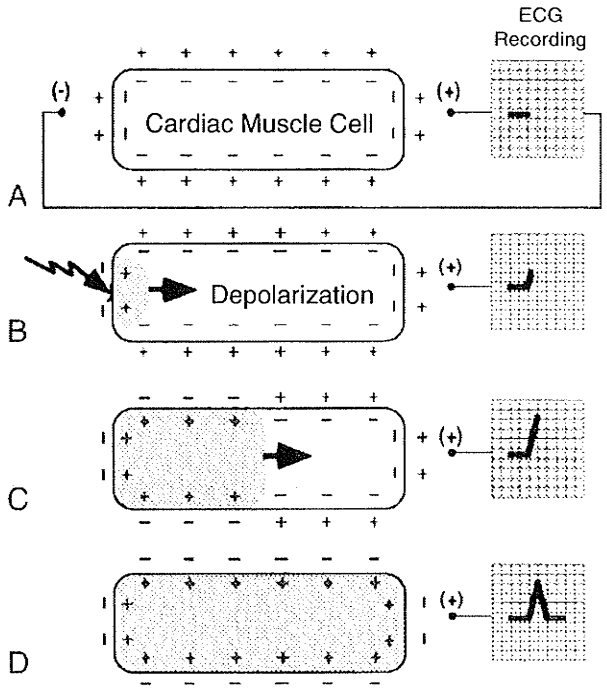


Fig. 2.22 Depolarization of a single cardiac muscle cell showing (a) steady state, (b) action potential, (c) spreading depolarization, and (d) completed depolarization. (From [Lily03]).

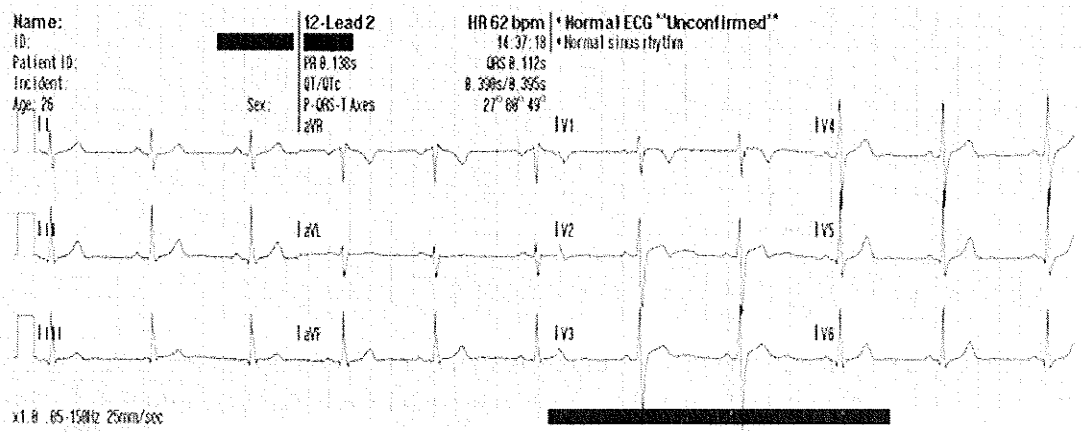


Fig. 2.23 A sample 12 lead electrocardiogram showing signals from the frontal plane (leads I-III, aVR, aVL, aVF) and transverse plane (leads V1-V6). (From [Wiki07]).

These features are sufficiently demonstrative of cardiac function that the ECG is the gold standard for the diagnosis of arrhythmias [Braun97]. In this context, the normal single-cycle heartbeat can be precisely defined as follows [Lily03]:

- every P wave is followed by a QRS;
- every QRS is preceded by a P wave,
- the P wave is upright in leads I, II, and III;
- the PR interval is greater than 0.12 sec; and
- the time between successive QRS complexes is 0.6 – 1 sec (corresponding to 60 to 100 beats per minute).

2.3.2 Analog Filtering and Digitization

Once the analog signal has been acquired it can be digitized in preparation for digital signal processing. Digitization entails discretizing a measurement of electrical potential in both time and amplitude. In time, this involves filtering in the frequency domain followed by sample-and-hold. Filtering, at a minimum, must consist of a low pass filter with a cutoff less than Nyquist's threshold. For ECGs, this low pass cutoff is typically 40% of the sampling rate [BBGH90]. Many times, high pass filtering is included at 0.1 Hz to eliminate drift followed by notch filtering at both 50-60 Hz and 30-40 Hz to eliminate the electrical mains signals and muscle tremor, respectively.

It should be noted that, if the goal in processing the ECG is QRS detection (as it is in this work) an absolute minimum lowpass filter cutoff point is 30 Hz since this is the upper bound for QRS fundamental frequencies [BBGH90]. Reliability of QRS detection will still be influenced, however, unless the low pass cut off frequency and sampling rate are higher [BePi67] [BFBD77] [BLWP77]. The American Heart Association, referencing an earlier study published by the IEEE, has stated that it is "clear" that a minimum of bandwidth of 500 Hz is necessary to capture all high frequency content of

the QRS complex [GoWH73] [BBGH90]. Therefore, lowpass filtering with a cutoff lower than 500 Hz will result in information loss.

In the amplitude domain, the resolution, Q, of the conversion process will depend on the analog full-scale voltage, V_{FS}, and the number of bits, n_b, according to $Q = V_{FS}/n_b$. For a linear quantization scheme, Q is also referred to as the dynamic range. ECG recordings frequently use 12 bits over a $\pm$ 10 mV full scale.

2.4 Background on Feature Extraction

The bulk of feature extraction is typically performed in the digital domain and the intent is to minimize aspects of the digital signal which do not elicit the state of the system and enhance those which do. In the language of signal processing, feature extraction is meant to reduce redundancy, irrelevancy, and the dimensionality of the data set so the content approaches minimal sufficiency for classification. Feature extraction will be discussed in terms of BSMS, data set size reduction, the principles of feature selection, and the feature selection process as it will be performed in Ch. 3 of this work. This development of ideas is shown below in Fig. 2.24.

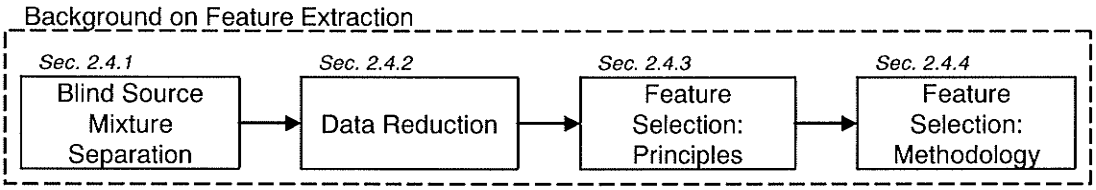


Fig. 2.24 The stages within the background development of feature extraction.

2.4.1 Blind Signal Mixture Separation

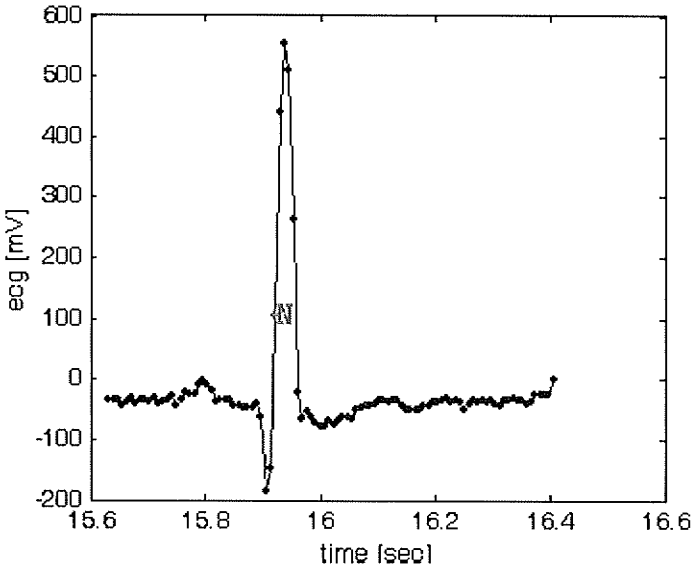
BSMS, which incorporates both denoising [Dono95] and demixing [Pott06], is not necessary for this work if a robust indicator of cardiac contraction onset can be found to

mitigate noise due to other biologically-created electrical potentials and low sampling rates. For an exposure, however, to demixing techniques applied to a problem in which they are critical see the fetal ECG separation work of Potter and Kinsner [Pott06].

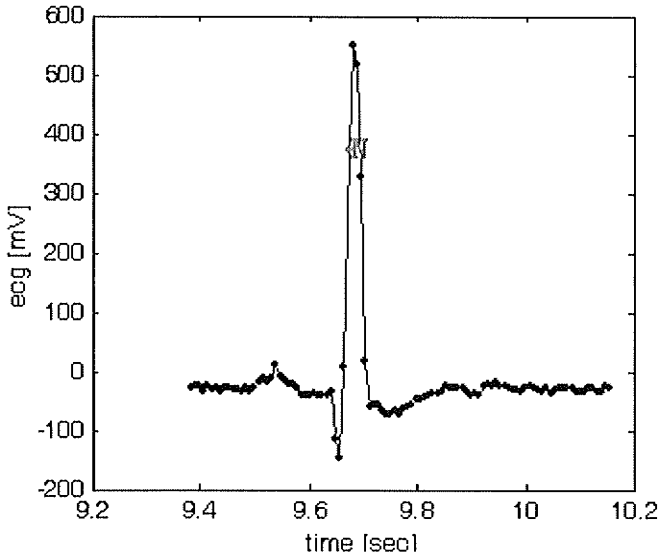
2.4.2 Data Reduction

Data reduction is a digital pre-processing step for dramatically reducing data set size. For this work, it entails the transformation of the ECG time series into the RR interbeat interval data. This occurs by first selecting a landmark on the ECG single cycle signal to symbolize the presence of a heart contraction. The most obvious and, consequently, robust of these landmarks is the landmark described the points Q, R, and S on the ECG waveform, shown earlier in Fig. 1.5. A robust, detectable landmark is of particular concern in the this application due to the typically slow sample rate used to acquire the digital ECG.

Given this landmark, an algorithm must be used to detect it. Many algorithms for this exist. See, for example, a review in Ref. [KoHO02]. In practice, all peak detection algorithms will incorrectly identify a sample as a peak some percentage of the time, introducing variability into the RR data set which is not physiological. Data in the Physionet database, for example was found by the author to have 1 to 2 samples of error. The presence of error is illustrated in Fig. 2.25 using data from Physionet Normal Sinus Rhythm Database record 16265 [GAGH00]. After detection, the RR interbeat interval data set may be constructed in a straightforward fashion, as shown in Fig. 2.26.



(a)



(b)

Fig. 2.25 Measurement noise caused by normal R-complex (detected location denoted by 'N') being assigned at (a) two samples, and (b) one sample from the actual peak.

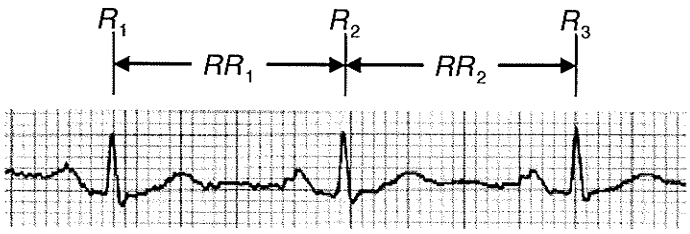


Fig. 2.26 Creation of RR interval number *i*, $RR(i)$, from the QRS complexes R_{i-1} and R_i. (After [Knu05]).

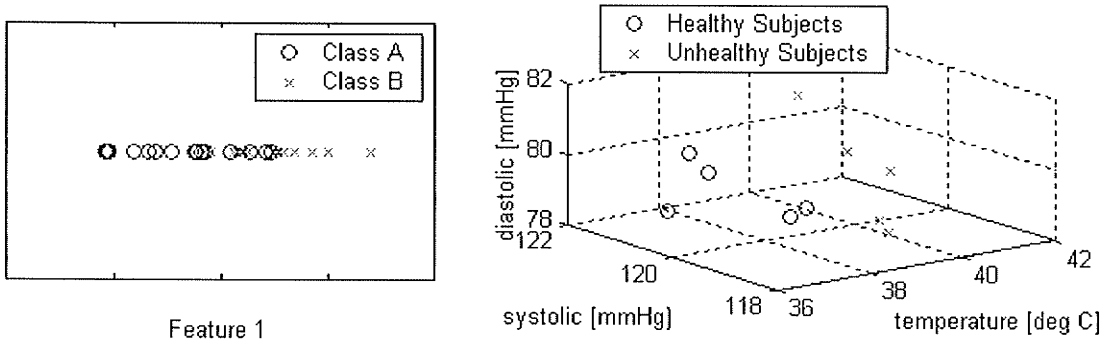
2.4.3 Feature Selection: Principles

Feature selection aims to clearly illuminate the state of the physical system underlying the data from the reduced data set – which ideally represents the minimal set necessary to do so.

Let the ingredients of feature extraction be formally defined so a precise basis exists for an intuitive and informal discussion of how they are used. First, feature extraction extracts state through the computation of M features from raw data whose real-valued magnitudes populate an M-dimensional subspace known as feature space, denoted $D \subset \mathbb{R}^M$. Let the notation $\{d_i \mid C_i\}$ represent the set of M-dimensional magnitudes, d_i, associated with class C_i. d_i, in turn, is composed of the set of magnitudes along dimension j, $d_i = \left\{d_i^j\right\}_{j=1}^{M}$, where $j \in \mathrm{I}$. For a physical system with N states, the set of all possible data points, d, can then be denoted $d = \{\{d_1 \mid C_1\},...,\{d_N \mid C_N\}\}$. Let the kth point within d_i be indexed as $d_{i[k]}$.

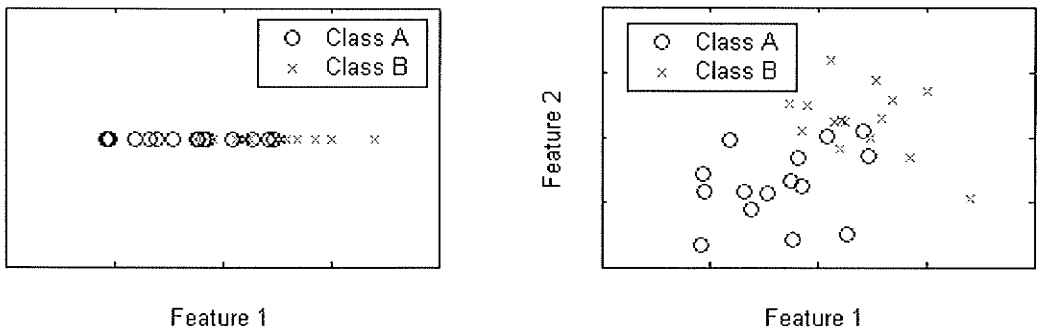
$d^0_{0[0]}$ and $d^1_{0[0]}$ could, for example, be the temperature and blood pressure of patient number 0 in a hospital. The feature space would, therefore, be two-dimensional feature space and be used for diagnosis. Intuitively, it is helpful for classification if body temperature in health is significantly different than body temperature in sickness. Note that, following commonly used terminology, the terms "class" and "classification" shall be synonymous with "physical state."

Note, also, that it is helpful if the number of variables is kept to a minimum for two reasons. First, it helps minimize the number of variables to consider simultaneously and, second, it minimizes the amount of data required. The latter is simply a consequence of the tendency of clusters of points to become less concentrated as they are embedded in higher dimensions. This diluting effect is illustrated below in Fig. 2.27. The cost of more data and, as a result, a more complicated decision boundary in feature space can be tolerated if higher dimensionality significantly increases discrimination power. This is illustrated in Fig. 2.28.



(a) (b)

Fig. 2.27 The dilution of data density as dimensions are added from (a) one dimension to (b) two dimensions.



(a) (b)

Fig. 2.28 Example in which the addition of a dimension increases discrimination power (a) from a single-feature feature space to a (b) two-feature feature space.

Pattern recognition in natural systems such as the cardiovascular system is potentially much more complicated and less reliable than in human engineered systems such as a digital transceiver. Performing pattern recognition in a natural system is, in a sense, like defining a receiver for an intricate digital transmitter for which no documentation exists. It is illustrative to briefly compare the two problem types to rationalize the conservative stochastic approach taken to feature extraction in this work as a result of cardiovascular complexity.

Take, for example, the problem of +3.3 V digital data contaminated by a considerable amount of additive noise. The expression for the received signal, $r(t)$, in this case is $r(t) = m(t) + n(t)$ where $m(t)$ is simply either 0 or 3.3 V, depending on whether the received value is a digital 0 or 1, respectively. Using voltage amplitude as the feature, feature space can be constructed as shown in Fig. 2.29, where each point represents a voltage measurement at point in time.

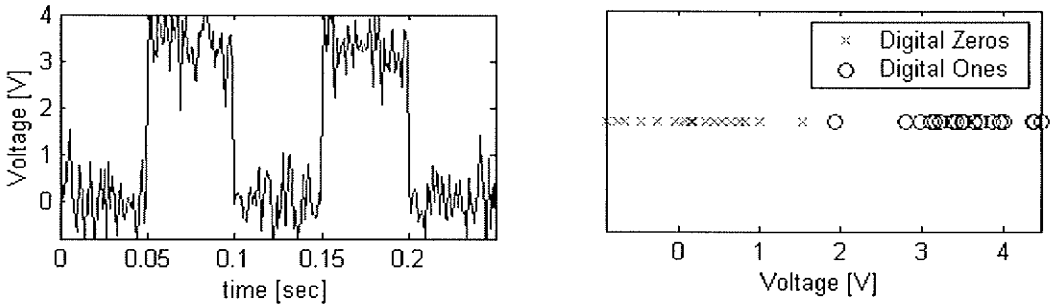


Fig. 2.29 *Amplitude shift keying* (ASK) digital transmission showing (a) the received signal contaminated with noise, and (b) the resulting feature space.

Alternately, one may consider information in the phase domain, in which case the feature of interest is the phase. For *quadrature phase shift keying* (QPSK), $r(t)$ is given by $r(t) = \sin\left(wt + m(t)\right)$ where $m(t)$ may take on one of 0, $\pi/2$, π, or $3\pi/2$ radians.

The feature space for an FPGA-based QPSK decoder the author has implemented in industry is shown in Fig. 2.30.

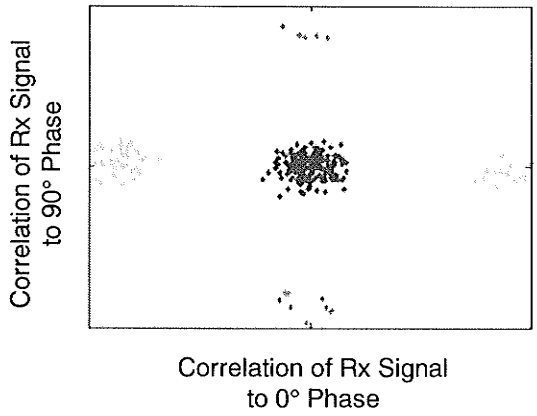


Fig. 2.30 Feature space of an FPGA-based QPSK decoder implemented by the author in industry.

These examples illustrate that if the unknown transmitter is ASK then only $m(t)$ must be decoded and it is available directly. In the case of QPSK, however, the received signal must first be interpreted in terms of a sinusoid before examining its parameters – one of which contains $m(t)$ which, in turn, must be decoded. This implies a very large range of potential $r(t)$ given a physical system with obscure inner workings.

This complexity is particularly true of human cardiovascular signals. This is likely due to multiplicative, nonlinear feedback loops [Gold92] [IAGH99] [BaLW994]. The significance of this is that even simple non-linear feedback systems are capable of demonstrating very complicated dynamics. A simple illustration of this can be seen with the discrete version of the so-called logistic equation $x_{n+1} = \mu x_n(1-x_n)$, $0 \le \mu < 4$ which, depending on the value of μ selected can exhibit either a very small or an infinite number of system states. This behaviour is demonstrated in Fig. 2.31. The promise of such a possibility is that the cardiovascular system, for all its complexity, may, in fact, have a very eloquent mathematical description. Such a formulation, however, remains

beyond the grasp of science to date [IAGH99] [GGCS87] [Duec04]. Thus, the lack of even linearity (and therefore the principle of superposition) in the cardiovascular system and the infeasibility of performing a truly in-depth physiological study for this work motivates an "Occam's Razor" model of stochasticity for this work.

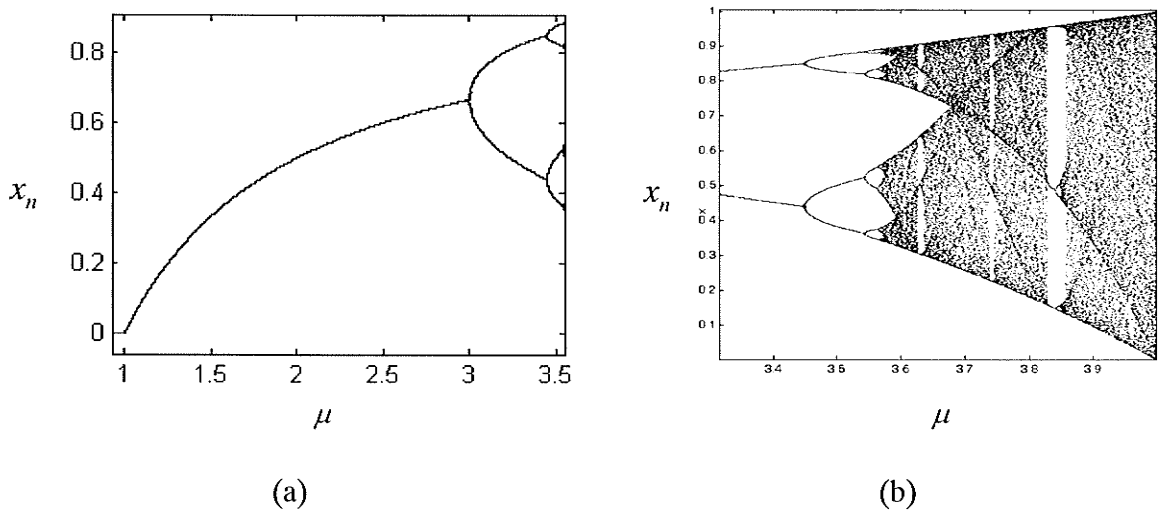


Fig. 2.31 The fully deterministic (non-stochastic) logistic equation demonstrating (a) a small number of steady states, and (b) infinite, non-repeating number of steady states.

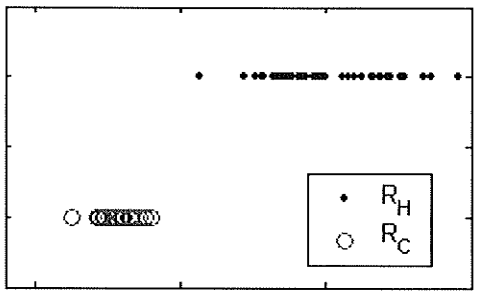
2.4.4 Feature Selection: Methodology

It is clear, from the examples of Sec.2.4.3, that data clusters which are amenable to classification possess a spread (i.e. variability) which is less than the separation between the clusters. To obtain this result by design, it is useful to visually inspect the variability and separability of the data as a prelude to a formal, quantitative analysis. The visual inspection process – essentially an assessment of feasibility – will be discussed below. Its intent will be to inspect raw data for a feature which is separable and consistent. A promising feature, if found, will be assessed quantitatively using the techniques discussed in Sec. 2.5.

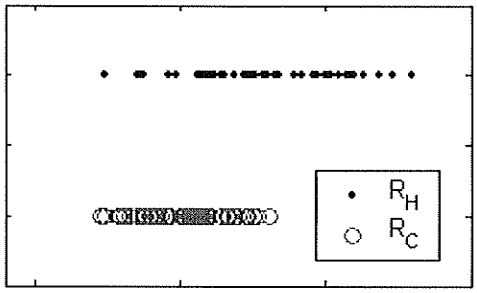
Consider, first, some definitions so that feature assessment methodology may be specified succinctly. Let the class of all healthy subjects, S_H, be defined as $S_H = \bigcup_{i=1}^{N_H} S_{H[i]}$ where N_H is the number of healthy subjects and $S_{H[i]}$ is subject i of the healthy class. Let the class of all CHF subjects, S_C, be defined as $S_C = \bigcup_{j=1}^{N_C} S_{C[j]}$ where N_C is the number of CHF subjects and $S_{C[j]}$ is subject j of the CHF class. Let, also, $R_{H[i]}$ represent the set of all RR interbeat intervals for the ith healthy subject where $i = 1..N_H$ and, similarly, $R_{C[j]}$ the set of all RR interbeat intervals for the jth CHF subject where $j = 1..N_C$. Consequently, the set of all healthy RR interbeat intervals may be defined as $R_H = \bigcup_{i=1}^{N_H} R_{H[i]}$, the set of all CHF RR intervals as $R_C = \bigcup_{j=1}^{N_C} R_{C[j]}$, and the set of all RR intervals as $R = R_H \bigcup R_H$. As a result, both $R_{H[i]}$ and $R_{C[j]}$ may be alternately defined as $R_{H[i]} = \{R \mid S_{H[i]}\}$ and $R_{C[j]} = \{R \mid S_{C[j]}\}$. Also, it is then possible to define the set of all healthy RR interbeat intervals, R_H, as $R_H = \{R \mid S_H\}$, the set of all CHF RR interbeat intervals, R_C, as $R_C = \{R \mid S_C\}$. Lastly, let the set of all RR interbeat intervals, R, be defined as $R = R_C \cup R_C$.

Separability, in this work, considers whether, at the population level, the general potential for discrimination at the individual level exists. This sense can be obtained by plotting R_H and R_C. Potential outcomes of this exercise are shown in order of ascending discrimination complexity in Fig. 2.32a, Fig. 2.32b, and Fig. 2.32c.

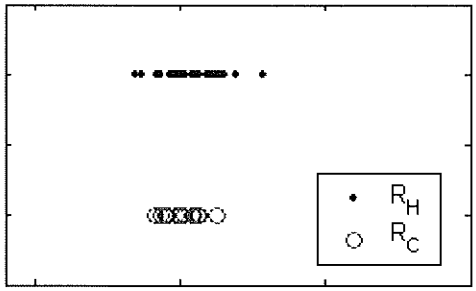
Inter-subject variability reflects how well feature separation at the population level will generalize and requires the inspection of $R_{H[i]}$ and $R_{C[j]}$. For example, both Fig. 2.33a and Fig. 2.33b are capable of producing the separation shown in Fig. 2.33c. The difference is that if the experiment of Fig. 2.33a were replicated on a different data set, it would have a much lower chance of producing the same results than the experiment of Fig. 2.33b. Consequently, it is desirable to have features which exhibit consistency from sample to sample.



(a)

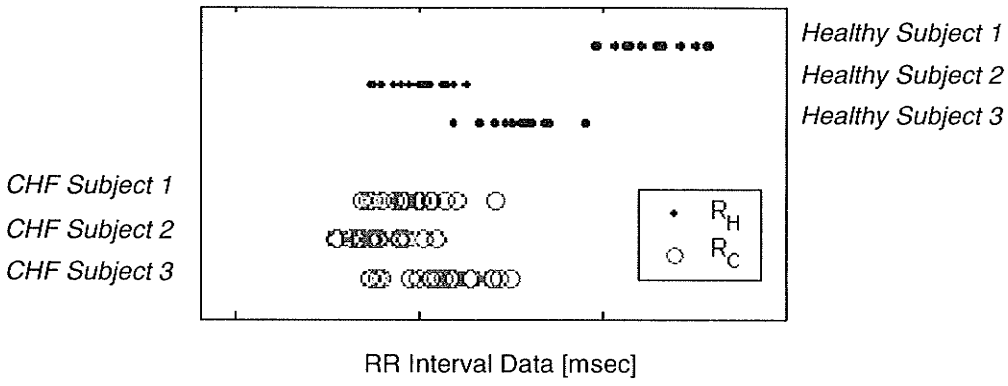


(b)

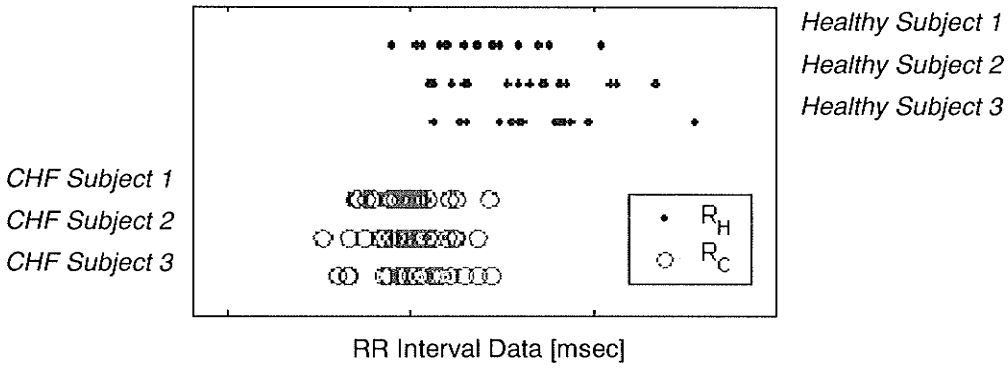


(c)

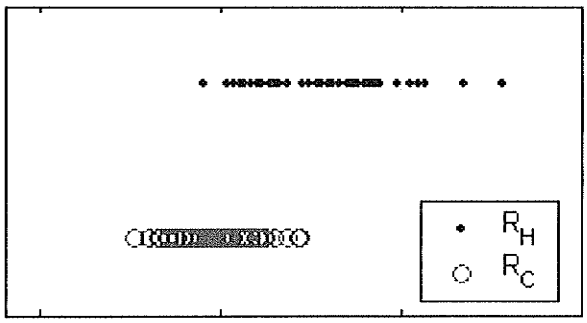
Fig. 2.32 Population-level plots of RR interval data showing examples of data which is generally (a) well-separated, (b) partially separated, and (c) poorly separated.



(a)



(b)



(c)

Fig. 2.33 Illustration with one-dimensional features of how both (a) high inter-subject variability and (b) low inter-subject variability (c) can create the same point density distributions when taken cumulatively as R_H and R_C.

The ultimate outcome, given a separable, consistent dataset, is the derivation of a HPV measure, $\mu(\cdot)$, which will assign a single measure, $M_{H[i]} = \mu\left(R_{H[i]}\right)$ or $M_{C[j]} = \mu\left(R_{C[j]}\right)$, to a single long-term record, $R_{H[i]}$ or $R_{C[j]}$, respectively, in manner which allows linear discrimination. Feature clusters are said to be separable via linear discrimination if the simplest decision boundary between them follows a straight line.

Common time domain HPV metrics appearing in literature include the following [Task96]:

- the mean of all RR interbeat intervals;
- the standard deviation of all RR interbeat intervals;
- the *root mean square of successive difference* (rMSSD) which is intended as a measure of standard deviation for data with a non-stationary mean [NKBH41];
- the number of RR interbeat interval changes exceeding 50 milliseconds (a time interval chosen empirically), a metric often denoted NN50 [EwNT84];
- the percentage of RR interbeat interval changes exceeding 50 microseconds, often denoted pNN50 [EMYC85]; and
- the percentage of RR interbeat interval changes exceeding x milliseconds, often denoted pNNx [MPHG02].

2.5 Background on Detection

A method of clearly visualizing the distribution of point densities of Sec. 2.4 is the probability distribution, which will be discussed in Sec. 2.5.1. The probability distribution forms the foundation of all three tests of HPV metric performance described in this section: the t test (Sec. 2.5.2), *coverage* via ROC analysis (Sec. 2.5.3), and both *efficiency* and *amplification* via Bayesian analysis (Sec. 2.5.4). The first two tests both possess a relationship to the third and these relationships are discussed in Sec. 2.5.6 and Sec. 2.5.7. This evolution of ideas is diagrammed in Fig. 2.34.

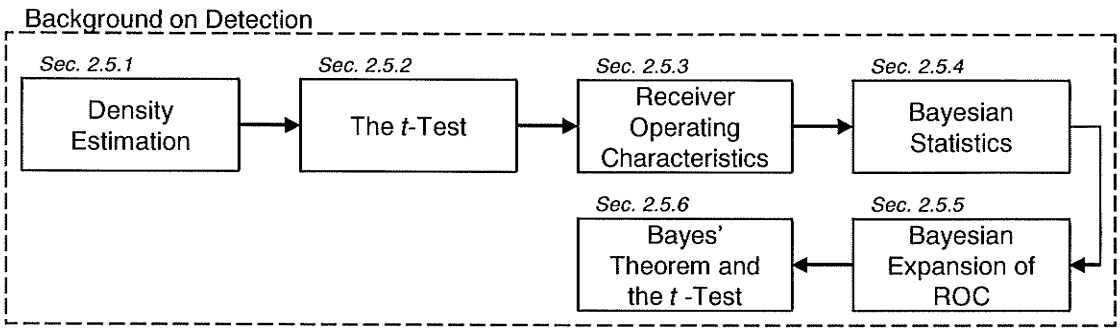


Fig. 2.34 Overview of the detection sub-section.

2.5.1 Probability Density Function Estimation

There is a need to clearly visualize the process which creates the distributions of feature point density in feature space. This is accomplished via the construction of a density function which plots, on the dependent axis, a curve which is proportional to the density of the discrete or continuous support beneath it on the independent axis. The support of a function is the smallest closed set, $S = \text{supp}\, P \subset \mathbb{R}$, such that the values of the function are zero everywhere on the complement of S. If the support, x, is continuous then the density function is referred to as a *probability density function* (pdf) and denoted $p(x)$. If the support, x, is discrete then the density function is referred to as a *probability mass function* (pmf) and denoted $f(x)$.

Formal probability theory considers a physical process which produces a set of observable events, E, and $\varepsilon \subset \Omega$, $\forall \varepsilon \in E$ for the sample space of all possible outcomes, Ω, of the physical system. These three quantities define a probability space denoted (Ω, E, P) in which P represents both a probability measure and the functional mapping $P: E \to \mathbb{R}$, $P \in [0,1]$. P can, conceptually, be thought of as the set of all measures performed on the pmf, $f(x)$, and the probability, $P(E)$, of an event, E, is defined as $P(E) = \sum_{x \in E} f(x)$. In addition, $\sum_{x \in \Omega} f(x) = 1$. A random variable, x, is defined as a

mapping into Ω with frequency P. Thus, a data set constructed from multiple observations of x will possess the probability characteristic P. Observations on the random variable are distinguished via capitalization and, thus, for a series of N observations, x is given by $x = \{X_1, X_2, ..., X_N\}$. It is frequently said that X is drawn from x.

It should be emphasized there is a distinction between the data points one collects in practice and the underlying physical process which generates them. This is to say that an attempt to calculate the pmf for a discrete set of real-valued features will always be an approximation, $\hat{f}(x)$, to the true pmf, $f(x)$, but that $\hat{f}(x)$ should converge to $f(x)$ in the limit of sample size as long as $f(x)$ does not change with time (i.e. is stationary). $\hat{f}(x)$ is sometimes referred to as the *sample pmf.*

The estimation of $f(x)$ can be approached either by assuming *a priori* knowledge of $f(x)$, called parametric estimation, or by making no assumptions, which is called non-parametric estimation. The more computationally intensive non-parametric approach to estimation, first illuminated in detail by the noted statistician Parzen [Parz62], shall be pursued to eliminate the risk of making an invalid *a priori* assumption. The treatment below shall follow Parzen [Parz62] with slight modification.

For non-parametric estimation, consider a feature space consisting of N independently drawn data samples, $d = \{d_i\}_{i=1}^{N}$, drawn from a distribution, $f(x)$, with a *cumulative mass distribution function* (cmdf), $F(x)$, given as

$$F(x) = \sum_{x'=-\infty}^{x} f(x') \tag{2.1}$$

It is assumed $F(x)$ is continuous but varying with x and there is, therefore, motivation to perform a succession of local estimates of $F(x)$ to capture this variation. Let the data samples within a small, local region $\Re$ be denoted as $\{d_i\}_{i \in \Re}$ and the sample cmdf for this region be denoted as $\hat{F}_{\Re}(d)$. $\hat{F}_{\Re}(d)$, in effect, represents the number of data samples, k,

less than or equal to d visible within $\Re$ as a fraction of the total number of data samples, N. This can be expressed as

$$\hat{F}_{\Re}(d) = \frac{k}{N} \tag{2.2}$$

which is essentially a binomially distributed random variable. This can, in turn, be defined as

$$\hat{F}_{\Re}(d) = \frac{1}{N}\binom{N}{k} F(d)\left(1 - F(d)\right)^{N-k} \tag{2.3}$$

This gives a mean value, $E[\cdot]$, for $\hat{F}_{\Re}(d)$ given by

$$E\left[\hat{F}_{\Re}(d)\right] = F_{\Re}(d) \tag{2.4}$$

Note that it can be shown, in the limit of sample size, that $\hat{F}_{\Re}(d)$ converges to $F_{\Re}(d)$ [Parz62].

For large sample sizes, $\hat{f}(d)$ can be considered a real-valued continuous density function and, as a result, an alternative expression to Eq. 2.2 for $\hat{F}_{\Re}(d)$ can be rendered as follows:

$$\hat{F}_{\Re}(d) = \hat{f}_{\Re}(d)V \tag{2.5}$$

where V is a volume measure on $\Re$. Note that N is dependent on V and this dependence will serve the purpose of growing or shrinking the granularity of the approximation region, $\Re$.

Combining 2.2 and 2.5 yields the final estimate

$$\hat{p}_{\Re}(x) = \frac{k/N}{V} \qquad = \frac{k/N}{V} \tag{2.6}$$

The technique of nearest neighbor non-parametric estimation approaches the calculation of Eq. 2.6 by fixing k and measuring volume according to a volume metric. The simplest volume metric for a zero-mean array is the sum of the array elements. Based on simulations by the author [Duec04], Duda and Hart [DuHa73] appear to use a metric similar to the root mean square distance.

Kernel-based non-parametric estimation, on the other hand, fixes V and counts (in the sense of convolution) k according to some counting metric. The simplest form of a counting metric is geometrically represented as a box function. More complicated counting functions can be imagined, however. The effect of non-box functions is to count points in a more exotic way than "one, two, three, ...". In the text by Bishop [Bish95], for example, counting is performed according to $\varphi(1), \varphi(2), \varphi(3)...$ where $\varphi(x)$ is the analytical form of a Gaussian distribution.

Kernel-based estimates are preferred, in general, because analytical expressions for estimator bias and variance can often be obtained. Nearest-neighbor estimation does not lend itself as easily to analytic expressions for convergence. Therefore, the type of kernel used for estimation in this report shall now be defined.

For kernel estimation, a continuous window function, $\Psi(x)$, must be defined which will define $\Re$ and a measure of its volume, V. The window, k will be given by

$$k = \sum_{i=1}^{N} \varphi\left(\frac{d - d_i}{h}\right)\Delta \tag{2.7}$$

using the approximation $\Psi(x) = \varphi(x)\Delta$ where $\varphi(x)$ is discrete and Δ is the sampling interval used to make the discrete conversion. The continuous pdf $\hat{p}(d)$, after substituting Eq. 2.7 into Eq. 2.6, is given by

$$\hat{p}(d) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{V}\varphi\left(\frac{d-d_i}{h}\right)\Delta \tag{2.8}$$

If a Gaussian window is chosen for its infinite support and smooth properties, then definition of $\hat{p}(d)$ in Eq. 2.8 is interpreted as the summation of a series N Gaussians, each centred at a point in the data set, $d = \{d_i\}_{i=1}^{N}$. An example of this estimation is shown in Fig. 2.35 below.

It is possible to define the average or *expectation*, $E[x]$, of a random variable, x, with distribution $\hat{p}(x)$, a statistic which will be used later, in terms of the distribution estimate as

$$\hat{E}[x] = \int_{-\infty}^{+\infty} x \cdot \hat{p}(x)\,dx \tag{2.9}$$

2.5.2 The t*-Test*

The *t*-test, the first of two tests to be performed on the HPV metric of Ch. 3, yields the *p* value mentioned in Sec. 1.2 and has the advantage of being both familiar to medical researchers and being featured in many publications Its limited assessment power [TvKa71] [Rose93] [Cohe94], however, will restrict its assessment capability in this work to providing a general impression of how separable the average HPV measure of R_C from the average HPV measure of R_H.

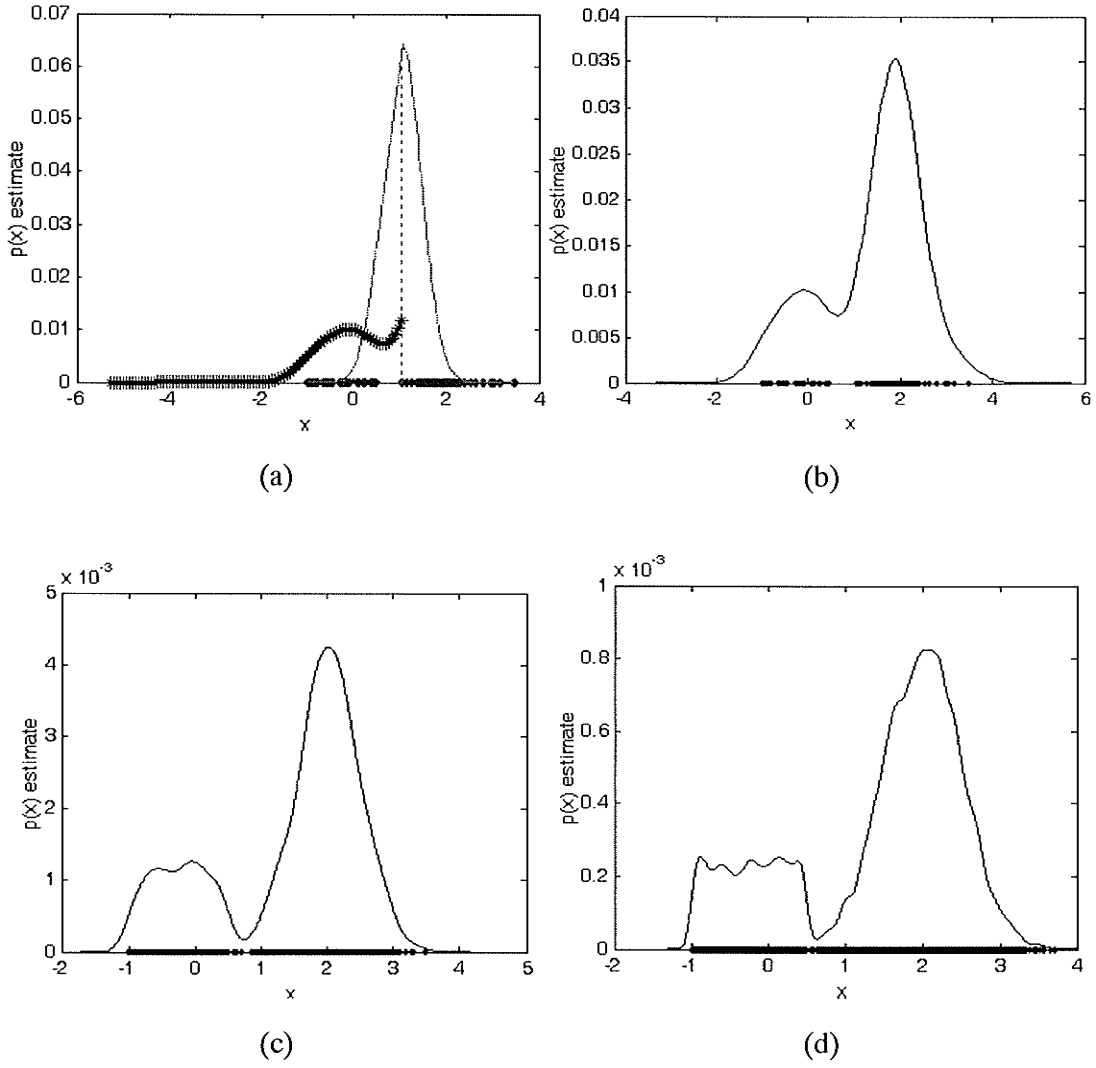


Fig. 2.35 Kernel-based estimation using a Gaussian kernel (a) in progress, and (b) for Gaussian data with 100 points, (c) 1000 points, and (d) 5000 points.

The nil hypothesis *t*-test, as it is used in this work, exists to test the likelihood that an observed difference between two normally distributed means has occurred by chance. If this chance is small, it is generally accepted that there is reason to belief a difference exists and the HPV metric has discrimination power. This alternative choice is commonly referred to as the alternative hypothesis and denoted H_1 and the claim to be refuted is termed the nil hypothesis, denoted H_0. The result of the test, the so-called *p*

value, represents the confidence in the null hypothesis that no difference exists between the means. These hypotheses may be formally stated as follows:

H_0: Sufficient variability exists that it cannot be established if the observed difference is real or accidental.

H_1: There is a low chance the difference is accidental and, as a result, the alternative hypothesis that the difference is real is plausible.

To understand the mechanism of the *t*-test, consider the plot of two normally distributed means calculated from a healthy population and CHF population in Fig. 2.36. The variability of the sample means is the *standard error on the mean* (SEM), which is given by

$$\hat{\sigma}_N = \frac{\hat{\sigma}}{\sqrt{N}} \tag{2.10}$$

where $\hat{\sigma}$ is an estimate of the sample standard deviation and N is the number of subjects in the sample. The *t*-test combines these two separate distributions into a single zero-mean distribution with pooled *standard error* (SE), $\hat{\sigma}_{P[N]}$. Using Student's *t*-distribution, the area, p, is measured in the part of the tail of the pooled distribution after the distance, d, representing the separation between the means.

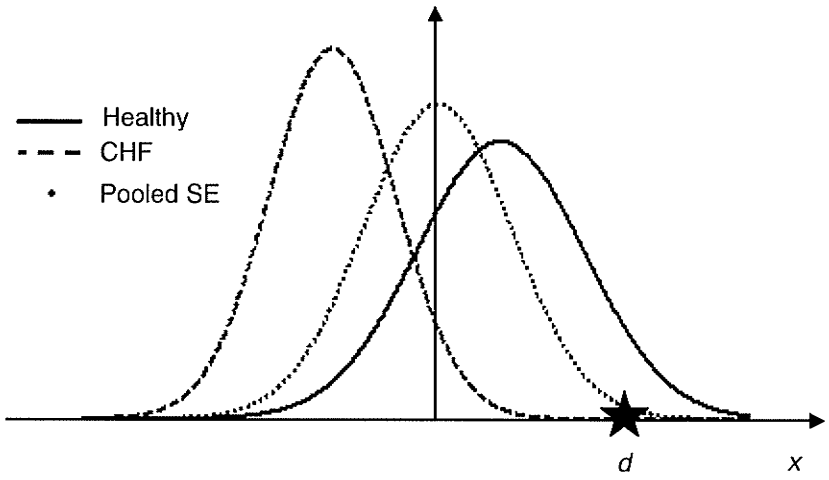


Fig. 2.36 Conceptual interpretation of the *t*-test showing the point under test, d (defined in Eq. 2.11), relative to a distribution with pooled SE (dot).

This distance, known as Cohen's effect size [Rose93], is given as

$$d = E[x_H] - E[x_C] \tag{2.11}$$

for a feature drawn from a healthy population, x_H, and from an imagined CHF sample, x_C. The pooled variance, $\hat{\sigma}_p^2$, upon which the pooled SE, $\hat{\sigma}_{P[N]}$, is based, is given as

$$\hat{\sigma}_p^2 = \frac{\sum(x_C - E[x_C])^2 + \sum(x_H - E[x_H])^2}{\delta_p} \tag{2.12}$$

where δ_p, the pooled degrees of freedom, is expressed as $N_C + N_H - 2$ and N_C and N_H are the number points in the CHF and healthy dataset, respectively. The pooled SE, $\hat{\sigma}_{P[N]}$, is then given as

$$\hat{\sigma}_{P[N]} = \hat{\sigma}_P \sqrt{\frac{1}{N_C} + \frac{1}{N_H}} \tag{2.13}$$

and the statistic to be tested, the distance in terms of pooled SEs, t, is given as

$$t = \frac{d}{\hat{\sigma}_{P[N]}} \tag{2.14}$$

Finally, the value, p, is given by

$$p = \frac{1}{\delta_p B\left(\frac{1}{2}, \frac{\delta_p}{2}\right)} \int_{-t}^{t} \left(1 + \frac{x^2}{\delta_p}\right)^{-\frac{\delta_p + 1}{2}} dx \tag{2.15}$$

where $B(a,b)$ is Euler's Beta function

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\, dt \tag{2.16}$$

As mentioned earlier, a small value of p reflects a small confidence in the null hypothesis that distance between the means is small in comparison to the pooled variability. The presence of p values below $p = 0.05$ are widely taken to mean that a "statistically significant" difference exists. It should be noted that this value of p has historical importance and a contemporary importance in the literature but, in reality, its scientific significance should be carefully scrutinized [Cohe94].

2.5.3 Receiver Operating Characteristics

As alluded to in Sec. 2.4, the act of making a decision in the two class case regarding an unclassified (undiagnosed) point is equivalent to setting a threshold in feature space and examining on which side of the threshold the point lies. In Fig. 2.29, this may involve setting the threshold as $T = 1.75$ V and classifying a received signal, $r(t)$ as digital zero if $r(t) < T$ and a digital one if $r(t) > T$. Sample thresholds are set for distributions in Fig. 2.37, which represent the distribution of digital ones, $p(x|C_1)$, and zeros, $p(x|C_0)$, previously encountered in the example of Fig. 2.29a. The probability representation $p(x|C)$ is often identified as an *a priori* probability. Figure 2.37a shows

a threshold which achieves a balance between the two distributions. If the two classes had unequal importance, however, an imbalanced decision threshold such as the one in Fig. 2.37b may be set.

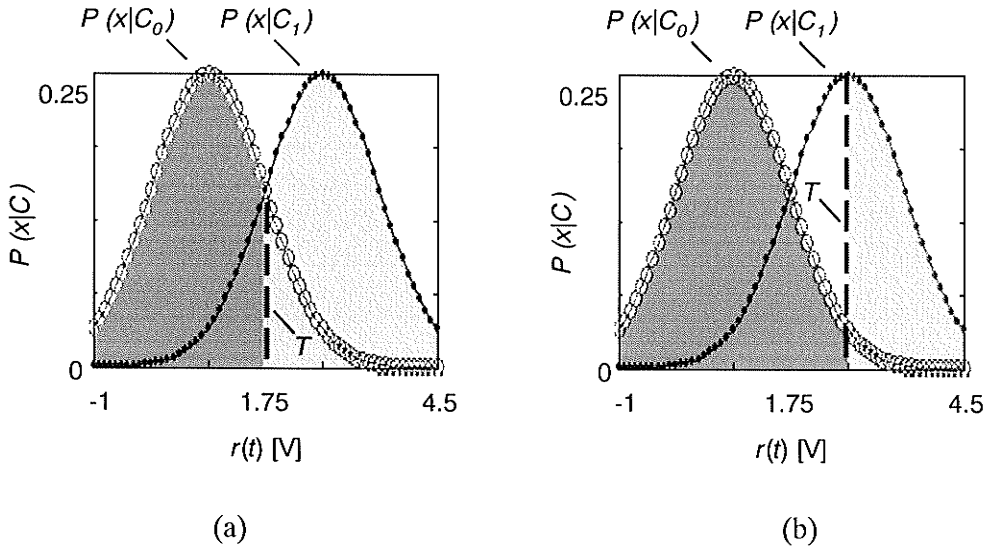


Fig 2.37 *A priori* distributions with thresholds which are (a) balanced to due equal class importance, and (b) imbalanced to the importance of $p(x|C_0)$ over $p(x|C_1)$.

In the case of medical diagnosis, being placed in the class of illness is often termed "testing positive" and the being placed in the class of health "testing negative." The probability of possessing an illness and testing positive for it is termed a "true positive" and the probability of being ill and testing negative is termed a "false negative." These two probabilities sum to 1. Likewise, the probability of being healthy and being classified as such is termed "true negative" and the chance of being mistakenly diagnosed as ill while healthy is termed "false positive." These two probabilities also sum to 1. The various proportions just discussed will henceforth be termed *true positive proportion* (TPP), *false negative proportion* (FNP), *true negative proportion* (TNP), and *false positive proportion* (FPP). These stages are labeled on two a priori distributions in Fig. 2.38.

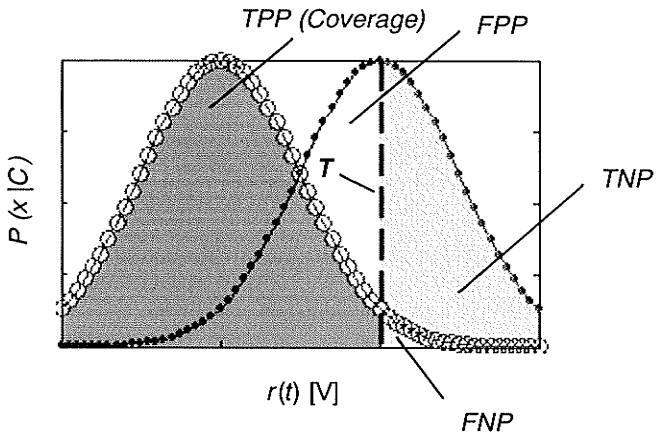


Fig. 2.38 Regions of interest with the "ill", $p(x|C_I)$, and "healthy", $p(x|C_{II})$, class densities as a result of defining a decision threshold, T.

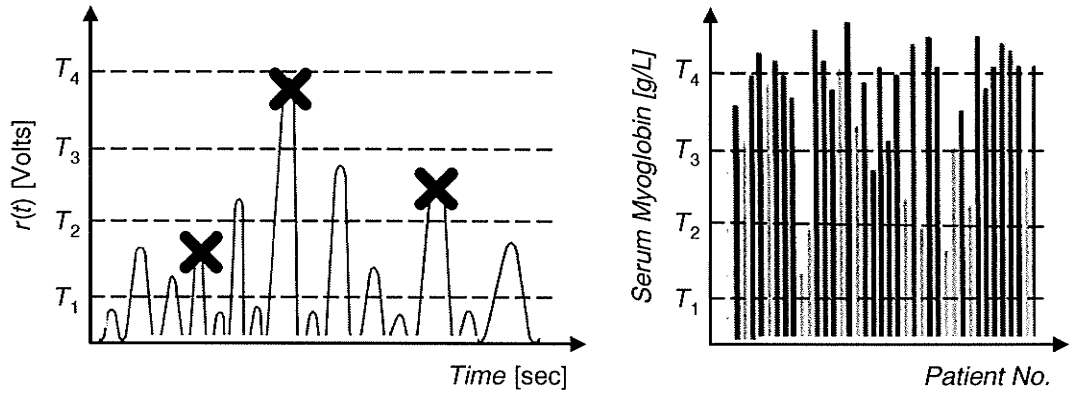
For this work, the TPP will be synonymous with the *coverage* of the test, γ, is defined as

$$\kappa = \int_{x=-\infty}^{T} P(x|C_I)\,dx \tag{2.17}$$

In the two class case for one dimension, T can, in general, take on any real scalar (for two classes in two dimensions this becomes a curve, and so forth). Studies reporting results for only a single threshold value are of extremely limited use [RoZw81] and, therefore, a method of summarizing the rate at which the areas under the *a priori* distributions of Fig. 2.38 change with T is required. Sample problems in which the tuning of this threshold is important are given in Fig. 2.39.

A common method is the *receiver operating characteristic* (ROC) curve, which was developed by the U.S. Air Force for radar applications during World War II and later, extensively, to problems of human judgment (particularly radiology) [Swet92]. The ROC operates by plotting the TPP against the FPP, as shown in Fig. 2.40. In this context, a test which is no better than a random guess produces a diagonal line, a test with some

discrimination power produces a curve with begins and terminates on the diagonal, and a perfect test produces a step response.



(a) (b)

Fig. 2.39 Four threshold possible levels indicated for (a) radar returns where peaks marked ✖ are aircraft and other peaks are noise, and (b) serum myoglobin (a marker for myocardial infarction) concentrations for patients admitted with chest pain to a coronary care unit (After [RoZw81]).

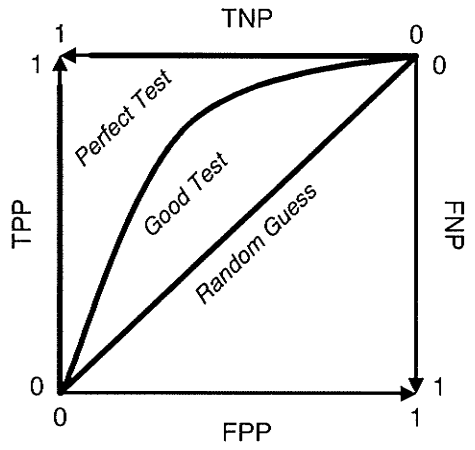


Fig. 2.40 Diagram showing ROC curves for three different tests: a test no better than a random guess (diagonal line); a test with some discrimination power (curve); and a perfect test with no false positives or negatives (angular).

2.5.4 Bayesian Statistics

The most significant weakness of ROC analysis is that it lacks context – that is to say it discusses proportions but not decision error rates. Decision error rates depend not only on the relative location of classes in feature space (which is captured by the ROC curve) but also on the percentage of time each class is present at that location. For a set of classes $\{C_k\}$ and a point x', Bayesian statistics will state: "C_1 resides at x' P_1 proportion of the time, C_2 resides at x' P_2 proportion of the time..." and so forth such that the sum of all P_i is 1. So given a new point, x'' - which represents, say, a measurement on a diseased patient – then it may be reasonable to classify the state x'' based on the most frequent past disease performance. The *decision rule,* in this case, would be "most frequent past performance" and, in general, is useful enough to be called *Bayes optimal* decision rule. The Bayes optimal decision rule is guaranteed to yield the highest percentage of correct decisions, a rate which can also be interpreted as proportional to the efficiency of the decision process and the P_i just discussed is often denoted as the posteriori probability $P(C_i|x)$. If the distributions of Fig 2.37a exist an equal proportion of the time then Fig 2.29a can be transformed into the *posteriori* probability plot of Fig. 2.41 below.

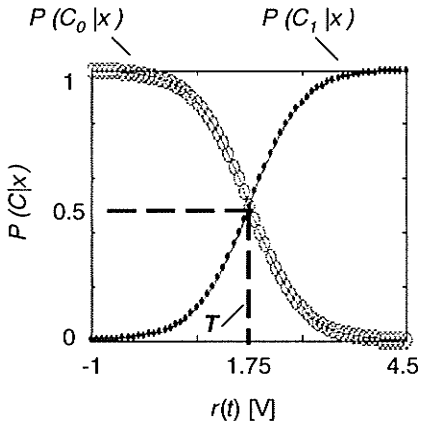


Fig. 2.41 Figure 2.29a interpreted in terms of *posteriori* probabilities if $p(x|C_1)$ and $p(x|C_0)$ occur with equal frequency.

The Bayes decision rule assumes, however, that the reward for correct classification is the same for all classes. This important point will be discussed further in Sec. 2.5.4. But assume, for example, that one is a general physician performing yearly check-ups while on the alert for a serious disease with subtle symptoms and a low prevalence. The most efficient decision rule in this case may simply be to ignore the existence of the disease by diagnosing everyone as healthy, where *efficiency* is defined as the percentage of positive test results which are true positives. In this instance, the most efficient decision is not the most effective as many people with a serious disease will leave the office under the impression they are well.

Consequently, the principle of sacrificing efficiency in favour of coverage is now introduced. If Bayes' rule is used, this rate is guaranteed to exceed 50%. *Coverage* will be the percentage of the diseased population which is correctly diagnosed and is, in general, the classification statistic to consider first when diagnosing disease. Efficiency concerns itself with correct decisions in a context while coverage is without context. For low prevalence diseases, both efficiency and coverage must be considered.

Bayesian theory may be formally developed by first considering a feature space populated by data points representing both disease and health. The pdf of the union of these two sets is the *total pdf,* $p(x)$. If, from this pdf, one would like to select either the health or sick distributions then the *a priori* distributions $p(x|C_H)$ and $p(x|C_S)$ are considered, respectively. A subset of the support of healthy features $\{x|C_H\}_{\Re}$ will occupy a region $\Re$ of feature space with in proportion to $p(x|C_H)/p(x)$ and, similarly, $\{x|C_C\}_{\Re}$ will occupy $\Re$ in proportion to $p(x|C_S)/p(x)$. At least one of the two classes must occupy $\Re$ and, therefore, it will be required

$$\frac{\frac{1}{2}p(x|C_H)+\frac{1}{2}p(x|C_S)}{p(x)}=1 \tag{2.18}$$

This assumes, however, that sickness occurs as frequently as health. Therefore, Eq. 2.18 is modified with their respective probabilities of occurrence, $P(C_H)$ and $P(C_S)$

$$\frac{P(C_H)\,p(x|C_H)+P(C_S)\,p(x|C_S)}{p(x)}=1 \tag{2.19}$$

It is clear, from Eq. 2.19, that $p(x)$ may be defined from the independently calculated *a priori* distributions as

$$P(C_H)\,p(x|C_H)+P(C_S)\,p(x|C_S)=p(x) \tag{2.20}$$

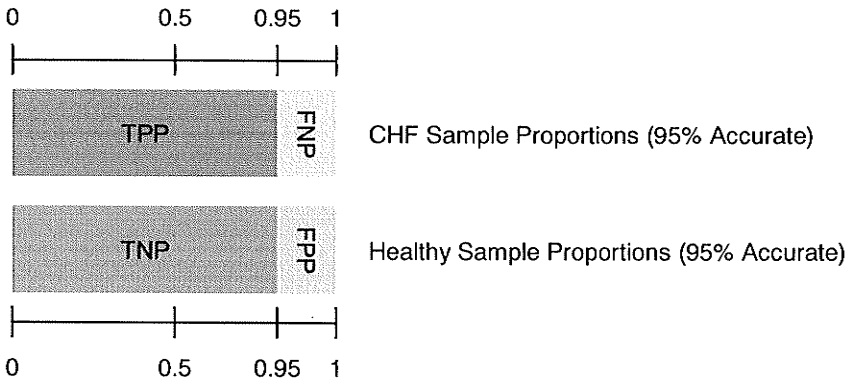
Recalling the proportions of sickness and health in region $\Re$ made earlier, Bayes' rule evolves naturally. Let $P(C_H|x)$ represent the proportion of healthy features at a point x and $P(C_H|x)$ the proportion of sick features. The "most frequent past performance" decision rule – Bayes' Rule – is effected by simply choosing the larger proportion from the following

$$P(C_H|x)=P(C_H)\frac{p(x|C_H)}{p(x)} \tag{2.21}$$

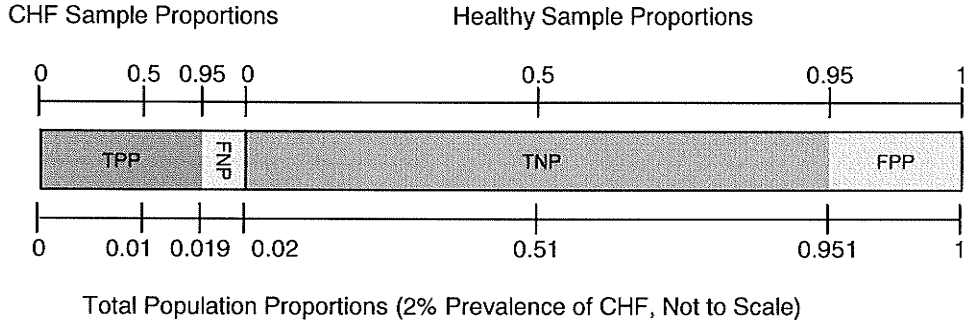
$$P(C_S|x)=P(C_S)\frac{p(x|C_S)}{p(x)} \tag{2.22}$$

2.5.5 A Bayesian Expansion of Receiver Operating Characteristics

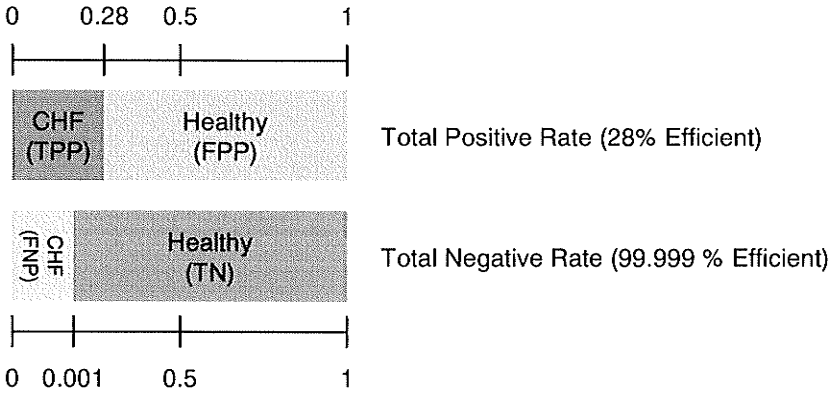
The ROC can be used for a number of coverage-based assessments (a summary is given in Appendix B) of a test but does not contain information about the usefulness or efficiency of the test. This is to say that a test may detect all members of an important but low-prevalence diseased population at the cost of having a tremendous number of false positives. In this instance, the *efficiency* – the number of true positives divided by the number of total positives – will be very low.



(a)



(b)



(c)

Fig. 2.42 Illustration of how a test for CHF which is 95% accurate can only be 28% efficient via (a) a rate (95%) on the sample, (b) placed in the context of the total population, and (c) a rate (28%) on the population.

It is generally known in advance whether the application requires a threshold with a low false positive rate (such as in industrial quality-control manufacturing), high true positive rate (such as in medicine screening), or a threshold which achieves a measure of balance (such as the Bayes optimal threshold). The approach in choosing a threshold, therefore, is to discover at what cost (in terms of efficiency) coverage may be purchased. Consequently, the ROC analysis should be expended by an analysis which considers coverage as the independent variable and efficiency as the dependent variable.

Efficiency will be defined simply as the average posterior probability in the positive diagnostic region of Fig. 2.38. Efficiency, η, can therefore be defined mathematically as

$$\eta = \frac{\int_{x=-\infty}^{T} p(x) \cdot P(C_I \mid x) dx}{\int_{x=-\infty}^{T} p(x) dx} \tag{2.23}$$

with the limits of the integral taken assuming the positive classification is closer to $-\infty$. It is similar to the traditional rate of TPP classifications, β, defined in the literature

$$\begin{aligned} \beta &= \int_{x=-\infty}^{T} p(x) \cdot P(C_I \mid x) dx \\ &= \int_{x=-\infty}^{T} p(C_I) \cdot p(x \mid C_I) dx \end{aligned} \tag{2.24}$$

In a medical sense, coverage is a test of the detector while efficiency is a test of the patient.

The relationship of η to the ROC curve is given as

$$P(C_I \mid x) = \frac{1}{\frac{P(C_{II})}{P(C_I)} \frac{p(x \mid C_{II})}{p(x \mid C_I)} + 1} \tag{2.25}$$

$$= \frac{1}{\frac{P(C_{II})}{P(C_I)} \frac{1}{\ell_{III}(x)} + 1}$$

where $\ell_{III}(x)$ is the slope of the ROC curve and is commonly referred to as a *likelihood ratio.*

Benefits and costs are notoriously difficult to assign when human lives are at stake - policy vs. patient and may be expressed financial and/or health costs from numerous perspectives (including the patient, health care providers, the insurers, dependents, and society) [ZwCa93] [Swet92]. Efficiency, in an absolute sense, is therefore a difficult metric to apply.

Efficiency, however, can also be interpreted as the new prevalence of the disease following the screening test. For example, a screening test may detect true positives at a rate of 5% (i.e. is 5% efficient) but the disease prevalence is only 1%. In this instance, the efficiency is 5 times the prevalence and represents a 400% increase. This idea of the percentage increase in efficiency relative to the prevalence, ρ, will be termed *amplification* and denoted α

$$\alpha = \frac{\eta}{\rho} - 1 \tag{2.26}$$

2.5.6 Relationship of Bayes' Theorem to the t*-Test*

The relationship between Bayesian statistics and the probability interpretation upon which the *t*-test is based, Frequentism is introduced with an example by noted statistician Jacob Cohen. Consider that the incidence of schizophrenia in adults is about 2% and that a proposed screening test is estimated to be 95% accurate in making the positive diagnosis and about 97% accurate in declaring normality.

Considering normality as the null hypothesis, this can be re-phrased as $f(x_N \mid H_N) = 0.97$ and $f(x_S \mid H_S) = 0.95$ .Consequently, $f(x_N \mid H_S) = 0.05$ or, in other words, $p = 0.05$ - the level of so-called statistical significance. It should be noted that what is evaluated by $p = 0.05$ is the test and not the patient. Charting these probabilities in a form recognizable to Bayesian statistics yields Fig. 2.43.

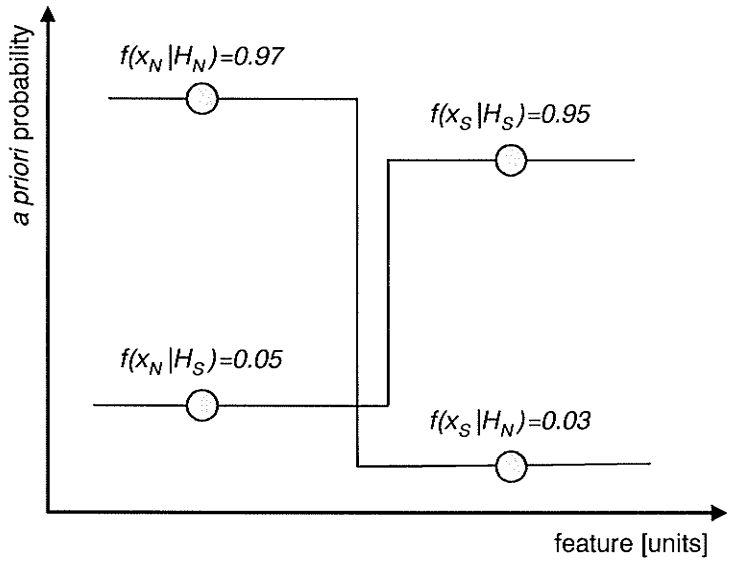


Fig. 2.43 Graphically representation of the hypothesis-based example schizophrenia diagnosis example.

Continuing with this example, it turns out that the probability of the patient being normal given a positive test for schizophrenia is about 0.6 despite possessing an a priori probability of 0.95. This can be seen in following:

$$f(H_N \mid x_S) = F(H_N)\frac{f(x_S \mid H_N)}{f(x)} = 0.98\frac{0.03}{f(x)} \tag{2.27}$$

where

$$\begin{aligned} f(x) &= F(H_N) f(x_S \mid H_N) \\ &\quad + F(H_S) f(x_S \mid H_S) \\ &= 0.98 \cdot 0.03 + 0.02 \cdot 0.95 \\ &= 0.1194 \end{aligned} \tag{2.28}$$

giving

$$f(H_N \mid x_S) = 0.607 \tag{2.29}$$

As a final note, when the prevalence of disease in a population is very low the rate of correct diagnosis for a screening test is given simply by the p value (expressed as a percentage) divided by the prevalence. So, in this instance, five subjects will be diagnosed with schizophrenia for every one subject who actually has the condition leading to a correct diagnosis rate of 0.4 and, therefore, a false diagnosis rate of approximately 0.6. This can be seen via an inspection of Eq. 2.27 and Eq. 2.28.

Another benefit of Bayesian analysis is the ability to include new *a priori* information as it becomes available. The latter is important if, for example, a patient's family was pre-disposed to schizophrenia and the prevalence rate for the patient could be adjusted from 2% to, say, 10%.

2.6 Summary

This chapter has discussed the anatomy and physiology of the cardiovascular system in health and congestive heart failure. It also discussed how biopotentials resulting from the depolarization of cells within the heart are transformed into digital electrocardiogram (ECG) data. Next, how the digital ECG is transformed into RR interval data via the R-complex of the lead III ECG is demonstrated. Feature extraction was discussed both in principle as is it will be applied in Ch. 3. Lastly, the techniques by which Ch. 3 will be evaluated in Ch. 4, Ch.5, and Ch. 6 were reviewed. This included the *t*-test, ROC curves, and Bayesian statistics and resulted in the creation of definitions for classifier coverage, efficiency, and amplification. The evolution of these ideas in the context of the thesis is depicted in Fig. 2.44.

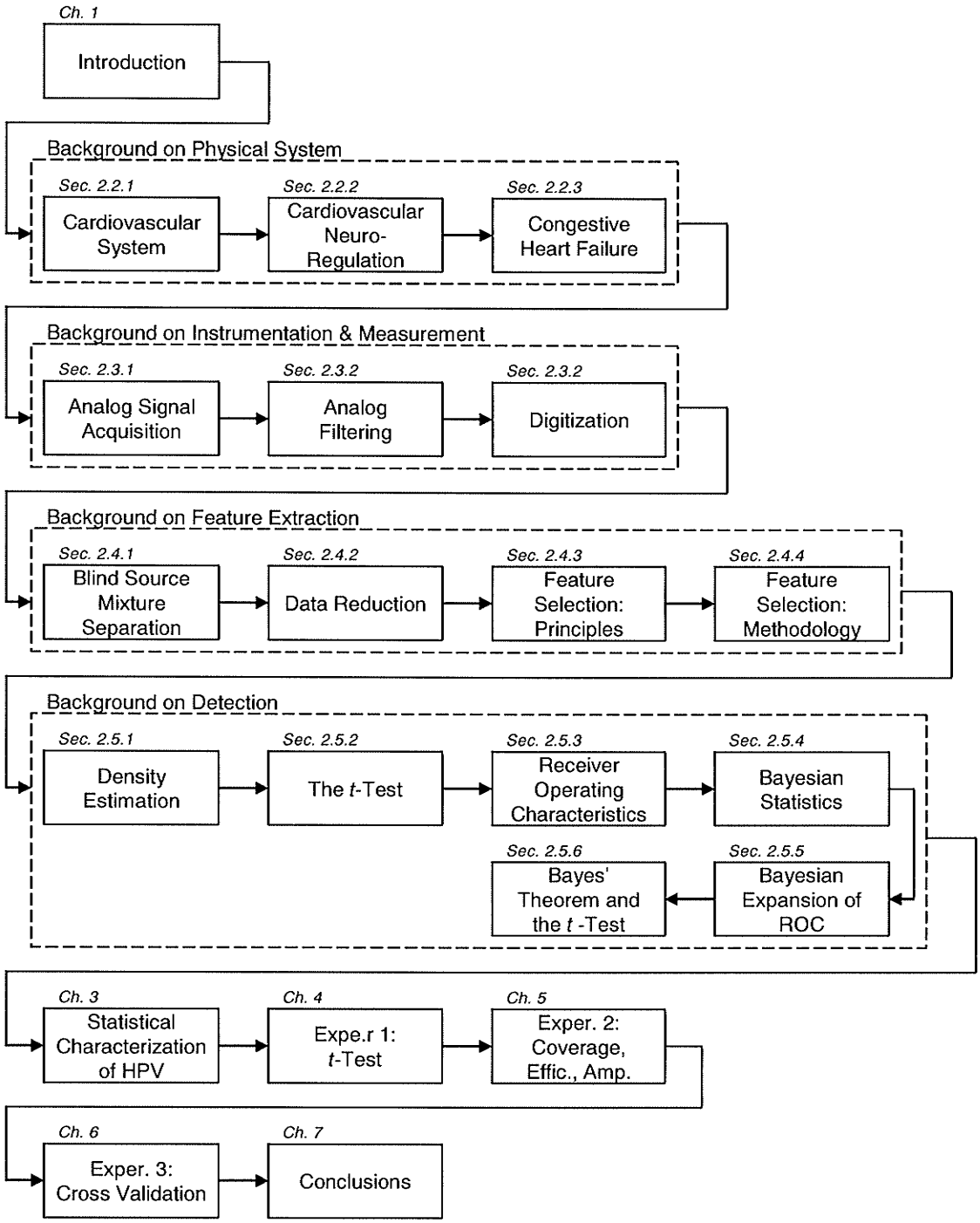


Fig. 2.44 Summary of ideas in the background in the context of the thesis.

Chapter 3

STATISTICAL CHARACTERIZATION OF HEART PERIOD VARIABILITY

3.1 Introduction

This chapter identifies the healthy and CHF data samples to be used in the search for a HPV metric capable of discriminating health from CHF. The procedure and its rationale have already been discussed in Sec. 2.4 but, to summarize, the search will first consider $\hat{p}(R|S_H)$ and $\hat{p}(R|S_C)$ in order to estimate whether the potential for a discriminatory metric exists in the sample. It then considers $\left\{\hat{p}\left(R|S_{H[i]}\right)\right\}_{i=1}^{N_H}$ and $\left\{\hat{p}\left(R|S_{C[j]}\right)\right\}_{j=1}^{N_C}$ on a subject-by-subject basis to establish whether there is enough consistency between subjects to warrant pursuit of the discriminatory metric of step one. These two considerations are performed on RR interbeat interval and successive differences on this time series.

3.2 Description of Data

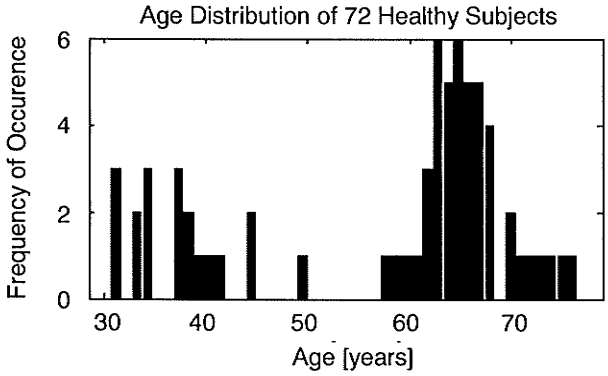
Healthy interbeat interval data were drawn from the MIT-BIH Normal Sinus Rhythm RR Interval (54 long-term ECG recordings, 30 men, 24 women) and MIT-BIH Normal Sinus Rhythm (18 long-term ECG recordings, 5 men, 13 women) Databases of Physionet [GAGH00]. The membership criterion for the former database was "normal sinus rhythm" and "no significant arrhythmias" for the latter [GAGH00]. Data was provided to Physionet by the Washington University School of Medicine, Columbia-Presbyterian Medical Center, and Israel Deaconess Medical Center. Average recording time was

23.1 ± 1.3 hours. Due to the name of the dataset, the word "normal' will henceforth be used interchangeably with "healthy." The data in this set has appeared in the Journal of the American College of Cardiology [GBSF92], Circulation (the journal of the American Heart Association) [BFSR95], American Heart Journal [SEDK99], and Heart (the journal of the British Cardiac Society) [MPHG02].

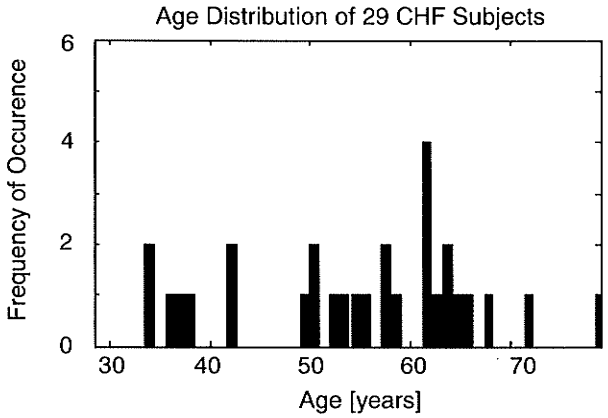
CHF interbeat interval data were drawn from the Congestive Heart Failure RR Interval Database (29 long-term ECG recordings, 8 men, 2 women) of Physionet [GAGH00]. Membership is graded on the NYHA CHF scale and it is assumed that a gold standard (see Sec. 2.2.3) method of diagnosis was performed prior to this purely functional grading. Data was provided to Physionet by Columbia-Presbyterian Medical Center. Average recording time was 22.2 ± 1.9hours. This data set has appeared in the Journal of the American College of Cardiology [KBGP95], American Journal of Cardiology [GBBK97], and Heart (the journal of the British Cardiac Society) [MPHG02]. While the number of CHF subjects is small, it is larger than the sample used in many other significant studies, including that of the Framingham Heart Study on HPV [HMPM97]. The Framingham Heart Study is a very well-known collection of longitudinal studies spanning 50 years.

All data have been sampled at 128 samples per second with an unknown filter frequency cut-off. Mathematically, it must, of course, be less than 128/2=64 Hz. The automated peak detection algorithm employed nor the extent of manual correction are reported by Physionet.

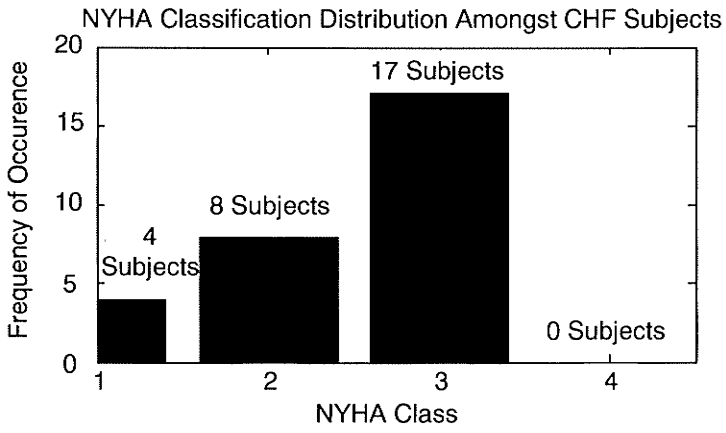
A comparison of Fig. 3.1a and Fig. 3.1b reveals the distributions of the normal and CHF samples with respect to age are homogenous. The exact distributions of both samples are provided in Appendix A. CHF severity is distributed with an emphasis on NYHA functional class III, as shown in Fig. 3.1c.



(a)



(b)



(c)

Fig. 3.1 Homogeneity analysis comparing (a) normal subject distribution with age, (b) CHF patient distribution with age, and (c) distribution of NYHA classifications.

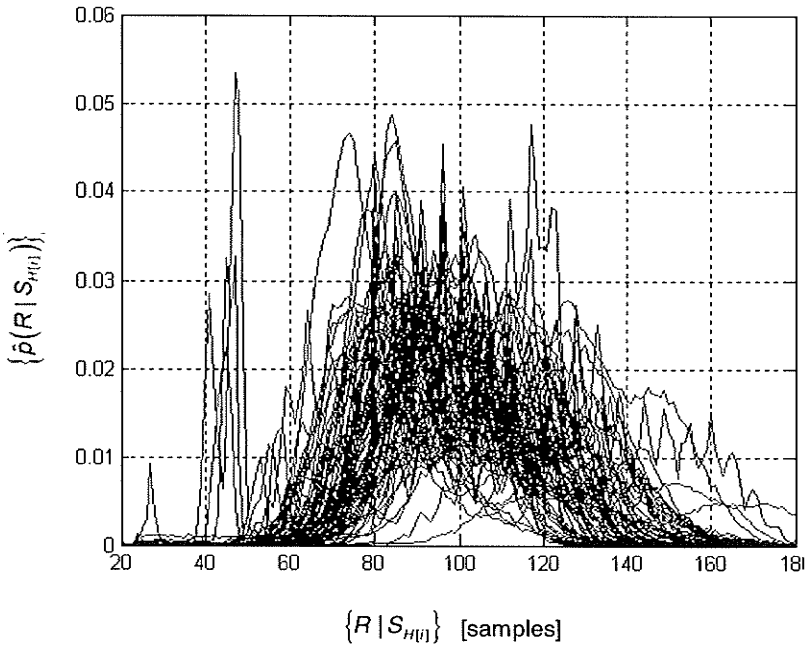
3.3 RR Interbeat Data

First, the distributions of RR interbeat interval data are plotted for every subject to check for consistency from individual to individual. This creates the plots of $\left\{\hat{p}\left(R\,|\,S_{H[i]}\right)\right\}_{i=1}^{72}$ and $\left\{\hat{p}\left(R\,|\,S_{C[j]}\right)\right\}_{j=1}^{29}$ shown in Fig. 3.2. It is apparent that there is a considerable degree of inter-subject variability in both health and CHF, making RR interbeat interval data an unattractive basis for a metric. Furthermore, the data are clearly non-Gaussian for both classes, indicating metrics based on standard deviation for these data are inadequate at best [PFTV92].

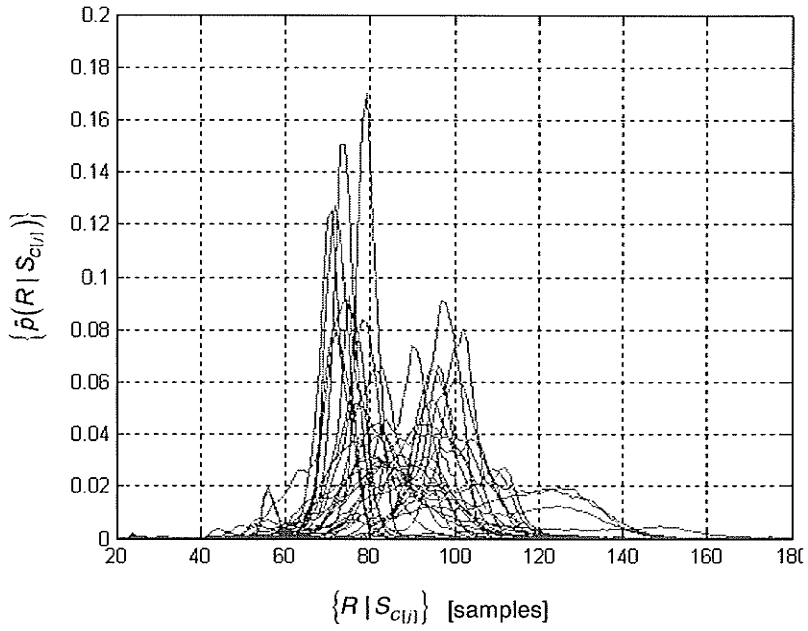
Next, the distributions of all data records concatenated, denoted $\hat{p}\left(R\,|\,S_H\right)$ and $\hat{p}\left(R\,|\,S_C\right)$, are plotted in Fig. 3.3 to achieve an estimate of discrimination potential for the entire sample as a whole. A comparison of these estimates reveals a complicated discriminatory relationship between the RR interbeat intervals of healthy subjects and CHF subjects. Complicated relationships are less attractive as the basis of reliable decisions. The extended distribution tails of the CHF distribution, in particular, make standard deviation an unpromising metric for HPV due to its sensitivity to outliers [PFTV92].

A direction with potential promise is to take the set of the above RR interbeat intervals, R, and repeat the analysis using successive differences, $\overline{R}$, on R. For an subset of R, $R' \subseteq R$, let the ith and (i-1)th RR intervals within R', $R'_{[i]}$ and $R'_{[i-1]}$, be transformed into $\overline{R}'_{[i]}$ via

$$\overline{R}'_{[i]} = R'_{[i]} - R'_{[i-1]} \tag{3.1}$$



(a)



(b)

Fig. 3.2 Plots showing poor inter-subject consistency of (a) $\left\{\hat{p}\left(R \mid S_{H[i]}\right)\right\}_{i=1}^{72}$, and (b) $\left\{\hat{p}\left(R \mid S_{C[j]}\right)\right\}_{j=1}^{29}$.

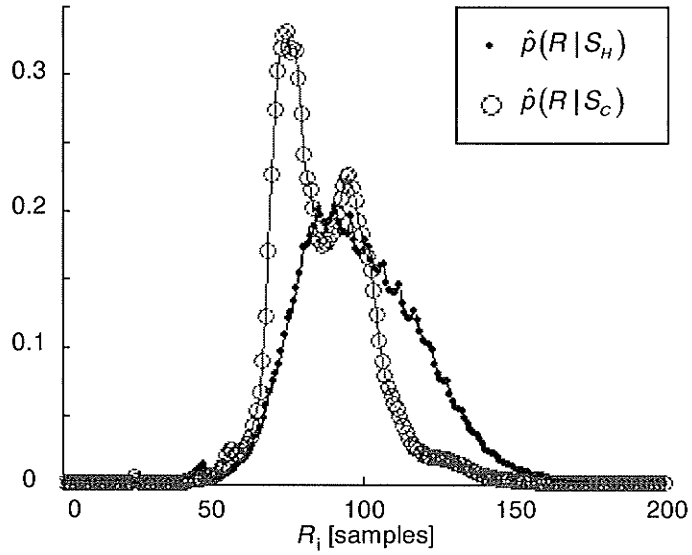


Fig. 3.3 Plots of $\hat{p}(R|S_H)$ and $\hat{p}(R|S_C)$ showing poor discrimination potential.

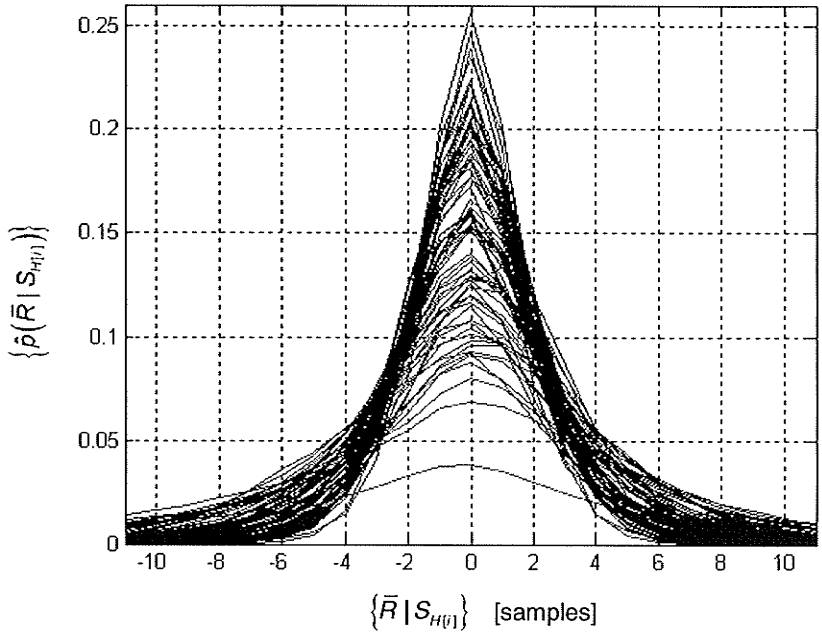
This approach, more commonly known as pre-emphasis, has long been used in linear predictive coding [Tier80] and radio broadcasting [MaGi87] as a simple method of reducing low-frequency wander. A non-stationary mean is the motivation behind the technique of rMSSD noted in Sec. 2.4.4. rMSSD, however, suffers from an assumption of Gaussianity which is invalid in this case. Successive differences do form the basis of the NN50, pNN50, and pNNx HPV metrics, though.

Given that the heart beat period is, over time, neither increasing or decreasing $\overline{R}$ is expected to have zero mean and, therefore, guaranteed to be stationary in at least one moment. It is expected, furthermore that, if the literature on HPV is correct, that successive differences of HPV in health should have larger variability than the successive differences of HPV in CHF subjects.

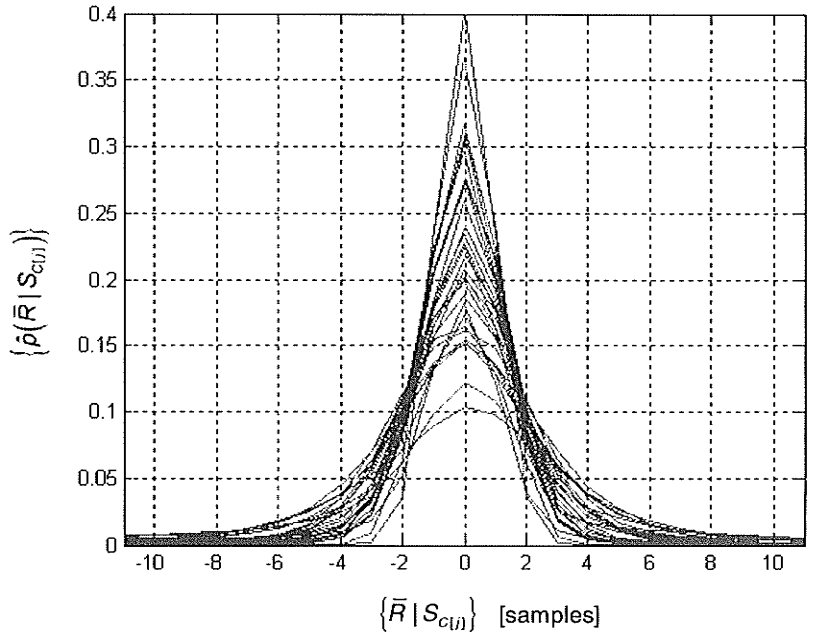
3.4 Successive Differences of RR Interbeat Data

The distributions $\left\{\hat{p}\left(\overline{R}\,|\,S_{H[i]}\right)\right\}_{i=1}^{72}$ and $\left\{\hat{p}\left(\overline{R}\,|\,S_{C[j]}\right)\right\}_{j=1}^{29}$ are plotted in Fig. 3.4 and, this time, demonstrate a very high degree of inter-subject consistency. The distributions of all data records concatenated together, $\hat{p}\left(\overline{R}\,|\,S_{H}\right)$ and $\hat{p}\left(\overline{R}\,|\,S_{C}\right)$, are plotted in Fig. 3.5 and a difference in the rate of decay of the distribution tails is observed between them. Specifically, a larger spread of the successive differences is seen for the healthy class, indicating that the healthy class experiences more diverse changes in heart period than the CHF patients.

It is preferable that a metric for capturing the differences in $\overline{R}_H$ and $\overline{R}_C$ observed in Fig. 3.5 does not assume anything about the form of the distributions and that it will adequately measure the tails of the distributions. It may be recalled that the Pareto distribution – a classicial heavy-tailed distribution – is often defined exclusively in terms of its tail as $P(Y>y)=\left(\frac{y}{y'}\right)^{-k}$ for some $y'\le y$, $k>0$. A heavy-tailed distribution is strictly defined as a distribution, $p(x)$, for which the measure $P(Y>y)$ scales as $y^{-\alpha}$, $0<\alpha<2$ as y tends to infinity. They may be distinguished from the exponential type distributions such as the Gaussian distribution, which decays more quickly. Practically, this model can be applied to any distribution which contains a significant amount of information in its tails.



(a)



(b)

Fig. 3.4 Plots showing good inter-subject consistency of (a) $\left\{\hat{p}\left(\bar{R}\,|\,S_{H[i]}\right)\right\}_{i=1}^{72}$, and (b) $\left\{\hat{p}\left(\bar{R}\,|\,S_{C[j]}\right)\right\}_{j=1}^{29}$.

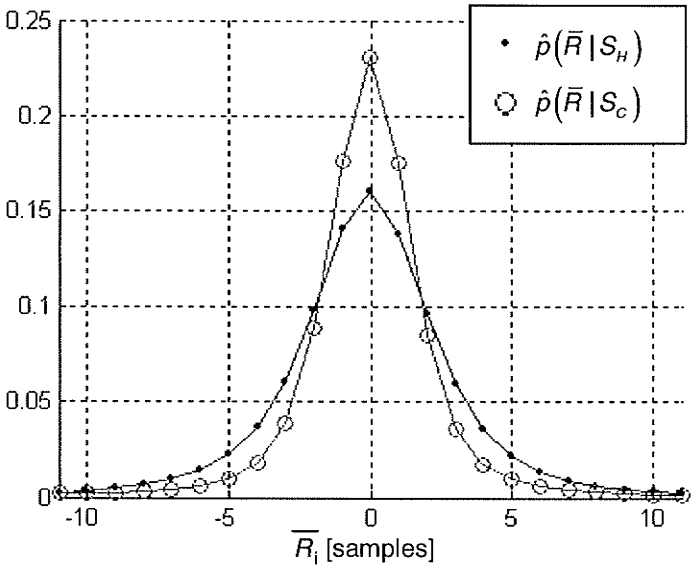


Fig. 3.5 Plots of $\hat{p}(\overline{R}|S_H)$ and $\hat{p}(\overline{R}|S_C)$ showing good linear discrimination potential.

Thus, a HPV measure of $\overline{R}$, $\mu(\cdot)$, can be defined. Let $\mu(\overline{R})$ be defined as the probability, for some threshold, x, of $\overline{R}$ amplitude that some choice of $\overline{R}$, $\overline{R}'$, is greater than x. In other words, let $\mu(\cdot)$ be a function of some subset, $\overline{R}' \subseteq \overline{R}$, of the successive difference data and of the threshold, x, and be denoted $\mu(\overline{R}', x)$. Also, let $\mu(\overline{R}, x)$ be defined as $\mu(\overline{R}, x) = P(\overline{R}' = \overline{R} > x)$. Note, also, that $\mu(\overline{R}', x)$ becomes equal to the pNNx metric due to noted researchers Mietus, Peng, Goldsmith, and Goldberger identified in Sec. 2.4.4 if $p(\overline{R}')$ is estimated using non-overlapping box kernel functions.

It is clear that potential for discrimination exists by plotting the distribution of average population HPV measure in health, $p\left(E\left[\left\{\mu(\overline{R}|S_{H[i]}), x=24\right\}_{i=1}^{72}\right]\right)$, alongside the distribution of the average population HPV measure in CHF, $p\left(E\left[\left\{\mu(\overline{R}|S_{C[j]}), x=24\right\}_{j=1}^{29}\right]\right)$ in Fig. 3.6.

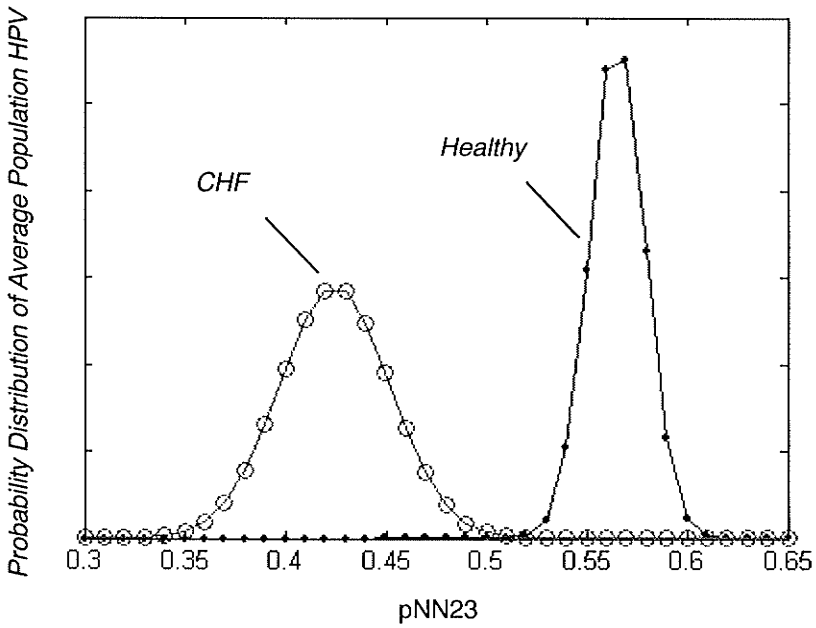


Fig. 3.6 Comparison of $p\left(E\left[\left\{\mu\left(\overline{R}\,|\,S_{C[j]}\right), x=24\right\}_{j=1}^{29}\right]\right)$ to $p\left(E\left[\left\{\mu\left(\overline{R}\,|\,S_{H[i]}\right), x=24\right\}_{i=1}^{72}\right]\right)$ (variability given by SEM) indicating potential of the derived HPV metric for linear discrimination.

3.5 Summary

This chapter has satisfied part (a) of the thesis statement by, first, demonstrating that RR interbeat intervals possess neither an obvious opportunity for simple discrimination between health and CHF or consistency between individuals. The set of successive differences of RR interbeat intervals, $\overline{R}$, however, possess both of these qualities and a metric denoted $\mu\left(\overline{R}',x\right)$ was created in Sec. 3.4. A plot of the distribution of $\mu\left(\overline{R}',x\right)$ indicates this $\mu\left(\overline{R}',x\right)$ is a potentially useful HPV metric. If the probability density of the successive differences is estimated using a non-overlapping box function kernel then $\mu\left(\overline{R}',x\right)$ is equivalent to pNNx, a metric which is found in literature. The first of three

tests of the discrimination power of $\mu(\overline{R}', x)$ is performed in the next chapter, as shown in Fig. 3.7.

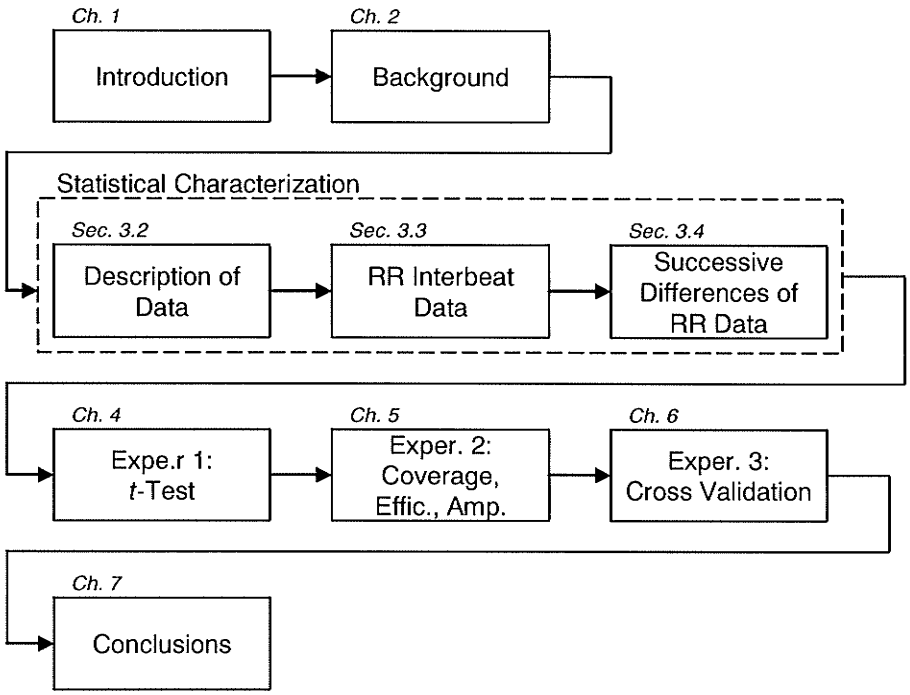


Fig. 3.7 Ideas within Ch. 3 shown in relation to the rest of the thesis.

Chapter 4

EXPERIMENTAL RESULTS AND DISCUSSION: *t*-TEST

4.1 Introduction

The experiment of this chapter measures average discrimination power using the p value (obtained via t-test). This test will be complemented by a more complete and precise analysis of discrimination power in Ch. 5. The measure is the p value via a nil hypothesis test and, as a result, supports the credibility of this work for the reasons discussed in Sec. 1.2.

4.2 Design

The statistic to be tested for both healthy and CHF subjects will the average pNNx value for the populations. This summary statistic provides a preview of the results to be encountered in Ch. 5. The averages for the healthy and CHF samples will be noted $E\left[\left\{\mu\left(R \mid S_{H[i]}, x\right)\right\}_{i=1}^{72}\right]$ and $E\left[\left\{\mu\left(R \mid S_{C[j]}, x\right)\right\}_{j=1}^{29}\right]$, respectively where x is in samples and $x \in \mathrm{I}$, $x \in [1..13]$. Moreover, the averages will be compared over a range of pNNx parameter, x, values to see if certain pNNx ranges are more promising than others. The specific hypotheses to be tested are as follows:

H_0: Sufficient variability exists that it cannot be established if the observed difference is real or accidental.

H_1: There is a low chance the difference is accidental and, as a result, the alternative hypothesis that the difference is real is plausible.

The samples possess sufficient size for the Central Limit Theorem to apply and, consequently, for the averages to be distributed as Gaussian random variables. As a result, the *t*-test can be used. Average differences may conceivably be either positive or negative and, as a result, a two-tailed test will be used. Observations between the healthy and CHF classes are uncorrelated and, therefore, the unpaired, two-tailed nil hypothesis *t*-test is the most adequate variant for this experiment.

The mechanics of the *t*-test have already been described in greater detail in Sec. 2.5.2.

4.3 Results and Discussion

The results of *t*-test comparisons of normal and CHF mean pNNx values over varying thresholds are plotted in Fig. 4.1 are in agreement with the results obtained by Mietus, *et al.* [MPHG02] using pNNx on different data.. The results of the *t*-tests indicate how confidently the normal and CHF pNNx means may be regarded as being the same. For visual confirmation of the *p* value trends, twice the SEM (defined in Sec. 2.4 and representing the envelope of 95% of all of the means) is also plotted in Fig. 4.1.

Figure 4.1 and the *p* value data on which it is based in Table 4.1 reveal that the desired level of significance is achieved for sample thresholds of 1, 2, 3, and 4 samples. These thresholds correspond to approximately x=8, 16, 24, and 32 msec, respectively.

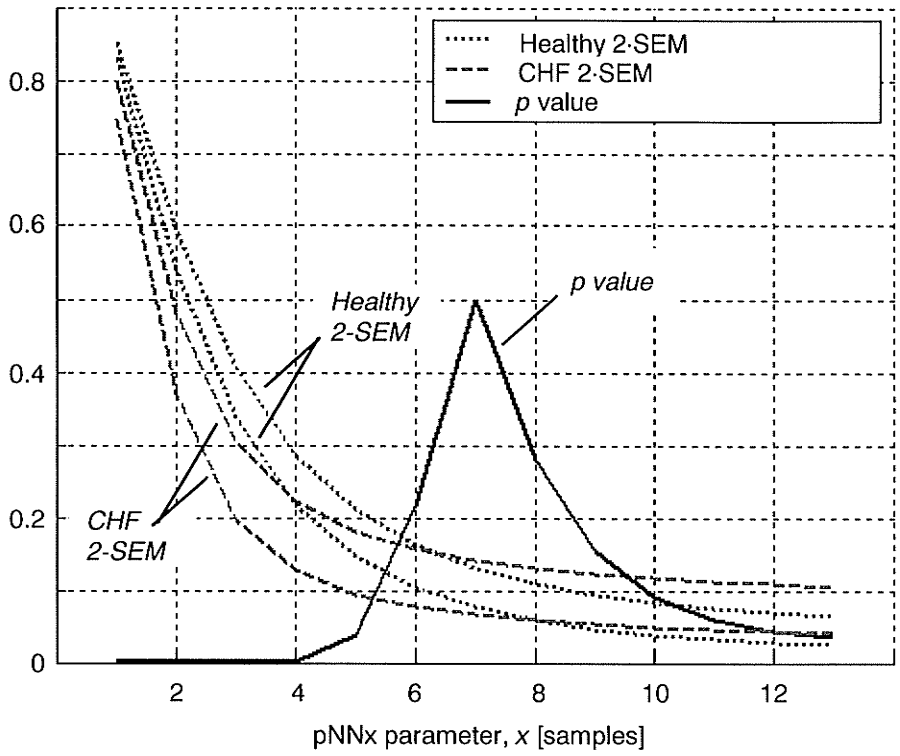


Fig. 4.1 Twice the standard error on the mean plotted alongside the confidence in the null hypothesis that no difference exists between the means (the *p* value).

Table 4.1 *p* value data used in Fig. 4.1.

pNNx Parameter, *x*		***p* value**
[samples]	**[msec]✞**	
1	8	2.05E-10
2	16	7.65E-09
3	24	6.76E-06
4	32	0.0023
5	40	0.0751
6	48	0.4362
7	56	0.9974
8	64	0.5606
9	72	0.3069
10	80	0.1795
11	88	0.1175
12	96	0.0876
13	104	0.0720

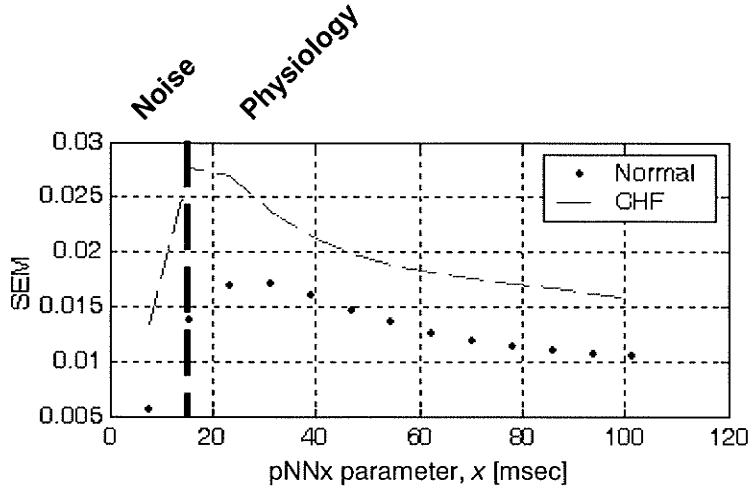
✞Threshold in msec calculated as the number of samples, *n*, divided by 128 seconds per sample.

A difference between normal and CHF samples appears to be observed for thresholds of x=8, 16, 24, and 32 msec. However, in inspecting Fig. 4.1, the SEM shrinks at x=32 msec and significantly for $x \leq 24$ msec. Plotting SEM vs. x for both samples in Fig. 4.2 exposes this more clearly but, more importantly, reveals the trend in the the mean also changes at these thresholds. This change in the character of the mean is most likely due to the 1-2 samples of data reduction noise encountered in Sec. 2.4.2. The implication of this observation is that pNNx estimates at x=8 msec should likely be discounted entirely.

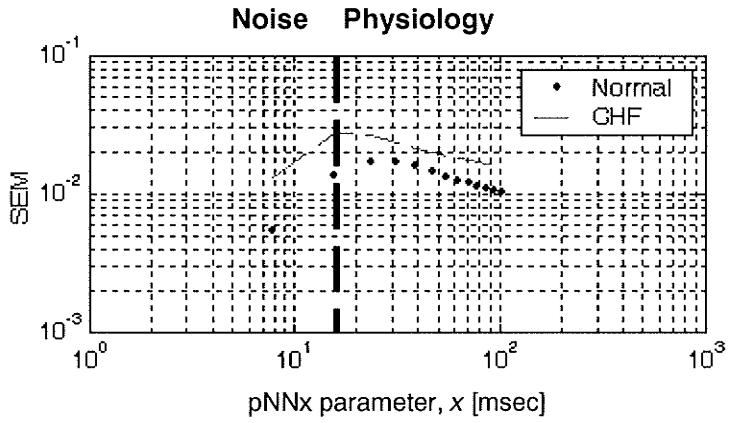
A surprising implication of Fig. 4.1, is that a threshold of 50 msec is inappropriate for characterizing the population of CHF patients. Consequently, it is clear that the pNN50 metric is inappropriate for this study. The prevalence of the pNN50 metric in literature may be due to success encountered by Ewing, *et al.* in characterizing patients with diabetes [EMYC85].

4.4 Summary

This chapter satisfies part (b) of the thesis statement by, first, comparing the HPV metric of Sec. 3, pNNx, on average for the healthy and CHF samples and finding sufficient evidence to warrant further, more detailed analysis of pNNx in the next chapter. Specifically, average pNNx values were considered over a range of x value using a t-test and the resulting p values were found to be both statistically significant ($p < 0.05$) for x=2, 3, and 4 samples and their trend, in general, was found to be consistent with results in literature generated with different data. The parameter value x=1 samples is regarded cautiously because its information on HPV physiology is contaminated with instrumentation noise; this is a new result not found in literature. This chapter is shown in relation to the rest of the thesis in Fig. 4.3.



(a)



(b)

Fig. 4.2 Standard error about the mean vs. threshold plotted (a) traditionally, and (b) log-log to illuminate small but important changes in data.

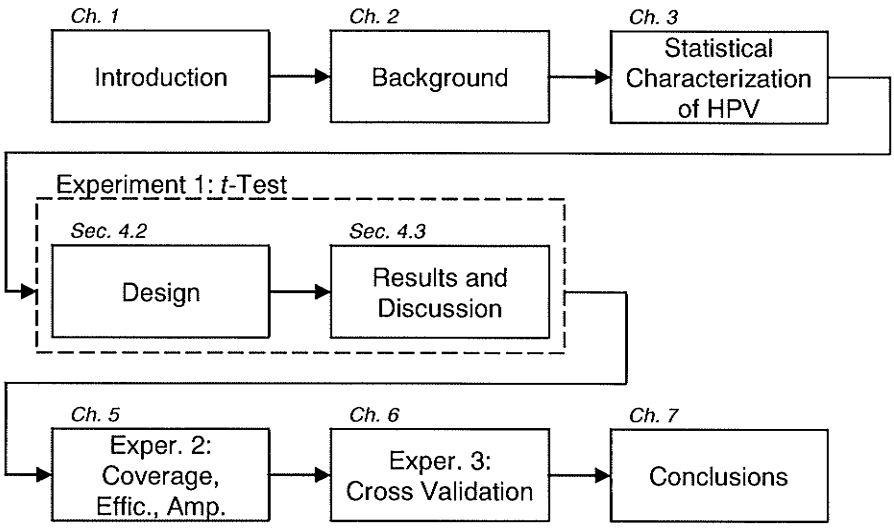


Fig. 4.3 Chapter 4 shown in relation to the rest of the thesis.

Chapter 5

EXPERIMENTAL RESULTS AND DISCUSSION: CLASSIFIER COVERAGE, EFFICIENCY, AND AMPLIFICATION

5.1 Introduction

This chapter extends the results of the previous chapter, which are necessary but not sufficient for discrimination ability to be demonstrated and for which results exist in literature for comparison. This extension fulfills the objectives corresponding to part (c) of the thesis statement. The generalizability of the results of Ch. 3 and this chapter are tested in the next chapter.

5.2 Design

5.2.1 Aims

The experiment of this chapter aims to acquire insight into

1. the cost at which coverage may be purchased by pNNx in terms of population-based misdiagnosis rates (efficiency) for an unscreened CHF population (2% prevalence,
2. how this cost scales for a population which is screened prior to the applying pNNx (thereby increasing the prevalence beyond 2%),
3. how much more visible (in terms of classifier amplitude) these screened populations (prevalence $\geq 2\%$) are after applying pNNx relative to their prevalence prior to detection, and

4. what opportunities exist for improving (in terms of coverage, efficiency, and amplification) detection based on pNNx.

Classifier efficiency is denoted η, classifier coverage is denoted κ, classifier amplification is denoted α, and all three are defined in Sec. 2.5.

These aims will be pursued in detail for the two special cases $\kappa = 95\%$ and $\eta = 50\%$ and, at a higher level of abstraction, for the general case. The intent of the special case analysis is to support inferences drawn in the general case.

5.2.2 Case 1 Method: High Coverage and Low Efficiency

The decision threshold for this case will be chosen by considering the cdf of the CHF distribution. The cdf of the CHF distribution will indicate the pNNx threshold which encompasses 95% of the pdf (to correspond with the intent of $p = 0.05$). The efficiency of the threshold – the expectation of the *posteriori* distribution, $p(C \mid x)$, in the thresholded region – will be considered as a consequence of this coverage choice. The total distribution, $p(x)$, and the *posteriori* distribution will be plotted along with their product to illustrate the character of the mathematical components of the efficiency computation. The results of Ch. 4 indicated that pNNx may not be reliable for x=1 samples and, consequently, analysis will be done here only for pNNx parameter values of x=2, 3, and 4 samples.

5.2.3 Case 2 Method: Low Coverage and High Efficiency

The second case is a test which is efficient but which lacks coverage. The efficiency will be set as 50% by the Bayes optimal threshold. *A priori*, cdf, total pdf, *posteriori*, and the total pdf multiplied by posteriori will be considered as in Case 1.

5.2.4 General Case Method

Exploration of the first aim will begin with an illustration of how the ROC curve, efficiency versus coverage curve, and the ROC slope (related to efficiency via *posteriori* probability) are related in practice. The ROC curve is a standard analytical tool (and is included for completeness) but the efficiency versus coverage curve is more illustrative of the underlying cost-benefit tradeoff present in HPV-based CHF detection. The ROC slope and *posteriori* probability estimate will be plotted to show how the two representations are related in practice. The assumption that pNNx is optimal for x=2 samples is explored within this. Comments on the efficiency versus coverage curves and ROC curves for x=1, 2 ,3 ,4, and 5 samples will be made.

The second aim is to inspect how this cost (in terms of efficiency) scales for a population which is screened prior to the applying pNNx (increasing the prevalence beyond 2%). The intent is to consider placements of pNNx within a stratification framework other than as a first stage or a modification of the 2% prevalence statistic (due to a change in assumptions, for example).

The third aim is to consider how much more visible (in terms of classifier amplitude) these screened populations (prevalence $\geq 2\%$) are after applying pNNx relative to their prevalence prior to detection. The intent of this, as discussed in Sec. 2.5.5, is that a disease with prevalence of 2% benefits from a two-fold increase in visibility to the healthcare system if the efficiency of pNNx reaches only 4%.

The fourth aim will be fulfilled via observations on how the pNNx may be improved and CHF be made more detectable in terms of coverage, efficiency, and amplification.

5.3 Results

5.3.1 Case 1 Results: High Coverage and Low Efficiency

The cdf estimate of Fig. 5.1a is used to choose the pNNx measure value $\mu(\bar{R},x)\big|_{x=16\text{msec}} \approx 0.7$ as the threshold to achieve a test coverage of $\kappa = 0.95$. The multiplication of the *posteriori* distributions and total distribution of Fig. 5.2a are used to compute the resulting efficiency for this test of $\eta = 0.022$. This implies that there will be slightly better than one true positive for every fifty false CHF test positives, a rate only slightly better than the prevalence. The results for pNNx parameter values of 24 msec (3 samples) and 32 msec (4 samples) do not fare any better.

It is apparent from the *a priori* distributions that pNNx measures for healthy subjects possesses a higher mean pNNx than the pNNx for CHF subjects. It is also clear that CHF subjects possesses more pNNx variability than healthy subjects. In other words, the heart period variability in CHF is more variable than in health. To interpret this, recall that pNNx is the area in the tail of successive RR interval differences larger than x. If the set of RR intervals is related to heart rate then pNNx can be thought of as a form of *cardiac acceleration* and pNNx as a measure of cardiac acceleration. The observations can, thus, be rephrased to state that CHF subjects occupy a larger range of cardiac acceleration but that this acceleration is higher, on average, in healthy subjects. Furthermore, a significant portion of CHF HPV indistinguishable form healthy HPV. This indicates pNNx is ineffective at discriminating health from CHF for many CHF subjects but that is capable of doing so if many false positives can be tolerated. Also, given the heterogeneity of the data set, Fig. 5.1 most likely indicates an opportunity for increasing efficiency if a pre-screening step to the application of pNNx in order to admit only a subset of CHF subjects to the pNNx test.

It should be expected, when the ROC curve is considered for the general case, that coverage will be obtained for a high FPP (i.e. low efficiency). It is also clear that the low

prevalence of CHF implies the total distribution is determined almost entirely by the healthy *a priori* distribution. This implies that, for efficiency, only the character of the healthy pNNx values is important and that the influence of CHF pNNx support will only be felt when it is at great distance from the healthy distribution. In other words, low prevalence increases mean separability requirements.

5.3.2 Case 2 Results: Low Coverage and High Efficiency

The posteriori probabilities of Fig. 5.4a are used to select a Bayes optimal threshold of $\mu(\overline{R},x)\big|_{x=16\text{msec}} \approx 0.2$. This achieves an efficiency of approximately 75% at the cost of a coverage of $\kappa \approx 0.15$ as indicated by the cdf of Fig. 5.3.a. Again, this result is better than the results obtained for pNNx parameter values of 24 msec and 32 msec.

It is clear by inspection Fig. 5.3a that the Bayes optimal approach is simply to assume CHF does not exist. In the region where it is not in conflict with the healthy support, CHF can, of course, be detected with great efficiency. This confirms the suggestion of Sec. 5.3.1 that there is a subset of the CHF subjects for whom pNNx is highly effective (in possession of both accuracy and efficiency). Recall the distribution of ages for the CHF subjects in the sample shown in Fig. 4.1. Reducing pNNx test subjects to, say, ages 50 and greater may be a low-cost way of effecting this subset. More data gathered in a carefully designed manner is required to verify this assumption (and related assumptions), however.

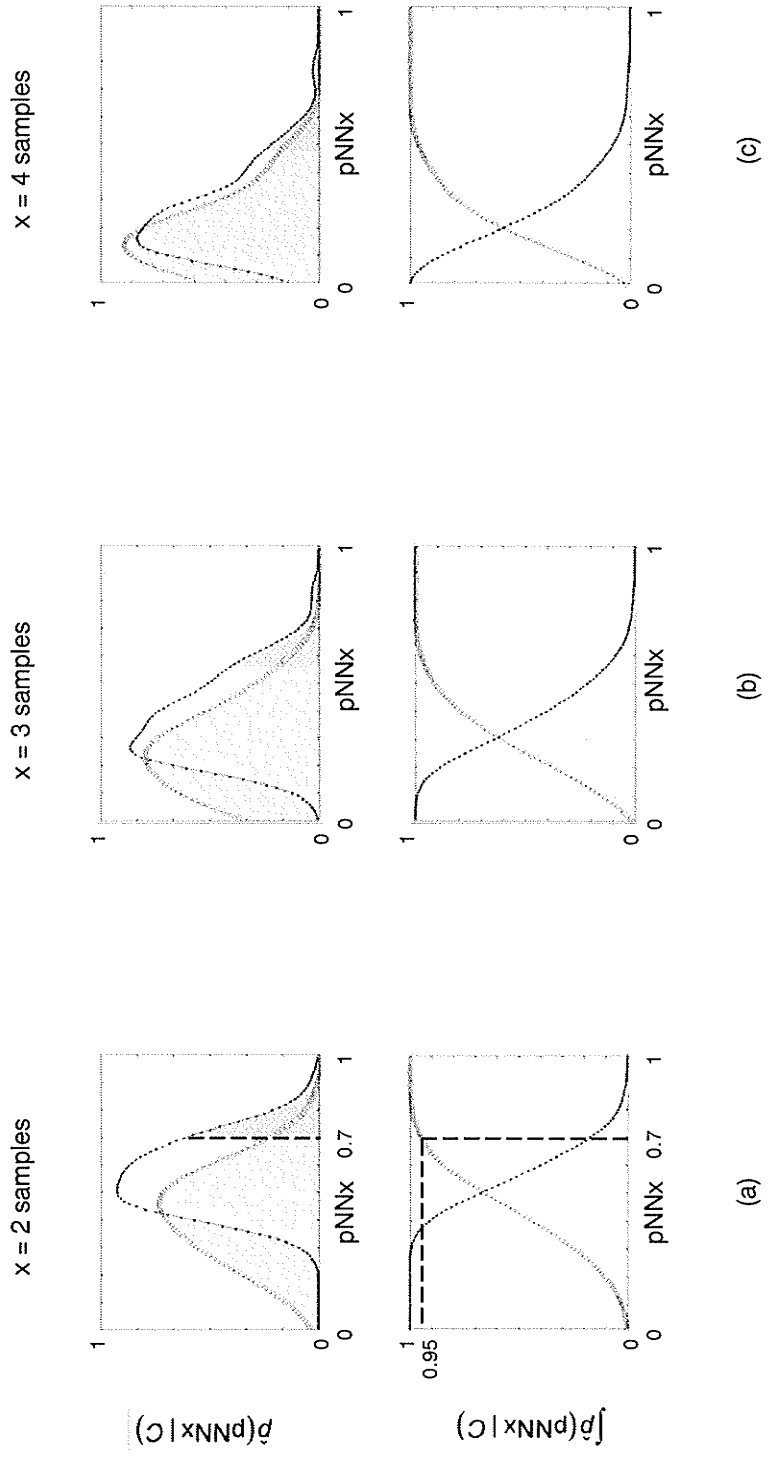


Fig. 5.1 Use of CHF (red circle) and healthy (blue dot) cdfs to establish decision threshold which achieves high accuracy (95% of the CHF *a priori* distribution) for pNNx parameter values of (a) 2 samples, (b) 3 samples, and (c) 4 samples.

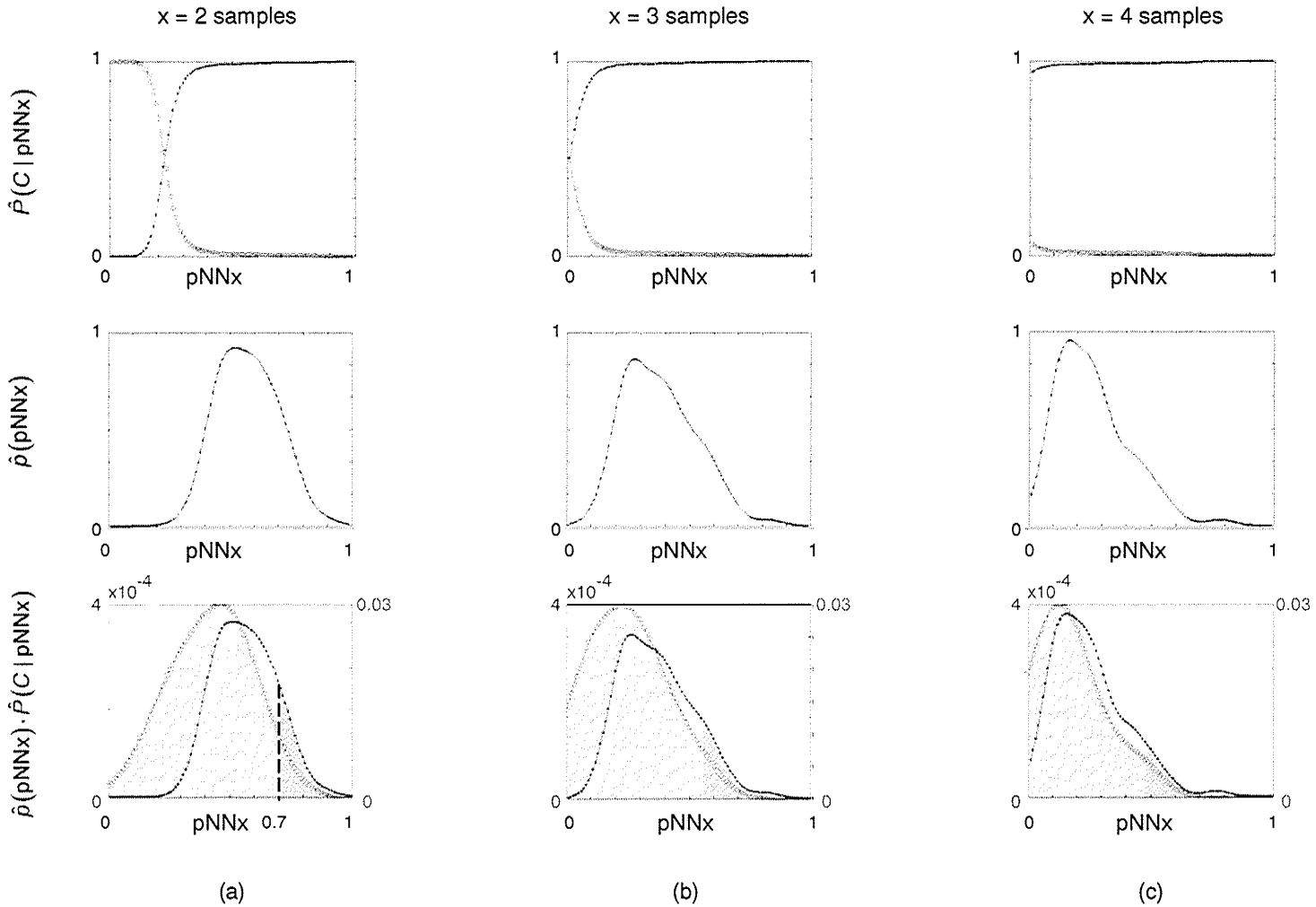


Fig. 5.2 Use of the product of the posteriori distribution and total pNNX distribution of CHF (red circle) and healthy (blue dot) samples to estimate effectiveness (2.2%) of a 95% accurate test for pNNx parameter values of (a) 2 samples, (b) 3 samples, and (c) 4 samples.

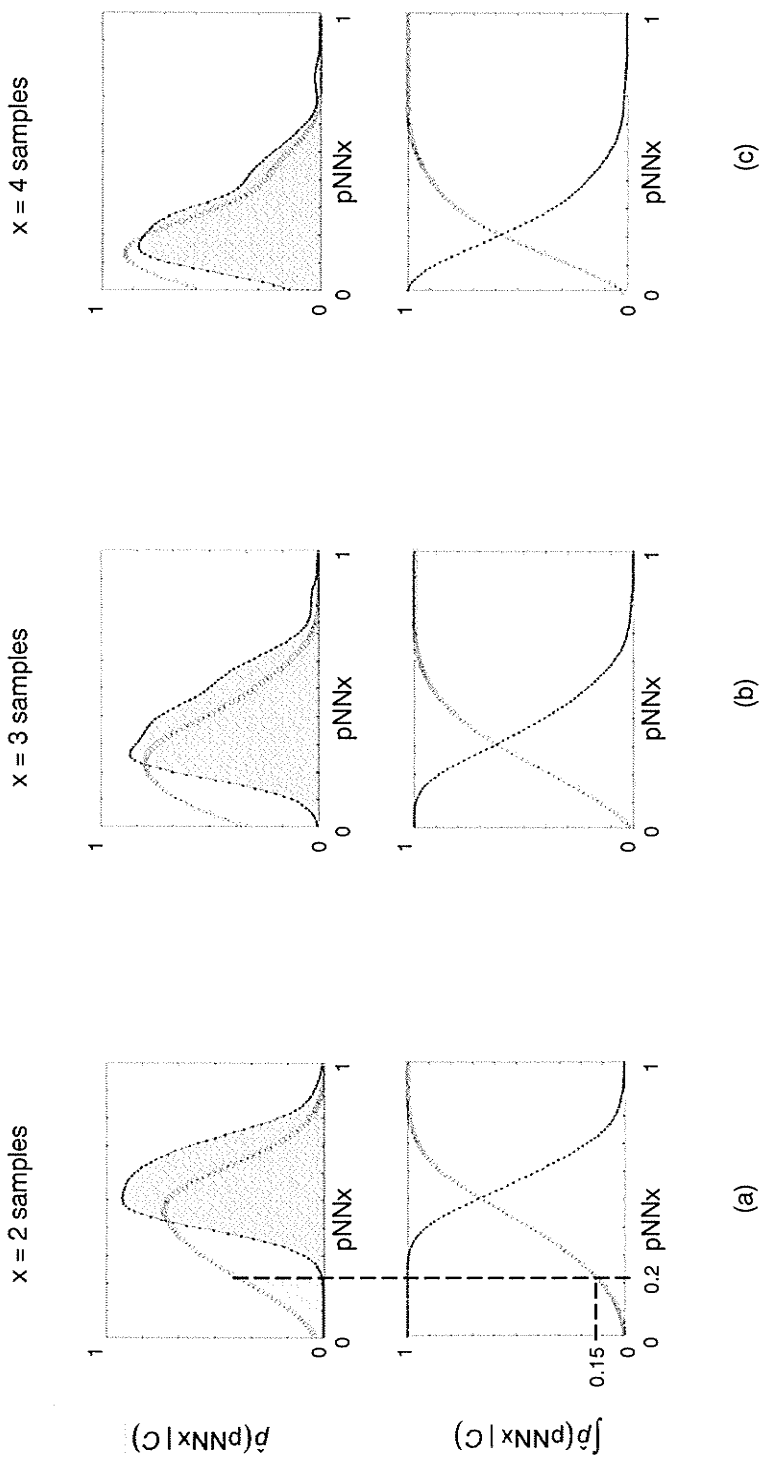


Fig. 5.3 Use of CHF (red circle) and healthy (blue dot) cdfs to establish decision threshold which achieves high effectivness (75%) at low efficiency (15% of the CHF *a priori* distribution) for pNNx parameter values of (a) 2 samples, (b) 3 samples, and (c) 4 samples.

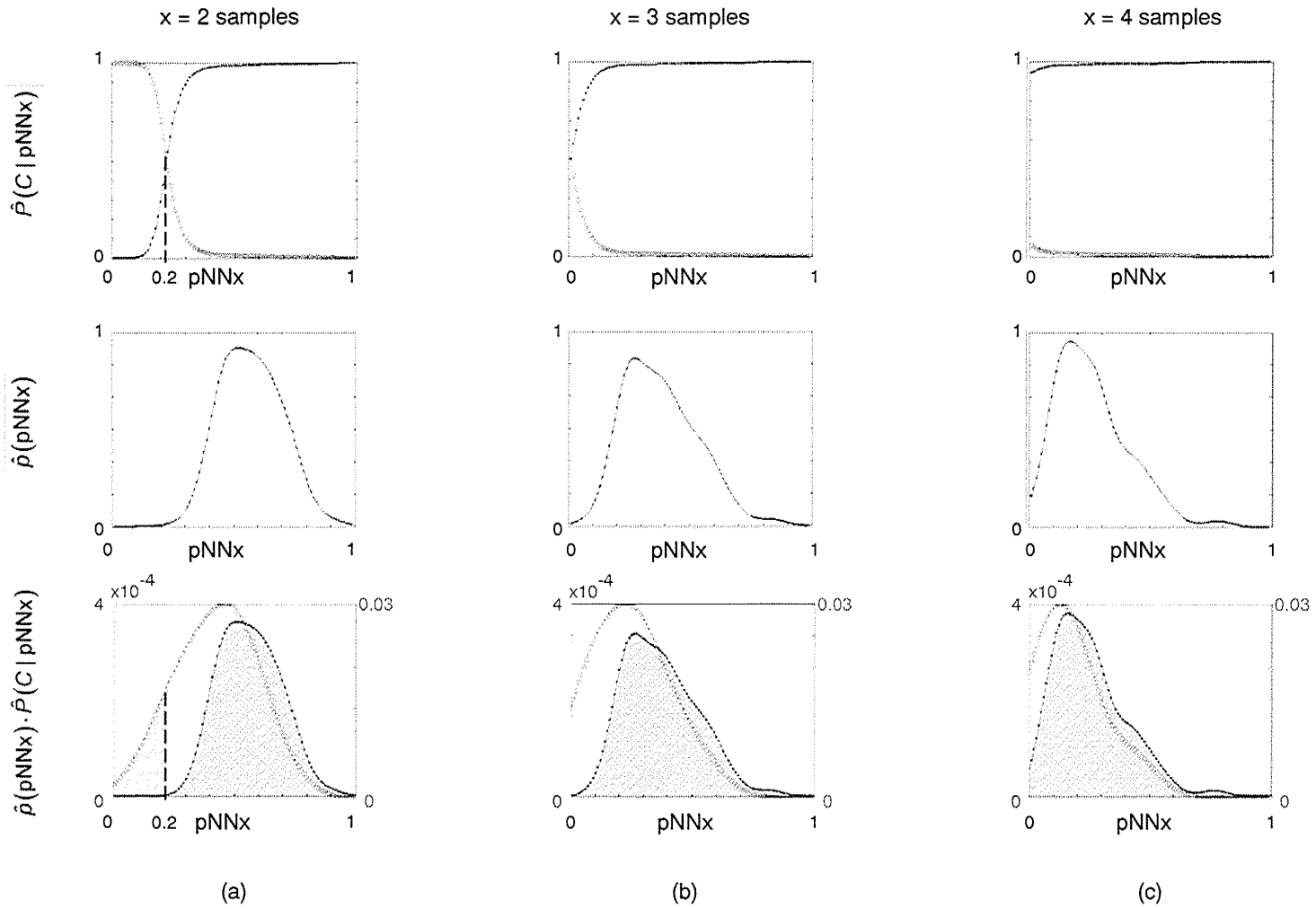


Fig. 5.4 Use of the product of the posteriori distribution and total pNNX distribution of CHF (red circle) and healthy (blue dot) samples to estimate effectiveness (75%) of a 15% accurate test for pNNx parameter values of (a) 2 samples, (b) 3 samples, and (c) 4 samples.

5.3.3 General Case Results

Recalling the ROC curve of Fig. 2.40, the ROC curve of Fig. 5.5a possesses moderate discrimination power over the range of its threshold. It must be emphasized, however, that only the high-coverage range of the threshold is of importance for detection problems in medicine. The relationship of the curve to the ROC curve for x=2 is given in Fig. 5.5b. This curve does not rise sharply during low values of FPP. This indicates that FPP will constitute a significant portion of the total positives for low values of FPP and that this proportion will grow as FPP increases. This is represented by the sharp decline in efficiency as a function of coverage. The relationship of efficiency to the ROC slope and likelihood ratio via *posteriori* probability is given in Fig. 5.5c. These relationship are illustrated for completeness more than for the purpose of empirical inference and were discussed theoretically in Sec. 2.5.5. Specific values are given for Fig. 5.5a and Fig. 5.5b in Table 5.1 and Table 5.2, respectively.

The results of Fig. 5.6 confirm the suggestion of Sec. 5.3.1 and Sec. 5.3.2 that x=16 msec is optimal for the digital sampling rate of 128 samples per second used in this work. It is clear, however, that, if sampling rate were increased, that the next most optimal curve would be found between x=8 msec and x=16 msec. This may be of considerable value because instrumentation noise due to inaccuracies in peak detection (discussed in Sec. 2.4.2) will also decrease as a result of an increase in sampling rate. It is also clear from the ROC curve of Fig. 5.6b that x=5 samples (40 msec) is essentially equivalent to a random guess. Thus, pNN50 metric is inappropriate in this instance.

Table 5.1 Values for Fig. 5.5a.

FPP	κ
0.000	0.000
0.025	0.303
0.050	0.360
0.075	0.401
0.100	0.421
0.125	0.463
0.150	0.485
0.175	0.507
0.200	0.529
0.225	0.551
0.250	0.573
0.275	0.595
0.300	0.616
0.325	0.638
0.350	0.638
0.375	0.660
0.400	0.681
0.425	0.702
0.450	0.722
0.475	0.742
0.500	0.761
0.525	0.780
0.550	0.797
0.575	0.815
0.600	0.831
0.625	0.846
0.650	0.861
0.675	0.874
0.700	0.887
0.725	0.899
0.750	0.910
0.775	0.920
0.800	0.929
0.825	0.938
0.850	0.952
0.875	0.959
0.900	0.969
0.925	0.978
0.950	0.984
0.975	0.993
1.000	1.000

Table 5.2 Values for Fig. 5.5b.

TPP	η
0.000	1.000
0.025	1.000
0.050	0.981
0.075	0.962
0.100	0.921
0.125	0.867
0.150	0.782
0.175	0.667
0.200	0.602
0.225	0.470
0.250	0.352
0.275	0.302
0.300	0.258
0.325	0.189
0.350	0.162
0.375	0.139
0.400	0.121
0.425	0.092
0.450	0.081
0.475	0.072
0.500	0.064
0.525	0.058
0.550	0.053
0.575	0.045
0.600	0.042
0.625	0.039
0.650	0.037
0.675	0.035
0.700	0.033
0.725	0.030
0.750	0.029
0.775	0.028
0.800	0.027
0.825	0.026
0.850	0.025
0.875	0.023
0.900	0.022
0.925	0.021
0.950	0.020
0.975	0.019
1.000	0.017

The plots of Fig. 5.7 reveal that, for a physician using pNNx in her office, that CHF (as it is defined by the present sample population) must have a prevalence of approximately 50% before an acceptable level (set here as 50% for demonstrative purposes) of efficiency is attained. This seems to bode poorly for the usefulness of pNNx unless the increase in disease visibility (the relative efficiency referred to here as classifier amplification) is considered. A potentially useful region of 95% coverage and 50% efficiency is shown in Fig. 5.7b as a classifier performance target. It should be noted that efficiency would improve significantly if CHF could be pre-screened to identify a sub-group with lower pNNx values (via some easy-to-assess risk factors). This would have the additional effect of raising the pre-pNNx prevalence.

Test efficiency is therefore, very dependent on prevalence. To remove this variable, test performance is plotted in terms of the percentage increase the test's efficiency achieves relative to its prevalence, termed classifier amplification in earlier discussions. This plot is shown in Fig. 5.8 and specific values drawn from this plot are given in Table 5.3. This shows that, despite poor efficiency, that pNNx is actually able to raise the visibility of CHF subjects relative to their unscreened detectability by 11.4% for 2% prevalence. A 34.1% increase can be obtained if 85% coverage is acceptable. In general, decreasing prevalence has the effect of reducing efficiency but increasing the benefit to the affected population.

Table 5.3 Percentage improvement relative to prevalence (classifier amplification).

PREVALENCE [%]	PERCENTAGE IMPROVEMENT		
	85% Coverage	90% Coverage	95% Coverage
2	34.1	23.1	11.4
10	30.4	20.8	10.4
25	24.1	16.8	8.5
50	14.9	10.6	5.5

5.4 Discussion

The results of the previous section found that CHF subjects occupy a larger range of cardiac acceleration (defined in Sec. 5.3.1) but that this acceleration is higher, on average, in healthy subjects. Also, it found that pNNx was optimal for x=2 samples (16 msec).

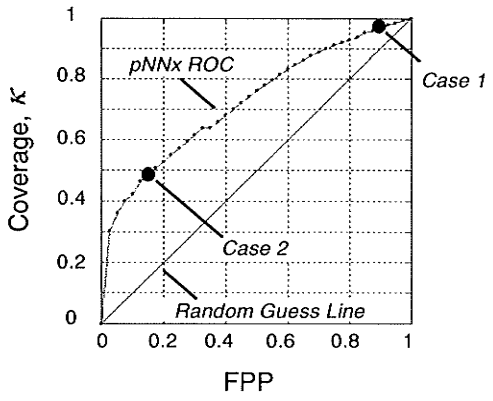
Since the size of pNNx distribution variability is on par with the separation of distribution means, CHF coverage can be made high (which is necessary for disease detection) but that a high price is paid in terms of efficiency. In fact, efficiency converges towards the CHF prevalence for a variety of assumed starting prevalences. This, therefore, implies that low prevalence increases mean separability requirements. Efficiency, however, is an absolute measure and will be low regardless of how good the detector is if the prevalence is low. Efficiency is, of course, important when tallying the costs of deploying a detector on a wide scale but does not tell the full story of pNNx.

The effect of prevalence is accounted for by considering an efficiency increase relative to prevalence – classifier amplification. The results of this analysis indicate that, the less prevalent the disease the higher the amplification and, therefore, relative benefit to the affected population. At the 2% level of prevalence (which is the actual level of CHF prevalence in practice), pNNx raised the visibility of the CHF sample by 11.4% and 34.1% while finding 95% and 85% of CHF subjects, respectively.

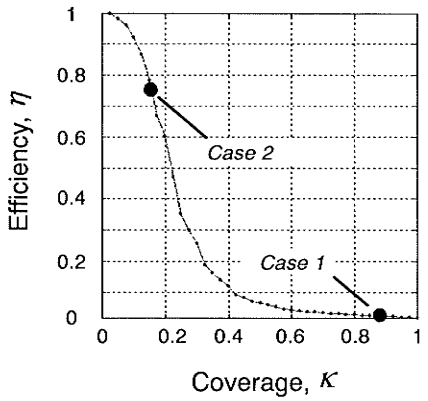
The above results considered pNNx as the first tier of a risk stratification tree (refer back to Fig. 1.2). If it is paired with a first tier stage which detects a key 15% of CHF individuals then efficiency can be raised to 75%. It should be noted that some inefficiency must be tolerated because the most efficiency approach is to simply assume CHF does not exist.

These results and discussion imply three potential future courses of action. First, pNNx could be used on an unscreened CHF population, used to detect 95% of CHF subjects, and the low efficiency can simply be tolerated as a reasonable cost for

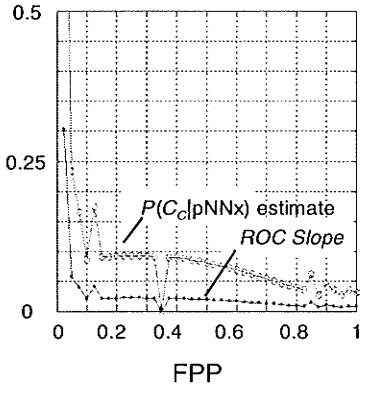
increasing CHF visibility via the amplification metric. Second, a search for a complementary second feature could be initiated to improve the mean separability of unscreened healthy and CHF pNNx distributions. Third, pNNx could be used as a second tier risk stratifier in combination with a first tier screening step which drastically increases pNNx distribution mean separation while reducing distribution variability. The outcome of this third option would be high efficiency at the cost of detecting only 15% of the present CHF population.



(a)

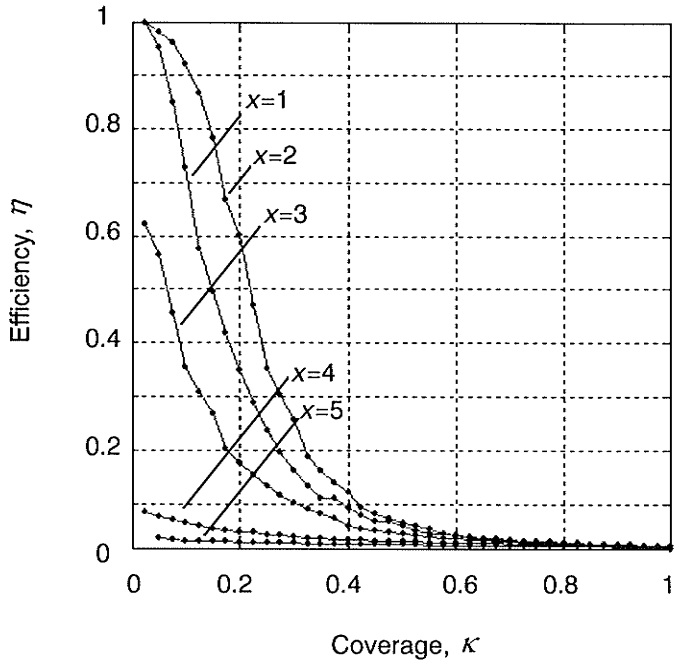


(b)

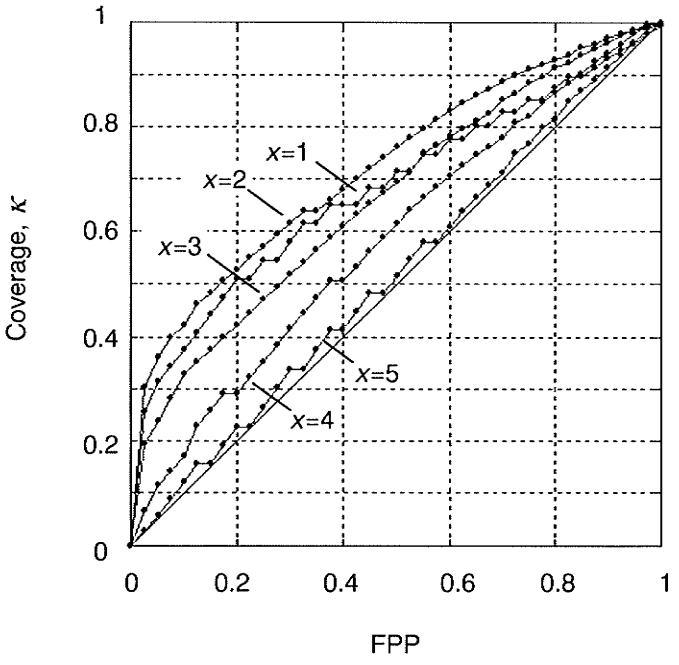


(c)

Fig. 5.5 Demonstration of relationship at x=2 of a) efficiency via Bayes' Theorem, b) the ROC curve, and c) Bayes posteriori probability calculated from the ROC curve.

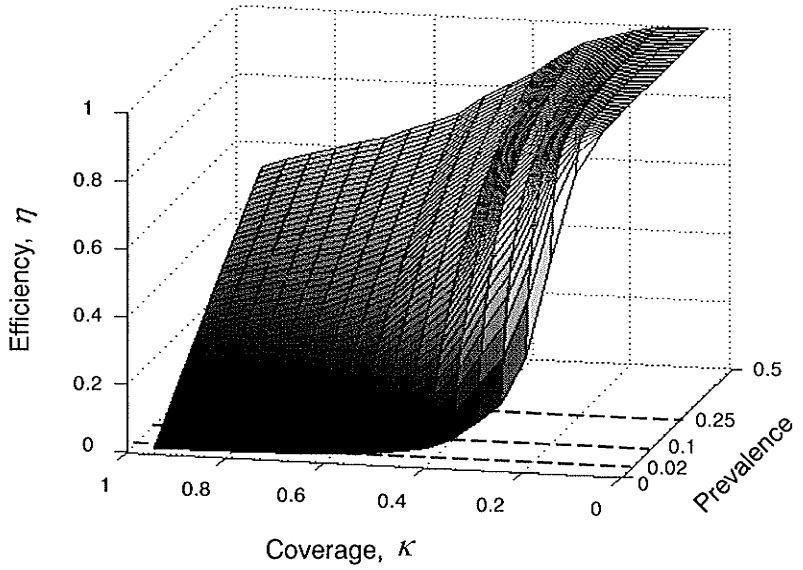


(a)

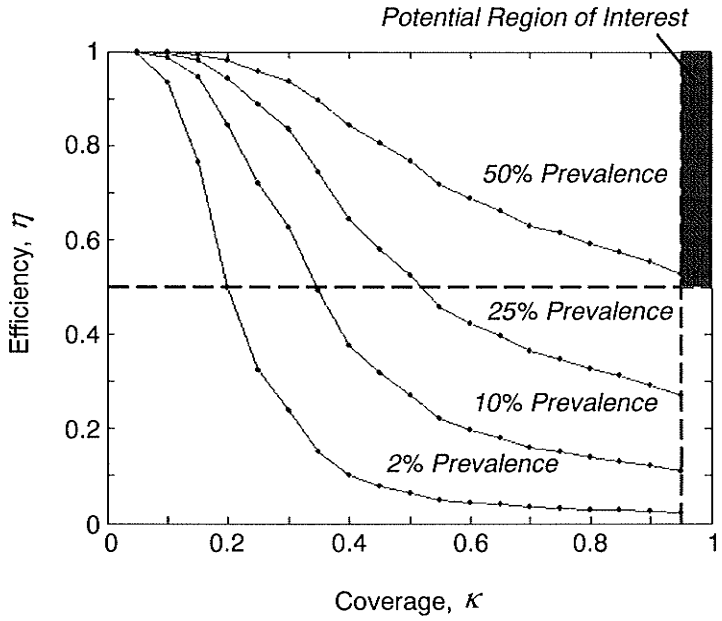


(b)

Fig. 5.6 Evolution with pNNx parameter x of a) efficiency, and b) the ROC curves.

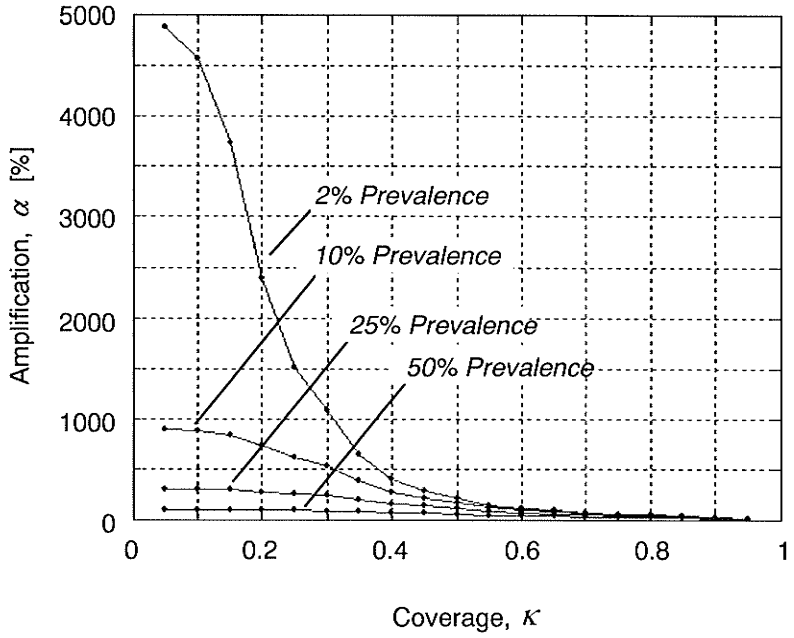


(a)

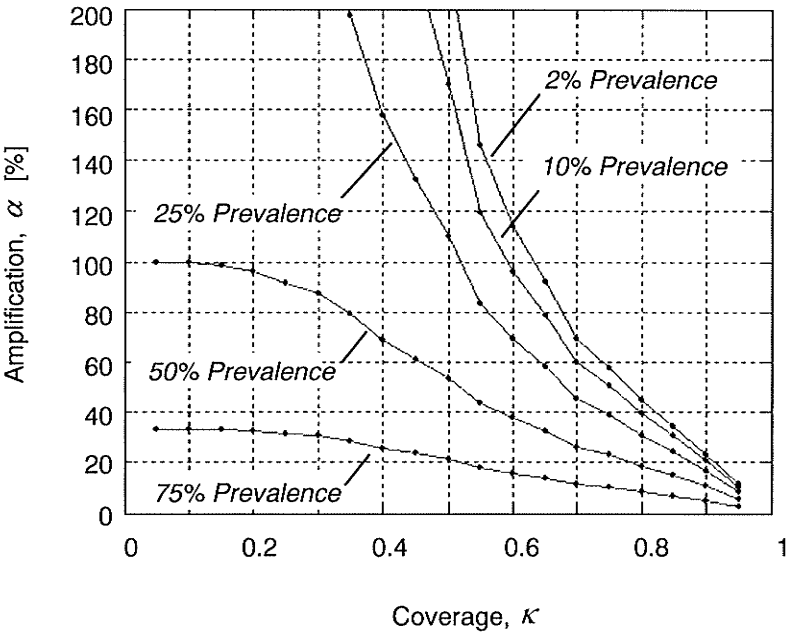


(b)

Fig. 5.7 Trend of efficiency versus coverage for a variety of prevalences showing (a) general case as a surface, and (b) specific instances with a potential target region for classifier performance (shown shaded grey).



(a)



(b)

Fig. 5.8 Efficiency improvement relative to prevalence
(a) for the full range of amplification, and
(b) in the high-coverage portion of the graph.

5.5 Summary

This chapter satisfies part (d) of the thesis statement by expanding the discrimination assessment of the previous chapter with the assessment measures of classifier coverage, classifier efficiency, and classifier amplification. The results of this chapter indicate a CHF sample can benefit from the use pNNx as an assessment of HPV with important limitations. The generalizability of the results of this and the previous chapter are tested in the next chapter. This chapter in the context of the rest of the thesis is shown in Fig. 5.9 below.

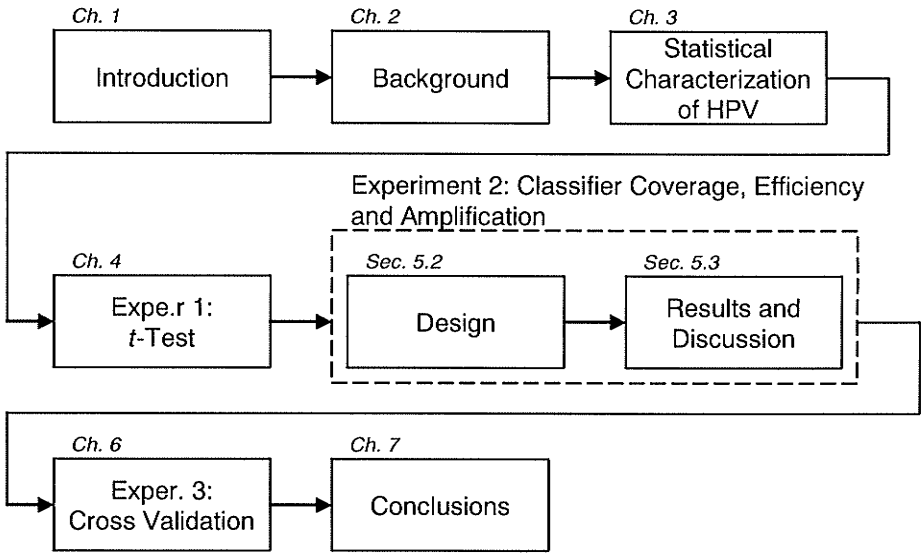


Fig. 5.9 Chapter 5 in the context of the rest of the thesis.

Chapter 6

EXPERIMENTAL RESULTS AND DISCUSSION: CROSS VALIDATION

6.1 Introduction

The previous chapter obtained results for the cost at which coverage is obtained using the pNNx metric. These results may, however, change given a different sample of CHF even if the statistical character is similar. To address this concern, this section performs cross validation.

6.2 Design

Cross validation is an estimate of generalization error achieved by partitioning the healthy and CHF data sets and examining the consistency with which the decision rule is able to classify CHF data which did not participate in its creation [Hior94]. Cross validation begins by partitioning data into a set, G – called the training set – which will be used to create the decision rule and a set, H - called the test set – which will be classified according to the decision rule. The results of Sec. 5 will be said to generalize if the cross validation of this section is consistent and in agreement with Sec. 5.

H will consist of a healthy test set, and a CHF test set. In general, it is desirable for each test and training set to consist of half of the available data. When data samples are small, as is the case with the CHF sample (only 29 subjects), it is common to use only 1 subject in the test set. This is called *leave one out* (LOO) cross validation in the literature [Hior94].

The prevalence of CHF, 2%, should be preserved in the test set and, therefore, the healthy test set will consist of 49 subjects. This leaves 23 healthy subjects and 28 CHF subjects for the test set.

The LOO approach allows 29 different training/test set pairs to be evaluated. The 49 subjects will be selected at random from the 72 for each of the 29 training/test set pairs. The outcome of each training/test set evaluation will be the success with each the single CHF subject is classified (it will either classify the single subject correctly or it will not) as well as false positives. The average rate at which the subject is classified correctly will be the efficiency. The decision threshold for pNNx will be set, as it was done in Sec. 5, to achieve a range of coverage.

The random selection of the 49 subjects is accomplished by selecting the first 49 locations from a randomized version of the 72 element array of the healthy subjects. Let a represent the array of healthy subjects sorted in order of record number (see Appendix C). Randomization is accomplished by generating 1000 numbers distributed uniformly between 1 and 72 using the Matlab rand(1,1000) function. Let this array be b. Randomization occurs by swapping the first element of a with the element of a at location $b(i)$ for i monotonically increasing where $b(i)$ is the element of b at location i.

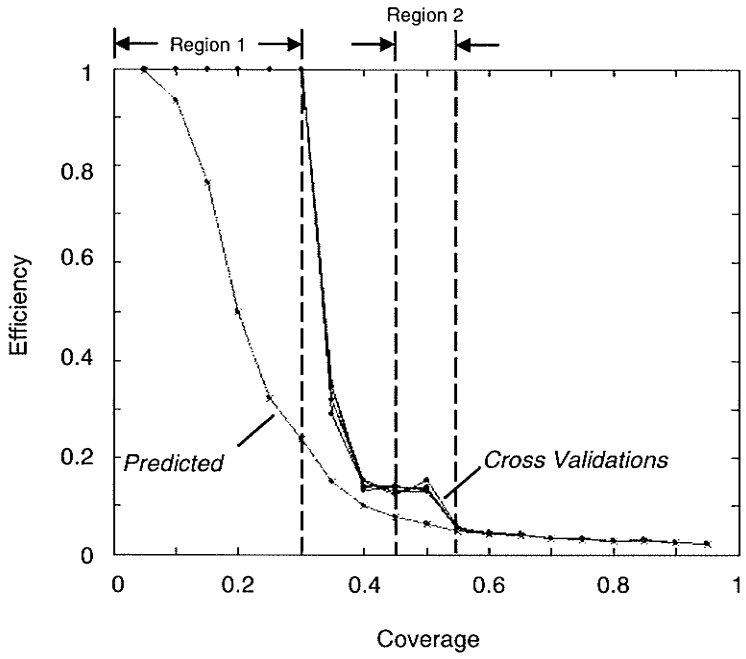
6.3 Results and Discussion

The results of Fig. 6.1 indicate the power of the results of Sec. 5 to generalize. The region of discrepancy for lower efficiencies is due to the sparseness of the support for healthy subjects in that region. This sparseness is visible when the healthy pmf is estimated using a non-overlapping box kernel (the so-called histogram approach). Sparseness does not exist in the densities used in Ch. 5 because a continuous Gaussian kernel function was used to estimate the density functions resulting in a smooth density estimation. It is more likely that the sudden density changes in Region 1 and Region 2 of

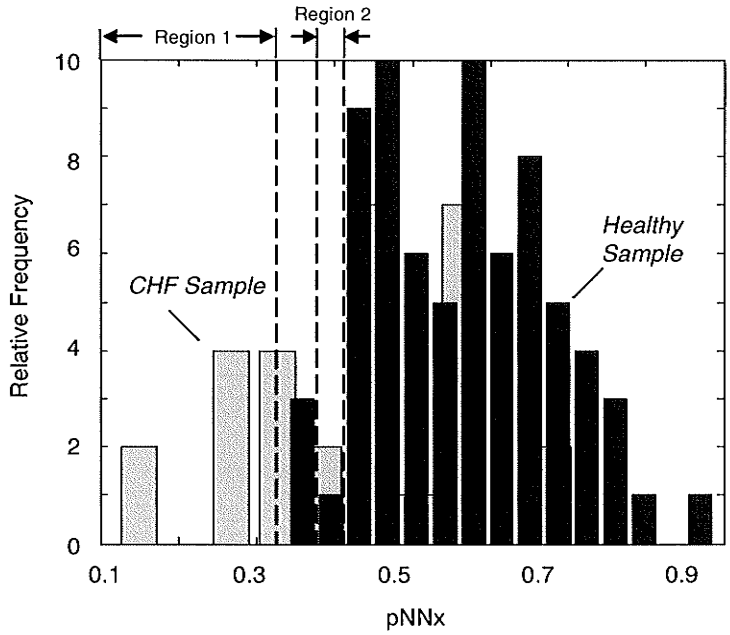
Fig. 6.1b are due to insufficient data than to physiology and, consequently, the use of a kernel which performs smoothing is appropriate.

6.4 Summary

This chapter fulfills part (d) of the thesis statement by finding that the performance of pNNx in Ch. 5 will likely generalize to other comparable CHF and healthy samples due to the high consistency of efficiency versus coverage curves obtained while performing LOO cross validation. This chapter is shown in relation to the rest of the thesis in Fig. 6.2.



(a)



(b)

Fig. 6.1 Results of cross validation showing (a) general agreement with predicted values, and (b) discrepancies explained by sparse data via density estimation of subjects using non-overlapping box kernels.

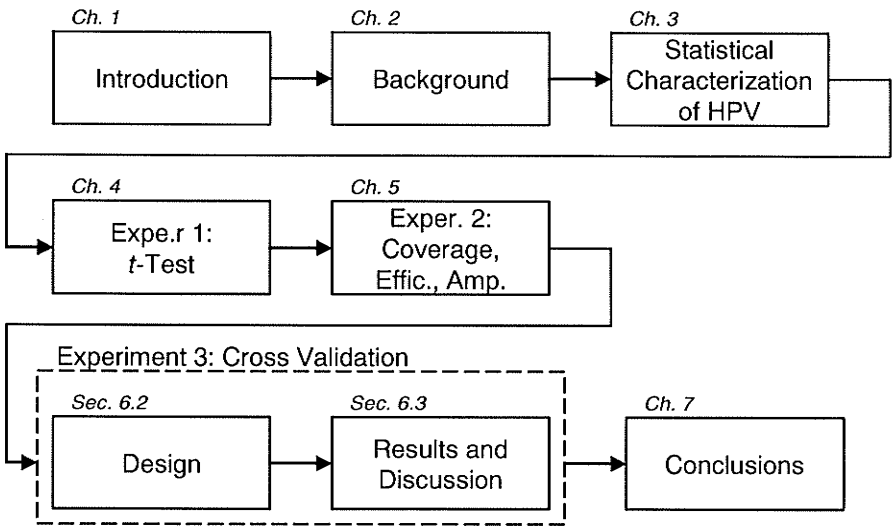


Fig. 6.2 Chapter 6 in relation to the rest of the thesis.

Chapter 7

CONCLUSIONS

7.1 Thesis Summary

7.1.1 Aims

This work advances the thesis that (a) a metric with the power to statistically discriminate individuals with *congestive heart failure* (CHF) from healthy individuals can be derived from RR interval data, (b) its discrimination power and its limitations can be characterized consistently with results existing in literature (namely via the *t*-test), (c) the results in literature can be expanded on with more powerful discriminant assessments such as error rates per class (termed classifier *coverage*), per decision (termed classifier *efficiency*), and per decision relative to CHF prevalence (termed classifier *amplification*), and (d) these results will generalize given a different sample of comparable data.

7.1.2 Work Done: Classification

Work done in satisfaction of part (a) of the thesis statement derived a data-driven metric of *heart period variability* (HPV). This metric considered physiological causes for variability in RR interbeat interval data but relied on inspections of data-driven *probability distribution function* (pdf) estimates to establish their distinct statistical character. The distinct statistical character of the pdf estimate is significant considering the standard approach in literature is to replace pdf-driven HPV metric derivation with other metrics that are supposed to be trivial variations on a theme. In contrast to this assumption, this work found only one logical form for a HPV metric which suited the

sample populations of the *electrocardiogram* (ECG) data used in this work. This form relies on information in the tails of the distribution of RR interval successive differences, $\overline{R}$, and is equivalent to the pNNx metric found in literature if a non-overlapping box function is used to estimate the distribution of $\overline{R}$. $\overline{R}$ can be thought of as a form of cardiac acceleration and pNNx a measure of this form of cardiac acceleration.

Work done for part (b) of the thesis statement found pNNx for x=8 msec is contaminated with instrumentation noise – an effect not reported in literature. It is presumed an increase in sampling rate would reduce errors in ECG R-complex annotation and, consequently, more fine structure within the ECG could be revealed. The discrimination power of low values of x implies, in general, that the HPV usefulness in discriminating CHF from health resides in low amplitude cardiac accelerations.

Additional work done for part (b) of the thesis statement obtained p values from a t-test as a function of the pNNx parameter, x, which were in agreement with results produced in literature using different data. The p value is widely used in medical literature.

Work done for part (c) of the thesis statement found pNNx produces optimal results for x=16 msec when a sampling rate of 128 samples per second is used. pNNx, when used in literature, is often used with the parameter value x=50 msec, a result chosen approximately 20 years ago in the analysis of HPV in diabetics. pNNx with x=50 msec, for the sample population of this work, is no better than a random guess.

These results were expanded by an analysis of accuracy (via *receiver operating characteristic* (ROC) curves) and efficiency (via Bayesian statistics) – techniques which are common in clinical chemistry, radiology, and radar signal processing but have been found in this work to be remarkably absent in HPV literature. It was found that CHF coverage can be made high (which is necessary for disease detection) but that a high price is paid in terms of efficiency and that, at the 2% level of prevalence, pNNx raised the visibility of the CHF sample by 11.4% and 34.1% while finding 95% and 85% of

CHF subjects, respectively. If pNNx is elevated from a first to a second tier risk stratifier (refer back to Fig. 1.2) efficiency can be made as high as 75% if the first tier stage detects a key 15% of CHF individuals.

7.1.3 Work Done: Generalization

The above outcomes of part (a), part (b), and part (c) of the thesis statement were found to have the power to generalize by cross validation, fulfilling the fourth and final part of the thesis statement, part (d). It is important for the results to generalize to be repeatable in an engineering prototype or for other researchers to successfully replicate the results.

Part (b) was fulfilled throughout via the use of publicly available data with a history of publication in peer reviewed medical journals.

7.2 Conclusions

The first conclusion of this work is that it is possible to use the characteristics of HPV data to derive a metric which suits the data. Alternately, the characteristics of the data can be a guide to the selection of a HPV metric from a set of existing metrics. The alternative often practiced in literature is to choose a HPV metric and then measure the data in terms of this metric. Only one metric of those used in literature, pNNx, was found to suit the general form of the metric derived in Ch. 3. An additional advantage of metric derivation is the relative ease with which meaning may be attached to the results. For example, it was clear that pNNx measured a form of cardiac acceleration.

The second conclusion of this work is that low amplitude cardiac accelerations hold the HPV information which allows CHF to be discriminated from health but that the lowest of these accelerations is contaminated by variability due to instrumentation and not to physiology.

The third conclusion is that pNNx may be applied with benefit in two regards. It may be applied to an unscreened population (represented by the present sample) with good coverage, a non-trivial increase in the visibility of the diseased population, but at the cost of a detection efficiency which may be low enough to prevent its implementation in practice. Alternately, pNNx may be applied to a screened CHF population (sacrificing coverage) but at very reasonable efficiency. Specifically, this corresponds to either 11.4% amplification for 95% coverage or 75% efficiency for 15% coverage. These numbers are significant because pNNx was applied on a simple premise (a model of stochasticity) and, therefore, they represent a performance baseline which must be surpassed in order to justify the use of more complex metrics.

The forth conclusion of this work is that the need for an effective characterization of discrimination power which enhances results found in literature is met in this work. This need is identified in Sec. 1.2 and is achieved through the creation of the metrics of classifier coverage, classifier efficiency, and classifier amplification and their use as above.

The fifth conclusion of this is work is that risk stratification remains a necessary approach to medical decision based for HPV for CHF. This is due to the cardiovascular system which, as a natural physical system, produces significant "bit error rates." in comparison to human-engineered systems.

The sixth conclusion of this work is that the performance of pNNx reported here will be accepted as credible from the standpoint of medicine. Medicine requires a physiologically motivated quantitative marker of HPV whose definition is described using accessible mathematical and statistical language. This has been achieved here via data-driven metric derivation (which captures the physiology via statistics) and the use of a powerful marker of ANS performance: the depolarization frequency of the Purkinje complex as it is manifest in the R-complex of lead III of the ECG. A p value analysis of the separability of the population average pNNx values was in agreement with literature

and the theory on which the p value rests was connected to the more descriptive metrics of classifier coverage, classifier efficiency, and classifier amplification in the theoretical development of Sec. 2.5.

The seventh conclusion of this work is that the performance of pNNx reported here will be accepted as credible from the standpoint of engineering. This is achieved, first, by deriving HPV metric requirements from ECG data in Ch. 3 and commenting, as the work progressed, on the consequent suitability of common HPV metrics found in literature. Credibility in the eyes of engineering is achieved, secondly, through the extension of a qualitative visual analysis of discrimination power (identified in Sec. 1.2.4 as the norm in engineering literature) through the metrics of classifier coverage, classifier efficiency, and classifier amplification.

The eighth and final conclusion of this work is that the above conclusions will have the power to generalize. This was found by cross validation in Ch. 6. It is important for the results to generalize to be repeatable in an engineering prototype or for other researchers to successfully replicate the results.

7.3 Contributions

This section will be divided into two parts. The first part will be *contributions*, which are defined here as ideas which the author has not encountered during an extensive review of literature and which required significant development in this work. The second part will be *innovations*, which are defined here as ideas which the author has not encountered during an extensive review of literature and which did not entail significant development.

Contributions made by this work include the following:

- creation of a statistical framework in which to judge the suitability of a given HPV metric, a framework not seen in literature despite the multiple and recent reviews of HPV metrics;
- the definition of pNNx in terms of this framework and resulting connection between pNNx and the tail of the of the RR interval successive differences, a result not seen in literature;
- formulation of probability density estimation and Bayesian probability which distinguishes between the true pdf, estimated pdf, true pmf, and estimated pmf at the various states of kernel-based non-parametric estimation;
- a new analysis of existing data;
- the distinction between variability due to instrumentation and variability due to the physiology including an estimate of the magnitude of the contribution of both to pNNx, a result not reported in literature;
- the definition of classification efficiency and classification amplification on *posteriori* distributions, the term coverage, and the application of all three as a unified approach to discriminant analysis; and
- the software and computation approaches developed to produce the results of this work;

Innovations made in this thesis include the following:

- efficiency measurements are performed for a range of coverage, expanding the scalar values typically reported in ROC literature;
- the first ROC analysis on HPV data encountered by the author is performed; and
- the first Bayesian analysis performed here on either untransformed HPV data or on CHF HPV data.

7.4 Future Work

The results of this work can be taken in three important directions. The first entails the careful design of the algorithm by which healthy and CHF samples are selected. This work has suggested a subset of the present CHF sample is important for efficient detection. This subset may be identified as simply as restricting the age of the CHF sample. Risk stratifiers such as aging, arrhythmogenesis, myocardial infarction, sleep apnea, and valvular disease have been identified elsewhere and their inclusion or exclusion should be considered. In addition, it would be helpful for generalizability if the sample was larger than the sample of this work which, though adequate, was only 29 subjects in size. Lastly, it is presumed an increase in sampling rate would reduce errors in ECG R-complex annotation and, consequently, more fine structure within the ECG could be revealed.

The second major promising direction entails the use of ECG recording equipment which extends the sampling rate of the data in this work. The extent to which an increase in sampling rate would reduce instrumentation noise is unknown but the optimal value of x for pNNx would shrink to x=4 msec if the sampling rate were increased to 512 samples per second and the principal error magnitude remains at 1 sample. More fine structure in the ECG would be revealed at this filtering level, however, so further empirical investigation is required.

The third major promising direction is pairing pNNx, the best known untransformed HPV metric, with a HPV metric which operates in a transform domain. Linear transforms should be considered before nonlinear transforms and, in particular, previous results produced using the Fourier transform should be repeated on HPV data which has been transformed such that its time domain sampling rate is uniform. Non-uniform time domain sampling makes the interpretation of the physical cause of the resulting "frequencies" difficult. The importance of nonlinear transforms is their consideration for the time-ordering of data.

REFERENCES

[ABSS89] M. Appel, R. Berger, J. Saul, J. Smith, and R. Cohen, "Beat to beat variability in cardiovascular variables: Noise or music?, *Journal of the American College of Cardiology,* vol. 14, no. 5, 1989, pp. 1139 – 1148.

[ACC93] American College of Cardiology, "Heart rate variability for risk stratification of life-threatening arrhythmias," *Journal of the American College of Cardiology,* vol. 22, no. 3, pp. 948-950, 1993.

[ACM05] ACM. Computing Curricula 2005 -- The Overview Report. [Online]. http://www.acm.org/education/curric_vols/CC2005-March06Final.pdf (as of July 4, 2007), 2005.

[AHA07] American Heart Association Statistics Committee and Stroke Statistics Subcommittee, "Heart Disease and Stroke Statistics — 2007 Update," *Circulation.* vol. 115, pp. e69-e171, 2007.

[Atki07] A. Atkielski, "Schematic diagram of normal sinus rhythm for a human heart as seen on ECG," [Online]. http://en.wikipedia.org/wiki/Image:SinusRhythmLabels.svg (as of July 4, 2007).

[ATR05] Aging and Technology Research. *Congestive Heart Failure.* [Online]. http://agingresearch.buffalo.edu/chf/general.htm (as of July 6, 2007), 2005.

[BAGC86] R. Berger, S. Akselrod, D. Gordon, and R. Cohen, "An efficient algorithm for spectral analysis of heart rate variability," *IEEE Transactions on Biomedical Engineering,* vol. 33, no. 9, pp. 900 – 904, 1986.

[BaLW994] J. B. Bassingthwaighte, L. Liebovitch and B. J. West. *Fractal Physiology.* Oxford, UK: Oxford University Press, 1994.

[BBGH90] J. J. Bailey, A. S. Berson, A. Garson, L. G. Horan, P. W. Macfarlane, D. W. Mortara, C. Zywietz, "Recommendations for standardization and specifications in automated electrocardiography: bandwidth and digital signal processing: A report for health professionals by an ad hoc writing group of the Committee on Electrocardiography and Cardiac Electrophysiology of the Council on Clinical Cardiology, American Heart Association," *Circulation*, vol. 81, no. 2, pp. 730-739, 1990.

[BePi67] A. Berson and H. Pipberge, "Electrocardiographic distortions caused by inadequate high-frequency response of directwriting electrocardiographs," *American Heart Journal,* vol. 74, pp. 208-218, 1967.

[BFBD77] A. Berson, T. Ferguson, C. Batchlor, R. Dunn, and H. Pipberger, "Filtering and sampling for electrocardiographic data processing," *Computers and Biomedical Research,* vol. 10, pp. 605-610, 1977.

[BFSR95] J. Bigger, L. Fleiss, R. Steinman, L. Rolnitzky, W. Schneider, and PO. Stein, "RR variability in healthy, middle-age persons compared with patients with chronic coronary heart disease or recent acute myocardial infarction," *Circulation,* vol. 91, pp. 1936-1943, 1995.

[BiDu90] G. E. Billman and J. P. Dujardin, "Dynamic changes in cardiac vagal tone as measured by time-series analysis," *The American Journal of Physiology,* vol. 258, no. 3, pt. 2, pp. H896 - H902, 1990.

[BKSR89] L. Bernardi, F. Keller, M. Sanders, P. Reddy, B. Griffith, F. Meno, and M. Pinsky, "Respiratory sinus arrhythmia in the denervated human heart," *Journal of Applied Physiology,* vol. 67, no. 4, pp. 1447–1455, 1989.

[BLWP77] A. Berson, F. Lau, J. Wojick, and H. Pipberger, "Distortions in infant electrocardiograms caused by inadequate high frequency response," *American Heart Journal,* vol. 93, pp. 730-734, 1977.

[Braun97] E. Braunwald, D. Zipes, and P Libby, Eds. *Heart Disease: A Textbook of Cardiovascular Medicine,* 5th Ed., Philadelphia, PA: W.B. Saunders Co., 1997.

[Cann32] W. Cannon. *The Wisdom of the Body.* New York, NY: W. W. Norton & Co., 1932.

[CaSG78] W. Casscells, A. Schoenberger, T. B. Graboys, "Interpretation by physicians of clinical laboratory results," *New England Journal of Medicine,* vol. 2, no. 18, pp. 999 – 1001, 1978.

[CBDJ84] J. Conway, N. Boon, C. Davies, J. V. Jones, and P. Sleight, "Neural and humoral mechanisms involved in blood pressure variability," *Journal of Hypertension,* vol. 2, no. 2, pp. 203 – 208, 1984.

[CGPK99] J. Croft, W. Giles, R. Pollard, N. Keenan, M. Casper, and R. Anda, "Heart failure survival among older adults in the United States: A poor prognosis for an emerging epidemic in the Medicare population," *Archives of Internal Medicine,* vol. 159, pp. 505 – 510, 1999.

[ChKa01] H-W. Chiu and T Kao, "A mathematical model of autonomic control of heart rate variation," *IEEE Engineering in Medicine and Biology Magazine*, pp. 69 – 76, March/April 2001.

[Clar97] J. W. Clark, Jr., "The origin of biopotentials," in J. G. Webster, ed. *Medical Instrumentation: Application and Design.* 3rd Ed. New York, NY: John Wily and Sons, 1997.

[Cohe94] J. Cohen, "The earth is round ($p < 0.05$)," *American Psychologist,* vol. 49, no. 12, pp. 997 – 1003, 1994.

[CoYW86] P. Corr, K. Yamada, and F. Wikowski, "Mechanisms controlling cardiac autonomic function and their relation to arrhythmogenesis," in H. Fozzard, E. Haber, R. Jennings, and A. Katz, Eds. *The Heart and Cardiovascular System.* New York, NY: Raven Press, 1986.

[Dake05] Dake, *Neuromuscular junction.* [Online]. http://en.wikipedia.org/wiki/Image:Synapse_diag3.png (as of July 6, 2007), August 2005.

[DDFH05] L. Dueck, S. Dueck, P. Furgale, M. Hudek, N. Juzkiw, J. Majewski, S. McAnn, S. Schärer, and G. StewartU, "The Team Manitoba Unmanned Aerial Vehicle: Design Goals and Performance," *2005 Student AUVSI UAV Competition,* [Online]. http://uav.navair.navy.mil/seafarers/papers/Manitoba.pdf (as of August 4, 2007), 2005.

[DGKK03] S. Dueck, N. Gadhok, M. Kinsner, J. Kraut, T. Tessier, and W. Kinsner, "YES2 DHS: A Space Tether Control Subsystem," *IEEE Canadian Conference on Electrical and Computer Engineering*, vol. 2, pp. 1287 – 1292, 2003.

[Dono95] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613 – 627, 1995.

[Duec04] S. Dueck, "Evidence of Chaos in Cardiology," *Unpublished.* 9 pp., 2004.

[DuHa73] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis.* New York, NY: John Wiley and Sons, 1973.

[DyFi03] G. Dyer and M. Fifer, "Heart failure," in L. S. Lily, Ed. *Pathophysiology of Heart Disease.* Philadelphia, PA: Lippincott Williams & Wilkins, 2003.

[Eckb80] D. L. Eckberg, "Parasympathetic cardiovascular control in human disease: a critical review of methods and results," *The American Journal of Physiology,* vol. 239, no. 5, pp. H581 – H593, 1980.

[EwNT84] D. J. Ewing, J. M. Neilson, and P. Travis, "New method for assessing cardiac parasympathetic activity using 24 hour electrocardiograms," *British Heart Journal*, no. 42, pp. 396-402, 1984.

[FGCD90] R. Furlan, S. Guzzetti, W. Crivellaro, S. Dassi, M. Tinelli, G. Baselli, S. Cerutti, F. Lombardi, M. Pagani, and A. Malliani, "Continuous 24-hour assessment of the neural regulation of systemic arterial pressure and RR variabilities in ambulant subjects," *Circulation,* vol. 81, no. 2, pp. 537 – 547, 1990.

[GBBK97] R. L. Goldsmith, J. T. Bigger, D. Bloomfield, H. Krum, R. Steinman, J. Sackner-Bernstein, and M. Packer, "Long-term carvedilol therapy increases parasympathetic nervous system activity in chronic congestive heart failure," *American Journal of Cardiology,* vol. 80, pp. 1101-1104, 1997.

[GBSF92] R. Goldsmith, J. Bigger, R. Steinman, and J. Fleiss, "Comparison of 24-hour parasympathetic activity in endurance-trained and untrained young men," *Journal of the American College of Cardiology,* vol. 20, pp. 552-558, 1992.

[GGCS87] L. Glass, A. L. Goldberger, M. Courtemanche, and A. Shrier, "Nonlinear dynamics, chaos, and complex cardiac arrhythmias," *Proceedings of the Royal Society of London*, Series A, vol. 413, pp. 9-26, 1987.

[Glas99] L. Glass, "Chaos and heart rate variability," *Journal of Cardiovascular Electrophysiology,* vol. 10, pp. 1358 – 1360, 1999.

[Gold92] A. L. Goldberger, "Applications of chaos to physiology and medicine," in *Applied Chaos.* J. H. Kim and J. Stringer, Eds. New York, NY: John Wiley & Sons, 1992.

[GoWH73] D. P. Golden Jr, R. A. Wolthius, G. W. Hoffler, "A spectral analysis of the normal resting electrocardiogram," *IEEE Transactions on Biomedical Engineering*, vol. 13, pp. 366-372, 1973.

[GrBW90] P. Grossman, J. Van Beek, and C. Wientjes, "A comparison of three quantification methods for estimation of respiratory sinus arrhythmia," *Psychophysiology,* vol. 27, no. 6, pp. 702-714, 1990.

[HCGR99] G. Haldeman, J. Croft, W. Giles, A. Rashidee, "Hospitalization of patients with heart failure: National Hospital Discharge Survey, 1985 to 1995," *American Heart Journal,* vol. 137, pp. 352-60, 1999.

[Hior94] J. Hjorth. *Computer Intensive Statistical Methods Validation, Model Selection, and Bootstrap*. London, UK: Chapman & Hall, 1994.

[HLOM99] M. Herrera, J. Llanos, J. Olivera, R. Martinez, and M. Valentinuzzi, "Processing pressure-volume diagrams (DPV) using virtual instrumentation," P*roceedings of the First Joint BMES/IEEE EMBS Conference*, vol. 1, pp. 228, 1999.

[HMPM97] K. Ho, G. Moody; C.-K. Peng, J. Mietus, M. Larson, D. Levy, and A. Goldberger, "Predicting Survival in Heart Failure Case and Control Subjects by Use of Fully Automated Methods for Deriving Nonlinear and Conventional Indices of Heart Rate Dynamics," *Circulation,* vol. 96, pp. 842 – 848, 1997.

[HoHu52] A. Hodgkin and A. Huxley AF, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *Journal of Physiology*, vol. 117, no. 4, pp. 500-544, 1952.

[HSSL03] J. Ho, W. Stenson, G. Strichartz, and L. S. Lily, "Mechanisms of cardiac arrhythmias," in *Pathophysiology of Heart Disease*. L. S. Lily, Ed. Philadelphia, PA: Lippincott Williams & Wilkins, 2003.

[IAGH99] P. C. Ivanov, L. A. N. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, Z. R. Struzik, and H. E. Stanley. "Multifractality in human heart rate dynamics," *Nature,* vol. 399, pp. 461–465, 1999.

[IEEE03] IEEE. *Instrumentation and Measurement Society homepage.* [Online]. http://ewh.ieee.org/soc/im/imnew/ (as of July 4, 2007), 2003.

[KaBe91] W. B. Kannel and A. J. Belanger, "Epidemiology of heart failure, " *American Heart Journal,* vol. 3, no. 1, pp 951 – 957, 1991.

[KaHA94] J. K. Kanters, N. H. Holstein-Rathlou, and E. Agner, "Lack of evidence for low-dimensional chaos in heart rate variability," *Journal of Cardiovascular Electrophysiology*, vol. 5, no. 7, 591-601, 1994.

[KaSc98] H. Kantz and T. Schreiber, "Human ECG: nonlinear deterministic versus stochastic aspects," *IEE Proceedings on Science, Measurement and Technology*, vol. 145, no. 6, pp. 279 – 284, 1998.

[KBGP95] H. Krum, J. T. Bigger, R. L. Goldsmith, and M. Packer, "Effect of long-term digoxin therapy on autonomic function in patients with chronic heart failure," *Journal of the American College of Cardiology*, vol. 25, pp. 289-294, 1995.

[KlSB05] R. E. Kleiger, P. K. Stein, J. T. Bigger, "Heart rate variability: measurement and clinical utility," *Annals of Noninvasive Electrocardiology*, vol. 10, no. 1, , pp. 88 – 101, 2005.

[Knu05] Knutux, "ECG lead 1," [Online]. http://en.wikipedia.org/wiki/Image:EKGI.png (as of July 6, 2007) May 2005.

[KoHO02] B.-U. Kohler, C. Hennig, and R. Orglmeister, "The principles of software QRS detection," *IEEE Engineering in Medicine and Biology Magazine*, vol. 21, no. 1, pp. 42 – 57, 2002.

[LeMo07] Le Moyne College, "Nervous system," [Online]. http://web.lemoyne.edu/~hevern/psy340/lectures/psy340.04.2.ns.structure.html (as of July 6, 2007).

[Lily03] L. S. Lily, "The electrocardiogram," in L. S. Lily, ed. *Pathophysiology of Heart Disease*. Philadelphia, PA: Lippincott Williams & Wilkins, 2003.

[LJGH06] D. S. Lee, H. Johansen, Y. Gong, R. Hall, J. Tu, and J. Cox, "Regional outcomes of heart failure in Canada," in *Canadian Cardiovascular Atlas: A Collection of Original Research Papers Published in The Canadian Journal of Cardiology*. J. V. Tu, W. A. Ghali, L. Pilote, and S. Brien, Eds.. Pulsus Group, Inc. and the Institute for Clinical Evaluative Sciences, Toronto: Canada, 2006.

[LoVe76] B. Lown and R. L. Verrier, "Neural activity and ventricular fibrillation," *New England Journal of Medicine*, vol. 294, pp. 1165-70, 1976.

[MaEL03] R. Malhotra, E. Edelman, and L. S. Lily, "Basic Cardiac Structure and Function," in *Pathophysiology of Heart Disease*, L. S. Lily, Ed. Philadelphia, PA: Lippincott Williams & Wilkins, 2003.

[MaES83] N. M. Magid, D. L. Eckberg, and J. M. Sprenkle, "Low dose atropine reduces ventricular vulnerability in normal and ischemic hearts," *Journal of Clinical Investigation,* vol. 31, pp. 261-264, 1983.

[MaGi87] J. Marino and B. Gilbert, "Improving AM sound: The NRSC standard," *IEEE Transactions on Consumer Electronics*, vol. CE-33, no. 3, pp. x – xvii, 1987.

[MPHG02] J. E. Mietus, C-K. Peng, I. Henry, R. L. Goldsmith, and A. L. Goldberger, "The pNNx files: re-examining a widely used heart rate variability measure," *Heart*, vol. 88, pp. 378-380, 2002.

[Neum98] M. R. Neuman, "Biopotential electrodes," in J. G. Webster, ed. Medical Instrumentation: Application and Design. 3rd Ed. New York, NY: John Wiley and Sons, 1998.

[NKBH41] J. von Neumann, R. H. Kent, H. R. Bellinson, and B. I. Hart, "The mean square successive difference," *Annals of Mathematical Statistics*, vol. 12, no. 2, pp. 153-162, 1941.

[OCBr93] J. B. O'Connell and M. R. Bristow, "Economic impact of heart failure in the United States: time for a different approach," *The Journal of Heart and Lung Transplant,* vol. 13, no. 4, pp. S107-S112, 1993.

[Parz62] E. Parzen, "On estimation of a probability density function and mode," *Annals of Mathematical Statistics*, vol. 33, pp. 1065-1076, 1962.

[PBGH72] T. Powell, L. E. Baker, L. A. Geddes, and H. E. Hoff, "Observations on cardiac energetics: 1. Theoretical introduction," *Bulletin of Mathematical Biology,* vol. 34, no. 4, pp. 533-546, 1972.

[PFTV92] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge, UK: Cambridge University, 1992.

[Pier06] E. Pierce, "Diagram of the human heart," [Online]. http://commons.wikimedia.org/wiki/Image:Diagram_of_the_human_heart_%28cropped%29.svg (as of July 6, 2007), June 2006.

[Pott06] M. Potter. *Feature Convergence of ECG Signals Processed by Independent Component Analysis*. Ph.D. Candidacy Dissertation, University of Manitoba; Winnipeg, Manitoba, January 2006.

[RKDL06] B. Rapatski, P. Klepac, S. Dueck, M. Liu, and L. Weiss, "Mathematical Epidemiology of HIV/AIDS in Cuba During the Period 1986-2000", *Mathematical Biosciences And Engineering*, vol. 3, no. 3, pp. 545 – 556, 2006.

[Rose93] R. Rosenthal, "Cumulating evidence," in G. Keren and C. Lewis (ed.), *A Handbook for Data Analysis in the Behavioral Science,* Hillsdale, NJ: Erlbaum, pp. 519 – 559, 1993.

[RoZw81] E. Robertson and M. Zweig, "Use of receiver operating characteristic curves to evaluate the clinical performance of analytical systems," *Clinical Chemistry*, vol. 27, 1569-1574, 1981.

[SaBe95] J. P. Saul, L. Bernardi, "Heart Rate Variability after Cardiac Transplantation" in *Heart Rate Variability,* M. Malik M and A. J. Camm, Eds. Armonk, NY: Futura, 1995.

[SALS89] K. Sands, M. Appel, L. S. Lilly, F. Schoen, G. Mudge, R. Cohen, "Power spectrum analysis of heart rate variability in human cardiac transplant recipients," *Circulation,* vol. 79, no. 1, pp. 76 – 82, 1989.

[ScVS00] G. Schwarzer, W. Vach, and M. Schumacher, "On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology," *Statistics in Medicine,* vol. 19, pp. 541-561, 2000.

[SEDK99] P. Stein, A. Ehsani, P. Domitrovich, R. Kleiger, and J. Rottman, "The effect of exercise training on heart rate variability in healthy older adults," *American Heart Journal,* vol. 138, pp. 567-576, 1999.

[Selk07] Selket, "Diagram of neuron with arrows but no labels," [Online]. http://en.wikipedia.org/wiki/Image:Neuron-no_labels.png (as of July 6, 2007), February 2007.

[SMMW88] D. Singer, G. Martin, N. Magid, J. Weiss, J. Schaad, R. Kehoe, T. Zheutlin, D. Fintel, A. Hsieh, M. Lesch, "Low heart rate variability and sudden cardiac death," *Journal of Electrocardiology,* vol. 21, pp. S46-S55, 1988.

[StJu07] St. Jude Medical, Inc., "Valve disease," [Online]. http://www.onevalveforlife.com/images/cndctn.gif (as of July 6, 2007).

[StSm01] J. Sterne and G. Smith, "Sifting the evidence – what's wrong with significance tests?," *British Medical Journal,* vol. 322, pp. 226 – 231, 2001.

[Swet92] J. Swets, "The science of choosing the right decision threshold in high-stakes diagnostics," *American Psychologist,* vol. 47, no. 4, pp. 522 – 532, 1992.

[Task96] Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *European Heart Journal*, vol. 17, pp. 354–381, 1996.

[Tier80] J. Tierney, "A study of LPC analysis of speech in additive noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 28, no. 4, pp. 389 – 397, 1980.

[TuZh99] J. Tu and H. Zhang, "Congestive heart failure outcomes in Ontario," in *Cardiovascular Health and Services in Ontario: An ICES Atlas*, C. Naylor and P. Slaughter, Eds. Toronto, Canada: Institute for Clinical Evaluative Sciences, 1999.

[TvKa71] A. Tversky and D. Kahneman, "Belief in the law of small numbers," *Psychological Bulletin,* vol. 76, no. 2, pp. 105 – 110, 1971.

[Wiki07] Wikipedia, "Electrocardiogram," [Online]. http://en.wikipedia.org/wiki/Ecg (as of July 6, 2007).

[ZWBS95] I. Zucker, W. Wang, M. Brändle, H. Schultz, and K. Patel, "Neural regulation of sympathetic nerve activity in heart failure," *Progress in Cardiovascular Diseases,* vol. 37, no. 6, pp. 397 – 414, 1995.

[ZwCa93] M. Zweig and G. Campbell G, "Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine," *Clinical Chemistry,* vol. 39, no. 4, pp. 561 – 577, 1993.

Appendix A

SAMPLE DEPOLARIZATIONS IN FRONTAL AND TRANSVERSE PLANES

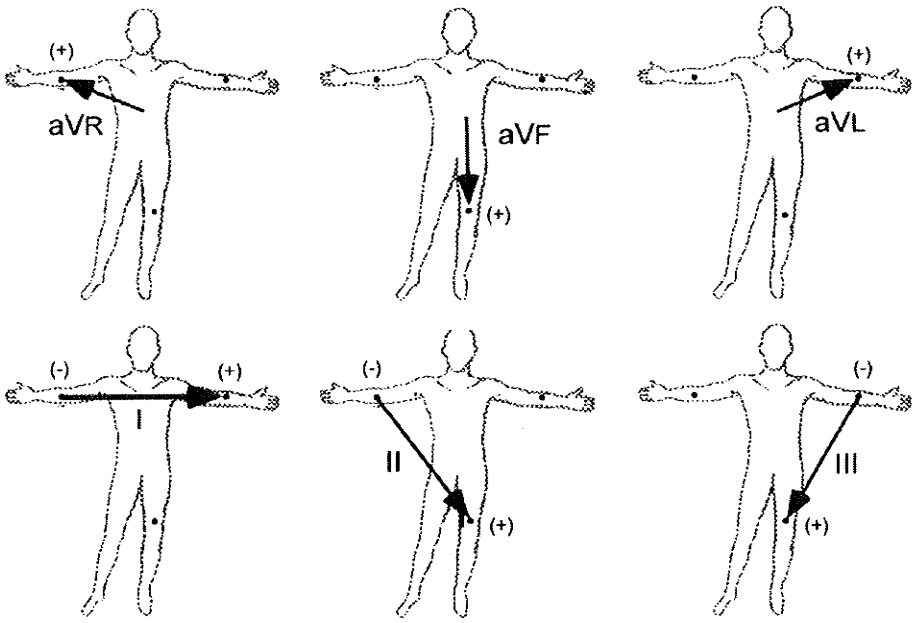


Fig. A.1 Direction of conduction vector measurements for the six frontal leads. (From [Lily03]).

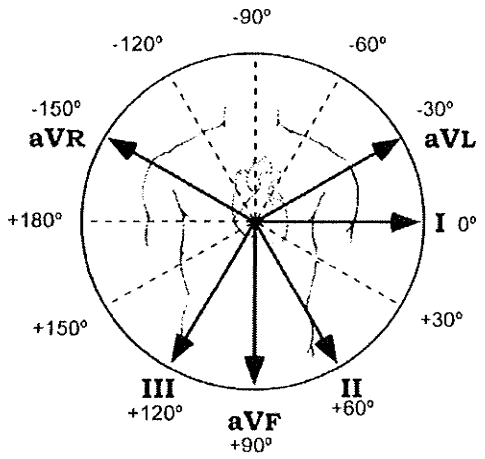


Fig. A.2 Radial coordinate system accomplished within the heart as a result of the placement of Fig. A.1. (From [Lily03]).

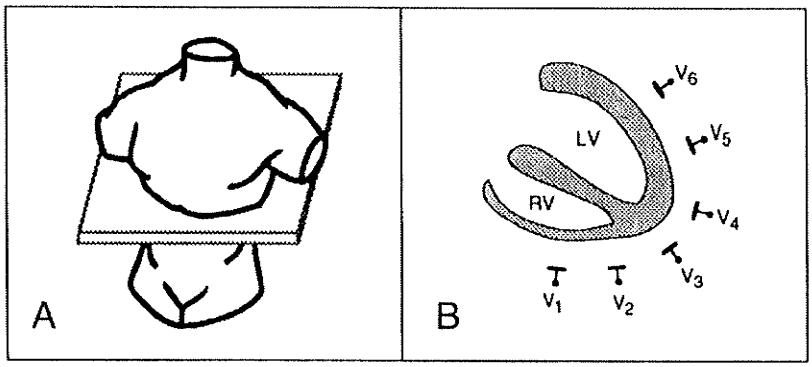


Fig. A.3 Transversal ECG acquisition showing (a) plane of acquisition, and (b) lead placement. (From [Lily03]).

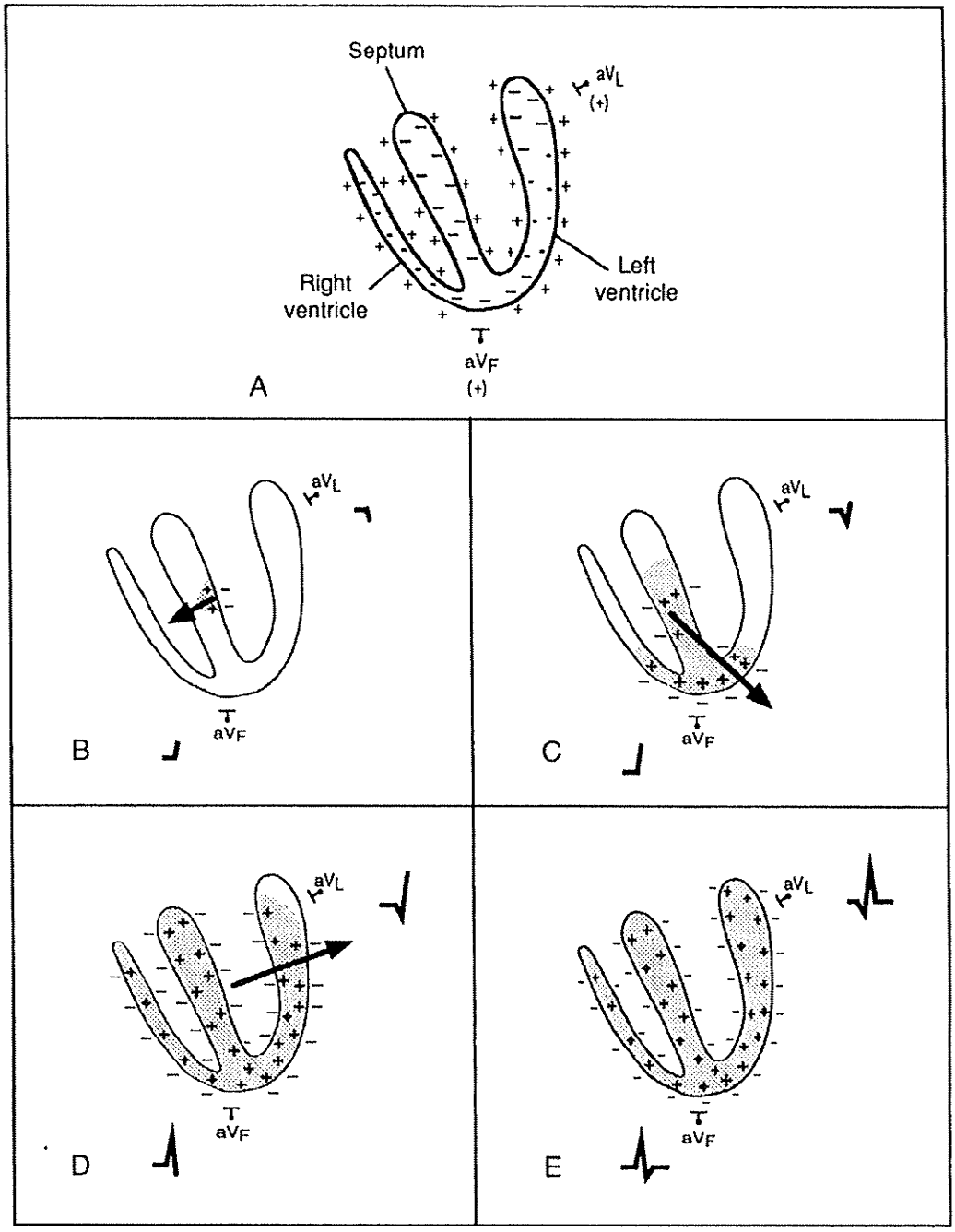


Fig. A.4 Normal ventricular depolarization as recorded as recorded in the frontal plane showing progression from (a) steady state to (d) complete depolarization. (From [Lily03]).

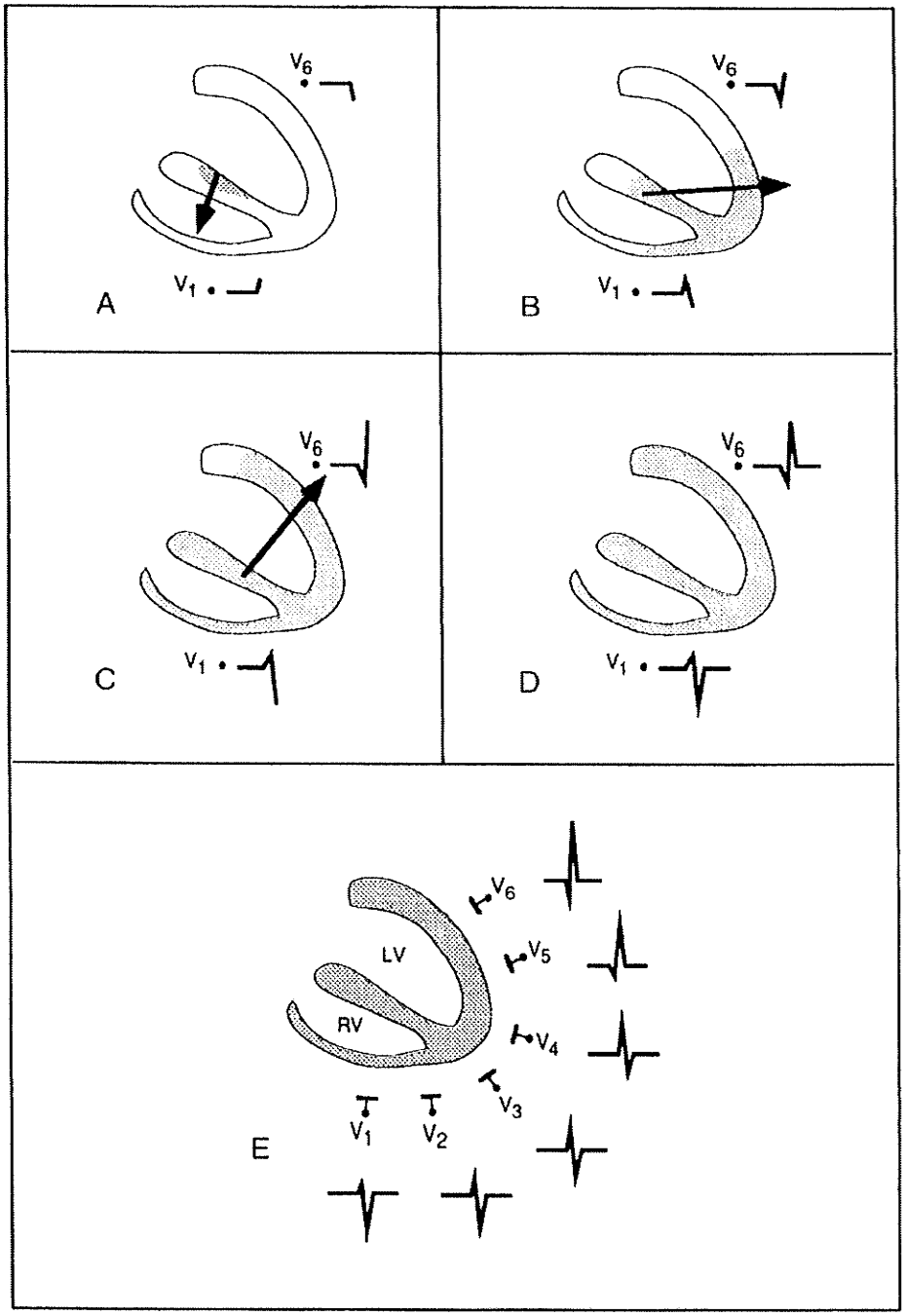


Fig. A.5 Normal ventricular depolarization as recorded as recorded in the transverse plane showing progression from (a) steady state to (d) complete depolarization. (From [Lily03]).

Appendix B

UTILITY OF RECEIVER OPERATING CHARACTERISTICS IN TEST PERFORMANCE EVALUATION

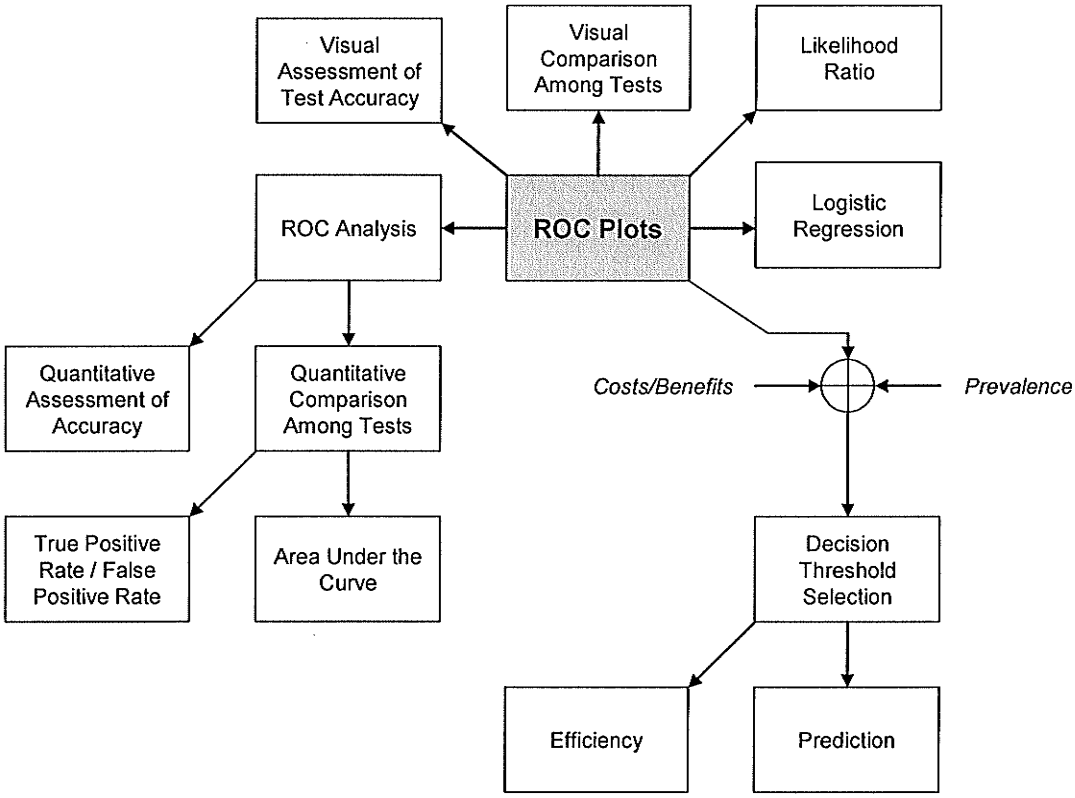


Fig. B.1 Diagram showing utility of ROC plots. (After [ZwCa93]).

Appendix C

SAMPLE DATA CHARACTERISTICS

Table C.1 Mean age with gender of normal and CHF samples.

	Number of Normal Subjects (mean age ± SD)	CHF Patients (mean age ± SD)
Men	35 (55.2 ± 16.0 years)	8 (57 ± 10.6 years)
Women	37 (54.0 ± 16.3 years)	2 (48.5 ± 14.8 years)
Unknown	NA	19 (54.3 ± 12.1 years)
Overall	54.6 (61.4 ± 16.0 years)	29 (55.3 ± 11.6 years)

Table C.2 Age distribution of normal sinus rhythm sample.

Record Name	Male	Female
nsr001		64
nsr002	67	
nsr003		67
nsr004		62
nsr005		62
nsr006	64	
nsr007	76	
nsr008		64
nsr009	66	
nsr010		61
nsr011		65
nsr012	66	
nsr013		63
nsr014		65
nsr015	74	
nsr016		73
nsr017		71
nsr018	68	
nsr019		65
nsr020		58
nsr021	59	

nsr022	68	
nsr023		66
nsr024		63
nsr025	75	
nsr026	72	
nsr027	64	
nsr028	65	
nsr029	63	
nsr030		70
nsr031	67	
nsr032	68	
nsr033	65	
nsr034	67	
nsr035	66	
nsr036		60
nsr037	63	
nsr038	62	
nsr039		70
nsr040		63
nsr041		64
nsr042		68
nsr043	66	
nsr044		65
nsr045		67
nsr046		63
nsr047	28.5	
nsr048	38	
nsr049	39	
nsr050	29	
nsr051	40	
nsr052	39	
nsr053	35	
nsr054	35	
16265	32	
16272		20
16273		28
16420		38
16483	42	
16539		35
16773	26	
16786		32
16795		20
17052		45
17453		32
18177		26

18184		34
19088		41
19090	45	
19093	34	
19140		38
19830		50

Table C.3 Age distribution of CHF sample.

Record Name	Male	NYHAC	Female	NYHAC	Unknown	NYHAC
chf201	55	3				
chf202			59	3		
chf203	68	3				
chf204	62	3				
chf205	39	3				
chf206			38	3		
chf207	62	3				
chf208	62	3				
chf209	65	3				
chf210	43	3				
chf211					34	2
chf212					54	2
chf213					53	1
chf214					79	2
chf215					43	2
chf216					58	2
chf217					50	1
chf218					72	1
chf219					62	3
chf220					64	2
chf221					37	1
chf222					63	3
chf223					56	3
chf224					35	2
chf225					66	3
chf226					51	2
chf227					64	3
chf228					51	3
chf229					58	3

Appendix D

SOURCE CODE

D.1 *t*-Test, *A priori, Posteriori* Distributions

```
% Author: Stephen Dueck
% Purpose:
%   1. Perform t-test
%   2. create a priori and posteriori
%      probabilities for range of pNNx parameter x.

close all;
newdata=0; % 1 load new data, 0 skip loading process

if newdata==1
    clear all;
    N=72; % 72 healthy subjects
    threshvec=1:13;

    normname(1,1:13)='normal\nsr001';
    normname(2,1:13)='normal\nsr002';
    normname(3,1:13)='normal\nsr003';
    normname(4,1:13)='normal\nsr004';
    normname(5,1:13)='normal\nsr005';
    normname(6,1:13)='normal\nsr006';
    normname(7,1:13)='normal\nsr007';
    normname(8,1:13)='normal\nsr008';
    normname(9,1:13)='normal\nsr009';
    for i=10:54,
        normname(i,1:13)=['normal\nsr0' num2str(i)];
    end;
    normname(55,1:13)='normal2\16265';
    normname(56,1:13)='normal2\16272';
    normname(57,1:13)='normal2\16273';
    normname(58,1:13)='normal2\16420';
    normname(59,1:13)='normal2\16483';
    normname(60,1:13)='normal2\16539';
    normname(61,1:13)='normal2\16773';
    normname(62,1:13)='normal2\16786';
    normname(63,1:13)='normal2\16795';
    normname(64,1:13)='normal2\17052';
    normname(65,1:13)='normal2\17453';
    normname(66,1:13)='normal2\18177';
    normname(67,1:13)='normal2\18184';
    normname(68,1:13)='normal2\19088';
    normname(69,1:13)='normal2\19090';
    normname(70,1:13)='normal2\19093';
    normname(71,1:13)='normal2\19140';
    normname(72,1:13)='normal2\19830';

    for i=1:N,
        fname=normname(i,1:13);
        disp(fname)
        heasig=readheader([fname '.hea']);
        if i<=54
            t=[1 1e9];
            annot=readannot([fname '.ecg'],t);
```

```
        else
        t=[1 1e9]; %avg
        annot=readannot([fname '.atr'],t);
    end;
    samples=annot.time;
    normDuration(i)=samples(end)-samples(1);
    samples=samples(find(annot.anntyp~='~')); %exclude signal quality change annotations

    RR = diff(samples);
    Ncount(i)=length(RR);

    dRRdn = diff(RR);
    j=1;
    for thresh=threshvec,
        NN(i,j) = sum( abs(dRRdn)>=thresh );
        j=j+1;
    end;
end;
normNN=NN;
normpNN=zeros(size(NN));
for i=1:N,
    normpNN(i,:)=NN(i,:)/Ncount(i);
end;
normSEM=std(normpNN)/sqrt(N);
normCount=Ncount;

%=====================================================
Ncount=[];
NN=[];
N=29;
chfname(1,1:10)='chf\chf201';
chfname(2,1:10)='chf\chf202';
chfname(3,1:10)='chf\chf203';
chfname(4,1:10)='chf\chf204';
chfname(5,1:10)='chf\chf205';
chfname(6,1:10)='chf\chf206';
chfname(7,1:10)='chf\chf207';
chfname(8,1:10)='chf\chf208';
chfname(9,1:10)='chf\chf209';
for i=10:29,
    chfname(i,1:10)=['chf\chf2' num2str(i)];
end;
chfname(30,1:10)='chf2\chf01';
chfname(31,1:10)='chf2\chf02';
chfname(32,1:10)='chf2\chf03';
chfname(33,1:10)='chf2\chf04';
chfname(34,1:10)='chf2\chf05';
chfname(35,1:10)='chf2\chf06';
chfname(36,1:10)='chf2\chf07';
chfname(37,1:10)='chf2\chf08';
chfname(38,1:10)='chf2\chf09';
chfname(39,1:10)='chf2\chf10';
chfname(40,1:10)='chf2\chf11';
chfname(41,1:10)='chf2\chf12';
chfname(42,1:10)='chf2\chf13';
chfname(43,1:10)='chf2\chf14';
chfname(44,1:10)='chf2\chf15';

for i=1:N,
    fname=chfname(i,1:10);
    disp(fname);
    heasig=readheader([fname '.hea']);
    if i<=29
        t=[1 1e9];
        annot=readannot([fname '.ecg'],t);
        if length(annot.time)==1e9
            disp('err (pnn18): fix up didn''t work');
            pause;
        end;
    else
        t=[1 1e9]; %avg
```

```
        annot=readannot([fname '.ecg'],t);
    end;
    samples=annot.time;
    chfDuration(i)=samples(end)-samples(1);
    samples=samples(find(annot.anntyp~='~')); %exclude signal quality change annotations

    RR = diff(samples); %msec
    Ncount(i)=length(RR);

    dRRdn = diff(RR);
    j=1;
    for thresh=threshvec,
        NN(i,j) = sum( abs(dRRdn)>=thresh );
        j=j+1;
    end;
end;
chfNN=NN;
chfpNN=zeros(size(NN));
for i=1:N,
    chfpNN(i,:)=NN(i,:)/Ncount(i);
end;
chfSEM=std(chfpNN)/sqrt(N);
chfCount=Ncount;

end;

%=============================================================
% CALCULATE P VALUES
%=============================================================
x1=normpNN;
x2=chfpNN;
u1=mean(x1);
u2=mean(x2);
N1=72;
N2=44;
for i=1:length(u1),
    sp(i)=sqrt( ( sum( (x1(:,i)-u1(i)).^2 ) + sum( (x2(:,i)-u2(i)).^2 ) )/(N1+N2-2) );
%pooled variance
end;
p=[];
for i=1:length(u1),
    sep(i)=sp(i)*sqrt(1/N1+1/N2); %standard error between means
    t(i)=(u1(i)-u2(i))/sep(i); %larger the value, the more convincing the difference
                    % this value measures the differences between our mean
                    % mean values in terms of standard errors
     p(i)=stud_dist(N1,N2,t(i));
 end;

%=============================================================
% PLOT
%=============================================================
figure; hold on;
plot(threshvec,mean(normpNN)+2*normSEM,'b:',threshvec,mean(chfpNN)+2*chfSEM,'r--
',threshvec,p,'k','linewidth',2);
legend('normal pNNx','chf pNNx','confidence in similarity (p value)');
plot(threshvec,mean(normpNN)-2*normSEM,'b:',threshvec,mean(chfpNN)-2*chfSEM,'r--
','linewidth',2);
%plot(threshvec,mean(normpNN),'k','linewidth',2);
%plot(threshvec,mean(chfpNN),'k','linewidth',2);

xlabel('threshold [samples]'); grid on;
%title('mean pNNx (line) with 2SEM (dot)');

%=============================================================
% RISK ANALYSIS
%=============================================================
figure;
V=.5;
plotstep=1; ts=.01;
p_chf = .02;
p_norm = .98;
```

```
p_chf = .02;
p_norm = 1-p_chf;

for j=1:5,

    [xchf, pchf_xh]=gausest2(chfpNN(:,j),V,ts,0);
    [xnorm, pnorm_xh]=gausest2(normpNN(:,j),V,ts,0);
    [xtotal, ptotal]=gausest2([chfpNN(:,j); normpNN(:,j)], V, ts, 0);

    pchf=pchf_xh;
    pnorm=pnorm_xh;

    ptotal=pchf_xh*p_chf + pnorm_xh*p_norm;

    I=find(xchf>0 & xchf<1); pchf=pchf(I); xchf=xchf(I);
    I=find(xnorm>0 & xnorm<1); pnorm=pnorm(I); xnorm=xnorm(I);
    I=find(xtotal>0 & xtotal<1); ptotal=ptotal(I); xtotal=xtotal(I);

    pchf_hx=pchf./ptotal*p_chf; xchf_hx = xchf;
    pnorm_hx=pnorm./ptotal*p_norm; xnorm_hx=xnorm;

    xchf=round(xchf/ts)*ts; xnorm=round(xnorm/ts)*ts; xtotal=round(xtotal/ts)*ts;

    if length(pnorm)~=length(pchf)
        disp('err: vectors not the same length');
        pause;
    end;

    %============================================================
    % DETECT THRESHOLD
    %============================================================
    Ichf=find(pchf_hx>pnorm_hx); Inorm=find(pnorm_hx>pchf_hx);
    chfright=max(Ichf); normleft=min(Inorm);
    if ~isempty(chfright)
        Ipthresh=min(chfright, normleft);  % choose max to take most
                                           % pessimistic estimate of
                                           % risk for undetected CHF
    else
        Ipthresh=normleft;
    end;

    pchf=pchf/sum(pchf);
    pnorm=pnorm/sum(pnorm);
    ptotal=pchf_xh*p_chf + pnorm_xh*p_norm;

    pthreshvec(j)=xnorm(Ipthresh);
    p_norm_dchf(j)=p_chf*sum(pchf(Ipthresh:end));
    p_chf_dnorm(j)=p_norm*sum(pnorm(1:Ipthresh));

    plotthresh=mean([xnorm(chfright) xnorm(normleft)]);

    %============================================================
    % PLOT
    %============================================================
    if plotstep==1
        figure;
        plot(xnorm,pnorm,'b.',xchf,pchf,'ro'); %,xtotal); ,ptotal,'k^'); hold on;
        hold on;
        legend('p(x|C_N)','p(x|C_C)'); %,'p(pNNx)');
        plot(xnorm,pnorm,'b',xchf,pchf,'r'); %,xtotal,ptotal,'k');
        xlabel('x'); ylabel('p(x|C) estimate');
        title(['threshold=' num2str(threshvec(j)) ' samples']);

        figure;
        pxc_norm=[]; pxc_chf=[];
        Tn_vec=1:length(xnorm);
        for Tn=Tn_vec,
            pxc_norm(Tn) = sum(pnorm(Tn:end));
            pxc_chf(Tn) = sum(pchf(1:Tn));
        end;
```

```
        plot(xnorm,pxc_norm,'b.',xnorm,pxc_chf,'ro'); hold on;
        plot(xnorm,pxc_norm,'b',xnorm,pxc_chf,'r');
        legend('p(x|C_N)','p(x|C_C)'); %,'p(pNNx)');
        xlabel('x'); ylabel('area under p(x|C)');
        axis tight;

        % Plot p(x|Ck)
        figure;

        plot(xnorm_hx,pnorm_hx,'b.',xchf_hx,pchf_hx,'ro'); hold on;
        legend('p(normal|pNNx)','p(CHF|pNNx)');

        plot(xnorm_hx,pnorm_hx,'b',xchf_hx,pchf_hx,'r');
        xlabel('pNNx'); ylabel('p(Ck|pNNx) estimate');
        title{['threshold=' num2str(threshvec(j)) ' samples']};

        %Plot Bayesian threshold
        if Ipthresh==chfright
            ytop=pchf_hx(Ipthresh);
        else
            ytop=pnorm_hx(Ipthresh);
        end;

        figure;
        plot(xnorm,ptotal(1:99),'k'); hold on;
        plot(xnorm,ptotal(1:99),'kx'); hold on;
        xlabel('pNNX'); ylabel('p(x) estimate');

        figure;
        [ax,h1,h2]=plotyy(xchf,p_chf*pchf,xnorm,p_norm*pnorm); hold on;
        set(h1,'Color',[1 0 0]);
        set(h2,'Color',[0 0 1]);
        set(ax(1),'YColor',[1 0 0]);
        set(ax(1),'YTick',[0:1e-3/10,1e-3]);
        set(ax(1),'YLimMode','auto');
        set(ax(2),'YColor',[0 0 1]);
        set(ax(2),'YLimMode','auto');
        h_YLabel=get(ax(1),'YLabel');
            set(h_YLabel,'String','P(C_C)*p(x|C_C)');
        h_YLabel=get(ax(2),'YLabel');
            set(h_YLabel,'String','P(C_N)*p(x|C_N)');
            set(h_YLabel,'Rotation',270);
        axis tight;
        set(ax(1),'XTickLabel',[]);
        set(ax(1),'XTick',[]);
        h_XLabel=get(ax(1),'XLabel');
        set(h_XLabel,'Position',[0.498871 -3e-5 17.3205]);
        [ax,h1,h2]=plotyy(xchf,p_chf*pchf,xnorm,p_norm*pnorm); hold on;
        set(h1,'LineStyle','o');
        set(h2,'LineStyle','.');
        set(ax(1),'YColor',[1 0 0]);
        set(ax(1),'YLimMode','auto');
        set(ax(2),'YColor',[0 0 1]);
        set(ax(2),'YLimMode','auto');
        set(h1,'Color',[1 0 0]);
        set(h2,'Color',[0 0 1]);
        xlabel('pNNx');
        axis tight;

        pause; hold off; close all;
    end;
end;

figure;
plot(2:4,pthreshvec(2:4),'bo'); hold on; plot(2:4,pthreshvec(2:4),'b');
xlabel('threshold [samples]'); ylabel('Bayesian threshold [pNNx]');
grid on;

figure;
plot(2:4,p_norm_dchf(2:4),'b*',2:4,p_chf_dnorm(2:4),'rd'); hold on;
legend('P(Hnorm|Dchf)','P(Hchf|Dnorm)');
```

```
plot(2:4,p_norm_dchf(2:4),'b',2:4,p_chf_dnorm(2:4),'r');
xlabel('threshold [samples]'); ylabel('probability');
grid on;
pause; close all;
```

D.2 Student's Distribution

```
function p=stud_dist(N1,N2,t);
% Author: Stephen Dueck
% Purpose: Returns two-tailed p value
r=N1+N2-2;
p=betainc(r/(r+t^2),1/2*r,1/2);
```

D.3 Non-Parametric Estimation via Gaussian Kernel

```
function [stepvec,p]=gausest2(x,V,ts,plotstep);
% function [stepvec,p]=gausest2(x,V,ts,plotstep);
% Author: Stephen Dueck

N=length(x); %disp('Sorting'); x=sort(x);
range=max(x)-min(x);
lowx=mean(x)-range; upx=mean(x)+range;
lowx=0; upx=1;
lowx=round(lowx/ts)*ts; upx=round(upx/ts)*ts;
stepvec=lowx:ts:upx;

sigma=V/sqrt(N);

% disp(['Calculating ' num2str(length(stepvec)) ' point estimate (Gaussian kernel).']);
p=zeros(1,length(stepvec));
dispperc=-1;
for i=1:length(stepvec),
    k=sum( 1/sqrt(2*pi*sigma^2)*exp(-(x-stepvec(i)).^2/(2*sigma^2)) )*ts;
    p(i)=k/N;

    perc=round(i/length(stepvec)*100);
    if perc ~= dispperc
        newperc=1;
    end;

    if plotstep==1
        disp(['k=' num2str(k)]);
        disp(['p(i)=k/N=' num2str(k) '/' num2str(N)]);
        disp(['p(i)=' num2str(p(i))]);

        if N<1e6
            plot(x,linspace(0,0,N),'.','markersize',15); hold on;
        end;
        kern_wind=1/sqrt(2*pi*sigma^2)*exp( -(x-stepvec(i)).^2/(2*sigma^2) )*ts;
        plot(x,kern_wind,'r','markersize',1); hold on;
        plot(x,kern_wind,'r.');
        plot([stepvec(i) stepvec(i)],[0 max(kern_wind)],'r:');

        plot(stepvec(1:i),p(1:i),'k*'); plot(stepvec(1:i),p(1:i),'k');
        xlabel('x'); ylabel('p(x) estimate');
        pause; close all;
    end;
end;


I=find(p>1e-10);
```

D.4 Cross Validation

```
% Author: Stephen Dueck
% Purpose: Cross validation
close all;
newdata=0;

if newdata==1
clear all;
N=72; %72
threshvec=1:13;

normname(1,1:13)='normal\nsr001';
normname(2,1:13)='normal\nsr002';
normname(3,1:13)='normal\nsr003';
normname(4,1:13)='normal\nsr004';
normname(5,1:13)='normal\nsr005';
normname(6,1:13)='normal\nsr006';
normname(7,1:13)='normal\nsr007';
normname(8,1:13)='normal\nsr008';
normname(9,1:13)='normal\nsr009';
for i=10:54,
    normname(i,1:13)=['normal\nsr0' num2str(i)];
end;
normname(55,1:13)='normal2\16265';
normname(56,1:13)='normal2\16272';
normname(57,1:13)='normal2\16273';
normname(58,1:13)='normal2\16420';
normname(59,1:13)='normal2\16483';
normname(60,1:13)='normal2\16539';
normname(61,1:13)='normal2\16773';
normname(62,1:13)='normal2\16786';
normname(63,1:13)='normal2\16795';
normname(64,1:13)='normal2\17052';
normname(65,1:13)='normal2\17453';
normname(66,1:13)='normal2\18177';
normname(67,1:13)='normal2\18184';
normname(68,1:13)='normal2\19088';
normname(69,1:13)='normal2\19090';
normname(70,1:13)='normal2\19093';
normname(71,1:13)='normal2\19140';
normname(72,1:13)='normal2\19830';

for i=1:N,
    fname=normname(i,1:13);
    disp(fname)
    heasig=readheader([fname '.hea']);
    if i<=54
        t=[1 1e9];
        annot=readannot([fname '.ecg'],t);
    else
        t=[1 1e9]; %avg
        annot=readannot([fname '.atr'],t);
    end;
    samples=annot.time;
    normDuration(i)=samples(end)-samples(1);
    samples=samples(find(annot.anntyp~='~')); %exclude signal quality change annotations

    RR = diff(samples);
    Ncount(i)=length(RR);

    dRRdn = diff(RR);
    j=1;
    for thresh=threshvec,
        NN(i,j) = sum( abs(dRRdn)>=thresh );
        j=j+1;
    end;
end;
normNN=NN;
normpNN=zeros(size(NN));
for i=1:N,
```

```
    normpNN(i,:)=NN(i,:)/Ncount(i);
end;
normSEM=std(normpNN)/sqrt(N);
normCount=Ncount;
%=====================================================
Ncount=[];
NN=[];
N=29; %44
chfname(1,1:10)='chf\chf201';
chfname(2,1:10)='chf\chf202';
chfname(3,1:10)='chf\chf203';
chfname(4,1:10)='chf\chf204';
chfname(5,1:10)='chf\chf205';
chfname(6,1:10)='chf\chf206';
chfname(7,1:10)='chf\chf207';
chfname(8,1:10)='chf\chf208';
chfname(9,1:10)='chf\chf209';
for i=10:29,
    chfname(i,1:10)=['chf\chf2' num2str(i)];
end;
chfname(30,1:10)='chf2\chf01';
chfname(31,1:10)='chf2\chf02';
chfname(32,1:10)='chf2\chf03';
chfname(33,1:10)='chf2\chf04';
chfname(34,1:10)='chf2\chf05';
chfname(35,1:10)='chf2\chf06';
chfname(36,1:10)='chf2\chf07';
chfname(37,1:10)='chf2\chf08';
chfname(38,1:10)='chf2\chf09';
chfname(39,1:10)='chf2\chf10';
chfname(40,1:10)='chf2\chf11';
chfname(41,1:10)='chf2\chf12';
chfname(42,1:10)='chf2\chf13';
chfname(43,1:10)='chf2\chf14';
chfname(44,1:10)='chf2\chf15';

for i=1:N,
    fname=chfname(i,1:10);
    disp(fname);
    heasig=readheader([fname '.hea']);
    if i<=29
        t=[1 1e9];
        annot=readannot([fname '.ecg'],t);
        if length(annot.time)==1e9
            disp('err (pnn18): fix up didn''t work');
            pause;
        end;
    else
        t=[1 1e9]; %avg
        annot=readannot([fname '.ecg'],t);
    end;
    samples=annot.time;
    chfDuration(i)=samples(end)-samples(1);
    samples=samples(find(annot.anntyp~='~')); %exclude signal quality change annotations

    RR = diff(samples); %msec
    Ncount(i)=length(RR);

    dRRdn = diff(RR);
    j=1;
    for thresh=threshvec,
        NN(i,j) = sum( abs(dRRdn)>=thresh );
        j=j+1;
    end;
end;
chfNN=NN;
chfpNN=zeros(size(NN));
for i=1:N,
    chfpNN(i,:)=NN(i,:)/Ncount(i);
end;
chfSEM=std(chfpNN)/sqrt(N);
```

```
chfCount=Ncount;

end;

p_chf = .02;
p_norm = 1-p_chf;

j=2;
V=.5;
ts=.01;

    [xchf, pchf_xh]=gausest2(chfpNN(:,j),V,ts,0);
    [xnorm, pnorm_xh]=gausest2(normpNN(:,j),V,ts,0);
    [xtotal, ptotal]=gausest2([chfpNN(:,j); normpNN(:,j)], V, ts, 0);

    pchf=pchf_xh;
    pnorm=pnorm_xh;

    ptotal=pchf_xh*p_chf + pnorm_xh*p_norm;

    I=find(xchf>0 & xchf<1); pchf=pchf(I); xchf=xchf(I);
    I=find(xnorm>0 & xnorm<1); pnorm=pnorm(I); xnorm=xnorm(I);
    I=find(xtotal>0 & xtotal<1); ptotal=ptotal(I); xtotal=xtotal(I);

    pchf_hx=pchf./ptotal*p_chf; xchf_hx = xchf;
    pnorm_hx=pnorm./ptotal*p_norm; xnorm_hx=xnorm;

    pxc_norm=[]; pxc_chf=[];
    Tn_vec=1:length(xnorm);
    for Tn=Tn_vec,
        pxc_norm(Tn) = sum(pnorm(Tn:end));
        pxc_chf(Tn) = sum(pchf(1:Tn));
    end;

    zn=1;
    zvec=0.05:0.05:0.95;
    e=[]; xvec=[];
    for z=zvec,
        a=find(pxc_chf<z);
        xvec(zn)=xnorm(a(end));
        e(zn)=sum(ptotal(a).*pchf_hx(a))/sum(ptotal(a));
%         e(zn)
        zn=zn+1;
    end;

    plot(zvec,e,'r.'); hold on;
    plot(zvec,e,'r'); hold on;

for z=1:1,
acc_vec=0.05:0.05:0.95;
true_pos_vec=zeros(1,length(acc_vec));
false_pos_vecs=zeros(1,length(acc_vec));
nacc=1;
for acc=acc_vec,
    true_pos = 0;
    false_pos = 0;
    for i = 1:29,
        chfpNNz=[];
        for k=1:29,
            if k~=i,
                chfpNN_train=[chfpNNz chfpNN(k,j)];
            end;
        end;
        chfpNN_test = chfpNN(i,j);

        I=round(rand(1,1000)*(72-1))+1;
        normpNNz=normpNN(:,j);
        for k=1:1000,
            tmp=normpNNz(1);
            normpNNz(1)=normpNNz(I(k));
            normpNNz(I(k))=tmp ;
```

```
        end;
        normpNN_train = normpNNz(1:23);
        normpNN_test_set = normpNNz(24:72);

        [xchf, pchf_xh]=gausest2(chfpNN_train,V,ts,0);
        [xnorm, pnorm_xh]=gausest2(normpNN_train,V,ts,0);

        pchf=pchf_xh;
        pnorm=pnorm_xh;

        ptotal=pchf_xh*p_chf + pnorm_xh*p_norm;

        I=find(xchf>0 & xchf<1); pchf=pchf(I); xchf=xchf(I);
        I=find(xnorm>0 & xnorm<1); pnorm=pnorm(I); xnorm=xnorm(I);
        I=find(xtotal>0 & xtotal<1); ptotal=ptotal(I); xtotal=xtotal(I);

        pchf_hx=pchf./ptotal*p_chf; xchf_hx = xchf;
        pnorm_hx=pnorm./ptotal*p_norm; xnorm_hx=xnorm;


        I=find(pxc_chf<acc);
        thresh=xnorm(I(end));

        if chfpNN_test < thresh,
            true_pos = true_pos + 1;
        end;

        for k = 1:49,
            normpNN_test = normpNN_test_set(k);
            if normpNN_test < thresh,
                false_pos = false_pos + 1;
            end;
        end;

    end;  % i

    true_pos_vec(nacc)=true_pos;
    false_pos_vec(nacc)=false_pos;

    nacc=nacc+1;
end; % acc

    effic=true_pos_vec./(true_pos_vec+false_pos_vec);

    plot(acc_vec,e,'rx',acc_vec,effic,'b.'); hold on;
    legend('Estimated Efficiency','Cross Validation Efficiency');
    plot(acc_vec,e,'r',acc_vec,effic,'b'); hold on;

    xlabel('Accuracy');
    ylabel('Efficiency');

end;

figure;
plot(acc_vec,xvec,'k.',acc_vec,xvec,'k-');
xlabel('Accuracy'); ylabel('Decision Threshold in Terms of pNNx');

figure;
[nchf,xchf]=hist(chfpNN(:,j));
[nnorm,xnorm]=hist(normpNN(:,j),15);
bar(xchf,nchf,'r'); hold on;
bar(xnorm,nnorm,'b');
xlabel('pNNx');
ylabel('Relative Frequency');
```

D.5 Receiver Operating Characteristic Curves

```
close all;
newdata=0;

if newdata==1
clear all;
N=72; %72
threshvec=1:13;

normname(1,1:13)='normal\nsr001';
normname(2,1:13)='normal\nsr002';
normname(3,1:13)='normal\nsr003';
normname(4,1:13)='normal\nsr004';
normname(5,1:13)='normal\nsr005';
normname(6,1:13)='normal\nsr006';
normname(7,1:13)='normal\nsr007';
normname(8,1:13)='normal\nsr008';
normname(9,1:13)='normal\nsr009';
for i=10:54,
    normname(i,1:13)=['normal\nsr0' num2str(i)];
end;
normname(55,1:13)='normal2\16265';
normname(56,1:13)='normal2\16272';
normname(57,1:13)='normal2\16273';
normname(58,1:13)='normal2\16420';
normname(59,1:13)='normal2\16483';
normname(60,1:13)='normal2\16539';
normname(61,1:13)='normal2\16773';
normname(62,1:13)='normal2\16786';
normname(63,1:13)='normal2\16795';
normname(64,1:13)='normal2\17052';
normname(65,1:13)='normal2\17453';
normname(66,1:13)='normal2\18177';
normname(67,1:13)='normal2\18184';
normname(68,1:13)='normal2\19088';
normname(69,1:13)='normal2\19090';
normname(70,1:13)='normal2\19093';
normname(71,1:13)='normal2\19140';
normname(72,1:13)='normal2\19830';

for i=1:N,
    fname=normname(i,1:13);
    disp(fname)
    heasig=readheader([fname '.hea']);
    if i<=54
        t=[1 1e9];
        annot=readannot([fname '.ecg'],t);
    else
        t=[1 1e9]; %avg
        annot=readannot([fname '.atr'],t);
    end;
    samples=annot.time;
    normDuration(i)=samples(end)-samples(1);
    samples=samples(find(annot.anntyp~='~')); %exclude signal quality change annotations

    RR = diff(samples);
    Ncount(i)=length(RR);

    dRRdn = diff(RR);
    j=1;
    for thresh=threshvec,
        NN(i,j) = sum( abs(dRRdn)>=thresh );
        j=j+1;
    end;
end;
normNN=NN;
normpNN=zeros(size(NN));
for i=1:N,
    normpNN(i,:)=NN(i,:)/Ncount(i);
end;
```

```
normSEM=std(normpNN)/sqrt(N);
normCount=Ncount;
%=====================================================
Ncount=[];
NN=[];
N=29; %44
chfname(1,1:10)='chf\chf201';
chfname(2,1:10)='chf\chf202';
chfname(3,1:10)='chf\chf203';
chfname(4,1:10)='chf\chf204';
chfname(5,1:10)='chf\chf205';
chfname(6,1:10)='chf\chf206';
chfname(7,1:10)='chf\chf207';
chfname(8,1:10)='chf\chf208';
chfname(9,1:10)='chf\chf209';
for i=10:29,
    chfname(i,1:10)=['chf\chf2' num2str(i)];
end;
chfname(30,1:10)='chf2\chf01';
chfname(31,1:10)='chf2\chf02';
chfname(32,1:10)='chf2\chf03';
chfname(33,1:10)='chf2\chf04';
chfname(34,1:10)='chf2\chf05';
chfname(35,1:10)='chf2\chf06';
chfname(36,1:10)='chf2\chf07';
chfname(37,1:10)='chf2\chf08';
chfname(38,1:10)='chf2\chf09';
chfname(39,1:10)='chf2\chf10';
chfname(40,1:10)='chf2\chf11';
chfname(41,1:10)='chf2\chf12';
chfname(42,1:10)='chf2\chf13';
chfname(43,1:10)='chf2\chf14';
chfname(44,1:10)='chf2\chf15';

for i=1:N,
    fname=chfname(i,1:10);
    disp(fname);
    heasig=readheader([fname '.hea']);
    if i<=29
        t=[1 1e9];
        annot=readannot([fname '.ecg'],t);
        if length(annot.time)==1e9
            disp('err (pnn18): fix up didn''t work');
            pause;
        end;
    else
        t=[1 1e9]; %avg
        annot=readannot([fname '.ecg'],t);
    end;
    samples=annot.time;
    chfDuration(i)=samples(end)-samples(1);
    samples=samples(find(annot.anntyp~='~')); %exclude signal quality change annotations

    RR = diff(samples); %msec
    Ncount(i)=length(RR);

    dRRdn = diff(RR);
    j=1;
    for thresh=threshvec,
        NN(i,j) = sum( abs(dRRdn)>=thresh );
        j=j+1;
    end;
end;
chfNN=NN;
chfpNN=zeros(size(NN));
for i=1:N,
    chfpNN(i,:)=NN(i,:)/Ncount(i);
end;
chfSEM=std(chfpNN)/sqrt(N);
chfCount=Ncount;
```

```
end;


V=.5;
plotstep=1; ts=.01;
p_chf = .02;
p_norm = .98;

p_chf = .02;
p_norm = 1-p_chf;

for j=2:2,
    [xchf, pchf_xh]=gausest2(chfpNN(:,j),V,ts,0);
    [xnorm, pnorm_xh]=gausest2(normpNN(:,j),V,ts,0);
%     [xtotal, ptotal]=gausest2([chfpNN(:,j); normpNN(:,j)], V, ts, 0);

    pchf=pchf_xh;
    pnorm=pnorm_xh;

    I=find(xchf>0 & xchf<1); pchf=pchf(I); xchf=xchf(I);
    I=find(xnorm>0 & xnorm<1); pnorm=pnorm(I); xnorm=xnorm(I);
%     I=find(xtotal>0 & xtotal<1); ptotal=ptotal(I); xtotal=xtotal(I);

    pchf=pchf/sum(pchf); pnorm=pnorm/sum(pnorm);
    ptotal=pchf_xh*p_chf + pnorm_xh*p_norm; ptotal=ptotal(1:99);
ptotal=ptotal/sum(ptotal);

    pchf_hx=pchf./ptotal*p_chf; xchf_hx = xchf;
    pnorm_hx=pnorm./ptotal*p_norm; xnorm_hx=xnorm;

    pxc_norm=[]; pxc_chf=[];
    Tn_vec=1:length(xnorm);
    for Tn=Tn_vec,
        pxc_norm(Tn) = sum(pnorm(1:Tn));
        pxc_chf(Tn) = sum(pchf(1:Tn));
    end;

    zn=1;
    zvec=0:0.025:1;
%     fpr=[];
    e=[];
    for z=zvec,
        a=find(pxc_chf<z);
%         fpr(zn)=sum(pnorm(a));
        e(zn)=sum(ptotal(a).*pchf_hx(a))/sum(ptotal(a));
        zn=zn+1;
    end;
    e=e/max(e);
%     tpr=zvec;

    zn=1;
    fpr=0:0.025:1;
    tpr=[]; xval=[];
    for z=fpr,
        a=find(pxc_norm<z);
        if isempty(a)
            xval(zn)=0
        else
            xval(zn)=xnorm(a(end));
        end;
        tpr(zn)=sum(pchf(a));
%         e(zn)=sum(ptotal(a).*pchf_hx(a))/sum(ptotal(a));
        zn=zn+1;
    end;
%     tpr=zvec;

%     subplot(3,1,1);
    figure;
    plot(zvec,e); hold on; plot(zvec,e,'.');
    xlabel('Accuracy [%]'); ylabel('Efficiency [%]');
    grid on;
```

```
    axis equal; axis tight;

%     subplot(3,1,2);
    figure;
    plot(fpr,tpr,'b.',tpr,tpr,'k-'); hold on;
    legend('ROC Curve','TPR=FPR Line');
    plot(fpr,tpr);
    xlabel('FPR [%]'); ylabel('Accuracy [%]');
    grid on;
    axis equal; axis tight;

%     subplot(3,1,3);
    figure;
    tpr_diff=diff(tpr);
    %     post_est=1./(1+tpr_diff)./fpr(2:end);
    a=.98/.02; a=1;
    post_est=a*tpr_diff./(a*tpr_diff+1);
    I=2:length(fpr);
    post_est=post_est/max(post_est);

%     tpr_diff=tpr_diff/max(tpr_diff);
%     post_est=post_est/max(post_est);

    plot(fpr(I),tpr_diff,'b.',fpr(I),post_est,'ro'); hold on;
    legend('ROC Curve Slope','P(C_C|x) Estimate as Function of ROC slope');
    xlabel('FPR'); %ylabel('[%]');
%     axis equal; axis tight;
    plot(fpr(I),tpr_diff,'b'); hold on;
    plot(fpr(I),post_est,'r'); hold on;
    grid on;

%     subplot(3,1,3);
%     zn=1;
%     zvec=0:0.025:1;
%     fpr=[];
%     e=[];
%     for z=zvec,
%         a=find(pxc_chf<z);
%         fpr(zn)=sum(pnorm(a));
%         e(zn)=sum(ptotal(a).*pchf_hx(a))/sum(ptotal(a));
%         zn=zn+1;
%     end;
%     e=e/max(e);
%     tpr=zvec;
%


%     pause;


end;

pause; close all;
for j=1:5,
    [xchf, pchf_xh]=gausest2(chfpNN(:,j),V,ts,0);
    [xnorm, pnorm_xh]=gausest2(normpNN(:,j),V,ts,0);
%     [xtotal, ptotal]=gausest2([chfpNN(:,j); normpNN(:,j)], V, ts, 0);

    pchf=pchf_xh;
    pnorm=pnorm_xh;

    I=find(xchf>0 & xchf<1); pchf=pchf(I); xchf=xchf(I);
    I=find(xnorm>0 & xnorm<1); pnorm=pnorm(I); xnorm=xnorm(I);
%     I=find(xtotal>0 & xtotal<1); ptotal=ptotal(I); xtotal=xtotal(I);

    pchf=pchf/sum(pchf); pnorm=pnorm/sum(pnorm);
    ptotal=pchf_xh*p_chf + pnorm_xh*p_norm; ptotal=ptotal(1:99);
ptotal=ptotal/sum(ptotal);

    pchf_hx=pchf./ptotal*p_chf; xchf_hx = xchf;
    pnorm_hx=pnorm./ptotal*p_norm; xnorm_hx=xnorm;
```

```
    pxc_norm=[]; pxc_chf=[];
    Tn_vec=1:length(xnorm);
    for Tn=Tn_vec,
        pxc_norm(Tn) = sum(pnorm(1:Tn));
        pxc_chf(Tn) = sum(pchf(1:Tn));
    end;

    zn=1;
    zvec=0:0.025:1;
%     fpr=[];
    e=[];
    for z=zvec,
        a=find(pxc_chf<z);
%         fpr(zn)=sum(pnorm(a));
        e(zn)=sum(ptotal(a).*pchf_hx(a))/sum(ptotal(a));
        zn=zn+1;
    end;
    if max(e)>1
        e=e/max(e);
    end;
%     tpr=zvec;

    zn=1;
    fpr=0:0.025:1;
    tpr=[];
    for z=fpr,
        a=find(pxc_norm<z);
        tpr(zn)=sum(pchf(a));
%         e(zn)=sum(ptotal(a).*pchf_hx(a))/sum(ptotal(a));
        zn=zn+1;
    end;
%     tpr=zvec;

%     subplot(2,1,1);
%     figure;
%     plot(zvec,e); hold on; plot(zvec,e,'.');
%     xlabel('Accuracy'); ylabel('Efficiency');
%     grid on;
%     axis equal; axis tight;

%     subplot(2,1,2);
% %     figure;
    plot(fpr,tpr,'b.',tpr,tpr,'k-'); hold on;
    legend('ROC Curve','TPR=FPR Line');
    plot(fpr,tpr);
    xlabel('FPR [%]'); ylabel('Accuracy');
    grid on;
    axis equal; axis tight;

    pause;


end;

pause; close all;
```

D.6 Reading Data

```
function heasig=readheader(name);
% function heasig=readheader(name);
% READHEADER function reads the header of DB signal files
%       Input parameter: character string with name of header file
%       Output parameter: struct heasig with header information

% Salvador Olmos
% e-mail: olmos@posta.unizar.es

% Opening header file
fid=fopen(name,'rt');
```

```
if (fid<=0)
   disp(['error in opening file ' name]);
   keyboard;
end

pp=' /+:()x';

% First line reading
s=fgetl(fid);
% Remove blank or commented lines
while s(1)=='#'
  s=fgetl(fid);
end

[heasig.recname,s]=strtok(s,pp);
[s1 s]=strtok(s,pp);
heasig.nsig=str2num(s1);
[s1 s]=strtok(s);
if isempty(findstr(s1,'/'))
   heasig.freq=str2num(s1);
else
   [s1 s2]=strtok(s1,'/');
   heasig.freq=str2num(s1);
   %[s1 s]=strtok(s);
end

[s1 s]=strtok(s);
heasig.nsamp=str2num(s1);

if ~isempty(deblank(s))
   [s1 s]=strtok(s,':');
   hour=str2num(s1);
   [s1 s]=strtok(s,':');
   min=str2num(s1);
   [s1 s]=strtok(s,pp);
   sec=str2num(s1);
end

if ~isempty(deblank(s))
   [s1 s]=strtok(s,'/');
   month=str2num(s1);
   [s1 s]=strtok(s,'/');
   day=str2num(s1);
   [s1 s]=strtok(s,pp);
   year=str2num(s1);
end
%if exist('hour','var') heasig.date=datenum(year,month,day,hour,min,sec); end

% default values
for i=1:heasig.nsig
  heasig.units(i,:)='mV';
end
sig=1;

% Reading nsig lines, corresponding one for every lead
for i=1:heasig.nsig
  s=fgetl(fid);
  % Remove blank or commented lines
  while s(1)=='#'
    s=fgetl(fid);
  end

  [heasig.fname(i,:),s]=strtok(s,pp);
  [s1,s]=strtok(s,pp);
  if i==1  heasig.group(i)=0;
  else
     if strcmp(heasig.fname(i,:),heasig.fname(i-1,:))
        heasig.group(i)=0;
     else
        heasig.group(i)=heasig.group(i-1)+1;
     end
```

```
  end
  a=[findstr(s,'x') findstr(s,':') findstr(s,'+')];
  if isempty(a)
     heasig.fmt(i)=str2num(s1);
  else
    [s2,s]=strtok(s);
    a=[a length(s2)+1];
    for k=1:length(a)-1
      switch (s2(a(k)))
       case '+',
        heasig.fmt(i)=str2num(s1);
        heasig.offset(i)=str2num(s2(a(k)+1:a(k+1)-1));
       case ':',
        heasig.fmt(i)=str2num(s1);
        heasig.skew(i)=str2num(s2(a(k)+1:a(k+1)-1));
       case 'x',
        heasig.fmt(i)=str2num(s1);
        heasig.spf(i)=str2num(s2(a(k)+1:a(k+1)-1));
      end
    end
  end
  [s1,s]=strtok(s,pp);
  a=[findstr(s,'(') findstr(s,'/')];
  if isempty(s1)
      heasig.gain(i)=0;
      heasig.baseline(i)=0;
  else
      if isempty(a)
        heasig.gain(i)=str2num(s1);
      else
       [s2,s]=strtok(s);
        a=[a length(s2)+1];
        for k=1:length(a)-1
          switch (s2(a(k)))
           case '(',
               heasig.gain(i)=str2num(s1);
               a2=findstr(s2,')');
               heasig.baseline(i)=str2num(s2(1+a(k):a2-1));
           case '/',
               heasig.gain(i)=str2num(s1);
                f=s2(a(k)+1:end);
               heasig.units(i,1:length(f))=f;
          end
        end
      end
      [s1,s]=strtok(s,pp);
      heasig.adcres(i)=str2num(s1);
      [s1,s]=strtok(s,pp);
      heasig.adczero(i)=str2num(s1);
      [s1,s]=strtok(s,pp);
      heasig.initval(i)=str2num(s1);
      [s1,s]=strtok(s,pp);
      heasig.cksum(i)=str2num(s1);
      [s1,s]=strtok(s,pp);
      heasig.bsize(i)=str2num(s1);
      heasig.desc(i,1:length(s))=s;
  end
end

fclose(fid);
```