# THE DESIGN AND EVALUATION OF A BEDSIDE CARDIAC ARRHYTHMIA MONITOR

by

Paul Scott Schluter

B.S., California Institute of Technology (1973)
S.M., E.E., Massachusetts Institute of Technology (1976)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS OF THE
DEGREE OF

DOCTOR OF PHILOSOPHY IN
ELECTRICAL ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September, 1981



Signature of Author_____
Department of Electrical Engineering
September 1, 1981

Certified by_____
Roger G. Mark
Thesis Supervisor

Accepted by_____
Arthur C. Smith
Chairman, Departmental Graduate Committee

# THE DESIGN AND EVALUATION OF A BEDSIDE CARDIAC ARRHYTHMIA MONITOR

by

Paul Scott Schluter

Submitted to the Department of
Electrical Engineering and Computer Science
on September 1, 1981 in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy
in Electrical Engineering

## ABSTRACT

The primary objective of this research was the development and evaluation of an advanced instrumentation system to improve the diagnosis and treatment of patients with cardiac arrhythmias. The instrument is a self-contained bedside monitor that performs real-time arrhythmia analysis for hospitalized patients. It provides clinically useful summary statistical and trend displays which are immediately available to medical personnel during monitoring. The system automatically documents important arrhythmias by producing hard-copy rhythm strips.

A major component of the thesis effort was the development and evaluation of algorithms to detect and classify cardiac events. The arrhythmia algorithm compresses the incoming single-lead electrocardiogram (ECG) to permit real-time operation, and employs feature extraction and clustering to classify QRS complexes. Integral to this effort was the creation of an annotated data base for algorithm design and evaluation. A multifaceted evaluation using the data base, long term analog tape recordings, and clinical evaluation of the prototype instrument was also performed.

Thesis Supervisor: Dr. Roger G. Mark, M.D., Ph.D.
Title: Matsushita Associate Professor of
Electrical Engineering in Medicine

ACKNOWLEDGEMENTS

No research is the effort of a single person, and this thesis is no exception. It is with pleasure that I acknowledge the support and inspiration of Roger Mark, who has for the past five years provided me with ideas, friendship, and encouragement. I also thank the other members of my thesis committee, Stephen Burns, Lou Braida, Berthold Horn, Alan Willsky, and Walter Abelmann for their guidance.

Many thanks go to the members of the Arrhythmia Project at the Beth Israel Hospital and at the Biomedical Engineering Center for Clinical Instrumentation at MIT. My warmest thanks go to Dianne Bacon, who played a crucial role in the clinical and Holter tape evaluation of the bedside monitor, and for getting the author and monitor to function properly in the clinical environment. I would also like to thank Joe Mietus who was responsible for the design and fabrication of the later models of the bedside monitor, and who generally kept everything working around the lab. Thanks also go to Scott Peterson, George Moody, and Cheryl Jackson for their efforts in the development and dissemination of the MIT-BIH ECG database; and to the newer members of the group, Phil Devlin and Ted Baker, for the further contributions to the data base and evaluation software; and to Peter Dinhofer for his effort in implementing the Holter tape evaluation procedure.

I am also grateful to the gang at the BME center for making my stay at MIT a pleasant one. I would especially like to thank Gloria McAvenia, who has waited these many years to see the completion of my thesis, and Donald Wade. I will also certainly miss John (Judy, Jeremy, and JoJo) Tole, and Jon Valvano, as we each go our separate ways. Also, to my roomates and friends Mike Coln and Jon Teich, farewell.

I gratefully acknowledge the financial support from the National Aeronautics and Space Administration (NASA Grant No. NAS-9-14618), and the Matsushita Professorship of Electrical Engineering in Medicine.

My sincere thanks go to all my family who have provided me with encouragement during my stay at MIT.

"This technologic terror you have created, commander, is no match for the power of the Force."

Darth Vader

in "Star Wars"

Table of Contents

TITLE PAGE 1

ABSTRACT 2

ACKNOWLEDGEMENTS 3

TABLE OF CONTENTS 5

LIST OF FIGURES 8

LIST OF TABLES 11

1. INTRODUCTION 12

1.1 Sudden Death and Coronary Heart Disease.....12
1.2 Coronary Care Units.....13
1.3 The Need for Arrhythmia Monitoring.....14
1.4 Rationale for a Bedside Arrhythmia Monitor.....18
1.5 Thesis Objectives.....19

2. THE ELECTROCARDIOGRAM 23

2.1 Mechanism of Generation.....23
2.2 Cardiac Rhythms.....27
2.3 Ectopic Beats and Rhythms.....33

3. REVIEW OF AUTOMATED ARRHYTHMIA ANALYSIS 42

3.1 Overview.....42
3.2 Overall Structure of Computer Arrhythmia Analysis Systems.....43
3.3 ECG Aquisition and Preprocessing.....46
3.4 Data Compression Techniques.....48
3.5 Detection of ECG Waveforms.....51
3.6 Delineation of ECG Waveforms.....55
3.7 QRS Shape Description and Morphology Classification.....57
3.8 Rhythm Characterization.....71
3.9 Algorithm Evaluation.....75

4. THE BEDSIDE MONITOR 78

4.1 Overall Design Goals.....78
4.2 Features and Operation of the Bedside Monitor....80
4.3 Hardware and Software Development Environment....84
4.4 ECG Analysis Algorithm Overview.....89

5. DATA AQUISITION AND EVENT DETECTION 95

5.1 Analog Data Aquisition and Filtering.....95
5.2 Data Compression.....100
5.3 Triangular Morph Concatenation.....111
5.4 Event Detection.....118
5.5 Discussion and Summary.....124

6. MORPHOLOGY CLASSIFICATION 127

6.1 Feature Extraction and Distance Measure.....131
6.2 Morphology Clustering.....142
6.3 QRS Morphology Family Identification.....157
6.4 QRS Complex Morphology Classification.....176
6.5 Final QRS classification.....184

7. TIMING CLASSIFICATION 185

7.1 RR Interval Prediction.....186
7.2 Atrial Fibrillation Detection.....210
7.3 Markov Atrial Fibrillation Detector.....211
7.4 Predictive Atrial Fibrillation Detector.....217

8. ARTIFACT REJECTION 234

8.1 Handling Processor Overload.....234
8.2 Artifact and noise rejection.....237

9. SUMMARY AND EPISODE REPORTS 247

9.1 Bedside Monitor Trend Displays.....252
9.2 Statistical Displays.....257
9.3 Detection of Alarm Conditions.....262
9.4 Alarm and Episode Documentation Strategy.....268
9.5 Bedside Monitor Log Reports.....281

10. DATA BASE DEVELOPMENT AND EVALUATION 284

10.1 ECG Data Base: Development.....288
10.2 ECG Data Base: Description.....294
10.3 Beat by Beat Comparison.....302
10.4 Beat-by-Beat Evaluation Results.....306
10.5 Other Performance Measures.....315
10.6 Sensitivity for Sequential Events.....320
10.7 SVPB Detection Accuracy.....323
10.8 Rhythm and Noise Classification Accuracy.....324
10.9 Analysis of Algorithm Errors.....325
10.10 Comparison With Other Algorithms.....338
10.11 Discussion.....342

11. LONG-TERM ANALOG TAPE EVALUATION 344

11.1 Hour-by-hour comparison.....347
11.2 Episode Documentation.....356
11.3 Discussion.....363

12. CLINICAL EVALUATION 364

12.1 Clinical Application of the Bedside Monitor.....367
12.2 Interviews with the Clinical Staff.....375
12.3 Cost Comparison of Arrhythmia Monitoring Systems.....380

13. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK 386

13.1 Conclusions.....386
13.2 Directions for Future Work.....389

APPENDICES 394

A-1 ECG Data Base Format for 1/4" Tape Cartridges...394
A-2 Specification for FM Analog Data Base Tapes.....405
A-3 "Compressed" Annotation Format Description.....408

BIBLIOGRAPHY 410

BIOGRAPHICAL SKETCH 425

# List of Figures


| | | |
|---|---|---|
| 2-1 | Cardiac Electrical Conduction System | 26 |
| 2-2 | Normal QRS Complexes | 26 |
| 2-3a | Normal Sinus Rhythm | 31 |
| 2-3b | Sinus Tachycardia | 31 |
| 2-3c | Sinus Bradycardia | 31 |
| 2-3d | Sinus Arrhythmia | 31 |
| 2-4a | First Degree A-V Block | 32 |
| 2-4b | Second Degree A-V Block - Mobitz Type II | 32 |
| 2-4c | Complete A-V Block with Nodal Rhythm | 32 |
| 2-4d | Right and Left Bundle Branch | 32 |
| 2-5a | Atrial Premature Beats | 35 |
| 2-5b | Atrial Tachycardia | 35 |
| 2-5c | Atrial Flutter | 35 |
| 2-5d | Atrial Fibrillation | 35 |
| 2-6a | Ventricular Premature Beats | 40 |
| 2-6b | Ventricular Premature Beats | 40 |
| 2-6c | Bigeminy | 40 |
| 2-6d | Multiform VPBs | 40 |
| 2-7a | Short Burst of Ventricular Tachycardia | 41 |
| 2-7b | Early VPB and Run of Ventricular Tachycardia | 41 |
| 2-7c | Ventricular Flutter | 41 |
| 2-7d | Ventricular Fibrillation | 41 |
| 4-1 | The Bedside Cardiac Arrhythmia Monitor | 81 |
| 4-2 | Bedside Monitor Hardware Block Diagram | 85 |
| 4-3 | ECG Processing Steps | 90 |
| 5-1a | ECG Amplifier and Analog Filtering | 97 |
| 5-1b | Analog Filter Frequency Response | 97 |
| 5-2a | 60 Hz Notch Filter Frequency Response | 99 |
| 5-2b | Overall Frequency Response | 99 |
| 5-3a | Detailed ZOI Operation | 103 |
| 5-3b | Hysteretical Smoothing Performed by the ZOI | 103 |
| 5-4a | Initial Waveform Partitioning | 105 |
| 5-4b | Final Waveform Segmentation | 105 |
| 5-5a | Initial Segmentation of the ECG | 107 |
| 5-5b | Original and Compressed ECG | 108 |
| 5-5c | Original and Compressed ECG Shown Separately | 109 |
| 5-6 | Triangular Morph Concatenation | 113 |
| 5-7 | RQUAL Computation | 115 |
| 5-8 | Triangular Morph Detection and Labeling | 119 |
| 5-9a | T-wave Detection Thresholds | 123 |
| 5-9b | R-X-R Interval Decision Boundaries | 123 |
| 5-10 | ECG Processing Stages | 125 |
| 5-11 | Major Data Structures | 126 |

| | | |
|---|---|---|
| 6-1 | QRS Classification Procedure | 128 |
| 6-2 | QRS Morphology Feature Set and Distance Measure | 138 |
| 6-3 | Morphology Clustering Procedure | 143 |
| 6-4 | Adaptive Similarity Distance Threshold | 148 |
| 6-5 | Morphology Classes | 155 |
| 6-6 | Width and Prematurity-Ratio Histograms | 162 |
| 6-7 | Morphology Labeling Decision Boundaries | 167 |
| 6-8a | Decision Regions for New Morphology Classes | 171 |
| 6-8b | Decision Regions for Populated Classes | 172 |
| 6-8c | Decision Regions for Wide Normal Beats | 173 |
| 6-9 | Weighting Functions for QRS Complex Classification | 180 |
| 7-1 | RR Interval Predictor Weighting Functions | 195 |
| 7-2a | SVPB detection performance: 100 series with TW=1*8 | 200 |
| 7-2b | SVPB detection performance: 100 series with TW=2*8 | 201 |
| 7-2c | SVPB detection performance: 100 series with TW=6*8 | 202 |
| 7-2d | SVPB detection performance: 100 series with TW=10*8 | 203 |
| 7-3a | SVPB detection performance: 200 series with TW=1*8 | 204 |
| 7-3b | SVPB detection performance: 200 series with TW=2*8 | 205 |
| 7-3c | SVPB detection performance: 200 series with TW=6*8 | 206 |
| 7-3d | SVPB detection performance: 200 series with TW=10*8 | 207 |
| 7-4a | Markov AF Detector Performance Excluding PVCs | 215 |
| 7-4b | Markov AF Detector Performance Including PVCs | 216 |
| 7-5 | Predictive AF Detector Response | 220 |
| 7-6a | BIN AF Detector Performance Excluding PVCs | 224 |
| 7-6b | BIU AF Detector Performance Excluding PVCs | 225 |
| 7-6c | PIN AF Detector Performance Excluding PVCs | 226 |
| 7-6d | PIU AF Detector Performance Excluding PVCs | 227 |
| 7-7a | BIN AF Detector Performance Including PVCs | 228 |
| 7-7b | BIU AF Detector Performance Including PVCs | 229 |
| 7-7c | PIN AF Detector Performance Including PVCs | 230 |
| 7-7d | PIU AF Detector Performance Including PVCs | 231 |
| 8-1 | Transformed Distance for Artifact Rejection | 244 |
| 8-2 | Examples of Artifact Detection | 246 |
| 9-1 | The Bedside Monitor - Model II | 248 |
| 9-2 | Entering Patient Information | 251 |
| 9-3 | Annotated Real Time ECG Display | 251 |
| 9-4a | One Hour Heart Rate Trend | 254 |
| 9-4b | One Hour SVPB Rate Trend | 254 |
| 9-4c | One Hour PVC Rate Trend | 255 |
| 9-4d | One Hour PVC Run Length Trend | 255 |
| 9-5a | RR Interval Histogram | 259 |
| 9-5b | PVC Coupling Histogram | 259 |
| 9-6a | Running Average QRS Width Histogram | 260 |
| 9-6b | Running Average Prematurity Ratio Histogram | 260 |
| 9-7a | Classifier Decision Regions | 261 |
| 9-7b | Candidate Distance to NORMALs and PVCs | 261 |
| 9-8 | Modifying the Episode Documentation Rate | 273 |
| 9-9 | Strip Chart Episodes and Hourly Summary | 279 |
| 9-10 | Example of a Bedside Monitor Log Report | 282 |
| 9-11 | Log Buffer Status | 283 |

| | | |
|---|---|---|
| 10-1 | Evaluation Questions and Techniques | 285 |
| 10-2 | Strip Chart Recording of Comparator Output | 305 |
| 10-3 | PVC Detection Accuracy for Individual Trials | 314 |
| 10-4 | Cumulative PVC Sensitivity | 318 |
| 10-5 | Cumulative PVC Positive Predictivity | 319 |
| 10-6a | Tape 201: Aberrated Atrial Premature Beats | 330 |
| 10-6b | Tape 213: Fusion Beats | 331 |
| 10-6c | Tape 107: Wide Normal Beats | 332 |
| 10-6d | Tape 203: Variable NORMAL Morphology | 333 |
| 10-6e | Tape 223: Bidirectional Ventricular Tachycardia | 334 |
| 10-6f | Tape 207: Onset of Ventricular Flutter | 335 |
| 10-6g | Tape 207: Ventricular Fibrillation | 336 |
| 10-6h | Tape 207: Termination of Ventricular Flutter | 337 |

List of Tables


| | | |
|---|---|---|
| 6-1 | Decision Region Vertices | 168 |
| 9-1 | Bedside Monitor Trend Plot Parameters | 256 |
| 9-2 | Table of Alarm Conditions | 267 |
| 9-3 | Prioritized List of Alarm and Episode Conditions | 270 |
| 10-1a | MIT-BIH Data Base Annotation Codes | 295 |
| 10-1b | Rhythm Onset Annotations | 296 |
| 10-2a | Predominant Rhythms for 100 Series Tapes | 298 |
| 10-2b | Predominant Rhythms for 200 Series Tapes | 299 |
| 10-3a | QRS Types for 100 Series Tapes | 300 |
| 10-3b | QRS Types for 200 Series Tapes | 301 |
| 10-4a | Beat-by-Beat Confusion Matrix for Tape 200 | 304 |
| 10-4b | PVC Run Length Confusion Matrix for Tape 200 | 304 |
| 10-5a | Evaluation Results for Entire Database | 308 |
| 10-5b | Summarized Results for Entire Database | 308 |
| 10-5c | Results for Atrial Flutter/Fibrillation only | 308 |
| 10-5d | Results for 100 series tapes | 308 |
| 10-5e | Results for 200 series tapes | 308 |
| 10-6 | Tabulated Results for Individual Trials | 312 |
| 10-7 | QRS and PVC Detection Accuracy for Individual Trials | 313 |
| 10-8 | PVC Run Length Comparison | 321 |
| 10-9 | Evaluation Results for Other Systems | 341 |
| 11-1a | Uncorrected Bedside Monitor vs Holter | 350 |
| 11-1b | Corrected Bedside Monitor vs Holter | 350 |
| 11-2a | Uncorrected BAM versus BAMC+HOLTER | 351 |
| 11-2b | Corrected BAM versus BAMC+HOLTER | 351 |
| 11-3a | Corrected Bedside Monitor vs Gold Hour | 353 |
| 11-3b | Holter vs Gold Hour | 353 |
| 11-4a | Uncorrected Bedside Monitor versus Data Base | 355 |
| 11-4b | Corrected Bedside Monitor versus Data Base | 355 |
| 11-5 | Bedside Monitor Documention Episodes | 359 |
| 11-6 | Episode Documentation for Holter Tape Evaluation | 360 |
| 11-7a | Distribution of Strip Production Rate | 362 |
| 11-7b | Poor Signal Duration | 362 |
| 12-1 | Patient Connect Time to Monitor | 367 |
| 12-2a | Episode Documentation Summary for All Patients | 370 |
| 12-2b | Distribution of Strips per Hour for All Patients | 371 |
| 12-2c | Poor Signal Duration for All Patients | 371 |
| 12-2d | Keyboard Requests per Hour for All Patients | 372 |
| 12-2e | Keyboard Requests by Type for All Patients | 372 |
| 12-3 | Individual Patient Summaries | 374 |
| 12-4a | Estimated Budget for the Bedside Monitor | 382 |
| 12-4b | Estimated Budget for Holter Arrhythmia Lab | 383 |

# 1. INTRODUCTION

## 1.1 Sudden Death and Coronary Heart Disease

Coronary heart disease is a major problem in the United States. It is estimated that over 638,000 deaths occur annually due to coronary heart disease in the population at large (AHA,1980). Approximately one-half of cardiac deaths are sudden and unexpected, and often occur within an hour of onset of symptoms.

The principal mechanism responsible for coronary "sudden death" in the patient with coronary artery disease is ventricular fibrillation (Lown, 1971). Ventricular fibrillation is a derangement of heart rhythm, the result of non-synchronous, ch tic electrical activity. In this condition, no coordinated pumping activity is possible, and death soon follows.

It has been often suggested that many people who die of sudden cardiac death "have hearts too good to die". The injury caused by coronary arterial disease or myocardial infarction is usually not sufficient to compromise the mechanical integrity of the heart. It is more likely that the concomitant metabolic ischemia causes "electrical instablity" of the myocardium. This can lead to heart rhythm abnormalities, and possibly to death if ventricular fibrillation occurs.

## 1.2 Coronary Care Units

A major therapeutic advance in the treatment of heart attack victims was the coronary care unit (CCU). Several observations and innovations led to the development of the CCU. First was the observation that ventricular fibrillation was a reversible disorder in heart rhythm. As early as 1947 it was demonstrated that fibrillation could be stopped by electric shock delivered directly to the heart muscle (Beck,1947). It was later demonstrated that a massive electric shock across the unopened chest could also terminate ventricular fibrillation (Zoll,1956). Second was the use of "external cardiac massage" to maintain the viability of vital organs such as the brain and heart. A person whose heart had stopped pumping could be kept alive by rhythmically compressing the lower part of the sternum. This maneuver provided the necessary time to get physicians and equipment to resuscitate the patient.

These considerations led to the development of the coronary care unit (Brown,1963, Day,1963, Julian,1964, Meltzer,1964). Initially, the focus of the CCU was to provide continuous monitoring of patients with acute myocardial infarction (MI), and if required, to rapidly apply cardiac resuscitation. These units were equipped with bedside monitors and staffed by nurses trained in arrhythmia recognition and resuscitation.

As CCUs proliferated, it was observed that MI patients exhibited a wide variety of cardiac arrhythmias. Evidence accumulated that catastrophic arrhythmias were often preceeded by lesser, premonitory arrhythmias such as ventricular premature beats (VPBs). Thus, the emphasis of the CCU switched from resuscitation to aggressive prophylactic therapy (Lown,1969, Meltzer,1969). Minor ventricular arrhythmias were treated, with the aim of preventing ventricular fibrillation. The coronary care unit, with its extensive monitoring equipment and trained personnel has resulted in a 30% reduction in mortality among hospitalized patients with myocardial infarction (Lown,1971). Unfortunately, two-thirds of the patients who die from coronary heart disease never make it to the hospital (Lown,1971).

## 1.3 The Need for Arrhythmia Monitoring

Arrhythmia monitoring in the CCU performs two vital roles. First, it provides the basic function of detecting life threatening arrhythmias, and providing some type of alarm if they occur. Second, monitoring documents trends in rhythm abnormalities, and the effectiveness of arrhythmia suppression therapy. Early CCUs used large screen ECG displays and simple analog heart rate meters with alarms. These early systems, however, had a number of faults. Alarms designed to indicate high or low heart rate would sound inappropriately as a result of ECG artifacts from muscle noise or movement of electrodes. Continuous observation of oscilloscope displays is impractical,

and unreliable in detecting rare but important cardiac events. One study of thirty-one patients in a typical CCU setting showed that VPB's were detected by conventional means in only 64% of those patients exhibiting them (Romhilt,1973). Serious premonitory ventricular events were missed even more frequently. Other studies (Vetter,1975, Lindsay,1975) gave similar results, and indicate that relatively rare cardiac events are difficult to detect with conventional monitoring techniques.

Automated computer monitoring of the ECG offers several advantages over conventional monitoring technique. Sophisticated algorithms could reduce false alarms due to artifact, and offer tireless and more reliable detection of rare cardiac events. Automatic and consistent documentation of drug trials could be provided. The data processing task is poorly suited for humans since it requires prolonged concentration and detailed record keeping.

Despite the evidence demonstrating a need for automated arrhythmia detection in the CCU, the use of such systems is not as widespread as one would expect. Although most computer based systems detect many times the percentage of arrhythmias detected by conventional means, many are not robust enough for clinical use. Experience in a university teaching hospital led one researcher (Yanowitz,1974) to conclude that the system he evaluated was not acceptable for general clinical use. In another instance, trained technicians were hired to correct the errors of a commercially available arrhythmia monitoring system

(Steiner,1971).

Economic factors also have hindered the widespread use of computer arrhythmia monitoring systems. Many hospitals cannot afford the large dollar outlay for installation and maintenance of a multipatient system. Even though these systems can typically service 8 to 16 patients, most of these systems cost from fifty to one hundred thousand dollars, and represent a major investment. Multipatient monitoring systems are usually associated with considerable medical technology and highly skilled nursing care. The cost of beds in such units is, of course, quite high compared to ordinary medical and surgical beds.

In the past five to ten years, long term "Holter" tape recording of the ECG has gained enormous popularity. The technique is suitable for outpatients who carry the battery powered recorders with them as they go about their daily activities. It is also useful for patients who are in the hospital, but less frequently in intensive care units. The tapes are analysed by highly skilled technicians who use specialized audio/visual displays to scan tapes at 60 or 120 times real time. Sample ECG rhythm strips and crude statistical summaries are provided for the physician. Some centers use computers to assist in the scanning process, and thereby can provide more consistent analysis and more detailed reports.

Our experience at a major Boston hospital is that a large fraction (50% to 60%) of the Holter tape recordings are from patients in the hospital, but not in intensive care units. Holter tape recording and subsequent analysis does provide the physician with the required information but it is both time consuming and expensive. Results are not available until at least a day after the recording is finished (and two days after it was first started). Thus, clinical decision-making is also delayed, leading to longer hospital stays in many cases. Furthermore, dangerous trends in cardiac rhythm are not apparent until long after they have occurred. Multipatient monitors have the disadvantages of high costs and relative inflexibility of location. They are inappropriate for a large proportion of the patient population requiring long-term arrhythmia surveillance. There is a need for a system designed to be small enough and inexpensive enough to be compatible with single-patient monitoring in any bed in the hospital, yet sophisticated enough to provide complete and reliable reports which are available immediately upon demand by the clinician.

## 1.4 Rationale for a Bedside Arrhythmia Monitor

A standalone bedside monitor would have several advantages over conventional multipatient monitoring systems. As mentioned earlier, many hospitals cannot afford the large dollar outlay for installation and maintenance of a multipatient system. An attractive feature of the bedside monitor is that it can be installed with a minimum of procedure and cost. Maintenance can be done by replacement of the entire unit, reducing service costs and downtime.

The portability and low cost of a bedside monitor would encourage its use outside the coronary care unit. This would benefit the many patients who require monitoring to evaluate the efficacy of drug therapy, but who do not require the services of the CCU. Typical patients include those who have recovered from a myocardial infarction, have had heart surgery, or exhibit continued risk of arrhythmias.

Treatment of patients with ventricular arrhythmias can be quite difficult. Individualized drug regimens are often required (Jelinek,1974, Winkle,1976). Drug combinations are often developed in an iterative, trial and error manner. Often, what works for one patient does not work for another. Drug side effects are common, and sometimes serious. Presently, long term "Holter" tape recordings are often used to evaluate drug therapy for hospitalized patients outside the CCU. Skilled technicians are required to scan the tapes, and the results typically are

available to the physician one or two days later. The instantaneous summary displays available with the proposed system should significantly speed up drug trials. Physicians can review results as they make routine bedside rounds, and make appropriate therapeutic changes immediately. Furthermore, dangerous trends in rhythm could be detected immediately, thus increasing the safety of drug trials.

A distributed system of independent monitors would offer greater reliablity than a centralized, multipatient monitor. An inoperative bedside monitor could be easily replaced, while a failure in a centralized system could disable monitoring for all beds.

Finally, issues such as "program efficiency" and "execution speed" figure less prominently in algorithm design for a single patient monitor. Greater attention can be devoted to algorithm robustness, artifact recognition and rejection, and overall reliablity.

## 1.5 Thesis Objectives

The major objective of this research is the development and evaluation of an advanced instrumentation system to improve the diagnosis and treatment of patients with cardiac arrhythmias. The instrument is a self-contained, bedside monitor that performs real-time arrhythmia analysis for hospitalized patients. It provides clinically useful summary statistical and trend

displays, which are immediately available to medical personnel during monitoring. The system automatically documents important arrhythmias by producing hard-copy rhythm strips. The instrument is intended primarily for patients who require monitoring to evaluate the efficacy of drug therapy, but who are not at such high risk as to require the (expensive) services of a coronary care unit.

ECG Analysis Algorithm Development

A major component of the thesis effort was the development of algorithms to detect and classify cardiac events. A major design constraint was that the algorithms had to work in real time in a system whose cost is compatible with single patient monitoring. The current implementation runs on an 8080 microprocessor, and requires approximately 32K bytes of read-only memory (ROM) for program storage and 16K bytes of read-write memory (RAM) for variables and arrays.

The algorithm is designed to analyse a single channel of ECG in real time, and recognize supra-ventricular and ventricular ectopic beats. It employs a novel data compression scheme to reduce processor workload and perform preliminary feature extraction. A "top-down" waveform parser is used to detect and delineate cardiac events, which are subsequently classified using feature extraction and clustering as the basis for morphologic comparison. The algorithm is "self-learning", and can adapt to a variety of background rhythms and QRS complex morphologies in an unsupervised environment.

Central to this effort was the creation of software tools to aid in the development and evaluation of ECG analysis algorithms. Part of this work was the design and implementation of an annotated digital ECG data base that currently contains over 50 half-hour ECG segments. This data base has proven to be an invaluable aid for development and trial testing of analysis algorithms.

Algorithm and System Evaluation

Evaluation of arrhythmia detectors is a difficult task that concerns both system designers and prospective users, yet lacks a generally accepted methodology. Evaluation questions can vary from the microscopic ("How well does the QRS detector work?") to the macroscopic ("What is the chance that the device will mislead the physician?"). A multifaceted evaluation employing three different methods was used to measure performance: (1) a beat-by-beat evaluation of algorithm performance using the annotated ECG data base, (2) an hour-by-hour evaluation using long term (Holter) analog recordings and (3) clinical trials of the prototype instrument.

The "beat-by-beat" evaluation was considered the only method that adequately tests the first stages of the system: the QRS detector, classifier, and overall rhythm classifier. It permitted automatic measurement of detector sensitivity and specificity for QRS complexes, PVC's, SVPB's, couplets, runs, etc. Reproducable experiments were possible that documented the effects of algorithm changes during the development process.

The second phase of the evaluation used long-term (Holter) tape recordings as test data. The reports produced by the bedside monitor were compared hour-by-hour with the previously scanned Holter reports. Although this procedure was not as precise as a beat-by-beat evaluation, it expanded the scope of the evaluation in several areas: (1) it exposed the algorithm to far more data than was available with the annotated data base, (2) it compared the system with a clinically accepted technique (Holter scanning), and (3) it exercised the entire arrhythmia detection system - the detector, classifier, and episode documentation strategy.

Finally, the real-world interface between the system and the clinician was studied to assess the importance of the device in clinical decision making and the confidence of the users in its performance. Although this level of evaluation was more subjective and qualitative than laboratory trials, it was of great importance to see how useful the system was to clinical personnel.

# 2. THE ELECTROCARDIOGRAM

## 2.1 Mechanism of Generation

The normal heart has four pumping chambers, the left and right atria and left and right ventricles. Systemic venous blood enters the right atrium and then the right ventricle via the tricuspid valve. When the right ventricle contracts, blood is pumped into the pulmonary artery via the pulmonic valve. In a similar fashion, returning pulmonary venous blood enters the left atrium and then the left ventricle via the mitral valve. The left ventricle contracts, and forces blood through the aortic valve into the aorta which supplies the systemic circulatory system.

The heart's pumping action is initiated and synchronized by electrical events. There are specialized muscle cells in the heart which are devoted primarily to the generation and propagation of action potentials. These cells share the property of "autorhythmicity", and are found in the sinoatrial (SA) node, the atrioventricular (AV) node, and in the specialized conducting systems of the atria and ventricles (Figure 1). The autorhythmic cells in the sinoatrial node tend to beat the fastest, and normally assume the role of "pacemaker" for the heart. Sympathetic and parasympathetic fibers innervate this region, and may speed up or slow down the rate of impulse formation.

Each contraction of cardiac muscle is preceeded by a wave of electrical excitation known as the "action potential". The action potential generated in the SA node spreads throughout the atria by means of at least three recognizable bundles of conductive tissue. The impulse then reaches the atrioventricular (AV) node, normally the only conducting bridge between the atria and ventricles. The action potential is drastically slowed down in the AV node, effectively introducing a delay of 70-80 msec that allows complete atrial contraction to occur before ventricular contraction begins. After emerging from the AV node, the impulse enters a specialized ventricular conduction system consisting of the common Bundle of His, and the right and left bundle branches. The bundle branches terminate in the richly branched Purkinje network which distributes the action potential to the myocardium. The impulse is transmitted rapidly (2 meters/sec) to the muscle cells on the inner wall of the ventricles (the endocardium). The impulse then spreads more slowly (1 meter/sec) through the heart muscle from cell to cell until the entire ventricular muscle mass is depolarized. After about 200 msec, electrical recovery or "repolarization" of the ventricles commences. Repolarization is not a synchronous process like depolarization, and generally proceeds in a direction opposite to that of depolarization.

Potentials can be measured at the surface of the body which correspond to the major electrical events of the cardiac cycle. These potentials, when plotted as a function of time, comprise the electrocardiogram (ECG). A typical normal ECG is shown in Figure 2. The initial low amplitude wave, termed the "P-wave", is due to the depolarization of the atria. The "QRS complex" corresponds to the depolarization of the ventricles, and the "T-wave", to ventricular repolarization. The amplitude of the QRS complex is typically one millivolt.

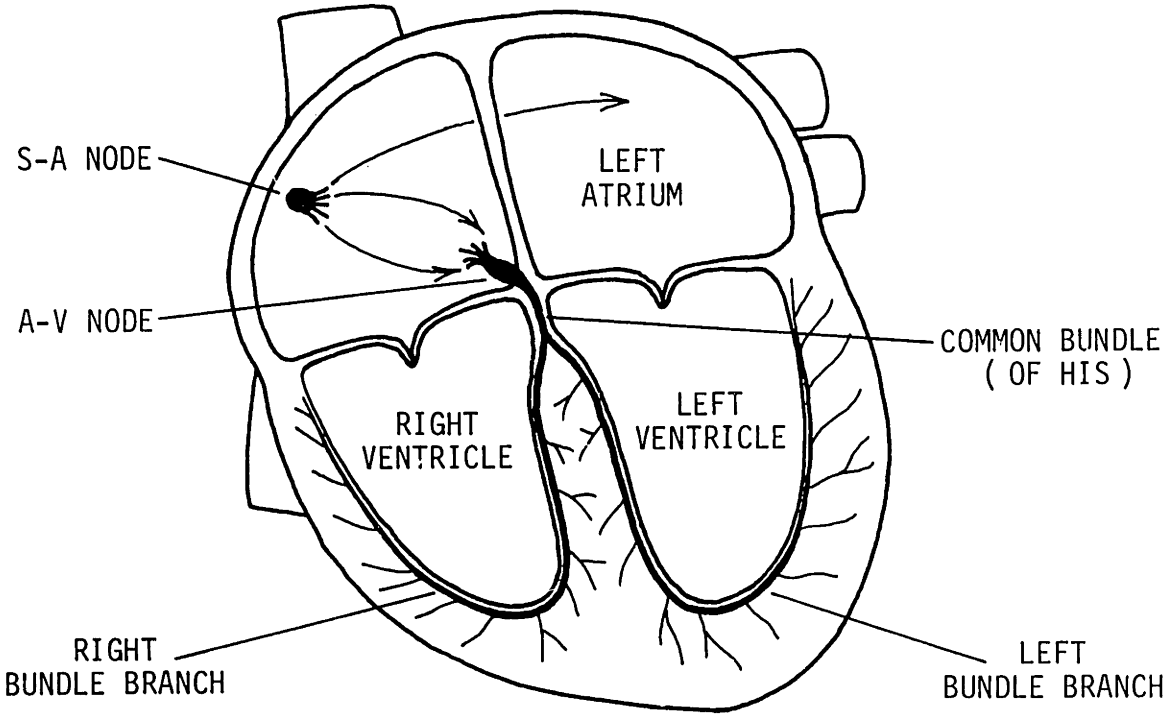



Figure 1: Cardiac electrical conduction system

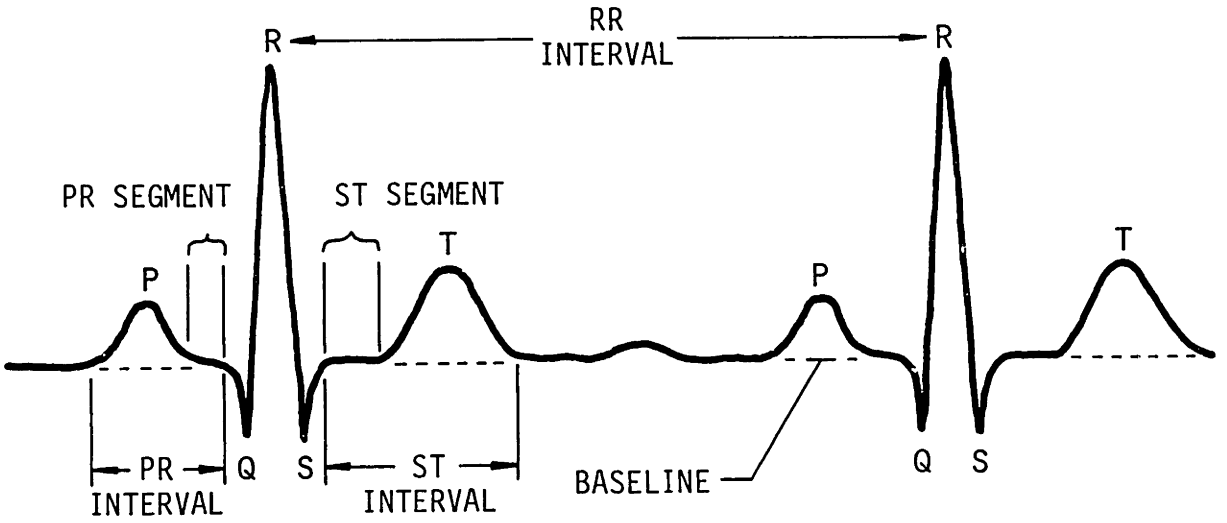



Figure 2: Normal QRS complexes; intervals and segments

## 2.2 Cardiac Rhythms

Disease can directly or indirectly damage portions of the heart's electrical system, and adversely affect the normal electrical and mechanical rhythm of the heart. Diseased tissue can cause abnormal electrical activity due to disturbance of impulse formation and/or conduction. These defects are reflected in the ECG as abnormal P-QRS-T morphology and rhythm. Some arrhythmias cause minimal hemodynamic abnormalities, while others may be life threatening, and in some cases (such as ventricular fibrillation) fatal.

### Sinus Rhythm

The sino-atrial node normally functions as the cardiac pacemaker, and typically establishes a rate of 60 to 100 beats per minute (BPM). This rhythm is termed "normal sinus rhythm", and is illustrated in Figure 3a. Autonomic nervous stimulation, drugs, circulating hormones, hypoxia, etc., can affect the rate of impulse formation. Slower and faster rates are termed "sinus bradycardia" (slower than 60 BPM) and "sinus tachycardia (faster than 100 BPM). A rate which varies, often periodically with respiration, is termed "sinus arrhythmia", and is quite normal, especially in younger people. A momentary lapse of pacemaker activity is termed "sinus pause", and complete cessation is termed "sinus arrest".

Arrhythmias Due to Abnormal Conduction

A conduction disturbance between the sinus node and the surrounding atrial tissue (S-A junction) can delay or block sinus impulses to the atria. The resulting arrhythmia is termed "sino-atrial block". Depending on the degree of block, escape rhythms of atrial, A-V nodal, or ventricular origin take over control of the heart, preventing cardiac arrest.

Abnormal conduction through the AV node or the Purkinje system can cause a number of rhythm disturbances by delaying or blocking atrial pacemaker impulses. In "first degree AV block", impulses are delayed by a long, but constant time. This is manifested in the ECG by a prolonged (greater than 200 msec) P-R interval. In "second degree AV block", some impulses are completely blocked while others are conducted to the ventricles with or without delay. There are two types of second degree A-V block. Type I (Wenckebach) is characterized by progressive prolongation of the P-R interval before a blocked P-wave. In type II, the P-R interval of the conducted beats is of constant duration, but not all beats are conducted.

The refractory period of the AV node also limits the passage rate of impulses to the ventricles. If impulses arrive at the AV node too frequently, some are blocked by the refractory period of the previous impulse. Extremely rapid supraventricular tachycardias may result in "2:1 AV block" (passage of every second impulse), "3:1 AV block" (passage of every third impulse), or "variable AV block".

In "third degree AV block" (or complete heart block), there is no conduction through the AV node, and atria and ventricles beat asynchronously at different rates. Ventricular depolarization may originate at either an AV-nodal or idioventricular pacemaker.

Faulty conduction in the ventricular conduction system can cause abnormal activation of the ventricles. "Bundle branch block" denotes the blockage of an impulse through one or more of the bundle branches. "Left bundle branch block" and "right bundle branch block" are more specific terms indicating the faulty branch. Bundle branch conduction defects usually produce wide QRS complexes with abnormal morphology, since excitation is by slow muscle-to-muscle conduction. The term "aberrant ventricular conduction" is often applied to transient failure of conduction in the bundle branches.

Escape Beats or Rhythms

The SA node is the normal pacemaker by virtue of having the fastest repetition rate of all autorhythmic cells in the heart. Potential pacemakers in the AV node and the Purkinje system (with natural repetition rates of 40-60 BPM and 15-40 BPM, respectively) are suppressed by impulses initiated by the SA node. These lower order pacemakers, however, can serve as a backup when the higher order pacemakers are depressed or blocked. This process is termed an "escape" mechanism. Isolated escape beats are termed "AV nodal (or junctional) escape" and "ventricular escape", depending on the site of the backup

pacemaker. A persistent escape rhythm is termed "AV nodal rhythm" or "idioventricular rhythm". The QRS complex of an idioventricular beat is often wide and bizarre, since excitation is not through the Purkinje system.

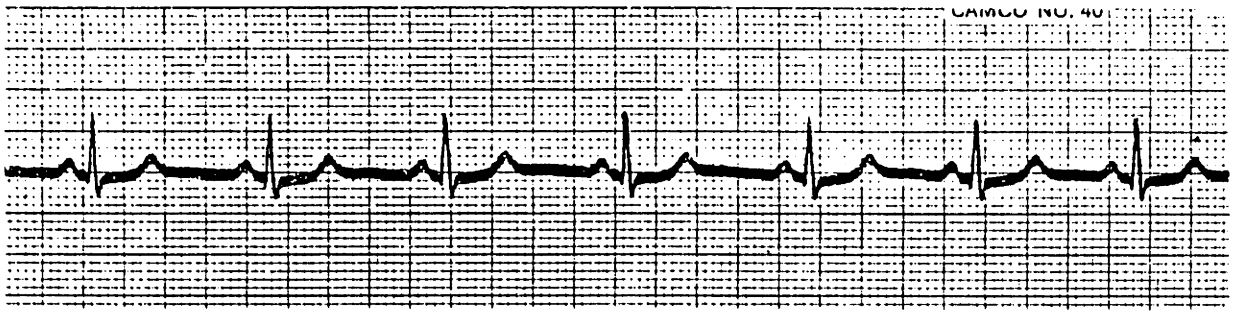


Figure 3a. Normal sinus rhythm. The heart rate is 69, and the PR interval is 0.15 sec.

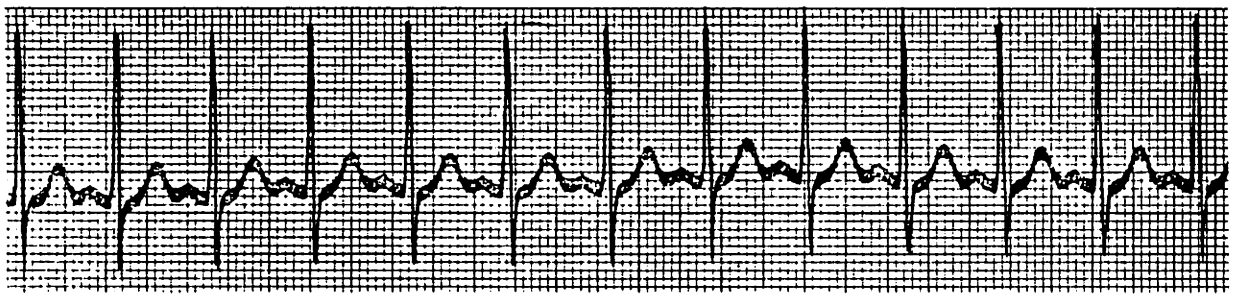


Figure 3b. Sinus tachycardia. The heart rate is 125, and each QRS is preceded by a P-wave.

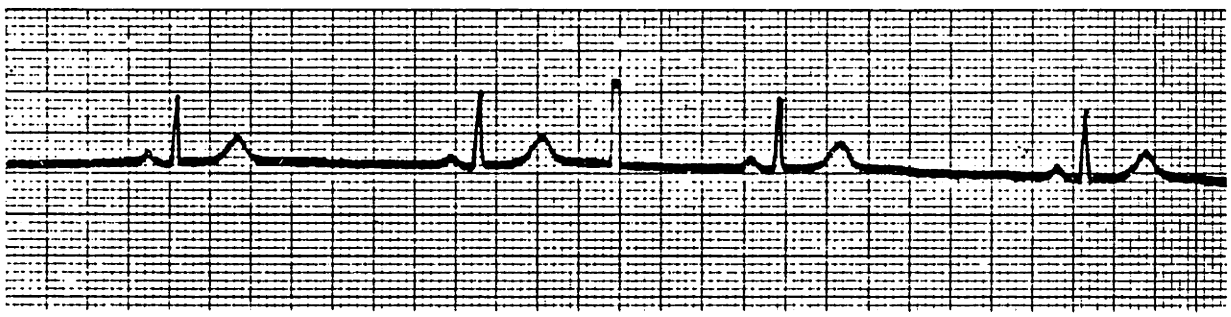


Figure 3c. Sinus bradycardia. The heart rate is 41, and each QRS is preceded by a P-wave.

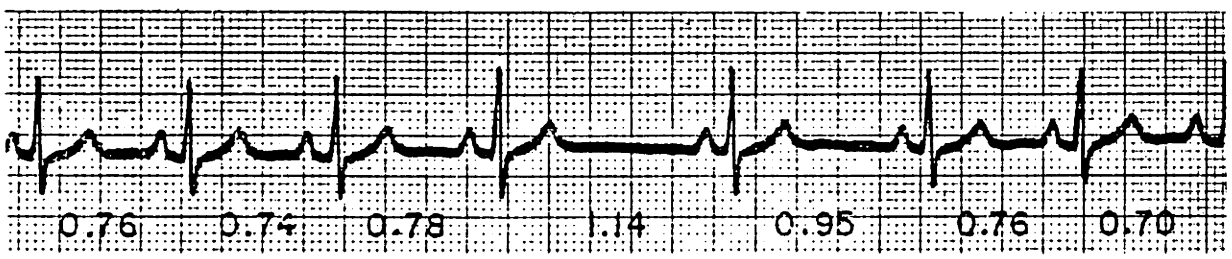



Figure 3d. Sinus Arrhythmia. There is a cyclic variation in heart rate, and constant PR interval.

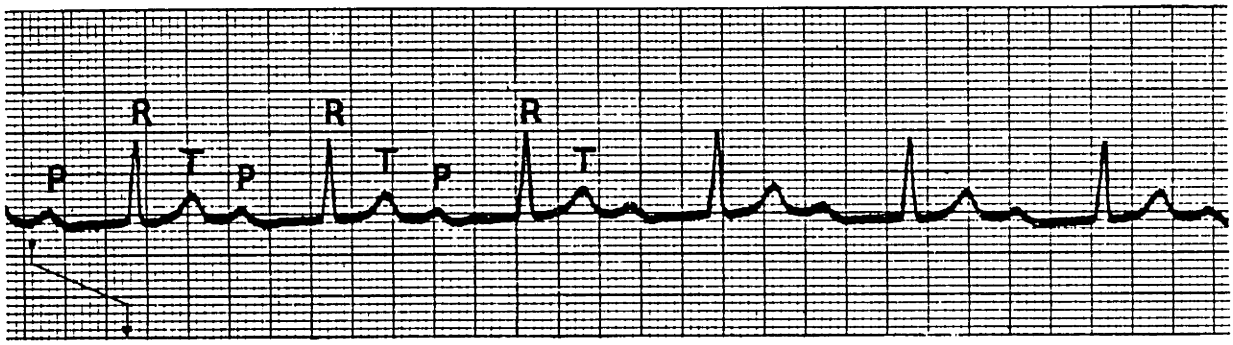



Figure 4a. First degree A-V block. Sinus rhythm at a rate of 66, with prolonged PR interval measuring 0.46 sec.

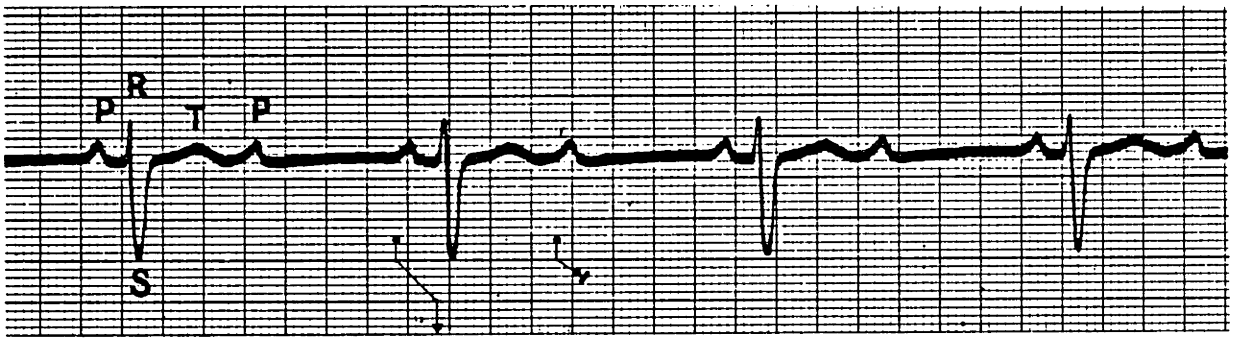



Figure 4b. Second degree A-V block, type II and bundle branch block. Every other sinus beat is conducted ( 2:1 block ).

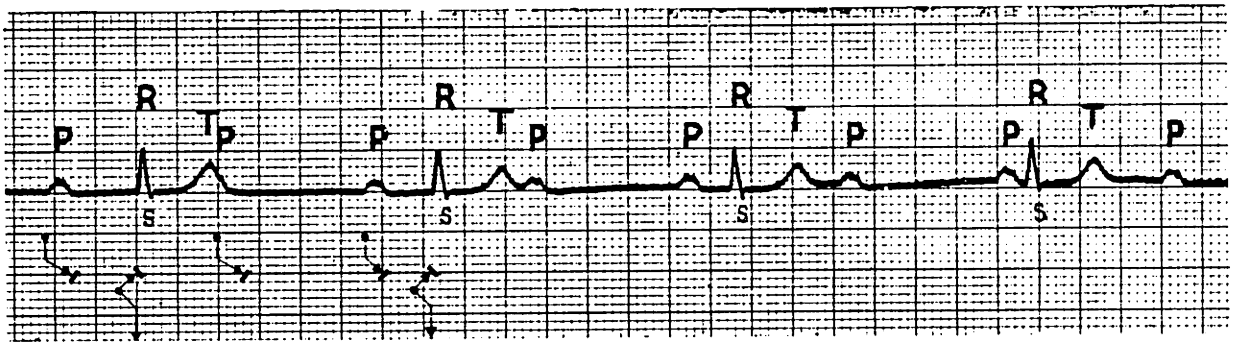



Figure 4c. Complete A-V block with idionodal escape rhythm at a rate of 42.

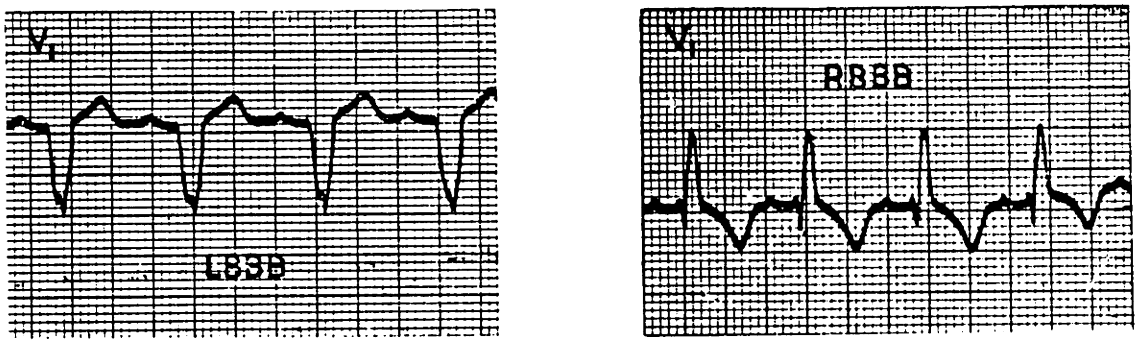



Figure 4d. Left and right bundle branch block.

## 2.3 Ectopic Beats and Rhythms

Pacemaker sites outside the SA node can compete for, or take over, the pacemaking role. These "ectopic" pacemakers prematurely activate the heart, and inhibit the normal activation sequence. Isolated ectopic beats are termed "atrial premature beats" (APBs) "nodal premature beats", or "ventricular premature beats" (VPBs), depending of the site of origin or "focus" of the ectopic beat. Normal and ectopic beats can alternate in a regular fashion, producing rhythms such as "bigeminy" (alternating ectopic and normal), "trigeminy" (ectopic and two normals), etc. Two ectopic beats together form a "couplet". Three or more ectopic beats in a row form a salvo of ectopic tachycardia (either supraventricular or ventricular, depending on ectopic site). Supraventricular premature beats may be abberantly conducted, particularly if they occur with a short coupling interval. These are often difficult to distinguish from VPBs, particularly if the ectopic P-wave is not clearly seen.

### Atrial Ectopic Rhythm

Ectopic mechanisms can lead to sustained arrhythmias. An atrial ectopic pacemaker can produce "paroxysmal atrial tachycardia", a rhythm with a highly regular rate of 160 to 200 BPM, or "atrial flutter", with a yet higher rate of 220 to 350 BPM. Variable or fixed AV block may be present. Hence, the ventricular heart rate may be irregular or regular.

Very rapid atrial activity of a chaotic, non-synchronous nature is termed "atrial fibrillation". This is accompanied by irregular transmission through the AV node, leading to a totally irregular ventricular response at a rate from 50 to 200 BPM. No P-waves are present: only irregular, low-amplitude baseline oscillations. Rapid ventricular rates may lead to decreased cardiac output. This arrhythmia is commonly associated with heart disease.

Nodal Ectopic Rhythms

A-V nodal (or junctional) ectopic rhythms result from the repetitive firing of an ectopic focus located within or in the immediate vicinity of the A-V node. Ectopic A-V nodal impulses may be conducted to the ventricles and retrogradely to the atria, producing inverted P-waves. Enhanced automaticity of the A-V node can result in "nonparoxysmal A-V nodal tachycardia" (60 - 160 bpm) or "paroxysmal A-V nodal tachycardia" (160 - 220 bpm).

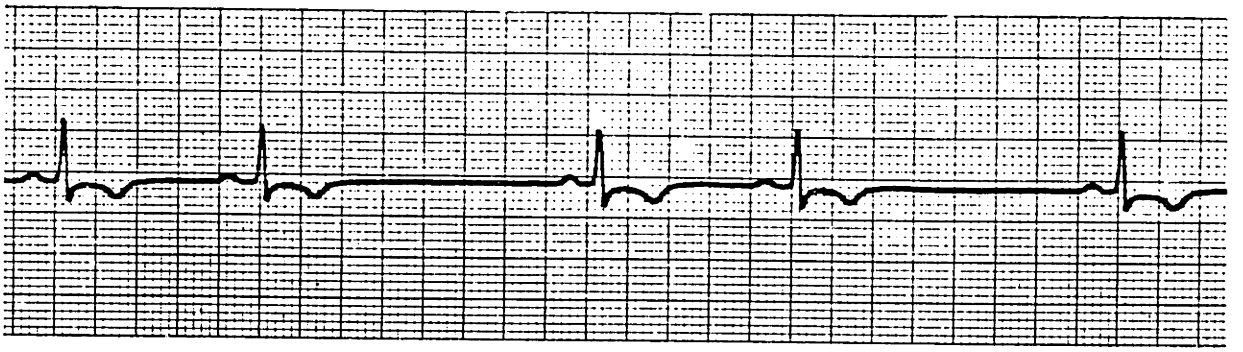


Figure 5a. Atrial premature beats in bigeminy and average ventricular rate of 45.

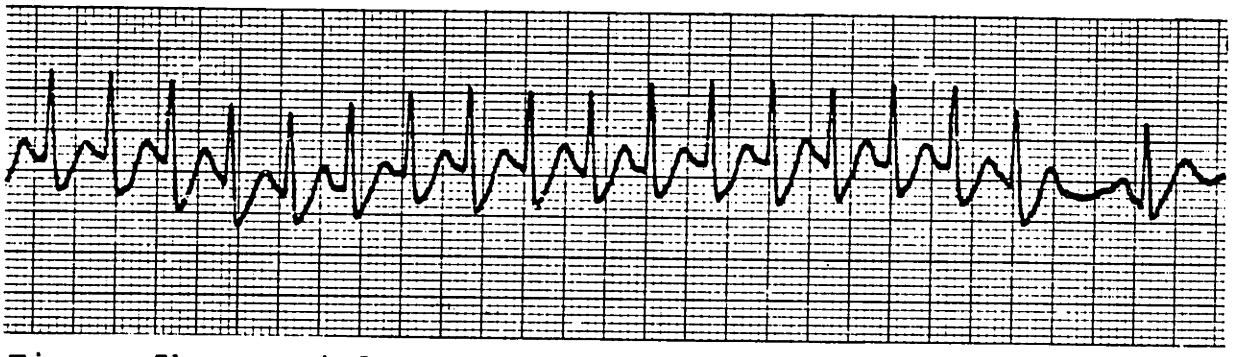


Figure 5b. Atrial tachycardia at a rate of 200, with 1:1 conduction.

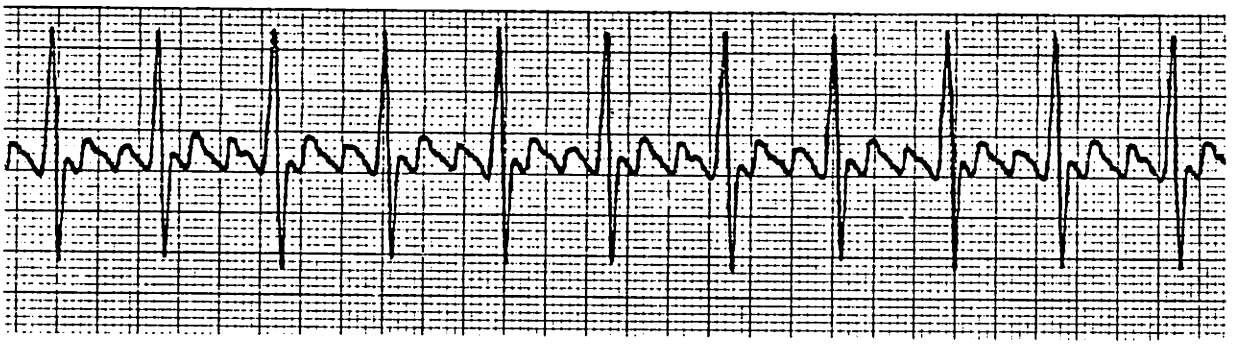


Figure 5c. Atrial flutter with 3:1 A-V block and ventricular rate of 110.

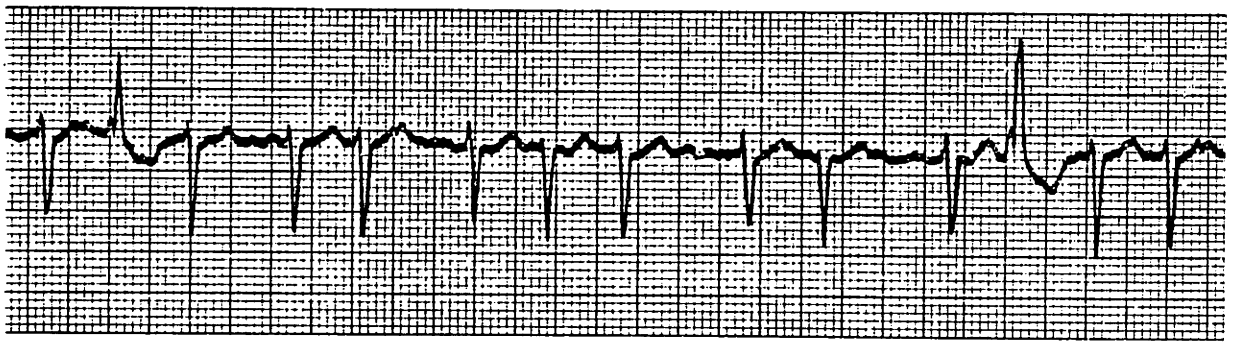


Figure 5d. Atrial fibrillation with average ventricular rate of 130, and two aberrantly conducted beats.

Ventricular Ectopic Rhythm

The "ventricular premature beat" (VPB) usually produces a wide, bizzare shaped QRS complex on the ECG. This is a consequence of the fact that the ectopic depolarization spreads more slowly and in a different manner from the normal activation through the high speed Purkinje system.

Ventricular premature beats may occur in otherwise healthy individuals, and probably have little significance if infrequent. If VPBs are present with heart disease, the outlook is more ominous. There is evidence to suggest that the risk of sudden death is greater in patients who show significant ventricular ectopic activity (Bleifer,1973, DeSoyza,1974). A variety of forms of ventricular ectopic activity may be defined in order of increasing severity: A grading system proposed by Lown (Lown,1971) can be used to rank the severity of ventricular ectopic activity:

Grade 1: Occasional "unifocal" VPBs

Unifocal VPBs originate from the single focus, and exhibit the same morphology. Isolated VPBs usually exhibit a "compensatory pause" -- the VPB inhibits what would have been the next normally conducted beat, but does not alter sinus activity. The interval between the two normal beats surrounding the VPB is twice the normal R-R interval. An "interpolated VPB" occurs early enough so that the next sinus beat is conducted.

Grade 2: Frequent unifocal VPBs

These VPBs originate from the same ectopic site, but are more frequent (more than one per minute or 30 per hour).

Grade 3: "Multiform" VPBs

Multiform VPBs originate from different ectopic sites, and generally exhibit different morphologies and coupling intervals on the ECG.

Grade 4: Couplets, triplets, runs of VPBs, and ventricular tachycardia

This category includes VPBs that occur in pairs, triplets, or short runs of consecutive VPBs.

Grade 5: Early-cycle VPBs

The coupling interval is so short that the VPB falls on the T-wave of the preceeding beat, in the so-called "vulnerable period". The early cycle VPB has a greater probability of inducing repetitive firing and ventricular tachycardia or ventricular fibrillation. For this reason, a VPB with its "R on T" is also called a "malignant VPB".

"Complex" ventricular ectopy includes VPBs in grades 3, 4, or 5, and is felt to be considerably more dangerous than grades 1 or 2.

(6) Ventricular flutter and fibrillation

Ventricular flutter is caused by rapid, repetitive firing of one or more ectopic foci at a rate between 150 and 300 bpm, and with a fairly regular rhythm. Ventricular fibrillation

results from very rapid chaotic firing of multiple ectopic foci.

There is no coherent electrical activity during ventricular fibrillation-- the ECG shows an irregular baseline oscillation with no QRS complexes (Figure 7d). The exposed heart quivers with fine fibrillatory waves (very much like a "bag of worms"), and is incapable of pumping blood. This is a fatal arrhythmia if allowed to persist. Electrical countershock together with other resuscitation measures is required to restore normal cardiac activity.

Currently, there are two possible explanations for the development of ventricular ectopic beats; one relating to increased automaticity and the second based on reentry (Han,1969). In the former theory, an ectopic site with increased automaticity is protected against being reset by conducted activity. The focus generates impulses at its own independent rate, unaffected by the dominant cardiac rhythm. Only those impulses that fall outside the refractory period induced by the dominant rhythm can spread to the surrounding myocardium and produce ectopic beats. These beats are termed "parasystolic", and do not exhibit a fixed coupling interval.

Reentry is made possible by disparate repolarization rates of different ventricular regions, probably brought on by mixed degrees of cellular ischemia. This dispersion of repolarization between cells may lead to sufficient voltage differences to

re-excite the tissue and propagate beats. Another possible mechanism of reentry is based on some regions exhibiting "unidirectional block" which can be retriggered later in the cardiac cycle. These beats are coupled to the dominant cardiac rhythm.

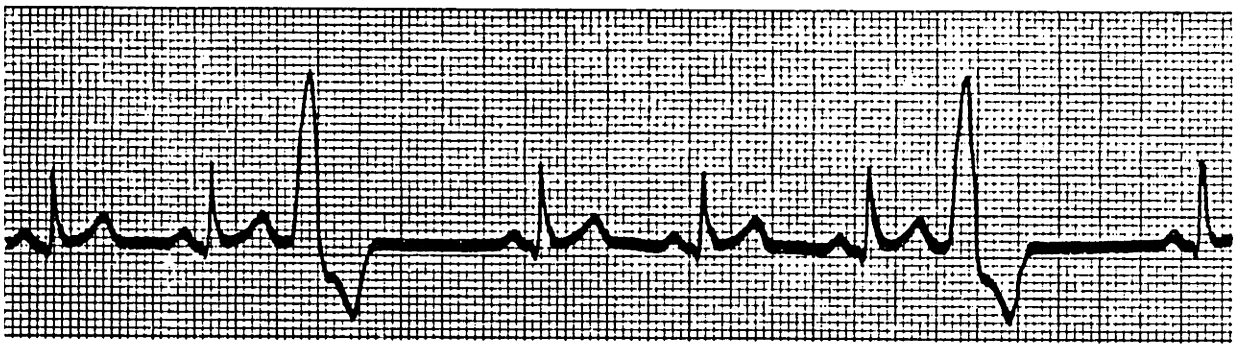


Figure 6a. Unifocal ventricular premature contractions.

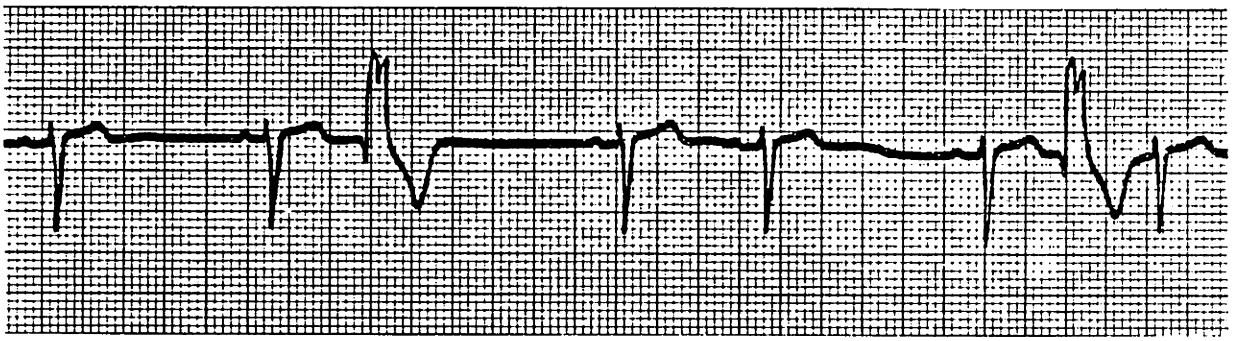


Figure 6b. Ventricular premature contractions
(with compensatory pause and interpolated VPC).

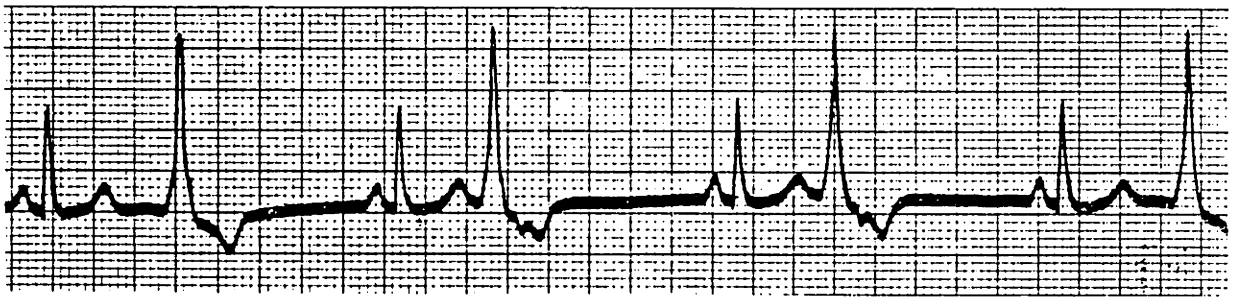


Figure 6c. Unifocal VPCs occurring in bigeminy

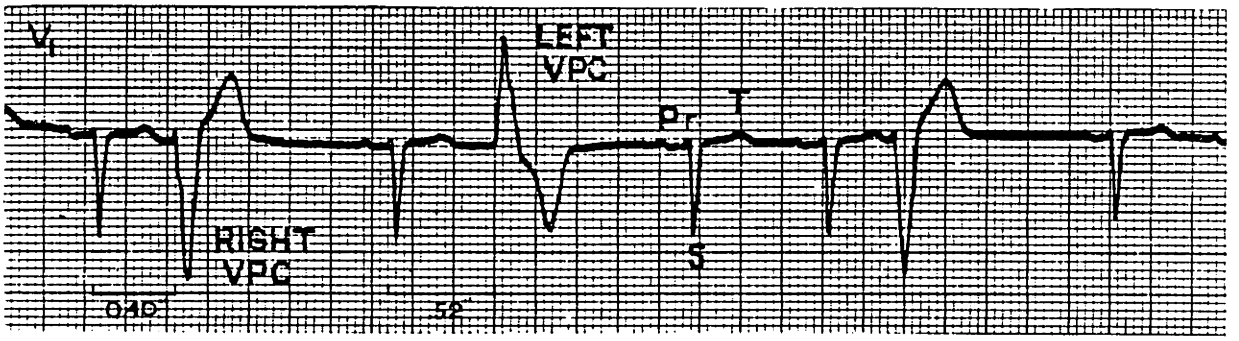



Figure 6d. Multifocal ventricular premature contractions

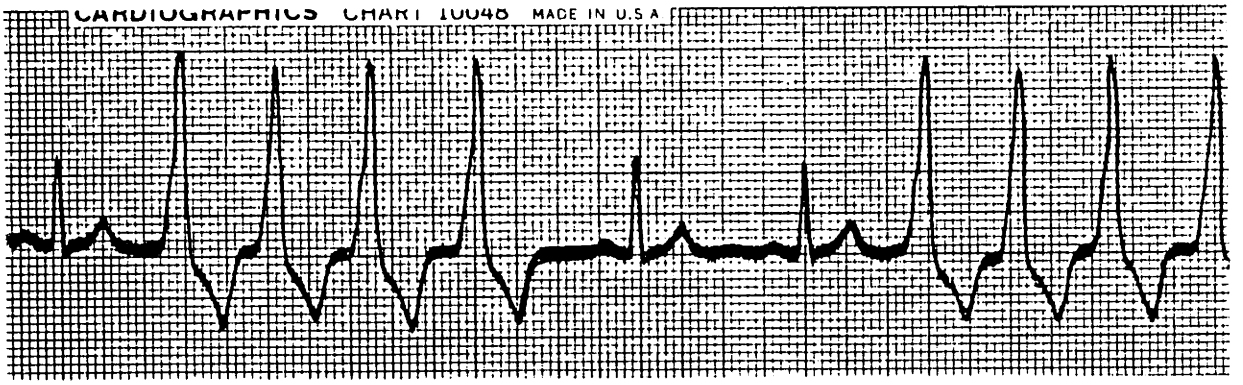



Figure 7a. Short burst of ventricular tachycardia.

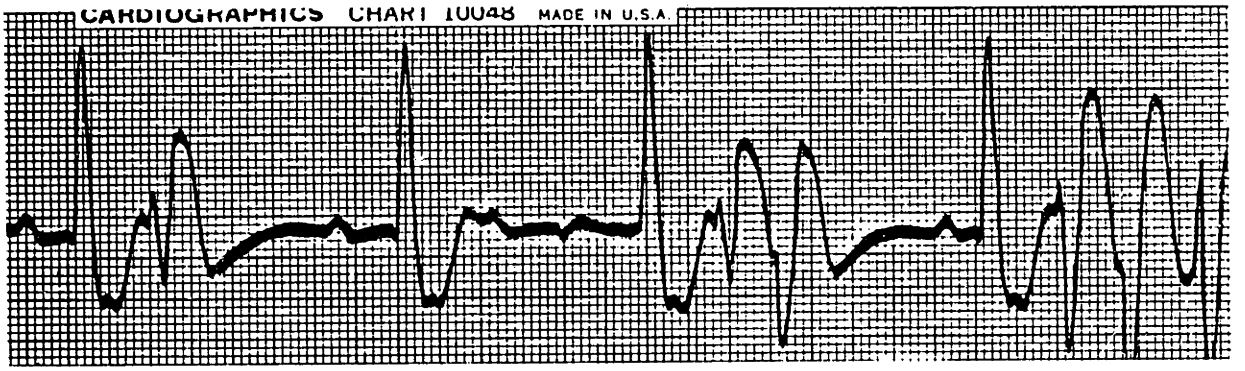



Figure 7b. Early coupling of VPBs and beginning of episode of ventricular tachycardia.

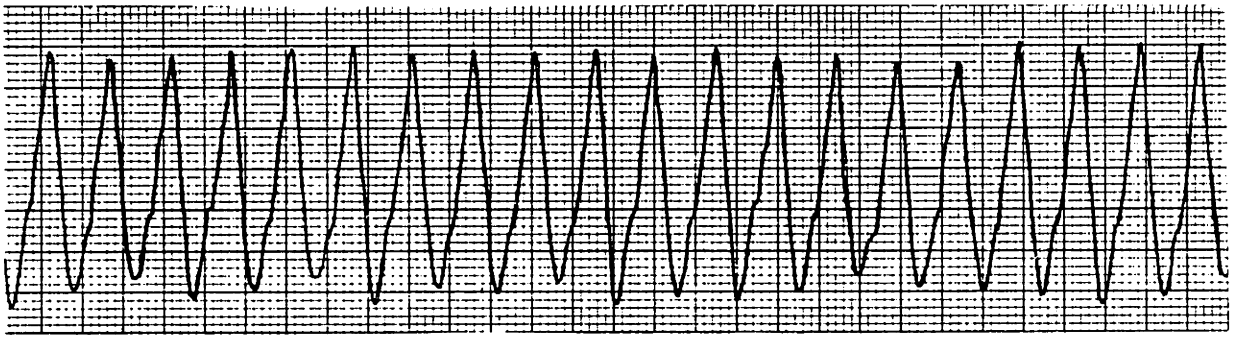


Figure 7c. Ventricular flutter at a rate of 200.

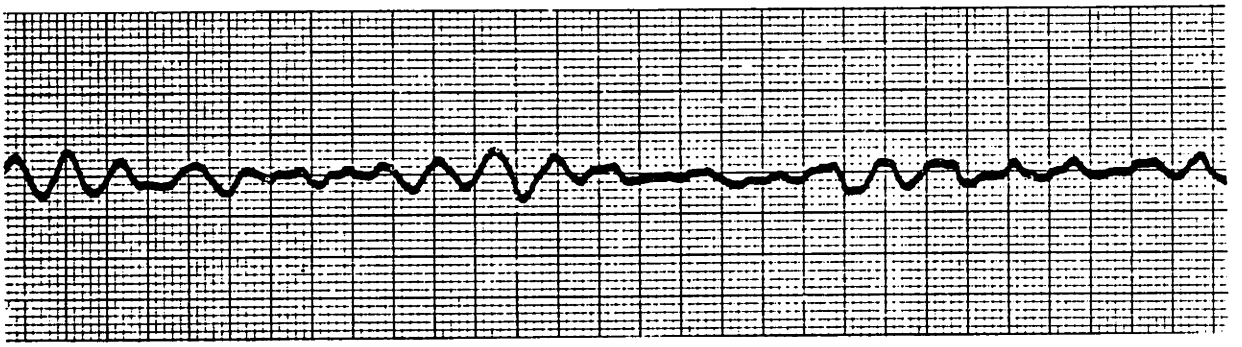


Figure 7d. Ventricular fibrillation

# 3. REVIEW OF AUTOMATED ARRHYTHMIA ANALYSIS

## 3.1 Overview

Automated monitoring of ECG arrhythmias has been a popular pursuit of biomedical engineers for well over a decade. Early systems employed analog techniques to detect the QRS complex, and only displayed the instantaneous heart rate and generated alarms when the rate was outside presettable limits. Hybrid systems employing analog QRS detection and a digital computer for data collection and display have also been developed. Currently, most investigators prefer to retain algorithmic control over QRS detection and classification, and rely on analog preprocessing only for simple filtering.

This review will cover primarily computer based real-time arrhythmia analysis techniques and algorithms. Less attention will be given to analog or hybrid arrhythmia monitoring systems, since their performance is limited by the rather simple algorithms embodied in hardware. "Heart station" rhythm diagnostic systems that perform an off-line analysis of a short (several second) ECG strip are only briefly described. Although important work has been done with heart station rhythm diagnosis, these systems enjoy a number of advantages usually not available to real-time monitoring systems:

1) access to multiple and sometimes simultaneous ECG leads,

2) signal quality control by human observers for each record, and

3) absence of real-time processing constraints.

Real-time monitoring systems, however, have at best infrequent human quality control over the ECG, and require special techniques to stay abreast of the incoming data.

There are several excellent articles that review automated ECG analysis. Many of the digital processing techniques used to analyse the electrocardiogram (and other biological signals) are reviewed in (Cox,1972), and several ECG interpretation and rhythm analysis systems are described in (LeBlanc,1973). More recent reviews (Thomas,1979, Oliver,1977, and Feldman,1977) discuss real-time and high-speed (Holter) arrhythmia analysis techniques, and their evaluation. A detailed discussion of early arrhythmia monitoring systems is presented in (Nolle,1972). Also, the conference proceedings Computers in Cardiology (years 1974 to 1981+) presents much of the current work done in applying computers to cardiovascular research.

## 3.2 Overall Structure of Computer Arrhythmia Analysis Systems

Most ECG analysis systems can be viewed as a cascade of processing stages. In the first stage, the ECG is amplified, filtered, and digitized. Most systems monitor a single ECG lead, although some use two or three. Typical sampling rates range from 120 to 500 samples per second, with 8 to 12-bit resolution. In some systems, the sample sequence is compressed to reduce the workload on subsequent processing stages and reduce storage requirements.

Common to all systems, however, is the detection of the QRS complex with a suitably defined amplitude and slope test. Search procedures of varying complexity, generally initiated by QRS detection, are used for detecting P and T waves (most systems do not attempt to detect the P-wave, however, since it is often obscured by noise). Waveform boundary delimitation may also be performed at this point to establish major features of the detected complexes. Waveform delimitation is a non-trivial task since it is often difficult to exclude artifact or portions of a different proximal wave.

Classification of ECG waveforms is usually based on the morphology and timing of the event. For example, ventricular ectopic beats are usually premature, wider than normal complexes, and have different morphology. The morphology of a waveform can be characterized in one of two ways: (1) as a time-ordered sequence (of sampled data, segment sequences, etc.), or (2) time-independent features (envelope measures such as duration, amplitude, polarity, area, maximum derivative, and other features extracted from the ECG waveform). Correlation with a stored normal beat is one technique that uses a time-ordered representation of the ECG. Feature extraction and clustering algorithms are an example of the latter. Some systems automatically determine the "normal" morphology for the patient based on timing, width, and relative cluster population, while others are "taught" the normal morphology by the operator.

After the morphology and timing of the QRS complex have been identified, the overall rhythm of the ECG must be determined. Most systems use simple beat-serial tests to identify ventricular rhythms such as bigeminy from the sequence of beat labels generated by the classifier. A number of systems, however, statistically analyse the RR interval sequence to identify atrial rhythms without relying on accurate P-wave detection (e.g. atrial fibrillation).

The final output (a stream of beat and rhythm labels) is summarized in trend plots of heart rate, PVC rate, SVPB rate, etc., and can be combined with trends of other physiologic variables (respiration, temperature, blood pressure) and trends that document drug dosage. Most systems generate alarms when life threatening events occur, and automatically activate a strip chart recorder to provide documentation. Some systems with disk storage allow the user to "edit" and correct the computer's mistakes, and continuously record the ECG for later review.

Signal quality control is a problem that plagues all real-time monitoring systems. Baseline shift due to poor electrode contact, muscle artifact, and powerline interference frequently corrupt the incoming ECG. Techniques that range from simple analog and digital filtering to complex strategies that analyse the artifact's "morphology" have been used to reduce errors. In addition, the patient's normal rhythm can make analysis very difficult. For example, aberrated supraventricular beats in a background rhythm of atrial fibrillation are

practically impossible to distinguish from PVCs. There are no set standards of prematurity and morphology that can be used to recognize PVCs: every patient is different, and requires that an analysis algorithm adapt to individual situations.

## 3.3 ECG Aquisition and Preprocessing

A single ECG "lead" (a pair of potential sensing electrodes) is commonly used for arrhythmia monitoring. Surface electrodes are the least invasive and subject the patient to minimal risk and discomfort. Special electrodes can be employed, however, if more detailed information is required. Atrial depolarization is usually difficult to sense with surface electrodes because the P-waves are often low in amplitude and variable in form. Esophageal electrodes can be swallowed by the patient and positioned in the esophagus near the atria to record large amplitude P-waves, and permit less ambiguous determination of rhythm. One researcher (Jenkins,1978) has developed a convenient form of this electrode suitable for long-term use.

Electrode placement can have a profound effect on the ECG signal quality. Skeletal muscle artifact has frequency components that overlap those of the ECG, and can be minimized by not placing electrodes over large muscles. Variations in the position of the heart and conductivity of the lungs (which occur when the patient moves or breaths) can cause variations in the surface ECG that can make analysis difficult. These effects can be minimized by selecting a lead vector that is less sensitive to

positional changes of the heart. Careful lead placement can also enhance the separability of PVC morphology from the patient's normal QRS. Electrode-coupling artifact can arise from poor contact with the skin, and can give rise to ECG like waveforms when the patient moves about. A number of techniques, including dual contact electrodes and specialized front-end processing (Feldman,1979) have been used to detect such artifacts and report loose leads.

The next step is to amplify the ECG, as the largest signal component (the QRS complex) has a typical amplitude of one millivolt. An electrically isolated amplifier is preferred for obvious safety reasons. AC-coupling is required at an intermediate amplification stage to remove DC electrode offset potentials, which can be as large as 300 mv. General recommendations for ECG data aquisition have been proposed, most notably those by the Committee on Electrocardiography of the American Heart Association (AHA,1975).

Practically all arrhythmia monitoring systems employ some sort of analog or digital filtering of the ECG. There are several additive noise sources that can corrupt the ECG: (1) power-line interference at multiples of 60 Hz, (2) muscle artifact, with a power spectrum that overlaps that of the ECG, and (3) baseline and motion artifact, with frequency components below several Hertz. Most of the energy of the QRS complex is within a two to twenty Hertz band. A bandpass filter with a similar passband suffices for detection purposes, but can remove

possibly important morphology information. The American Heart Association recommends a 100 Hz cutoff frequency for diagnostic quality ECGs. This admits too much noise for most arrhythmia monitoring systems, however, and typically a passband of 0.1 Hz to 50 Hz is used by systems that use waveform morphology for classification. Such a filter preserves ST-segment level and T-wave morphology, and has a sufficiently high frequency response so that waveform timing measurements are not distorted. Many systems also employ an analog or digital 60 Hz notch filter to reduce power line interference.

Reported ECG sampling rates range from 60 to 1000 samples per second, and resolution range of 8 to 12 bits. Most systems sample at 120 to 500 samples per second, using 10 bit resolution. In general, feature extraction and clustering systems use higher sampling rates to minimize temporal error in extracted features. Systems employing correlation for morphology comparison generally use lower sampling frequencies.

## 3.4 Data Compression Techniques

Several arrhythmia analysis systems compress the incoming ECG (Cox,1969, Sanders,1974, Yanowitz,1974). Data compression reduces the workload on subsequent processing stages, and reduces storage requirements. In some instances, data compression can be viewed as an initial feature extraction stage where the raw data is rendered into a more easily and accurately parsable representation. A fast, flexible preprocessor capable of

deleting features due to noise, yet retain information concerning the gross shape of the waveform is required.

One popular compression technique, termed AZTEC (Amplitude-Zone-Time-Epoch Coding) has been used in several ECG arrhythmia analysis systems. This technique (Cox,1968) compresses the data in two steps. The first, termed a zero-order interpolator (ZOI) performs a hysteresis smoothing of the data. The algorithm works as follows: an upper and lower bound Vmax and Vmin are set to an initial sample value Vo. Vmax and Vmin are adjusted to contain subsequent samples, as long as (Vmax - Vmin) is less than an experimentally determined aperture. When a sample necessitates separating the limits by more than the aperture, a new ZOI segment is started, and the duration and amplitude (Vmax - Vmin)/2 is stored. The final result of the ZOI is a sequence of flat segments with variable duration that approximates the original ECG. The ZOI compressor has been used (Walters,1976) in a microprocessor arrhythmia monitor for ambulatory subjects. The principal virtues of the ZOI are its computational simplicity and efficiency, and typical compression ratios of 5:1 to 10:1 for an ECG like waveform.

A signal with large amplitude and derivative such as the QRS will produce a series of short ZOI segments. Groups of consecutive ZOI segments of short duration that have monotonically increasing or decreasing values can be represented by a "slope". The final compressed stream of data is a sequence of ZOI segments and slopes of variable duration. This algorithm,

termed AZTEC (Cox,1968), has been widely used to compress biological signals. It is computationally simple, achieves reasonable compression, and provides a simple representation of the compressed data. However, the loss of low-voltage low-frequency events limits the clinical acceptance of AZTEC reduced data, and thus the raw data are usually saved for display.

The "split and merge" algorithm (Pavlidis,1973, Pavlidis,1974) can be used to partition and approximate the ECG with a series of first order (linear) segments. The waveform is initially divided into an arbitrary number of segments with coincident endpoints, and for each one, a linear approximation is found minimizing the orthogonal integral square error. Each iteration of the algorithm consists of three steps:

1) splitting segments with large error until the sum of segment errors (the total error) is brought below a given error tolerance,

2) merging adjacent segments which yield a small increase in total error, and

3) minimizing the total error by adjusting the segment endpoints, with the number of segments fixed by steps 2 and 3 (Pavlidis,1973).

Although this is a computationally demanding task, it partitions the data according to absolute voltage levels, rather than derivatives, making it somewhat more immune to noise.

Other techniques, more suited for data compaction rather than feature extraction, have been used to compress the ECG to reduce storage or data transmission requirements. In a clinical data aquisition system (Ripley,1976), the second differences of the ECG were encoded using a truncated Huffman code. This technique permitted exact reproduction of the ECG, using an average of 4 bits for every sample (10-bit samples, 250 Hz). In another effort (Ruttiman,1976), the error of a second-order linear predictor was Huffman encoded in an effort to reduce telephone transmission requirements.

A "turning point compressor" (Mueller,1978) has been developed to reduce the amount of storage required to save rhythm strips in an ambulatory patient monitor. The algorithm simply replaces an epoch of N samples with an extremum value, or the last value if the samples are monotonic in the epoch. The algorithm preserves major amplitude features of the ECG, and produces clinically acceptable strips at output rates of 50 to 100 Hz.

## 3.5 Detection of ECG Waveforms

Reliable detection and delineation of the QRS complex, and to a lesser extent the P-wave and T-wave, is the central problem of arrhythmia analysis. The subsequent morphology and rhythm classification stages are often at the mercy of the event detector. At the present time, most algorithms detect the QRS complex reliably. Search procedures of varying complexity,

generally initiated by QRS detection, are used to detect P and T-waves. The P-wave presents special difficulties, however, since it is often obscured by noise or other large ECG events. Specific detection algorithms for other waves such as flutter, fibrillation, and artificial pacemaker spikes have also been attempted.

Detection of the QRS complex

The detection of the QRS complex is a common element of all arrhythmia analysis systems. QRS detection may be done on the analog signal, raw digital data, first or second difference of the digital data, or the compressed version of the original signal. This is usually accomplished with suitably defined amplitude and slope tests based on fixed, adaptive, or user adjustable thresholds.

Large P and T-waves can present significant false detection problems. Frequently, post QRS detection lockout intervals are used to prevent false detection of T-waves as QRS complexes. However, such a strategy carries the risk of missing clinically important "early-cycle" (R-on-T) PVCs. Absolute amplitude and slope measures are inadequate to reject T-waves as QRS complexes. In some instances, a patient may have T-waves that are as large, or larger than the normal QRS complex. Likewise, ectopic beats may have slopes less than the slope of the T-wave. This slope threshold overlap imposes requirements for a joint decision based on amplitude, slope, and temporal relationships to correctly detect the QRS complex and reject P and T-waves

(McClelland,1976). The same system also employs digital filtering to enhance the high frequency QRS complex and suppress P and T-waves.

False positive QRS detections due to T-waves can also be reduced by comparing the morphology of the candidate QRS complex with the T-waves of previous QRS complexes. One high-speed Holter tape analysis system allows the operator to relabel PVC classes as "T-wave" classes, and all events with similar coupling intervals and features which are labeled PVCs were automatically relabeled "T-wave" (Mead,1975). In another system (Dillman,1978), a tentative beat is called a T-wave if it is within 350 msec of its predecessor and correlates with the predecessor's family's T-wave, but does not correlate with the predecessor family's QRS. An early beat is declared R-on-T if it fails to correlate with the T-wave of its predecessor's family.

Detection of the QRS complex becomes difficult in the presence of artifact that mimics a QRS-like shape. One system (Arnold,1975) requires that a complex present in a moderate amount of noise have been "seen" earlier, before it is classified as an ectopic beat. This reduces the number of artifacts that are detected as QRS complexes from being mis-classified as ectopic beats, since it is unlikely that one instance of artifact will be similar to another instance of artifact.

Detection of Atrial Activity

The detection of atrial activity is the most important limiting factor of current arrhythmia analysis programs. Although the P-wave is visually recognizable in most cases, it is often of low amplitude and barely distinguishable from background noise. Furthermore, since the timing of the P-wave relative to the QRS complex cannot be assumed a priori, a systematic search for a possible P-wave must cover the entire RR interval. It can be obscured entirely by the QRS complex or T-wave. Most programs that "detect" the P-wave search for a "bump-like" waveform in the vicinity of the QRS complex. One approach (Gustafson,1979) attempts P-wave recognition by first detecting and then removing the QRST complexes from the signal, followed by searches for possible P-waves on the QRST-free waveform. This technique, however, requires a relatively noise free ECG (which is often not the case in clinical practice).

Probably the most promising approach is to enhance the P-wave with a special lead system. In one system (Bernard,1974) an intra-atrial electrode and a peripheral electrode were used for arrhythmia monitoring. The system generates PP, PQ, and QQ intervals which are used to provide continuous on-line rhythm analysis. Unfortunately, venous catheterization is required. In a more recent investigation (Jenkins,1976), an esophageal lead consisting of a swallowable pill-electrode was developed and incorporated into an arrhythmia monitoring system.

Detection of Ventricular Fibrillation

Ventricular fibrillation (VF) is the asynchronous depolarization of the heart muscle fibers, and produces a signal varying in shape from a sinusoid to something resembling low-frequency baseline artifact. A number of systems utilize frequency domain information to identify VF. One system (Nygards,1977) estimates the power spectrum using a FFT, and declares VF if most of the waveform energy is between 2 and 9 Hz. Another approach (Kuo,1978) employs a time-domain approach to test for the likelihood of VF. Both systems require the absence of normal complexes to screen for false positives, and initial results look encouraging.

## 3.6 Delineation of ECG Waveforms

After the QRS has been detected, it is often necessary to delineate it to identify important features. Most feature extraction algorithms, for example, require accurate determination of QRS onset and termination so that features such as width and area of the QRS complex can be computed. Systems that measure parameters such as ST level also require accurate and repeatable delineation of the QRS. On the other hand, if correlation or linear transforms are used, calculation of a reliable fiducial is necessary.

Many delineation techniques have been tried, and hint at the difficulty of this processing task. Simpler strategies include backward and foward searches around the QRS detection point for baseline, or conversely, the absence of high-slope sections of the QRS complex. A more sophisticated approach (Mead,1979) builds up the QRS complex out of compressed AZTEC data by computing a "vector sum" as each segment is added to the candidate complex. Delineation of a given complex is linked to minimizing the vector sum, which favors a complex with equal onset and offset voltage values. Advanced work on rhythm diagnosis for the 12-lead ECG described by Bonner and Schwetman (Bonner,1968a) illustrate the potential complexity of the delineation task. Their system compressed the incoming data as a sequence of straight line segments. Adjacent line segments were then grouped together into simple upright, inverted, or tilted V or U shapes, and the overlapping shapes were scanned repeatedly to detect and measure ECG waves.

A more general and formal algorithm for waveform detection and characterization based on parsing techniques was explored by Horowitz (Horowitz,1975). The raw ECG data is represented as a sequence of straight line segments. Line segments are assigned tokens SLOPEi and BASEi that specify the slope (positive, negative, zero) and relationship to a fixed baseline. Context-free grammars (in Backus-Naur form) are constructed for strings that specify productions relating to peaks. The procedure generates a set of delineated ECG subwaveforms.

Event detection and delineation is a very difficult component of automated arrhythmia analysis, and can have a significant impact on overall system performance. In the ARGUS2H high-speed Holter tape analysis system (Mead,1979), greater than 95% of all false positives were traced to errors in either delineation or detection. In our work (Schluter,1980), over 25% of the false positive PVCs are due to event detector error. Many factors contribute to difficulty in these two stages. Muscle noise proximal to the QRS is often "added" to the delineated complex. Also, there is a wide variety of "normal" waveforms that make it difficult to devise a universal definition of what a real QRS looks like (and what does not).

## 3.7 QRS Shape Description and Morphology Classification

Once the QRS complex has been detected and delineated, its morphology and timing must be characterized for final classification. The major purpose of morphological classification is to distinguish beats with "normal" morphology from those of ectopic origin (most VPBs have wide, bizzare-shaped QRS complexes due to the abnormal temporo-spatial spread of depolarization through the myocardium). This information, combined with timing information, is used to determine the final classification of the beat and overall rhythm. Bigeminy, for example, is characterized by alternating beats of normal and ectopic morphologies, with the ectopic beats preceded by RR

intervals shorter than those preceding the normal beats.

Morphological characterization has other uses as well. In monitoring situations in which multifocal PVCs are observed, the set of ectopic beats can be subdivided into classes (according to the presumed site of origin) based on morphology. Morphological analysis has also been used to reduce false detection of T-waves as PVCs (Mead,1975, Dillman,1978), and also to reject artifacts that are non-physiological in nature (Arnold,1975).

Most absolute measures of QRS morphology have little value in rhythm monitoring, due to variation among patients and ECG lead placement. Perhaps the most useful absolute measure is the duration of the QRS complex, since the spread of depolarization from an ectopic focus is generally slower than for a normally conducted beat. However, a number of conditions preclude duration as a totally reliable morphology measure. The apparent duration of the QRS complex can be erroneously lengthened by noise that mimics Q and S-waves. The presence of bundle-branch block or other intraventricular conduction defects also results in wide "normal" beats. In other instances, the QRS of a VPB may appear to be narrow because its initial or terminal portions are isoelectric. Despite these difficulties, practically all monitoring systems rely to some extent on QRS width in deciding whether or not a beat is ectopic.

However, many monitoring systems employ higher dimensional descriptors of QRS shape. There are two widely used techniques of QRS morphological description which have been used in contemporary monitoring systems.

(1) as a time-ordered sequence (of sampled data, first differences, extrema, segment sequences, etc.). Correlation with a stored normal beat is one commonly used technique employing this type of waveshape comparison.

(2) time-independent (envelope measures such as duration, amplitude, polarity, area, maximum derivative, etc.). Most feature extraction and clustering algorithms fit in this category.

A third method of extracting features is to apply a linear transformation that maps the original sequence of time samples into a lower dimensional subspace. These techniques are discussed in greater detail in the following sections.

Correlation Coefficient

The cross-correlation coefficient is a popular measure for stored normal QRS morphology comparison. In one system (Feldman,1971, Hubelbank,1978), a 300 ms segment of 200 samples/s data centered about the QRS detection point is compared with a stored normal QRS by computing the cross-correlation coefficient. The waveform was adjudged abnormal if the correlation coefficient was below 0.875 or (2*Pavg)-1, where Pavg is a running average of the correlation coefficient over the last five beats. Beats classified as abnormal were also correlated with the first beat

which was premature and followed by a compensatory pause. The purpose of the latter test is to discover multifocal VPBs. Also, the computer operator is allowed selection of VPB classification criteria, as well as R-wave detection thresholds and other program parameters.

In a commercial multi-patient monitoring system manufactured by Electrodyne (Arnold,1975, McClelland,1976, Shah,1977), cross-correlation is used to group beats into morphological families, as well as to identify abnormal beats. If the candidate complex does not match any template with a sufficiently high correlation coefficient, a new "family" is created. The classification of an individual complex is determined by the classification of the template that it matches. The classification of each family template is based on the width of the template, and average prematurity of beats belonging to the family. Since the width measurement is based on smoothed average of at least two complexes, noise is reduced with a subsequent reduction of errors in width measurement. The use of average prematurity rather than individual prematurity of each beat minimizes classification error during atrial fibrillation and sinus arrhythmia.

Other efforts that have used correlation are described in papers by (Okajima,1963, Wigertz,1964, Balm,1967, Berni,1975, Spitz,1977, Dillman,1978, and Gradman,1978).

Cross-correlation, however, has a number of limitations as a measure of waveform similarity. The magnitude of the coefficient is very sensitive to the time alignment of the reference and test waveform. In one investigation (Nolle,1972, Appendix 6.5, Cox,1972, pg. 1156), it was found that a time skew of only 4 msec caused the coefficient to drop from unity to 0.9 for one patient's QRS when correlated with itself. In other records, VPBs had high peak correlations with the normal references, reaching a value of 0.9. Despite the apparent simplicity and ease of implementation of cross-correlation, it requires considerable computation, especially with refinements such as additional stored normal references, time shifting to find the peak coefficient, and increased sampling rate are incorporated. Furthermore, the coefficient can be relatively insensitive to substantial differences in durations of two waveforms. As a single measure of the shape of QRS complexes, cross-correlation may not be sufficient for differentiating VPBs from normally conducted beats.

Feature-Extraction and Clustering

The majority of arrhythmia analysis systems employ feature extraction and clustering to classify the morphology of the QRS complex. Initially, many of these systems employed heuristically chosen feature sets that were similar to those that a cardiologist might use to describe the shape of the QRS complex. Many of these measures used only specific points on the ECG waveform, and were not reliable and repeatable in the presence of noise. In more recent work, developers have incorporated "derived" features that are less sensitive to noise and artifact.

The ARGUS (ARrhythmia GUard System), developed at the Washington University School of Medicine (Nolle,1972, Nolle,1971, Cox,1969) is a notable example of a system that uses feature extraction and clustering. It employed a hierarchy of cascaded data transformations that produced a more compact waveform presentation of the EKG. The first pre-processing stage, termed AZTEC (Cox,1969) compressed the incoming data, producing a stream of flat lines and sloping segments of variable duration that caricatured the original ECG. The second processing stage, termed PRIMITIVE, provided detection, delimitation, and shape description of the ECG waves.

Each QRS complex was described by four parameters: 1) QRS width, 2) QRS peak to peak amplitude, 3) QRS offset (midpoint between lowest and highest excursions of the complex, measured from the baseline), and 4) rectified QRS area (about the baseline). QRS complexes were clustered into "families" by

placing the candidate complex in the family whose feature boundaries encompassed the candidate complex. If the complex did not fit into any existing family, but was "close" to one, it was incorporated into that family, and the feature boundaries of the family were extended to enclose the new beat. Otherwise, a new family was created. The classification distance metric was empirically based on A/D converter units, sampling intervals, etc., rather than a normalized measure. Final classification of the QRS complex was based on its prematurity, the relative population of complexes in the morphology class to which it belongs, and the individual beat's duration.

Alternative feature sets have been used by other researchers, and are listed below:

QRS width (Horth,1969),

QRS component wave onset and termination times, component wave heights, and location of component wave peaks (Haywood, 1970);

amplitude, magnitude, and duration of waves measured on various leads or lead combinations (Willems,1972);

peak to peak QRS amplitude (Whiteman,1974);

quantities related to the zero frequency component of the FFT (Burton,1975);

the minimum first derivative and a quantity related to the second derivative of the QRS (Frankel,1975);

the sum of all data points of the QRS from the beginning of the second change in sign of the first difference (Quinn,1975);

the QRS amplitude relative to baseline and a "TDIFF" parameter that represents the T-wave amplitude relative to the ST segment amplitude (Lovelace,1976);

the QRS area, height and mean deflection over the baseline (Birman,1978);

the ARGUS feature set plus the first spectral moment (the center of gravity of the amplitude or power spectrum from 5-25 Hz) combined with the 5-Hz and 10-Hz phase angles (Mead,1978).

Syntactic Methods

One criticism directed at heuristic feature sets is that they are often "incomplete"; that they do not totally specify the waveform. The problem, as observed in (Dillman,1978), was that occasionally, very dissimilar beats can be represented by the same, or very similar feature vectors. For example, a QR beat and a beat 180 degrees out of phase, an RS beat, have identical ARGUS feature set values. A few arrhythmia detection systems attempt to avoid this by parsing the ECG into a sequence of subwaveforms and using the resulting sequence to represent the shape of the waveform (Wartak,1970, Geddes,1971, Bussmann,1975). This approach retains more temporal information about the QRS complex than waveform "envelope" measures typically used in feature extraction and clustering systems.

This method of waveform description was used in a rhythm diagnostic system for the 12-lead ECG described by Bonner and Schwetman (Bonner,1968a). Their system compressed the incoming data as a sequence of straight line segments. Adjacent line segments were then grouped together into simple upright, inverted, or tilted V or U shapes. These overlapping shapes were scanned repeatedly to detect and measure conventional ECG waves. QRS waveshape comparison was performed by comparing the amplitude, duration, and relative configuration of the subcomplexes that comprise the QRS. Complexes that are not

"identical" or "similar" to the dominant type are re-examined for the presence of typical measurement errors. An interesting feature of the algorithm is that it will "add" or "delete" minor subcomplexes from the test waveform when trying to match with known waveform classes.

Linear Transformations

Another method of reducing feature space dimensionality is to perform a linear transformation of the ECG and map it into a lower dimensional space. Several of the linear transforms specifically minimize the average reconstruction error to compute the lower dimensional representation, and thus answer some of the issues of "inversibility".

A number of multiple-lead rhythm analysis systems describe the QRS complex as a linear combination of basis functions. In one system proposed by Young and Huggins (Young,1964), the three lead vectorcardiogram is expressed in terms of a series of 12 orthonormal exponentials, with a average reconstruction error of about five percent. They then determined a diagnostic matrix to relate the exponential basis function coefficients to a feature vector useful in establishing the subject's membership in a disease category.

Several other transformations have been applied to ECG data. The Fourier transform has been used to compute the first spectral moment of the QRS complex, which is used as an additional feature to the ARGUS feature set (Mead,1978). The Haar Discrete Cosine

(Ahmed,1975a), Walsh-Hadamard (Ahmed,1975b), and the Karhunen-Loeve (Halliday,1973, Womble,1977, Gustafson,1974, Gustafson,1975) transforms have also been used. The Karhunen-Loeve transform (KLT) has been shown to be the optimal orthonormal linear transformation (to the extent that it minimizes the expected value of the squared error of the reconstructed signal) (Fukunaga,1972). Results from Halliday's work (Halliday,1973) indicate that 6 to 8 coefficients are sufficient to represent the QRS with an average reconstruction error of 5%.

Morphology Classification

The majority of monitoring systems use an adaptive clustering approach to morphology recognition. Here the classifier "learns" the normal and abnormal morphologies for the patient by seeking naturally occurring modes in feature space. This is based on the assumption that beats of the same classification occur in tightly packed "clusters" in feature space, and that the clusters do not overlap. The general rule used by most algorithms is that if the "distance" of the candidate QRS complex to the nearest cluster is small, the candidate complex is given the classification of that cluster, and that the description (mean and variance) of the cluster is updated to include the features of the new member. If a candidate complex does not match an existing cluster, then it defines the initial mean for a new cluster.

There are several reasons why adaptive clustering is used in most real-time arrhythmia monitoring systems. First, there is no universal definition of normal and abnormal waveform morphology: the normal ECG waveform for one individual may closely resemble the abnormal waveform of another. Furthermore, an individual can exhibit a wide variation of normal morphology due to axis shifts of the heart when the patient breaths or moves about. A cardiologist, for example, usually studies the beats preceding and following the beat to be classified, and makes his decision based on the context of the surrounding beats.

A suitable measure of "distance" between individual complexes and clusters is required by any clustering procedure. Practically all of the traditional n-space norms have been used as distance measures: (1) the city block norm (Wheeler,1975), (2) the squared Euclidean distance (Otterstrom,1977, Swenne,1977, Higgins,1978), (3) the normalized Euclidean distance (Mark,1979), (4) the squared Mahalanobis distance (Gustafson,1975), and (5) the correlation coefficient (references listed earlier).

Procedures for cluster labelling are of paramount importance to any automatic classifier. Most morphology systems attempt to distinguish "normal" and "abnormal" morphologies, and a few systems also employ "T-wave" and "artifact" cluster labels. Some systems reported in the literature require the operator to identify the normal morphology during a learning phase; most high speed Holter tape analysis systems employ human "editing" of morphologies during the entire scanning period. Adaptive

clustering systems have relied primarily on the width and prematurity (usually running averages) of cluster members to label the cluster. Many systems also rely on the population of a cluster and its similarity to known normal clusters.

Unsupervised clustering algorithms are not without their theoretical and practical problems, however. There are no general guarantees of convergence: the suitability of most of the clustering techniques used is based on empirical evidence. Most clustering algorithms dynamically update the cluster description to follow long term variations in QRS morphology, but this can also lead to unwanted drift of the cluster description. This problem is especially acute when the normal and abnormal clusters overlap naturally, and has led to instances where normal and abnormal clusters were switched.

## Comparison of Features and Classifiers

The relative merits of different feature sets and classifiers have been the center of considerable discussion. Feature extraction and correlation have been widely used in commercial and university monitoring systems, with roughly the same success. It is very suprising, however, that very little work has been done to compare the discriminatory power of different features and classifiers. This situation will hopefully improve when standard evaluation databases are available that will permit meaningful comparisons.

Recent work at MIT (Kao,1980a, Kao,1980b) addressed several of the issues regarding classifier performance. Three methods for characterizing QRS morphology were compared by incorporating each method into three different supervised classifiers. The three QRS morphology feature sets were: (1) one hundred samples (of 360 sps data) centered around a fiducial, (2) six heuristic QRS features (QRS width, amplitude, offset, absolute area, signed area, and width estimate), and (3) six KLT transform features. The three classifiers were based on the: (1) correlation coefficient, (2) Mahalanobis distance, and (3) Parzen probability density estimate.

Each of the nine feature set - classifier combinations was evaluated based on the error rate of a normal/abnormal classification using a simple distance threshold (optimized for each patient). Each combination was trained on the first 500 beats of a half-hour annotated recording, and tested on the remainder. The Mahalanobis classifier / heuristic feature set performed the best, but overall error rates were surprisingly similar for seven best of the nine combinations. Individual trial performance, however, varied from 0 to 20 percent classifier error. Baseline artifact, delineation, and fiducial errors accounted for 42% of these errors, and fundamental Bayes error may account for the remainder. The scope of the study was limited in several respects, however. First, only a single normal cluster was used (only the Parzen probability density estimate was capable of representing multi-modal distributions),

and no clusters were used to represent abnormal morphology Second, only a single timing fiducial was used to compute the KLT coefficients and to align the time-sample waveforms for classification. It would have been reasonable to try sliding the waveform in time for the best match, thereby reducing the effects of fiducial alignment when comparing classifiers and feature sets.

To summarize, the major advantages and disadvantages of correlation, feature extraction, and linear transformations are listed below:

Correlation

++ uses sample values, and not noise sensitive point features,

+ does not require accurate delineation of QRS onset and termination,

+ is an integrative measure, and therefore less sensitive to noise,

+ conceptually straightforward and theoretically sound,,

- is sensitive to fiducial alignment (however, waveforms can be time shifted for peak correlation coefficient),

- is computationally intensive, especially if waveforms are time- shifted for peak correlation coefficient.

Heuristic Feature Extraction and Clustering

++ is a quick, one-shot calculation (waveform alignment is not critical),

+? can employ features that are similar to those used by the cardiologist, allowing one to borrow from human interpretation logic,

-? generally is not a reversible process (the waveform generally cannot be reconstructed from its feature set, and

that the same feature values can represent totally different waveforms),

-- often uses features that are sensitive to noise, artifact, and the contour of the waveform.

Linear Transforms

+ provide filtering against noise and artifact, since the predominant eigenfunctions tend not to correlate with noise and artifact,

+? are generally inversible (permit waveform reconstruction),

- require accurate determination of a fiducial to compute the reduced dimensionality coefficients,

-- are very computationally intensive, limiting its use in real-time systems.

## 3.8 Rhythm Characterization

Interwave interval measurements are an important part of ECG rhythm classification. Most algorithms employ "local" rhythm analysis (e.g. whether a beat was premature, compensatory, or missed) to detect and classify individual QRS complexes. A number of systems also use "global" rhythm analysis which deals with rhythm characteristics over a few seconds to a few minutes. In this category are rhythm statements such as "a run of superventricular tachycardia" or "ventricular bigeminy". Several systems indirectly deduce atrial activity from a careful statistical analysis of the RR interval sequence, and make rhythm statements (guesses?) such as "atrial fibrillation" or "sinus arrhythmia".

The value of timing data was recognized in the earliest ECG monitors. A number of systems that used only RR interval analysis to recognize ectopic beats were developed, but researchers soon recognized that classification is difficult without information about the width and shape of the QRS. For example, prematurity by itself is insufficient to identify VPBs, since atrial or junctional premature beats may be present.

Prematurity Criterion

Many systems use the average R-R or Q-Q interval of the most recent beats to estimate the atrial pacemaker rate. A beat is declared "premature" if it occurs within a pre-set percentage of the average R-R or Q-Q interval. Fixed and operator adjustable (10%, 20%, 30%, etc.) prematurity thresholds have been used.

Improvements over the interval-averaging predictor are possible. A number of systems only incorporate intervals that are distinctly regular into the average R-R interval estimate. A more general linear predictor that could track respiratory modulation of the heart rate in normal sinus rhythm was investigated (Karaian,1972). The interval prediction is based on a weighted average of the four preceeding intervals. It was found, however, that the best predictor coefficients varied from one individual to another, and may vary for a given individual under different conditions.

Interwave interval clustering can be used to characterize rhythm. For example, the R-R interval histogram demonstrates patterns which may be visually correlated with different arrhythmias. Automatic adjustment of prematurity criterion is possible by delineating normal and premature beat histogram peaks.

Markov Characterization of R-R Interval Sequence

A statistical characterization of rhythm has been investigated by several workers. Gersch et al. (Gersch,1975, Gersch,1970) transformed R-R interval sequences into a three symbol Markov chain sequence. Every R-R interval is classified as short (S), regular (R), or long (L) (90 and 110 percent average R-R interval thresholds are used). The probability that the observed sequence of S, R, and L states was generated by each of a set of prototype models characteristic of different cardiac disorders is computed. The prototype corresponding to the largest probability of the observed sequence is declared the disorder. In an evaluation, however, over 400 R-R intervals were required to accurately distinguish between atrial fibrillation and atrial fibrillation with occasional VPBs.

The commercially available monitoring system made by Electrodyne (Shah,1977) uses a similar approach to detect atrial fibrillation, which in turn modifies a number of decision thresholds in their algorithm. Three models were used: sinus arrhythmia, frequent supraventricular premature beats, and atrial

fibrillation. If the rhythm was classified as atrial fibrillation, beats were not identified as supraventricular prematures and some of the rules for VPB classification were modified.

In recent work by Gustafson and Willsky (Gustafson,1978a, Gustafson,1978b), R-R intervals are continuous quantities, rather than quantized, discrete sets (short, regular, long). The sequential behavior of the R-R intervals is modeled by a set of simple linear Markov processes, one for each arrhythmia category to be detected. A distinction is made between "persistent rhythms" that exist for over 6-10 heartbeats, and "transient rhythms" characterized by abrupt changes that take place over a period of 6 beats or less. The magnitude and likelihood of each of the possible rhythm categories is computed, allowing some measure of the severity of the arrhythmia.

Arrhythmia Interpretation Logic

Most arrhythmia monitor development efforts have concentrated upon the detection of premature ventricular contractions. Only two fairly complete arrhythmia interpretation algorithms have been published: the IBM system (Bonner,1968b) and one developed by Watanabe (Watanabe,1970). Both employ decision-tree structures that roughly correspond to textbook classification procedures, and contain most common diagnostic statements required for routine ECG analysis.

## 3.9 Algorithm Evaluation

A well-designed evaluation may be as critical to an algorithm's success as the original algorithm design, and may require even more effort than the original system development. Beat-by-beat evaluations of arrhythmia detectors are extremely tedious and expensive. Furthermore, the measured performance of an algorithm can be highly dependent on the data base used for the evaluation. The pattern recognition logic in many automatic arrhythmia detectors is often complex, and the number of algorithm pathways that must be tested can be very large. An exhaustive test of each pathway with all combinations of input would be prohibitive, and perhaps impossible in some cases.

Most algorithm evaluations have been a beat-by-beat comparison of the algorithm's and a cardiologist's interpretation of each beat. A commonly used performance measure is VPB detection. The better algorithms have yielded good results, especially when compared with human observers working in the CCU. The ARGUS system (feature extraction and clustering) developed at Washington University was evaluated using a data base of approximately 50,000 QRS complexes taken from 39 patients (Ripley,1975). Results demonstrated that of 3880 VPBs identified by the cardiologist, 3045 (78%) were correctly detected by the computer. 835 VPBs were missed (called normal), yielding a false negative rate of 21%. 174 non-VPBs were labeled VPBs by the algorithm, giving a false positive rate of 0.4% of total beats. The ARGUS evaluation was performed using a digital data base of

selected electrocardiographic episodes. These records were annotated by a cardiologist, and the comparison performed automatically by computer.

An evaluation of the Electrodyne arrhythmia monitor (which uses cross-correlation to classify QRS families) demonstrated extremely good performance (Shah,1977). In this study, a total of 146,638 QRS complexes were classified during a monitoring period of 1,170 minutes. The accuracy rate of computer QRS detection was 99.94%. During this period 67 artifacts were identified as beats. The computer correctly identified 5467 of 5729 VPBs (95.43%), and had a VPB false positive rate of 0.1%. Computer accuracy in detecting VPBs in the presence of atrial fibrillation was a respectable 93%. The data base used in this evaluation was 57 continuous 30 minute electrocardiographic episodes from 30 patients. "Blinded" interpretation of each beat was made by a cardiologist, and then compared with the computer's classification.

The performance of other automatic arrhythmia detectors was reviewed by Ripley (Ripley,1975). Although the general performance levels were quite encouraging, it is difficult to precisely compare evaluation results. Different operating protocols, adjustment of variable parameters, and amount of operator intervention can significantly affect the performance of a system in the clinical environment. Different performance measures were also utilized. Although most evaluators test VPB detection, some do not. Some systems detect only premature

beats, and some detect only abnormally shaped beats. Perhaps the most crucial factor that affects intrasystem comparison is the ECG data base. In several studies, the algorithm being evaluated worked very well for some patients, whereas in other cases its performance would be quite poor. The number and types of arrhythmias and the amount of noise can have a profound effect on algorithm performance.

A number of researchers have recognized the need to establish a standard digital data base to provide for uniform and more automated evaluations of arrhythmia detectors (Feldman,1974, Ripley,1975). The development of an annotated digital ECG data base has been started at MIT (Schluter,1978), and is currently available (MIT-BIH,1980). Recently, the Evaluation Group for Arrythmia Detectors (EGAD) at Washington University has been funded via the American Heart Association to create a standard digital data base (AHA,1978). Their data base consists of 160 3-hour segments of two channel ECG. The last half-hour of each tape will be annotated on a beat-by-beat basis. Thus, in two to three years the uniformity of algorithm evaluations should increase substantially.

# 4. THE BEDSIDE MONITOR

This chapter briefly describes the bedside monitor, its overall design constraints and goals, and a brief description of how it is used. The hardware and software architecture of the monitor is also discussed, and the environment in which it was developed. The main purpose of this chapter is to provide an introductory framework for the more detailed chapters that follow.

## 4.1 Overall Design Goals

The bedside monitor is intended primarily for patients in the hospital who require long term continuous monitoring of the ECG. They include post-operative patients, patients with intermittent symptoms due to arrhythmias, post myocardial infarction patients, and patients with neurological symptomatology. Quite frequently, many of these patients are located on general surgical and medical wards that often do not have central monitoring facilities. There is clearly a need for a system designed to be small enough and inexpensive enough to be compatible with single patient monitoring in any bed in the hospital, yet sophisticated enough to provide complete and reliable reports which are immediately available upon demand by the clinician.

This environment imposes a number of constraints on a single patient monitor:

(1) The monitor must be able to continuously analyze the electrocardiogram in real time. Unlike some systems that periodically analyze short segments of the ECG, the bedside monitor should never be in a condition where it could miss clinically important episodes.

(2) The monitor's cost should be compatible with single patient monitoring. Since the monitor will be used as a single patient monitor, its initial capital cost and operating costs should be competitive with conventional multipatient arrhythmia monitoring systems.

(3) The monitor should provide trend and statistical displays that indicate correlations between drug levels and ectopic activity. The displays should be on-line, and immediately available upon demand by the clinican.

(4) The monitor should be human engineered and easy to use. The general surgical and medical wards usually do not have specially trained personnel for operating sophisticated instrumentation. Thus, the bedside monitor should be easy to learn how to use. It should not use a typewriter keyboard, but instead a pushbutton console that gives instant feedback to the user. This system should almost be operable "by inspection" -- its command menu should be available at a glance. Also, the displays should be easy to read and interpret.

(5) The monitor should provide information about how well it is working. For example, the monitor should provide annotated real-time displays and strip chart recordings so that the user can make an "on the spot" estimate of the system's accuracy.

Many of the points discussed above have been incorporated in the design of the bedside monitor, and are briefly discussed in the following sections.

## 4.2 Features and Operation of the Bedside Monitor

The bedside monitor is contained in a 8 x 11 x 17 inch housing with a console for the operator (see Figure 4-1). The housing includes the processor (8085 microprocessor), memory (31K bytes of read-only memory) and 17K bytes of read-write memory), optically isolated ECG amplifier, A/D converter, and other peripherals. A storage oscilloscope (Tektronix 603) is used to display the incoming data and to present trend plots and statistical displays. The monitor also includes a strip chart recorder with two alphanumeric print heads for ECG documentation.

The incoming ECG is displayed in real-time, and each beat is labelled with its classification:

"N" for a normal QRS complex,
"V" for a PVC,
"S" for a supraventricular premature beat,
"W" for a wide, on-time beat,
"?" for complexes likely to be distorted by noise, and
"↑" for a very narrow complex likely to be artifact,
"L" for a learning mode beat (first 50 complexes only).

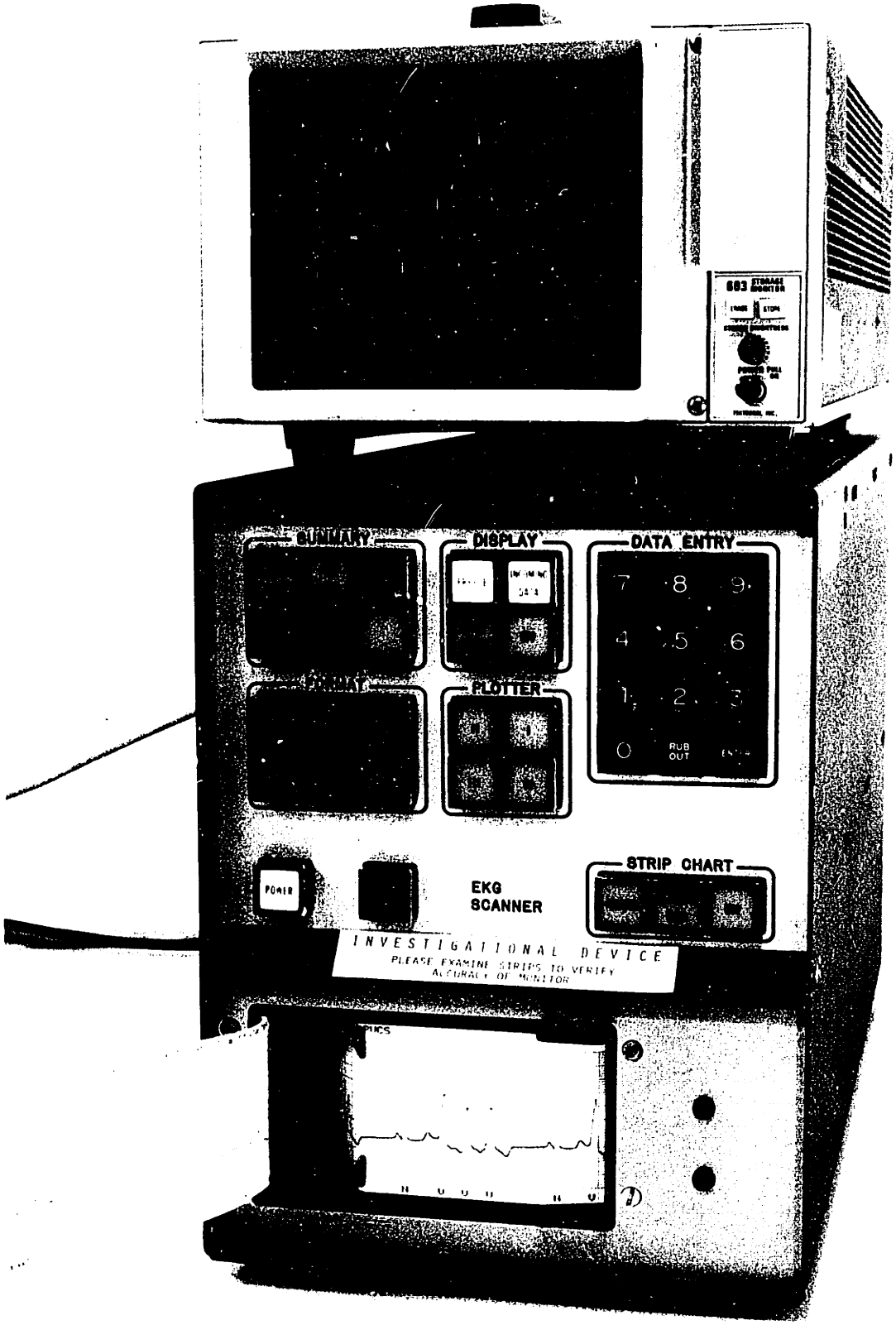



Figure 4-1: The Bedside Cardiac Arrhythmia Monitor

When a complex is detected and classified (usually one to two seconds behind the incoming data), the delineated complex is highlighted by overdrawing it with a straight-line approximation, and its classification is drawn just above it. Three traces are drawn, maintained for a few seconds, and then the screen is redrawn with the new incoming data.

Labelling the beats provides important feedback to the user about the monitor's accuracy, and helps establish the proper degree of confidence in the system. The clinican can get a reasonable "feel" for the monitor's accuracy by casual observation. Although such an estimate is crude, it is probably much better than no real-time feedback at all (most multipatient monitors do not provide an annotated real-time display). Similarly, all beats are labelled on all hardcopy produced by the strip chart recorder. Again, it enhances confidence of the user, and also documents errors made by the monitor.

When requested, the monitor can also display the following trend plots for the most recent 15 minute, 1 hour, 3 hour, or 12 hour period:

| | |
|---|---|
| Heart Rate | (beats per minute) |
| PVC Rate | (beats per minute) |
| PVC Run Length | (longest run of consecutive PVCs) |
| SVPB Rate | (beats per minute) |

The monitor also maintains R-R interval and PVC coupling-interval histograms. Strip chart recordings are automatically made whenever unusual ectopic rhythms are detected, and can also be manually requested. Each strip is labelled with the date, time,

patient identification, and reason for the strip. The classification of each beat is also written on the strip. The monitor produces hourly summary strips containing the lowest, average, and peak heart rate and PVC rate; total number of single PVCs, couplets, and runs of ventricular tachycardia for each hourly period.

More specialized functions are supported by the front console. Numeric data can be entered with a 10 digit numeric keypad. This is used for entering the patient identification number, the date, and the time. It is also used to tailor the number and type of strips made each hour. An optional plotter can be attached to the monitor to make hardcopy records of any of the trend or statistical displays.

All display, plot, and strip-chart processing is concurrent with the ECG analysis, which never ceases. For example, the operator can make a continuous strip chart recording of the incoming data while simultaneously making hardcopy plots of summary trend data.

The monitoring system has been designed to require a minimum of operator intervention. After the monitor is turned on, the user enters the patient's identification number, date, and time. The system then automatically "learns" the patient's normal complex, and after the first 50 beats, begins classification. The only user adjustable parameter is the gain of the ECG amplifier, which is set to produce a visually pleasing display.

The monitor thereafter adapts to changes in the QRS size, morphology, and average heart rate.

Likewise, the strip chart documentation protocol automatically adapts to the level of ectopic activity. The protocol is designed to record only the most severe ectopic activity in order to conserve paper. The bedside monitor will document most runs, but fewer couplets and isolated PVCs for a patient who has frequent episodes of all three categories. The level of documentation can be tailored for an individual patient using the numeric keypad. For example, the bedside monitor can be instructed to record all examples of ventricular runs and couplets.

## 4.3 Hardware and Software Development Environment

The bedside monitor is based on a bus-oriented microprocessor system developed at the Biomedical Engineering Center for Clinical Instrumentation at MIT (Burns&Schluter,1978). The overall hardware organization of the bedside monitor is shown in Figure 4-2. All system components identically connect to and communicate over a single bus. This bus provides all data, address, and control signals required by both memory and peripherals. The peripheral interfaces are "memory mapped", and no distinction is made between memory registers and the data, status, and control registers of a peripheral device.

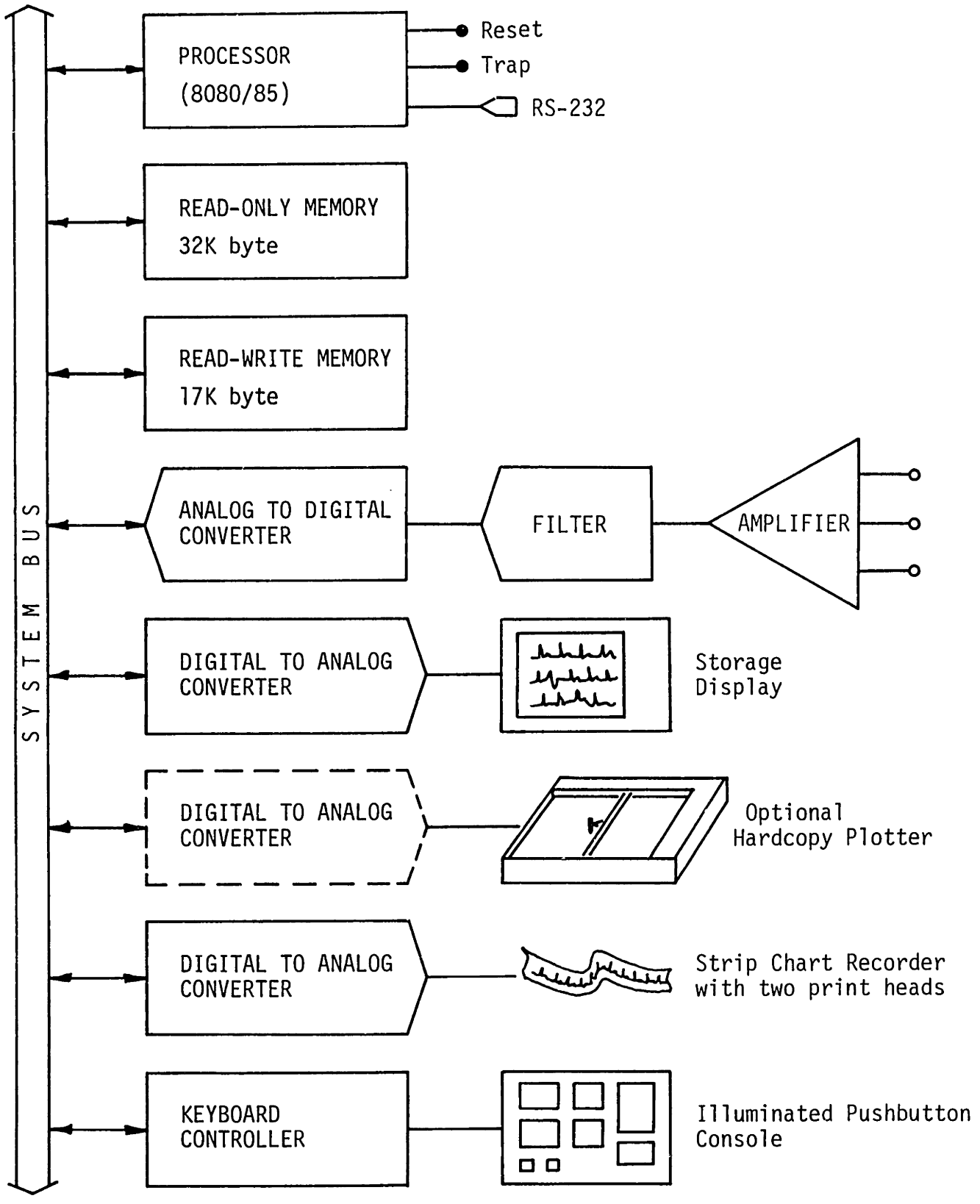



Figure 4-2: Bedside Monitor Hardware Block Diagram

The modular approach to the development of the bedside monitor has been extremely effective. First, most of the modules used in the monitor are "off the shelf" items, and only specialized modules such as the ECG amplifier and keyboard controller had to be developed. Second, larger systems that can support and use the same modules were available for developmental work. The larger "bench" system can contain up to 21 modules (the clinical prototype can hold only 10) and support additional peripherals and development aids.

The larger bench system provided a convenient development environment for algorithm development. It supports two 3M cartridge tape transports (to play back an annotated ECG data base), floppy disk, variable persistence (Tektronix 607) storage display, and a fully symbolic debugger with hardware breakpoint capability (Schluter,1976). Program preparation was done on a Nova 3 minicomputer, which supports an 8080 cross-assembler (Walters,1977) and a displaying editor. Programs assembled on the Nova were downloaded to the 8080 bench system for debugging and testing.

The annotated digital ECG data base played a crucial role in algorithm development. It provided a convenient source of data that could be played back at any speed, or even stopped for detailed review. The draft versions of the algorithm displayed the incoming data on the variable persistence display, and as events were detected and classified, the information was displayed as well. The variable persistence scope was an

excellent display device for this application. The slow fade out took care of deleting old graphical data, and made it very easy to add graphical indication of the algorithm's decisions. Of course, the image could be permanently stored for long-term visual inspection.

In addition to serving as a source of ECG data, the data base also provided a mechanism for automatically verifying the algorithm's beat classifications. This was done by using a "run-time comparator" which listed all disagreements between the "truth" annotations on the data base and the algorithm. Also, the comparator could automatically stop data playback if an error was made, and allow detailed inspection of the algorithm's state. This capability is hard to duplicate with an analog data source, or with evaluation techniques that provide only "after the fact" summaries of algorithm performance.

One crucial decision in a heavily software oriented system is the selection of processor and programming language. The choice of the Intel 8080 microprocessor was based primarily on the availability of hardware and software support for this processor at the Biomedical Engineering Center for Clinical Instrumentation (BMECCI) at MIT. There were other reasons as well. First, the 8080 is a popular microprocessor in its own right, and it was likely that industrial collaboration would be more likely with the 8080 than other processors. Also, the 8080 microcomputer family has a wide range of support chips that would make a very inexpensive, yet sophisticated monitoring system that

would require less than a square foot of printed circuit board area. Battery powered 8080 systems are quite feasible (and have been built), especially since there is a low-power CMOS version of this popular microprocessor.

Selection of the programming language was another crucial decision. In light of current trends, my choice (assembly language) can be rightly criticized. Assembly codes are not concise nor expressive languages, and programs written in assembler are notoriously hard to understand by personnel other than the original programmer (or the programmer himself a year later!). Assembly coded programs are not "portable" (can only run on a specific processor type), and thus cannot take advantage of improvements in processor architecture. Assembly language is poorly suited for complex, investigational algorithm design and development. A high level language (such as "C" or Pascal) would make it easier to experiment with different algorithms, and are supported by considerable extant software libraries.

Despite its disadvantages, assembly language was selected as the development language. Many of its alledged drawbacks can be minimized with the proper software tools. The issue of "non-portability" is somewhat mute, given the substantial acceptance of the 8080 by industry, and the relatively large number of programmers and engineers that are familiar with it. The other issue is software maintainability. Here, careful "top-down" design, structured control paths, and accurate, informative commenting of the assembly source can make it almost

as comprehensible as the same algorithm written in a higher level language. The basic structure of the program will be the same, no matter what language it is written in.

## 4.4 ECG Analysis Algorithm Overview

The algorithms used in the bedside monitor do not represent a fundamental departure from techniques used in other arrhythmia detection systems. The analysis program uses feature extraction and clustering as the basis for morphologic classification. The structure of the arrhythmia analysis algorithm can be viewed as a cascade of data processing stages, where each stage accepts a data stream as input, and produces another, more condensed data stream as output. The stages are relatively uncoupled, which results in more debuggable, reliable, and maintainable software. Figure 4-3 shows the overall software organization used by the algorithm.

### ECG aquisition and filtering

The ECG is amplified using an optically isolated amplifier, and filtered to suppress baseline and muscle artifact. The filtered waveform is sampled at 360 Hz, passed through a 60 Hz digital notch filter (to remove powerline intereference), and stored in a circular buffer. The circular buffer RDBUF contains the most recent 10 seconds of ECG data (3584 double-byte samples). It is used to buffer data for strip chart output (to show several seconds of ECG prior to an episode for context), and for algorithm stages which need to refer to the original samples.

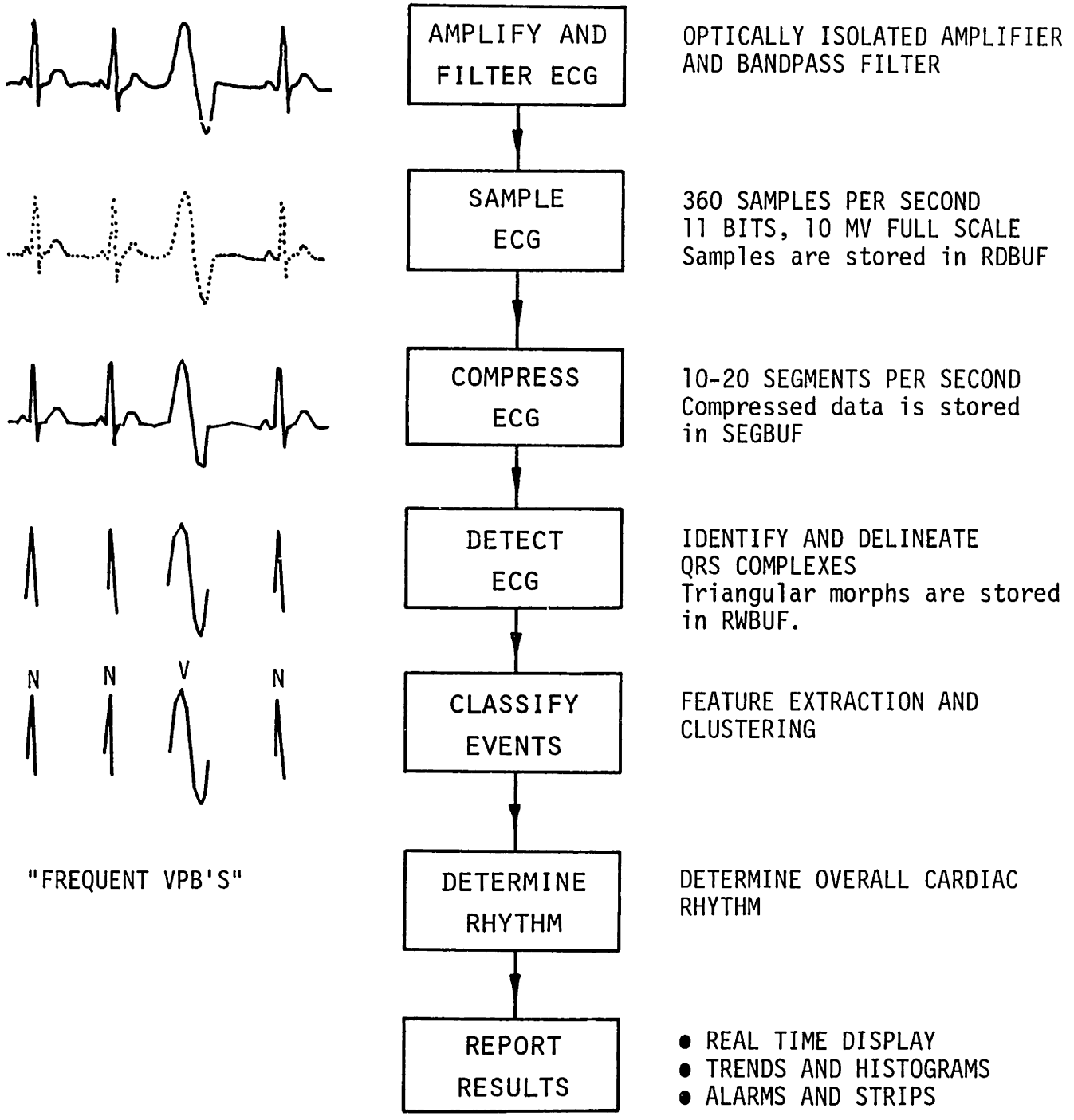



Figure 4-3: ECG Processing Steps

ECG compression

The next processing stage, termed SEGMENT, approximates the incoming ECG as a sequence of line segments that caricature the incoming data. This compression reduces the workload on subsequent processing stages, and serves as an initial feature extraction stage by identifying the low-level waveform structure. The output of SEGMENT is stored in a second circular buffer "SEGBUF", which contains 448 segments endpoint values. Each segment endpoint specifies the time (a pointer to an ECG sample in RDBUF), and waveform amplitude estimate at that time.

"V"-shaped event concatenation

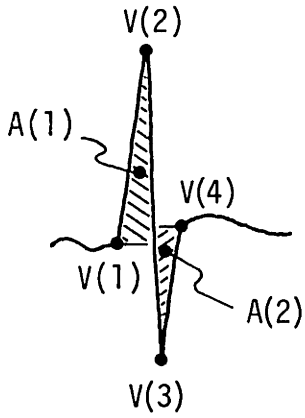
The next stage, termed TRIANGLE, concatenates segments into triangular morphs that may be possible components of QRS complexes. An "R-wave quality" value is computed for each triangular morph, and is a rough measure of how likely the morph is part of a QRS complex. The value is used in the event detection stage that follows, and for correct top-down parsing of major ECG waveforms. The triangular morphs are stored in the highest level circular buffer, termed "RWBUF". Each RWBUF entry contains SEGBUF pointers to the onset, extremum, and termination of the triangular entity, which in turn link to the original sample values in RDBUF.

QRS detection and delineation

The third processing stage concatenates triangular entities together as QRS complexes. The triangular morphs in RWBUF are searched in "R-wave quality" order (largest signal to noise ratio events are detected first). During the search, triangular entities are labelled according to the context of previously detected QRS complexes surrounding it. Subcomplexes within a larger complex are automatically recognized and identified when the triangle stream is scanned in this top-down fashion.

Feature extraction and morphology clustering

The bedside monitor's morphology classifier uses feature extraction and clustering to catalogue different waveform morphologies. Each QRS complex is characterized in terms of its triangular morphs (which roughly correspond to classical Q, R, and S waves). For example, a monophasic QRS complex is represented as a single triangular morph, whereas a biphasic complex is represented by two triangular morphs (see adjacent figure). The QRS is described by two sets of numbers: (1) the onset, triangular morph extrema, and termination voltage amplitudes, and (2) the areas of the triangular morphs that represent the complex. The time-ordered sequence of features preserves the temporal character of the QRS complex that would otherwise be lost if only "envelope" features such as height, area, and width were used. The other benefits of



feature extraction are still retained: the waveform comparison is a quick "one-shot" calculation, and does not require determination of a fiducial point.

After a QRS is detected, its morphology is compared with a catalogue of up to sixteen previously detected morphologies using a suitably defined distance measure. Each event is assigned to the morphology cluster that is closest to it, provided that the distance to this closest cluster does not exceed a "similarity" distance threshold. If the complex does not match any cluster, it is considered to be a new morphology, and is added to the catalogue.

Cluster and beat classification

Morphology clusters are classified as normal, abnormal, or unknown, according to the average prematurity and width of complexes that match the cluster, and the cluster population. The classification of an individual complex is determined by the classification of the cluster that it matches and its rhythmic context. If a beat matches a normal cluster, it is declared NORMAL or SVPB, depending on its prematurity. If a beat matches an abnormal cluster, and is premature, it is declared a ventricular ectopic beat. The end result is a stream of beat labels and times that drive the bedside monitor's trend displays and episode processor.

Reports and episode documentation

Once the QRS complexes have been identified, the remaining tasks are statistical summarization and indentification of sequences of events and other conditions likely to be of clinical interest. The stream of beat labels and times are transmitted to a variety of trending and statistical summarization routines, and to an episode documentation processor for the strip chart.

# 5. DATA AQUISITION AND EVENT DETECTION

This chapter discusses the aquisition and filtering of the ECG, compression and approximation by straight line segments, waveform parsing, and event detection. Some of the strategies used to detect and process high-frequency noise are also discussed. The importance of these stages cannot be emphasized enough, since subsequent classifier stages are at the mercy of the accuracy of the event detector and delineator.

## 5.1 Analog Data Aquisition and Filtering

The first processing stage of the bedside monitor is the amplification and filtering of the electrocardiogram. A simplified schematic of the bedside monitor's analog front end is shown in Figure 5-1a. A Burr-Brown model 3652 optically isolated amplifier is used to amplify the ECG, and includes a pair of 1.6 meg-ohm input/patient protection resistors that can withstand defibrillator pulses. This amplifier has other desirable properties as well: high differential and common mode input impedance, high common mode rejection ratio, and low input current leakage. The isolation amplifier was configured for a gain of 25, and thus will allow amplification of an ECG with up to a 400 mV DC differential electrode offset potential.

A single-pole high pass filter (with a 2 second time constant) provides AC coupling between the first two amplification stages, and removes any DC offset potential due to the electrodes. The diode network and 1K resistor are used to provide rapid recovery from defibrillator and other transient voltages that would normally saturate the input stage. During normal operation, however, the diodes do not conduct, and the network behaves as a linear R-C high-pass filter. The ECG is further amplified so that the overall gain is between 250 to 1000. The gain is adjustable from the rear of the monitor, and is usually set to produce a satisfactory display.

The fully amplified signal is then low-pass filtered with a linear-phase, two pole filter with a cutoff frequency of 72 Hertz. The primary function of the low-pass filter is to reduce high-frequency muscle artifact and serve as an anti-aliasing filter. However, the cutoff frequency was selected to be high enough to pass sufficient waveform detail for the analysis program and strip-chart recordings. Although more aggressive low-pass filtering could have been used, most effort was spent in designing algorithms to recognize and discard artifactual data, rather than try to process it. The frequency response for the analog filters is shown in Figure 5-1b.

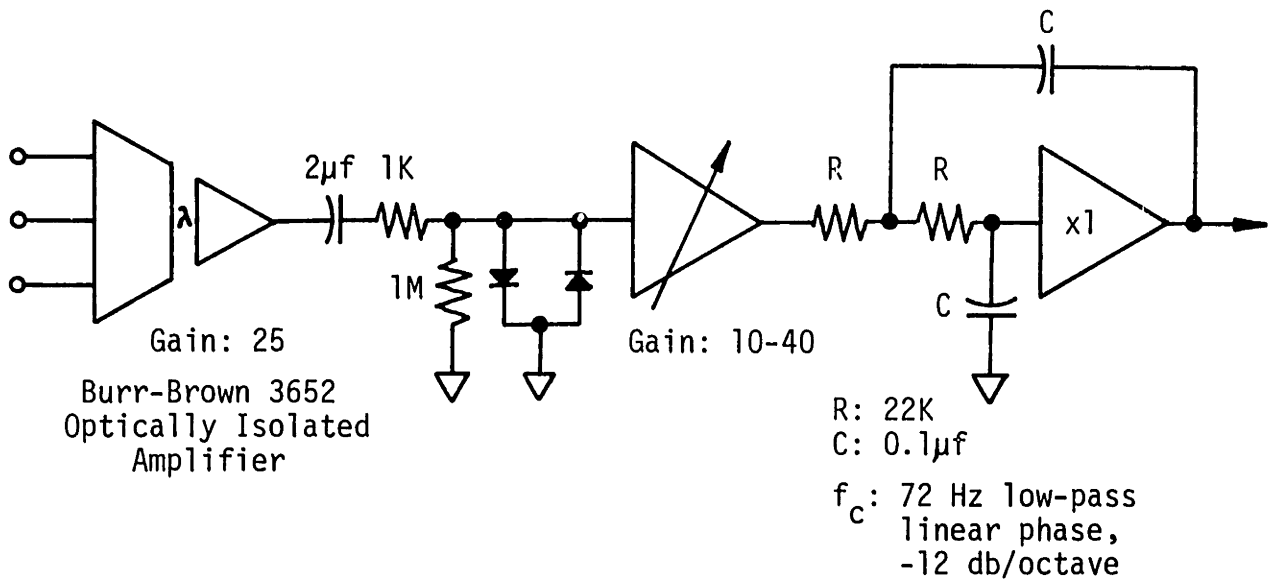



Figure 5-1a: ECG Amplifier and Analog Filtering

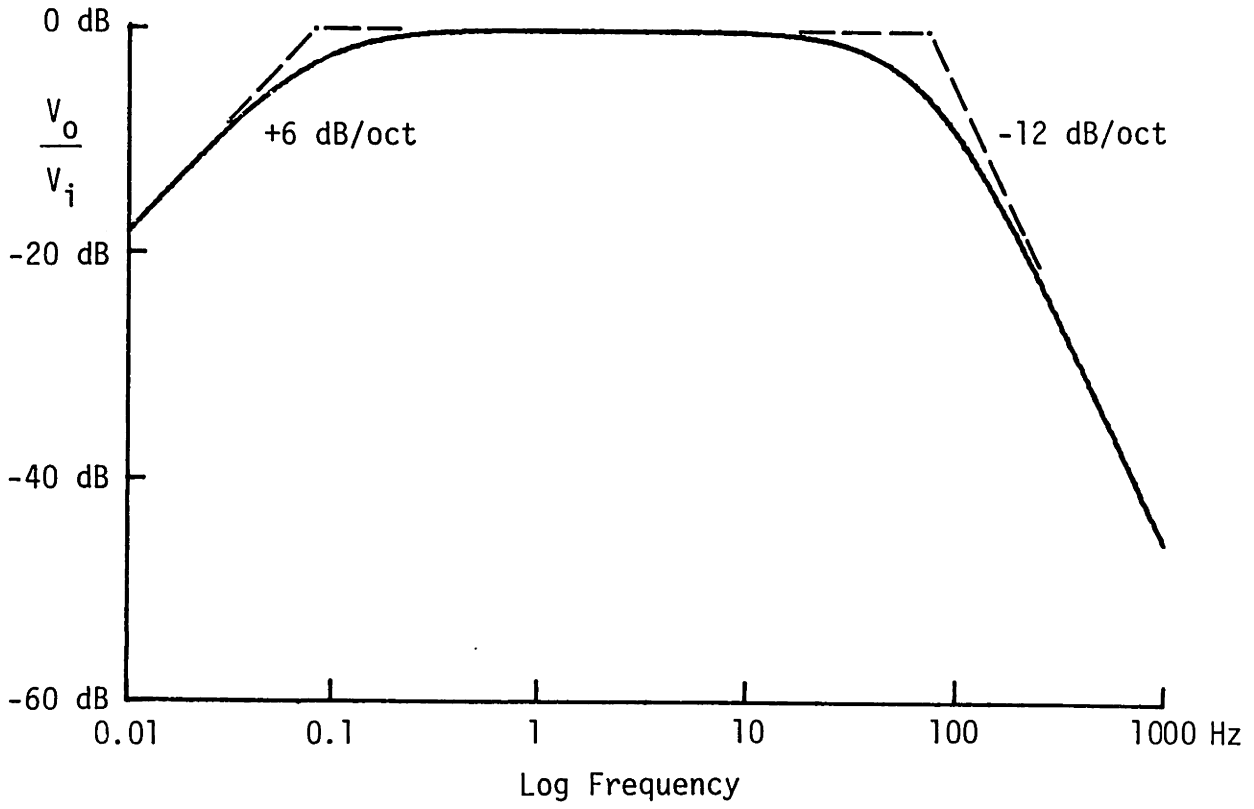



Figure 5-1b: Analog Filter Frequency Response

The filtered analog signal is then sampled at 360 samples per second, with a resolution of 12 bits over a +/- 5 mV range. The sampling rate was chosen to be sufficiently high so that temporal features such as QRS width would be accurately measured. The sampling rate is a multiple of 60 Hertz so that it would be easy to construct (in software) digital notch filters. Sampling rate selection represents a tradeoff between processor time and storage requirements versus possible improvements in algorithm performance due to more accurate waveform measurements.

Several finite impulse response (FIR) filters with transmission zeroes at multiples of 60 Hz were considered. A prime requirement was that they be computationally efficient, even at the cost of some filter performance. Given this constraint, two types of notch filters were considered, and the simpler (quicker) was selected. The first filter was the uniform average of the preceding six samples, so that its impulse response had a duration of 1/60 of a second. This filter has zeroes at 60, 120, and 180 Hz, and is quite effective at reducing power-line interference and harmonics (see Figure 5-2a). The second notch filter that was considered is even simpler, and is the average of the current sample (n) and sample (n-3). The major disadvantage of the second filter is that it doesn't have a zero at 120 Hz, but was selected anyway since it was easier to compute. The overall frequency response for the analog and notch filters is shown in Figure 5-2b.

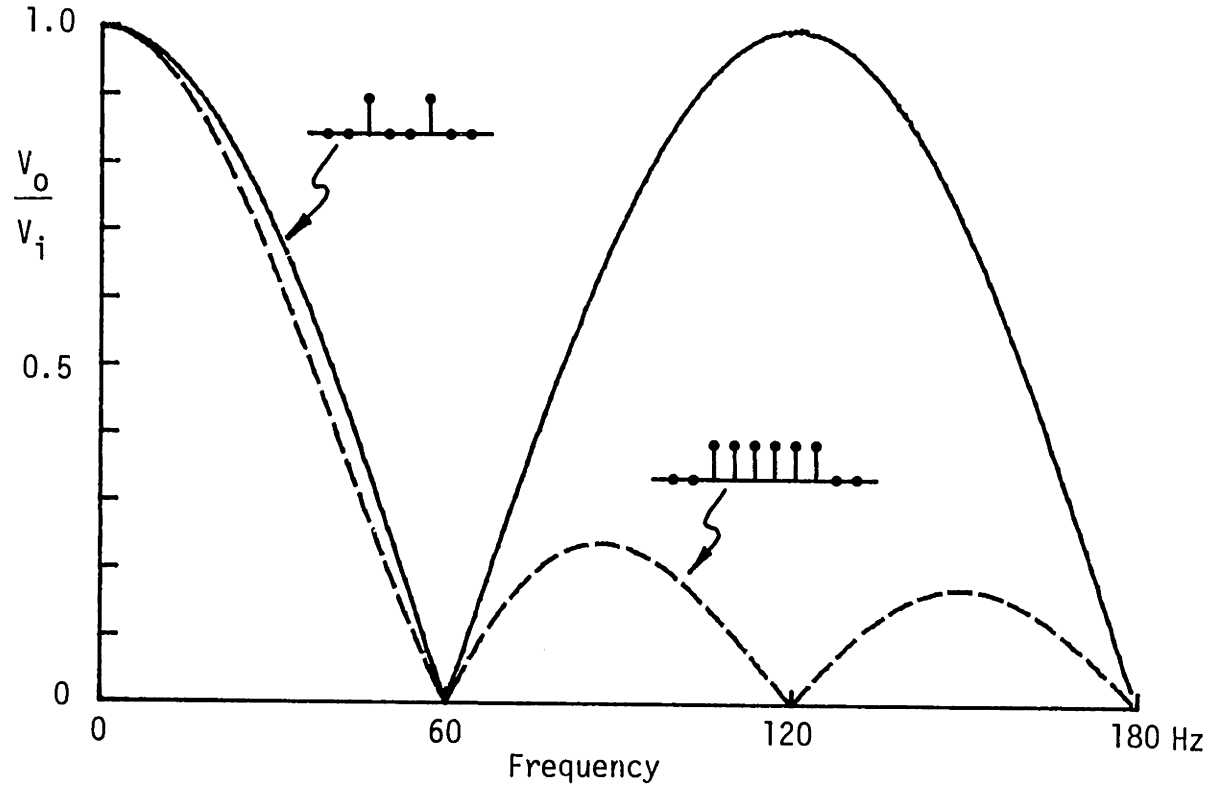



Figure 5-2a: 60 Hz Notch Filter Frequency Response

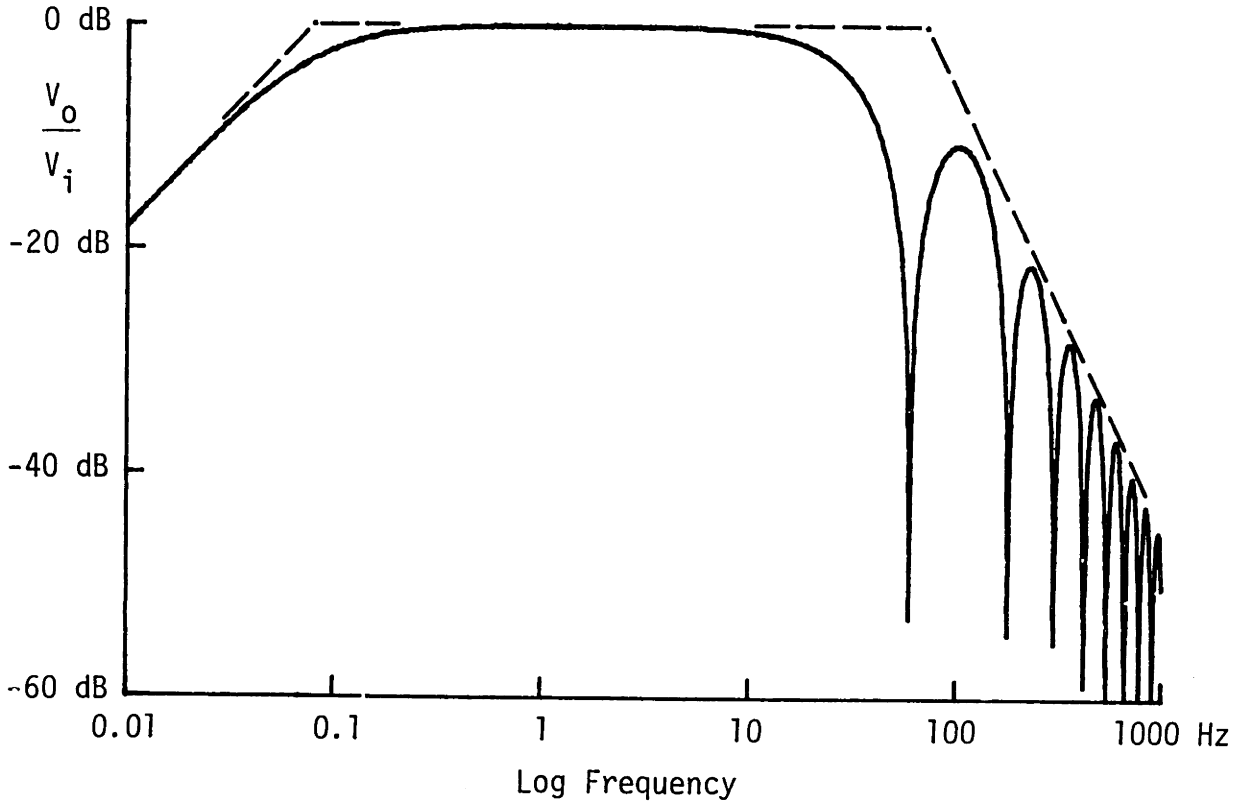



Figure 5-2b: Overall Frequency Response

The filtered samples are then stored in the raw data sample buffer RDBUF. The buffer is approximately 10 seconds long, and the samples are stored as 16-bit signed values. The buffer provides the basic raw data for stages of the algorithm that need to refer to the original data, and for context for strip chart recordings. The A/D sampling interrupt handler also updates the raw data display which shows the incoming data in real time.

## 5.2 Data Compression

The next ECG processing stage compresses the raw data samples into a more compact and easily parsable representation. The output of the compressor is a sequence of contiguous line segments that approximate the original data. The compressor also produces an estimate of the amplitude of high-frequency noise present in the incoming data.

### Rationale for compression

Data compression has been used in work by others for a variety of reasons. First, data compression reduces the computational load on subsequent processing stages. Scanning back and forth through the sampled data can become prohibitive, especially for sophisticated event detectors that may need to examine several seconds of data to make a decision about a questionable beat. Second, data compression can significantly reduce storage requirements (if the original data is discarded) especially in such applications as data transmission over phone lines, or compression of episode strip storage to reduce memory

requirements. Third, data compression permits convenient visualization of the data, and makes it easier for algorithm developers to view the detailed operation of the algorithm.

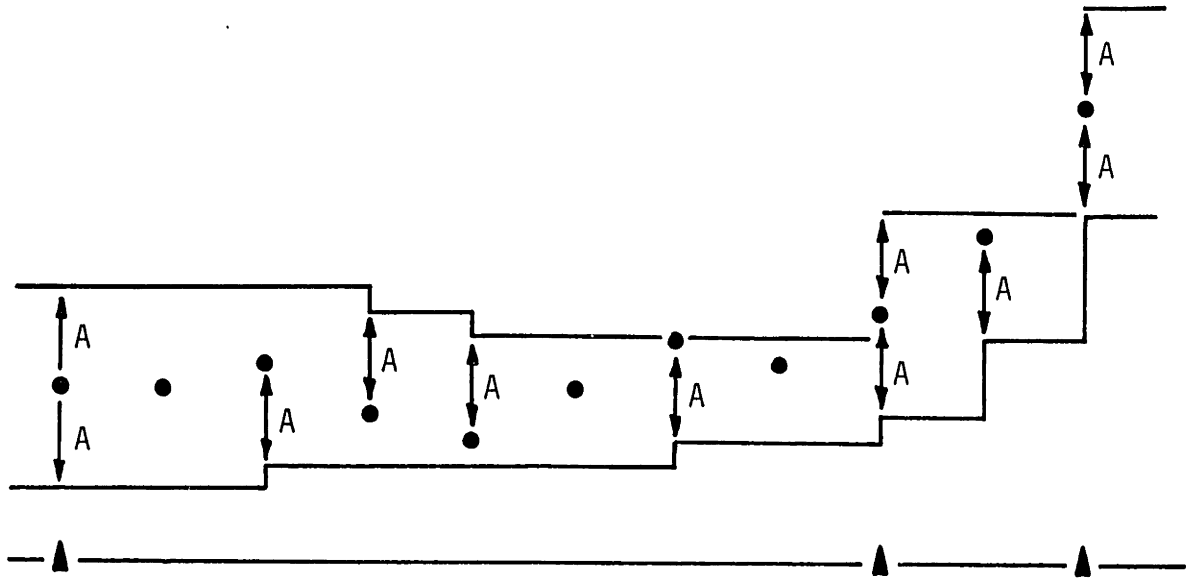
The major viewpoint taken in this work (and in others) is that data compression can also serve as the first of several feature extraction stages of a pattern recognition algorithm. It was decided early in the design of the bedside monitor that most compression techniques suitable for waveform preprocessing would be unsuitable for rhythm strips, and that the raw data would have to be saved in any case. Furthermore, future algorithm developments such as P-wave detection probably will need the raw data to detect low-amplitude complexes that would otherwise be "removed" by most data compression schemes.

Thus the prinicipal reasons for data compression in the bedside monitor are to reduce processor workload, and to serve as an initial feature extraction stage where the incoming data is rendered into a more compact and easily parsable representation. At the same time, the compressor should be computationally simple (it should not take more than 20% of processor time), achieve significant compression, and use a simple representation for the compressed output.

Initial waveform segmentation

The ECG data compression is performed in two stages. The first stage uses a Zero Order Interpolator (ZOI) to partition the data samples into subsets that are processed by the second stage, which does the final segmentation of the data. The ZOI is similar to the first stage of the Amplitude Zone Threshold Epoch Coding (AZTEC) compression algorithm first used by (Cox,1968). The AZTEC compressor has been shown to be useful in processing a variety of physiologic signals, especially the electrocardiogram. The ZOI operates by tracking the samples of the input waveform, and following the maximum and minimum value of the signal in an interval (see Figure 5-3a). As long as the signal does not change appreciably, no output is produced. When the signal changes such that the difference between the maximum and minimum values in the interval exceeds a specified aperture (ZOIAPER), the ZOI returns an estimate of the data (usually the average of the maximum and minimum values in the interval, and the duration of the interval). This representation is appropriate for the ECG, since a substantial fraction of the signal is baseline.

The principal function of the zero order interpolator is to identify flat baseline sections, and significant waveform extrema (see Figure 5-3b). The ZOI performs a "hysteretical smoothing" of the raw data that ignores local extrema due to noise, but does not smooth out significant extrema or inflection points. And unlike linear smoothing filters, the ZOI does not "smear" out detailed timing information about the waveform.




▲ indicates the beginning of a new ZOI segment

A indicates the value of the ZOI aperture

Figure 5-3a: Detailed ZOI Operation

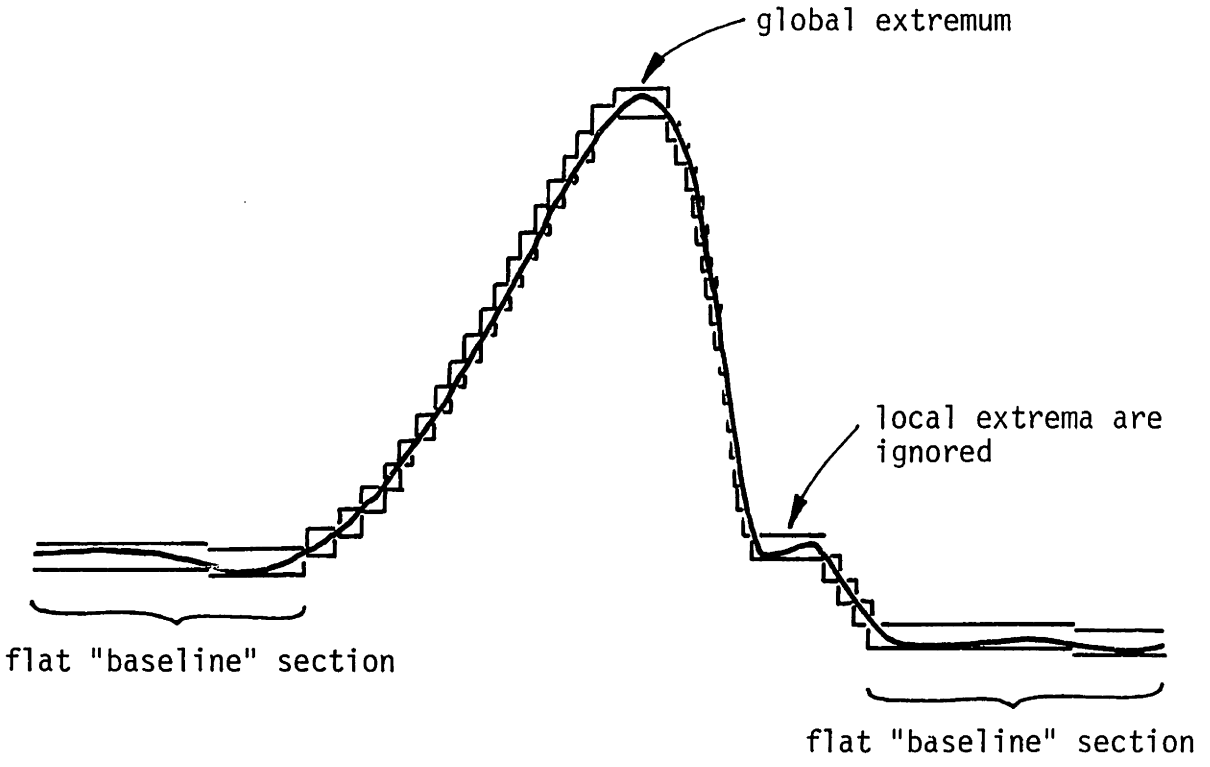



Figure 5-3b: Hysteretical Smoothing Performed by the ZOI

The zero order interpolator alone is inadequate as a data compressor, however. Although it reliably identifies waveform extrema and flat baseline sections, it doesn't compress gradual sloping sections in a consistent manner. Also, during steep sections of the waveform (such as the QRS complex), it doesn't compress the data at all. Thus, a second post-fitting procedure is used to represent sloping waveform sections.

Segment post-fit procedure

The post-fit procedure is applied to all sections of data bounded be an extremum or flat baseline. An extremum is declared if one is detected by the ZOI, and flat baseline is declared if a ZOI segment greater than 20 samples long is returned by the zero order interpolator. Every region of raw data bounded by a ZOI extremum or flat section is approximated by a straight line segment, as shown in Figure 5-4a. The error between the raw data and the straight line guess is computed point by point, and the largest positive and negative error is noted. If the error(s) exceed the ZOI aperture, the initial guess is broken into two or three new line segments (see Figure 5-4b).

Originally the post-fit procedure was repeated until the straight line approximation fit within the ZOI aperture for all data points. Later experience revealed that a single pass was sufficient, and set an upper bound on the time the compressor would take. Although it was not guaranteed that the fit would be precise in all instances, it did lead to somewhat more consistent compression of the data.

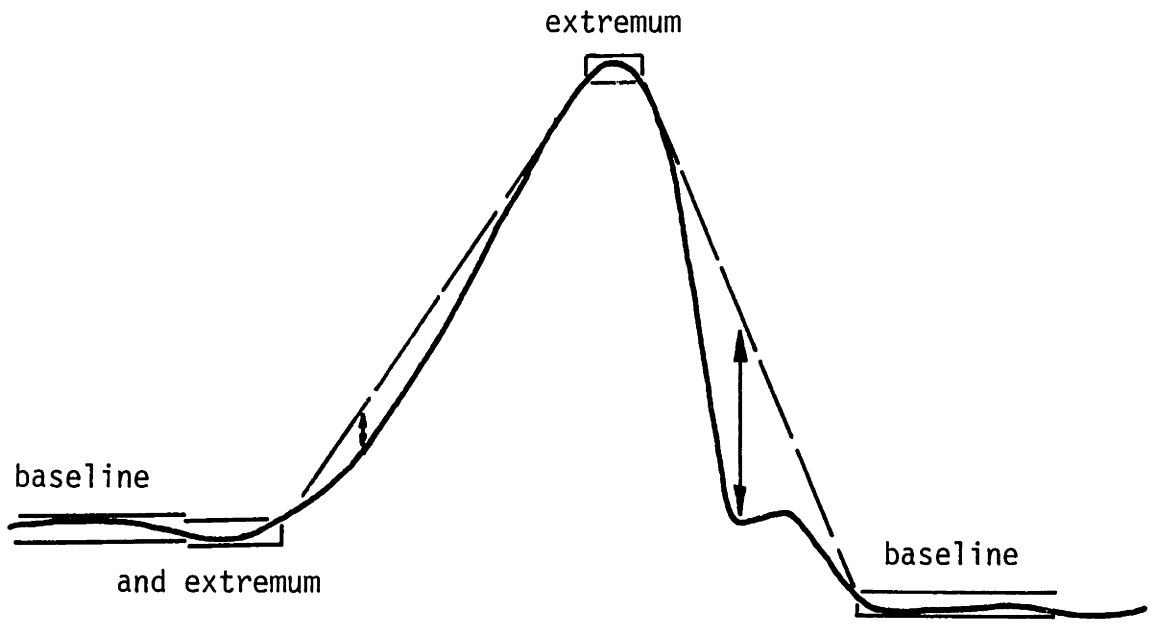



Figure 5-4a: Initial Waveform Partitioning

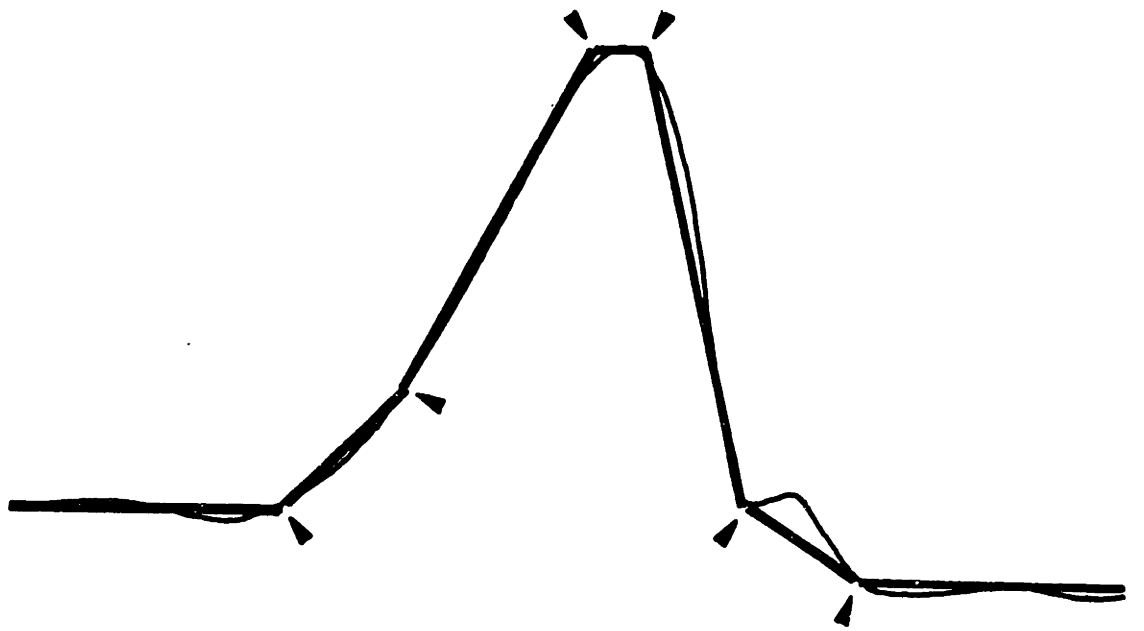
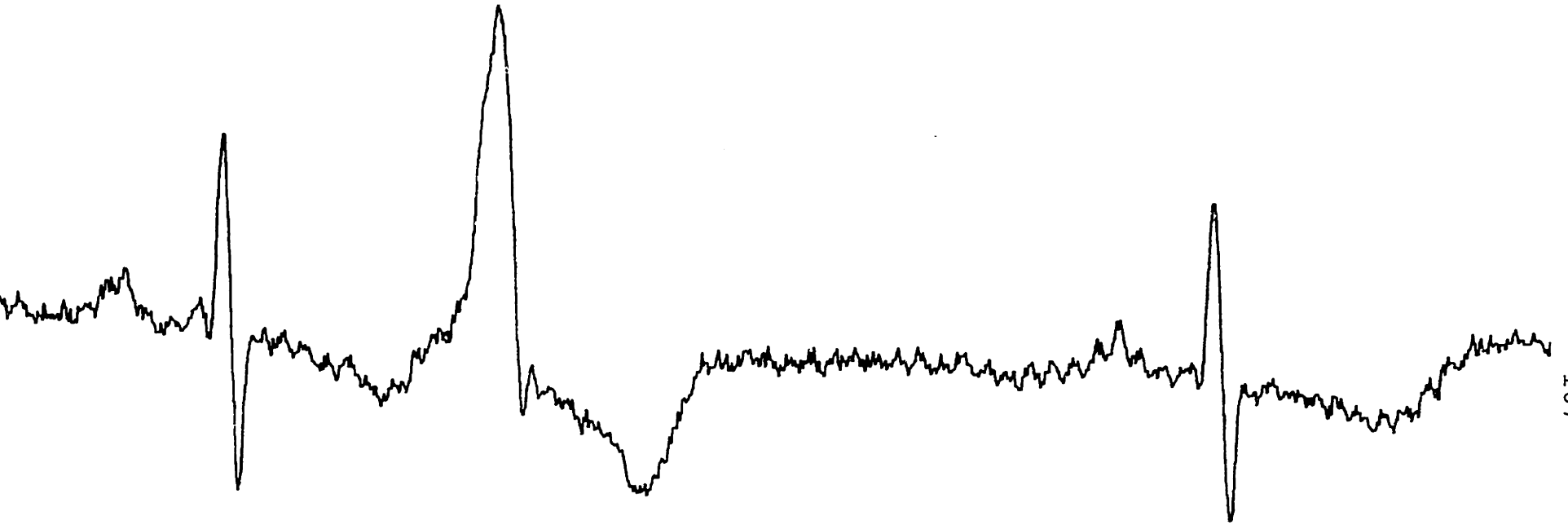


Figure 5-4b: Final Waveform Segmentation

The entire compression procedure is summarized in Figure 5-5. Figure 5-5a shows the original data and the initial segmentation (without the post-fit procedure). Note that for regions of gradual slope (such as the upslope of the PVC), the extrema/baseline representation is inadequate, and indicates the need for the post-fit procedure. The final compressed waveform and raw data are shown in Figure 5-5b, and is one example of the reasonably accurate fit that the compression algorithm is capable of. The orginal data and compressed waveform are shown separately in Figure 5-5c, and suggests that the compressed waveform retains most of the significant features of the waveform (note, however, that low amplitude P-waves would be "deleted" by the compressor if their amplitude is less than the ZOI aperture).

Adaptive ZOI aperture

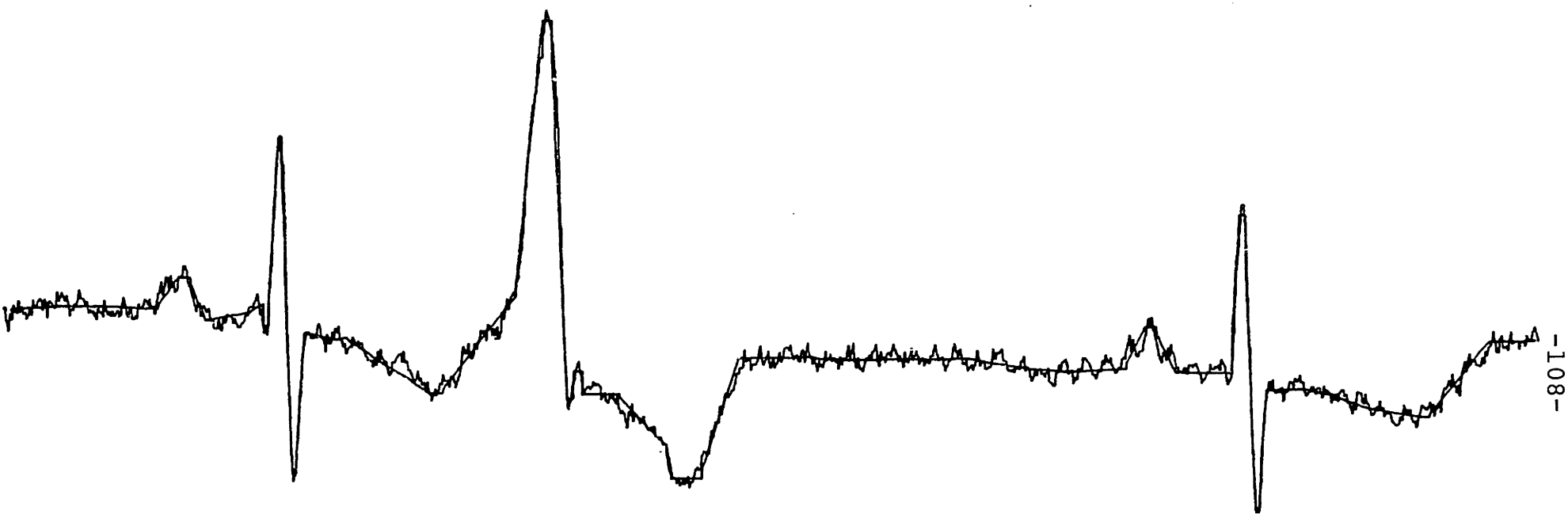
The value of the ZOI aperture has a significant impact on compressor performance. In the current version of the compressor, the ZOI aperture is dynamically adjusted to produce up to a maximum of 20 to 30 segments per second. This is done by computing the average number of segments produced per second (average over several seconds), and increasing or decreasing the size of the ZOI aperture if too many or too few segments are produced. Increasing the aperture increases the compression ratio, but at the cost of lower compression accuracy. This feature was added to the compressor so that it could adapt to different levels of low amplitude, high frequency noise. A fixed aperture is less desirable, since if the noise amplitude





INITIAL SEGMENTATION

Figure 5-5a





ORIGINAL AND COMPRESSED ECG

Figure 5-5b

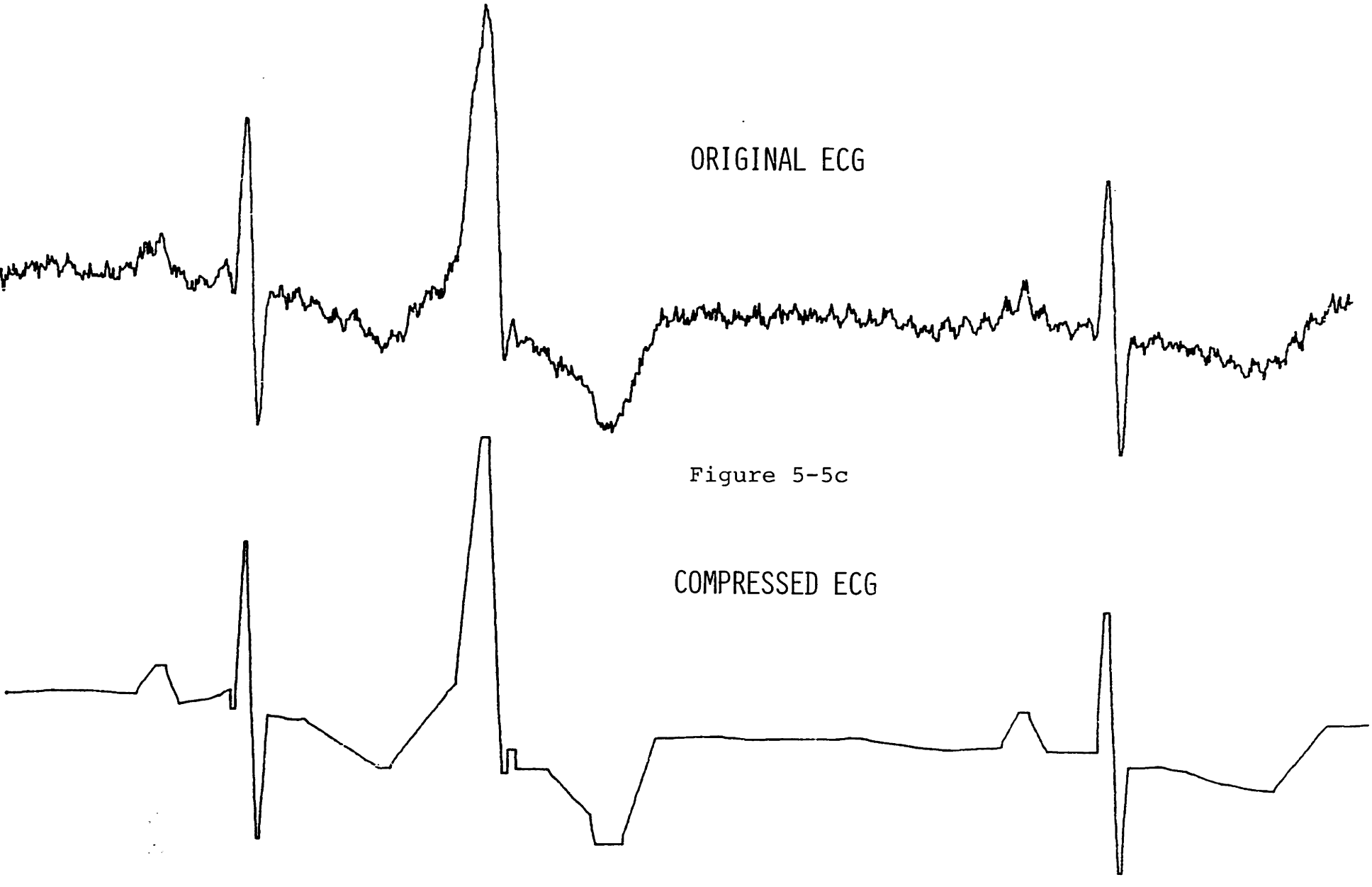



Figure 5-5c

increases every so slightly over the fixed aperture, the compression ratio is dramatically reduced. On the other hand, if the noise amplitude is very low, the aperture is reduced within a few seconds to a fixed lower limit of 5% of the nominal (1 mv) QRS amplitude. Normally the noise amplitude is low enough so that the aperture remains at the lower fixed limit.

Discussion

The compressor used in the bedside monitor has a number of advantages over previously report compression algorithms. It is relatively insensitive to noise (since it does not employ first or second derivatives to locate "points of interest"). It excellently corrects for any errors in estimating the endpoints of baseline (flat) sections of data (see Figure 5-5b,c) with the use of a second post-fit pass. This is an important feature for accurate QRS onset and termination measurement. The compressor is time efficient (typically 15 to 20 percent of an 2MHz 8080), since successive points on the approximating line segment are computed with an efficient incremental algorithm. The final data is in a convenient endpoint-value format (shown below), rather than a more complicated slope-endpoint or other format.

| RDBUF pointer (time) |
| --- |
| amplitude estimate |

## 5.3 Triangular Morph Concatenation

The next stage concatenates segments into triangular morphs that may be possible components of QRS complexes. For each triangular morph, an "R-wave quality" value is computed that is a rough measure of how likely the morph is a component of a QRS complex. The value is used in the event detection stages that follow, and plays an important role in parsing major ECG waveforms.

The triangular morph description of the ECG waveform offers several benefits for waveform analysis. First, only waveform structures likely to be components of cardiac events are passed to the subsequent event detector stages. Second, it is a more condensed, higher-level waveform description that makes it easier to perform sophisticated contextual scans. For example, the QRS complex can be easily characterized by three overlapping triangular morphs representing the Q, R, and S-waves. Ventricular tachycardia and flutter waves are characterized by a more prolonged sequence of alternating, overlapping triangular morphs.

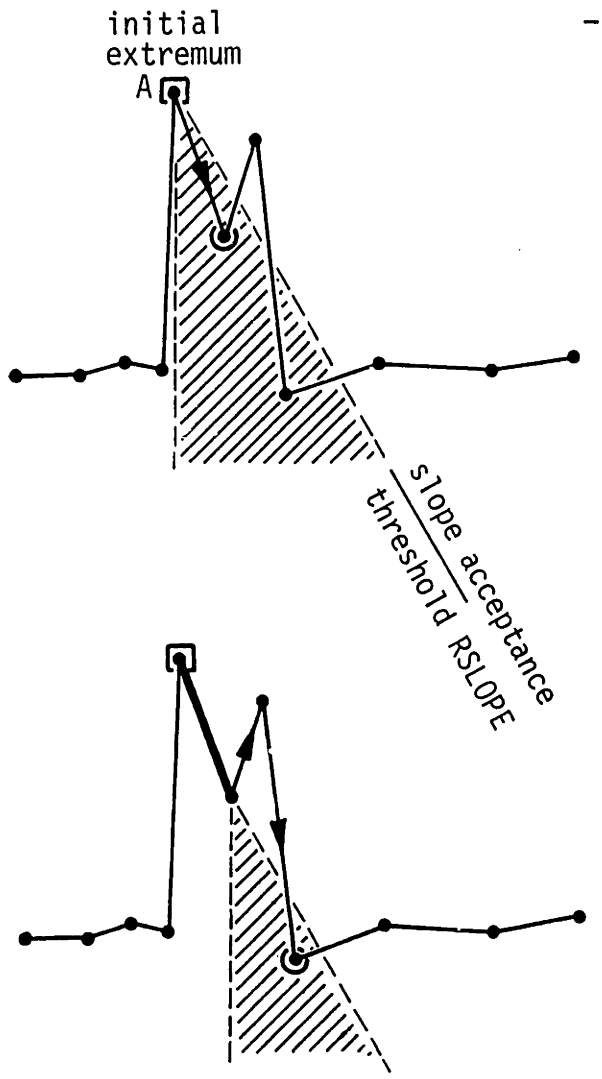
### Concatenation procedure

The procedure used to concatenate segments into triangular morphs is shown in Figure 5-6. The segment buffer is searched (going forwards in time) until a positive or negative extremum is found. This "initial extremum" becomes the starting point for forward and backward scans that attach segments that will

comprise the triangular morph associated with the initial extremum. Since the forward scan can be abnormally terminated due to insufficient data ahead of the initial extremum, the forward scan is conducted first.

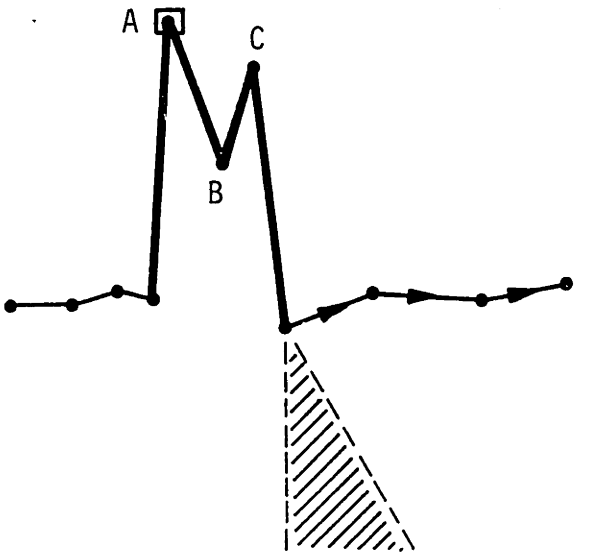
Starting at the initial extremum, the slope of each segment is compared with a slope acceptance threshold RSLOPE. Each segment that is within the slope acceptance threshold becomes a new origin for the slope test (Figure 5-6a), and the test is applied to the next segment. If one or more segments have insufficient slope, the scanning procedure "looks ahead" up to eight segments. If a segment is found within eight segments, a new slope test origin is declared, and the search continues (Figure 5-6b). If no segments are found, the search is terminated (Figure 5-6c). A similar procedure is used to scan backwards from the initial extremum to complete the delineation of the triangular morph.

The "look ahead" feature provides some immunity to premature termination of the forwards and backwards scans. Often there are small "notches" in the QRS complex of a physiologic origin, or due to artifact, that could terminate a scan with look-ahead of only one segment.




This figure illustrates how segments are concatenated to form a triangular morph. Every morph has an 'initial extremum' that is the starting point for forward and backward scans through the segment buffer. A slope acceptance threshold RSLOPE is used to test successive segments.

Each segment that is within the slope acceptance threshold becomes the new origin for the slope test. The scanning procedure 'looks ahead' up to eight segments, and if one is within the slope acceptance threshold, a new origin is declared.



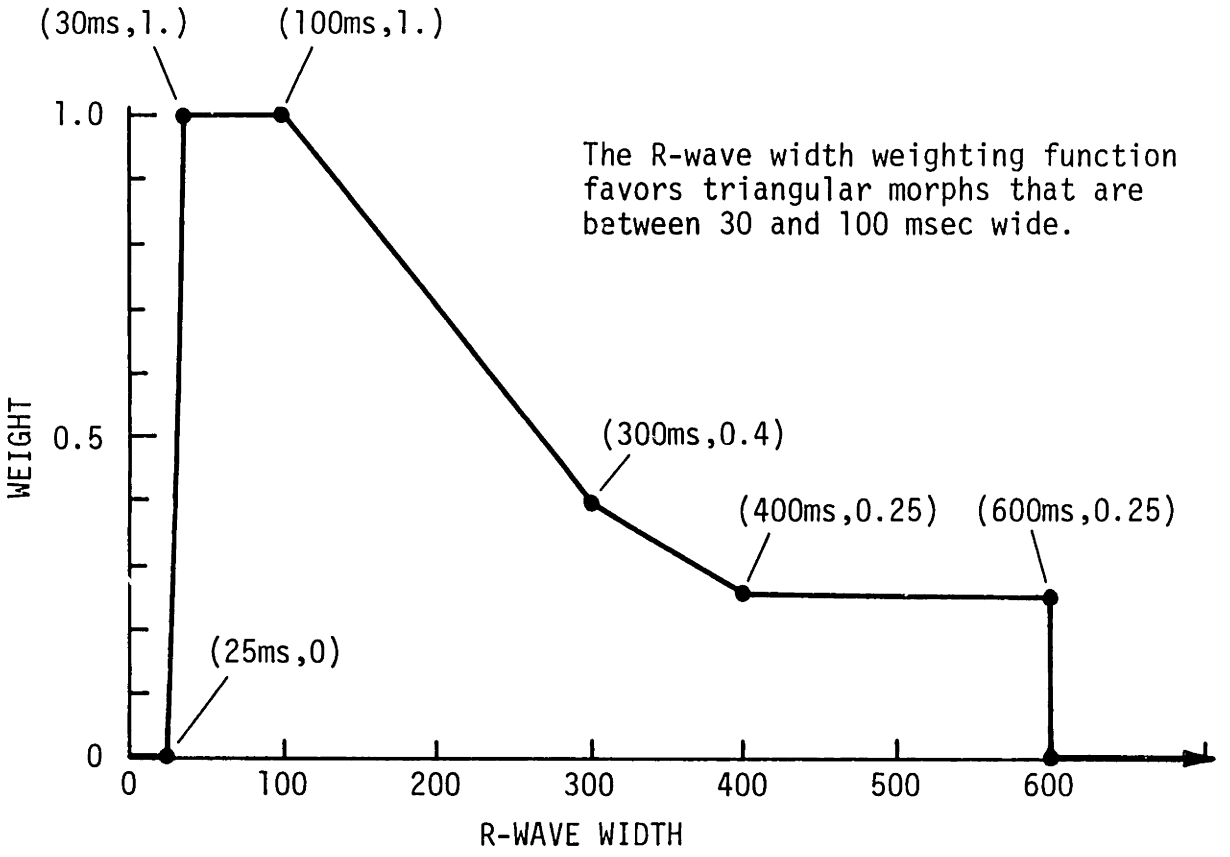

The concatenation stops if there are no segments within the slope acceptance threshold. A similar procedure is used to scan backwards through the data. After the triangular morph associated with extremum A has been delineated, the extrema at B and C will be processed in a similar fashion.

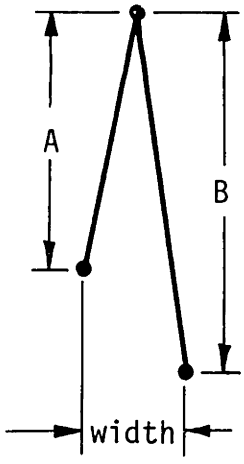
Figure 5-6: Triangular Morph Concatenation

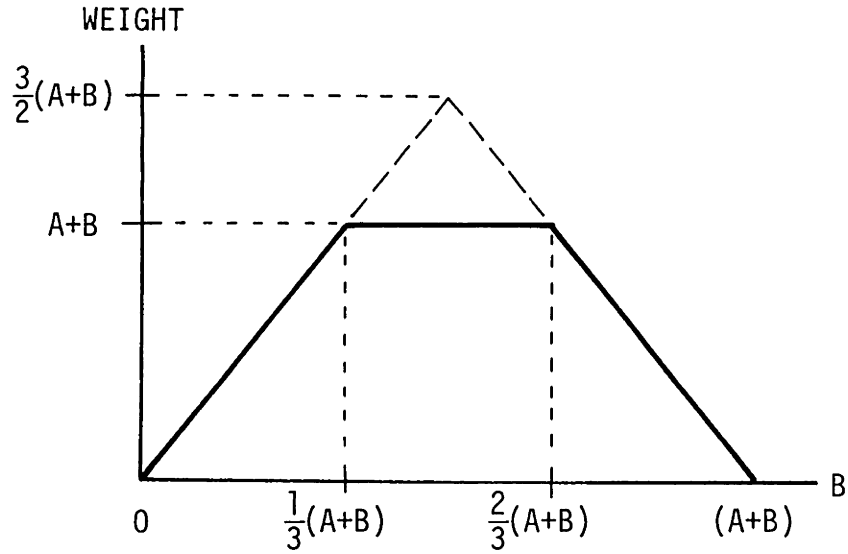
RQUAL computation

After the triangular morph is delineated, an "R-wave quality" value is computed that is a rough measure of how likely the morph is a component of a QRS complex. Its computation is illustrated in Figure 5-7. The value of RQUAL is the product of an amplitude dependent function and a width dependent function. The amplitude dependent function is simply the sum of the two heights of the legs of the triangular morph, and tends to favor symmetrical morphs. If one of the legs of the morph is less than half the other, the amplitude function drops off rapidly to zero, again favoring symmetrical morphs.

The width dependent function is designed to emphasize morph widths that are within the normal physiologic range for QRS complexes. Events that are 30 to 120 msec wide are favored, whereas much wider events (due to T-waves or large baseline excursions) are weighted less. Events less than 30 msec wide (that are likely to be artifact or 60 Hz powerline interference) are rejected by setting the width weight to zero.




The graph above illustrates how RQUAL favors triangular morphs where A≈B. The ordinate shows the value of B in terms of (A+B).

RQUAL = [ MIN( A+B , MIN(A,B)*3 ) * R-width_weight ]

Figure 5-7: RQUAL Computation

Ventricular tachycardia recognition

The triangular morph description of ECG complexes makes it relatively easy to test for overlapping, alternating polarity morphs characteristic of ventricular tachycardia and flutter. If the incoming data satisfies the test described below, the morphs that comprise it are labelled as "flutter" waves. The flutter label forces the QRS classifier (described in a later chapter) to classify detected complexes as PVCs, regardless of their timing and morphology.

The flutter wave test checks for a sequence of overlapping triangular morphs, and imposes several requirements on the individual morphs in the sequence. First, each morph must be at least 120 msec wide. If a narrow morph (normal QRS complex) is present in the sequence, the sequence is discarded as a possible run of ventricular tachycardia. Second, each morph must be less than 600 msec wide, which corresponds to a minimum run rate of 100 beats per minute. Third, the RQUAL value of all morphs in the sequence should be greater than one half the running average RQUAL of the previous 8 beats in the sequence.

The triangular morph buffer

After a triangular morph has passed through the preceeding stages, it is put into the triangular morph buffer RWBUF. Only morphs that have an RQUAL greater than 1/8 the running average RQUAL of previous eight QRS complexes are included, thereby discarding morphs likely to be artifact or due to 60 Hz powerline interference. It is important to note that RWBUF will have

morphs that are P-waves, Q-waves, S-waves, T-waves, and "noise events", in addition to the major R-waves of QRS complexes. The simple RQUAL test at this point only removes morphs that are not likely to be cardiac events.

Each RWBUF entry has the format shown below:

| | | |
|---|---|---|
| word 0 | Morph ONSET | (SEGBUF pointer) |
| word 1 | Morph EXTREMUM | (SEGBUF pointer) |
| word 2 | Morph TERMINATION | (SEGBUF pointer) |
| word 3 | LABEL | AUX INFO |
| word 4 | R-wave quality RQUAL | |

The first three words of a RWBUF entry point to the onset, extremum, and termination of the triangular morph. The byte LABEL is used by the event detector (described in the following section) to label morphs as Q, R, S, T-waves, or other labels. The AUX INFO byte is used for other information, including the eventual classification of the complex if it is labelled an R-wave.

The triangular morph buffer RWBUF is the highest level circular buffer. Its current size is about 140 entries, which permits it to cope with short bursts of high frequency artifact.

## 5.4 Event Detection

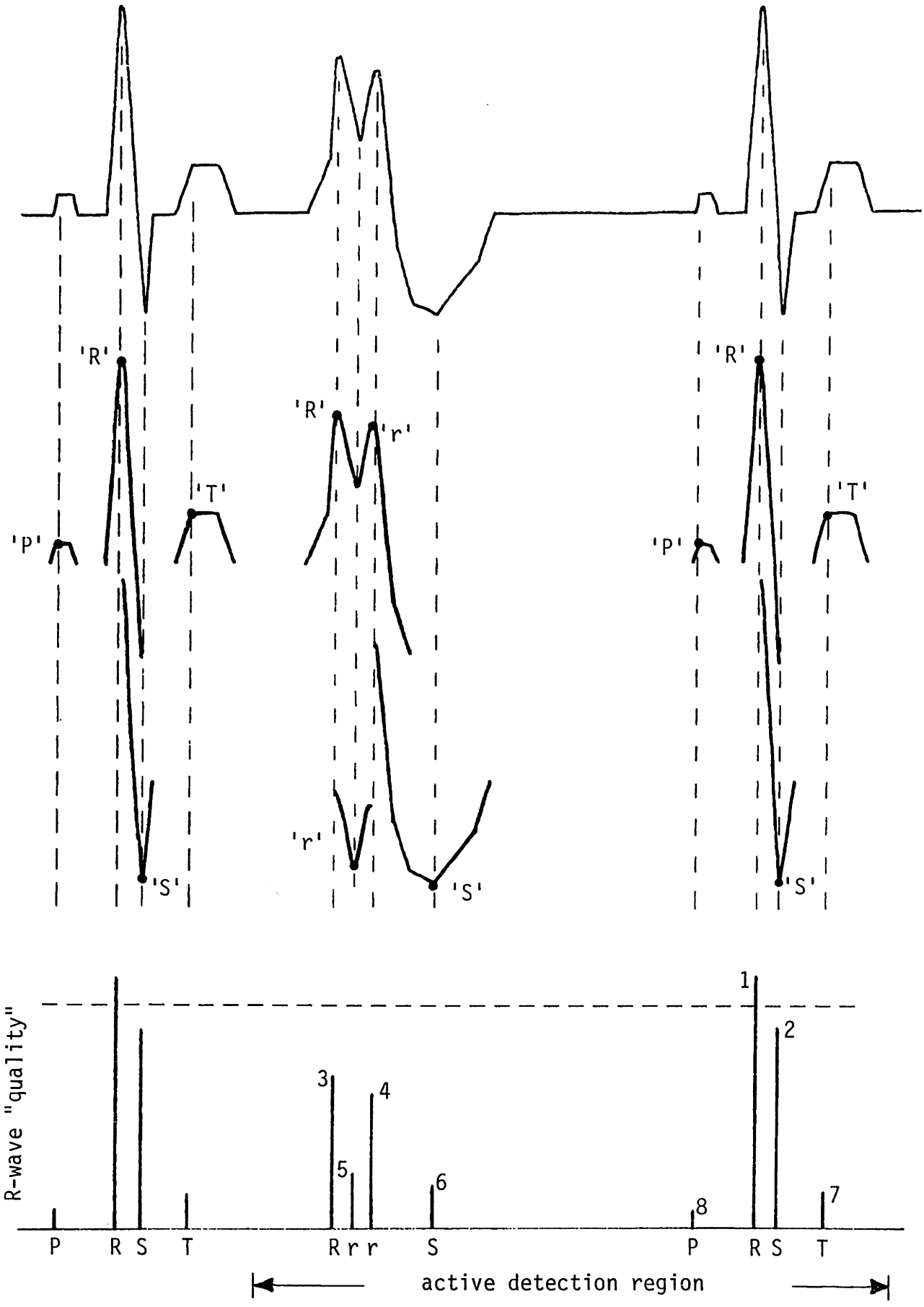
The next stage has the dual role of event detection (recognizing R-waves), and event delineation (by adding Q, S, and T-waves to the R-wave). Morphs in RWBUF are first tested sequentially (in time) to detect large amplitude QRS complexes, followed by a "parallel" search for possible low amplitude complexes. At the end of the detection scan most of the triangular morphs in RWBUF will be identified with labels such as P, Q, R, S, or T-waves, or the label "non-event".

### Morph search procedure

The overall event detection procedure is illustrated in Figure 5-8. The event detector starts (from where it had previously stopped), and sequentially (in time) compares the RQUAL of each morph with a very strict detection threshold (7/8 the average RQUAL of the eight previously detected QRS complexes). Any morph exceeding the threshold is automatically declared to be a R-wave.

Since the sequential detection threshold is very strict, it is likely that lower amplitude events will be missed. Thus, a second, more "parallel" detection procedure is invoked that detects low-amplitude QRS complexes, and also attaches subwaveforms to the main R-wave. An "active detection region" between the previously detected R-wave and the current (sequentially) detected R-wave is established, and all unlabelled morphs are scanned in RQUAL order (the largest first). During

Figure 5-8: Triangular Morph Detection and Labeling



the search, triangular morphs are labeled according to their contextual relationship with previously labeled morphs.

The primary benefit of the RQUAL ordered search is that morphs with the highest signal to noise ratio are detected and labeled first. Thus, lower signal to noise ratio morphs (lower RQUAL) can be labeled in the context of previously detected and labeled morphs, allowing one to make extensive use of relations between them.

Morph labelling

The rules for morph labelling are based on the relationship of the candidate morph "X" and the two nearest R-waves "R" that surround it.

E R I X I R E

where: X is the candidate morph
R are previously detected R-waves
I is an interval of time (if none, morphs overlap)
E is the end of the buffer

Eight different relationships between X, R, I, and E are possible:

| Case | Description and Action |
|---|---|
| RIXIR | Candidate X between two R-waves: Invoke RXRDET, and change "possible R-wave" to "non-event". |
| RIXE | Candidate X near end of buffer: Invoke RXRDET, and if "possible R-wave", wait for more data. |
| EXR | Candidate X overlaps R-wave after it: |
| RIXR | Candidate X overlaps R-wave after it: Label morph as a "Q-wave". |
| RXE | Candidate X overlaps R-wave before it: |
| RXIR | Candidate X overlaps R-wave before it: Label morph as a "S-wave". |
| EXE | No R-waves in the vicinity of candidate X, |
| EXIR | or candidate follows a long pause: Label X an R-wave if XRQUAL > (RQMIN), otherwise label it a "non-event". |

The routine RXRDET processes candidate morphs that might be potential R-waves. RXRDET labels the candidate morph with one of the five following labels:

(1) definite R-wave
(2) possible R-wave:
    if RIXIR case, the candidate is a "non-event"
    if RIXE case, need more data after candidate
(3) possible P-wave
(4) possible T-wave
(5) non-event

The test is performed in two steps. First, the candidate morph is tested according to the decision boundaries shown in Figure 5-9a. The principal function of this test is to determine whether the candidate is a T-wave, depending on how close the candidate morph is near the preceeding R-wave. Other possible final classifications at this point are "definite R-wave" or "non-event".

If the test in Figure 5-9a suggests a "possible R-wave", then the test depicted in Figure 5-9b is used to determine whether or not the candidate morph is an R-wave. This test is based on the interval of time between the two R-waves that surround the candidate morph. If the interval is longer than the predicted R-R interval, the detection threshold is lowered to accept lower amplitude complexes, such as low amplitude PVCs. If a "possible R-wave" classification is returned for reasons of insufficient data after the candidate morph, event detection is suspended until more data is available.

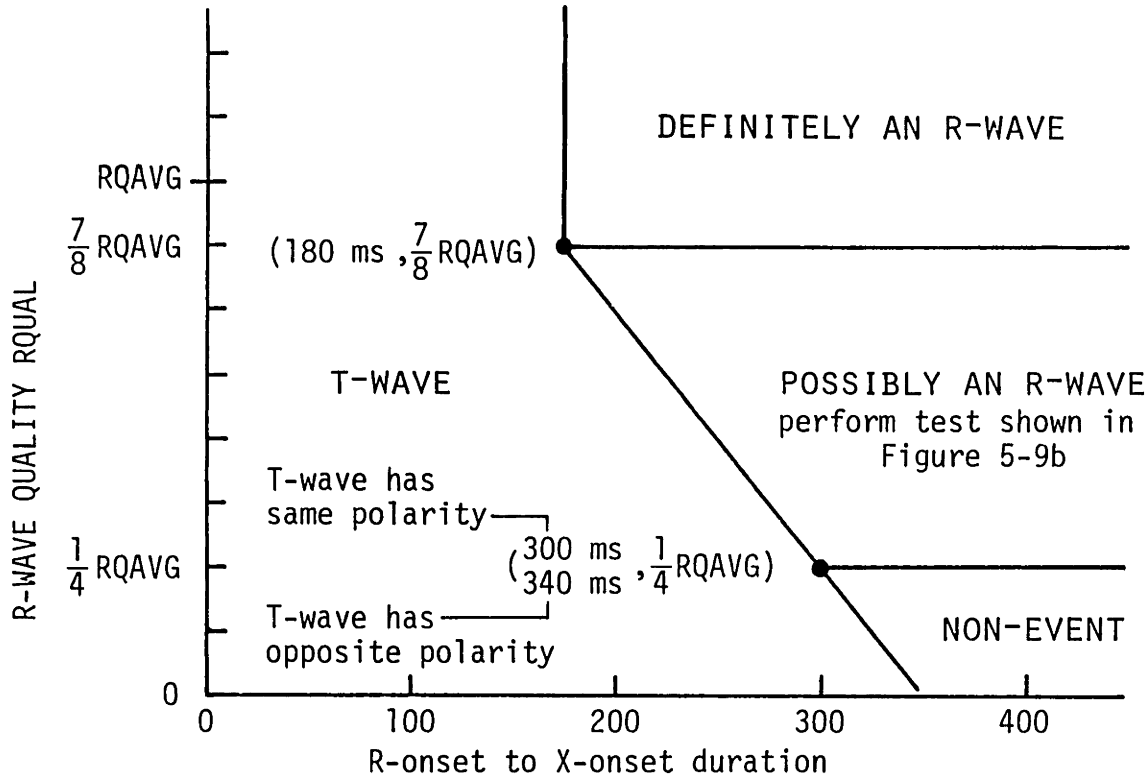



Figure 5-9a: T-wave Detection Thresholds

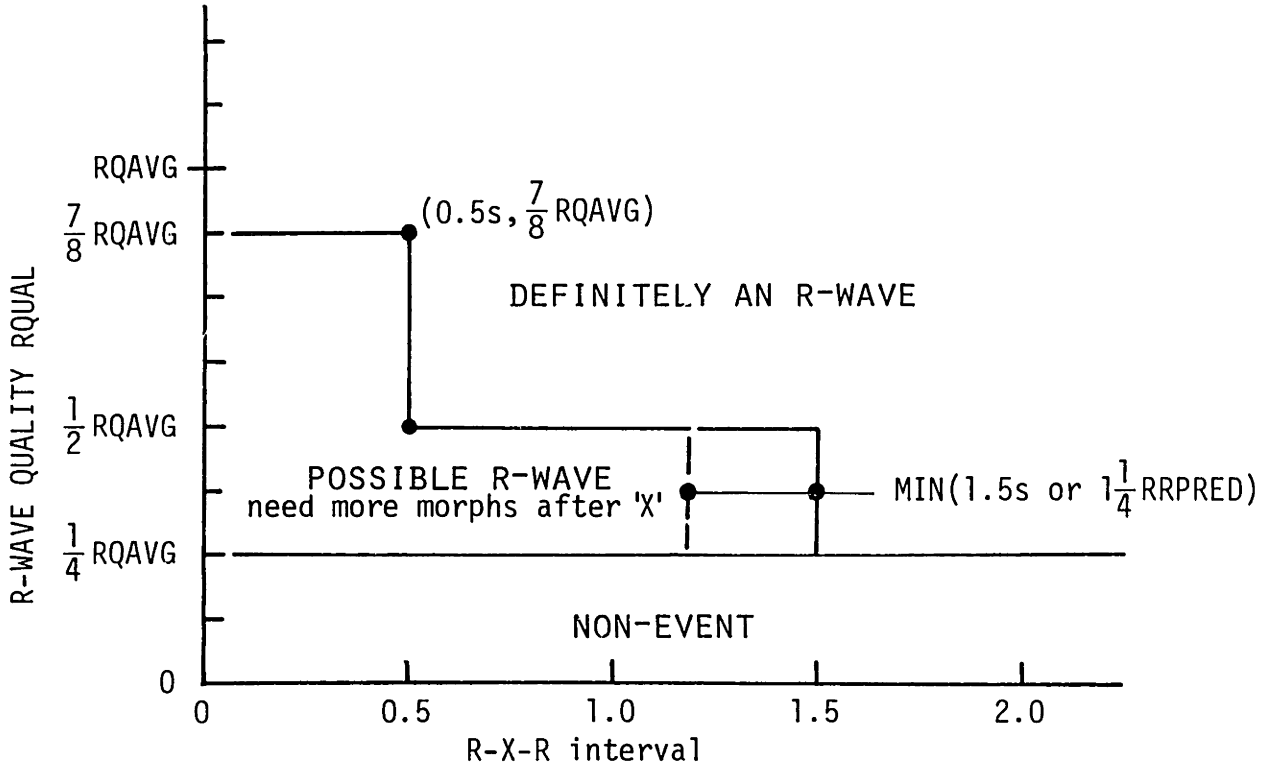



Figure 5-9b: R-X-R Interval Decision Boundaries

## 5.5 Discussion and Summary

The ECG processing steps that were detailed in the preceding sections are summarized here. Figure 5-10 reviews the major processing stages using a hardcopy plotter "snapshot" of the algorithm. The first three traces illustrate the compressor stage (a-c), followed by the parsing of the waveform into triangular morphs (d). Associated with each morph is an RQUAL value, shown in trace (e). The RQUAL sequence is then scanned by the event detector, producing the string of QRS complexes depicted in trace (f). It is interesting to note that the "connecting ligament" between the two PVCs of the couplet was rejected as a QRS complex. This is a result of the "top-down" morph search that first labeled the two PVCs of the couplet, followed by the labeling of the ligament. Here, the ligament is rejected as a QRS complex since it overlaps (in time) two QRS complexes that were already labeled.

The major data structures used in the monitoring algorithm are depicted in Figure 5-11. The highest level waveform representation is the triangular morph, which links to the segment buffer, which in turn links to the raw data. At any point of the chain, processing stages can deal with the waveform description that is the most suited for it. For example, various waveform displays must have access to the raw data and the compressed data. On the other hand, the event detector need only deal with the triangular morph representation.

Figure 5-10a
Incoming
ECG

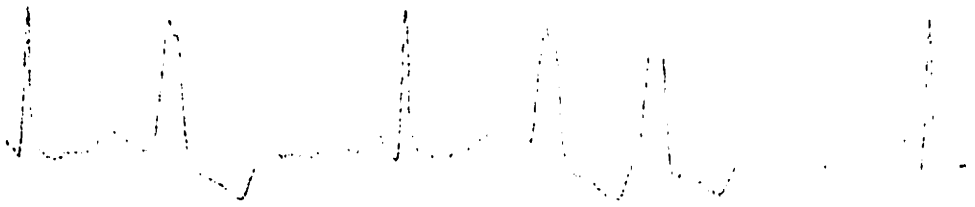


Figure 5-10b
ECG and
Segments

Figure 5-10c
Segments

Figure 5-10d
Triangular
Morphs

Figure 5-10e
R-wave
"quality"

Figure 5-10f
QRS
Complexes

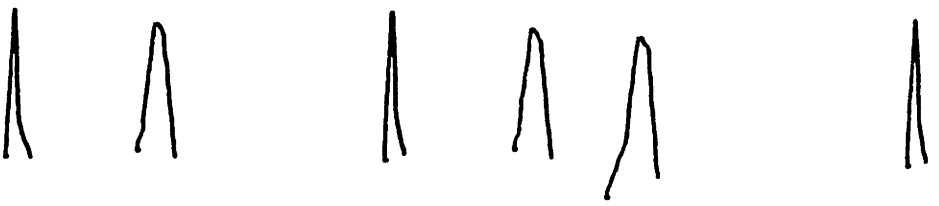
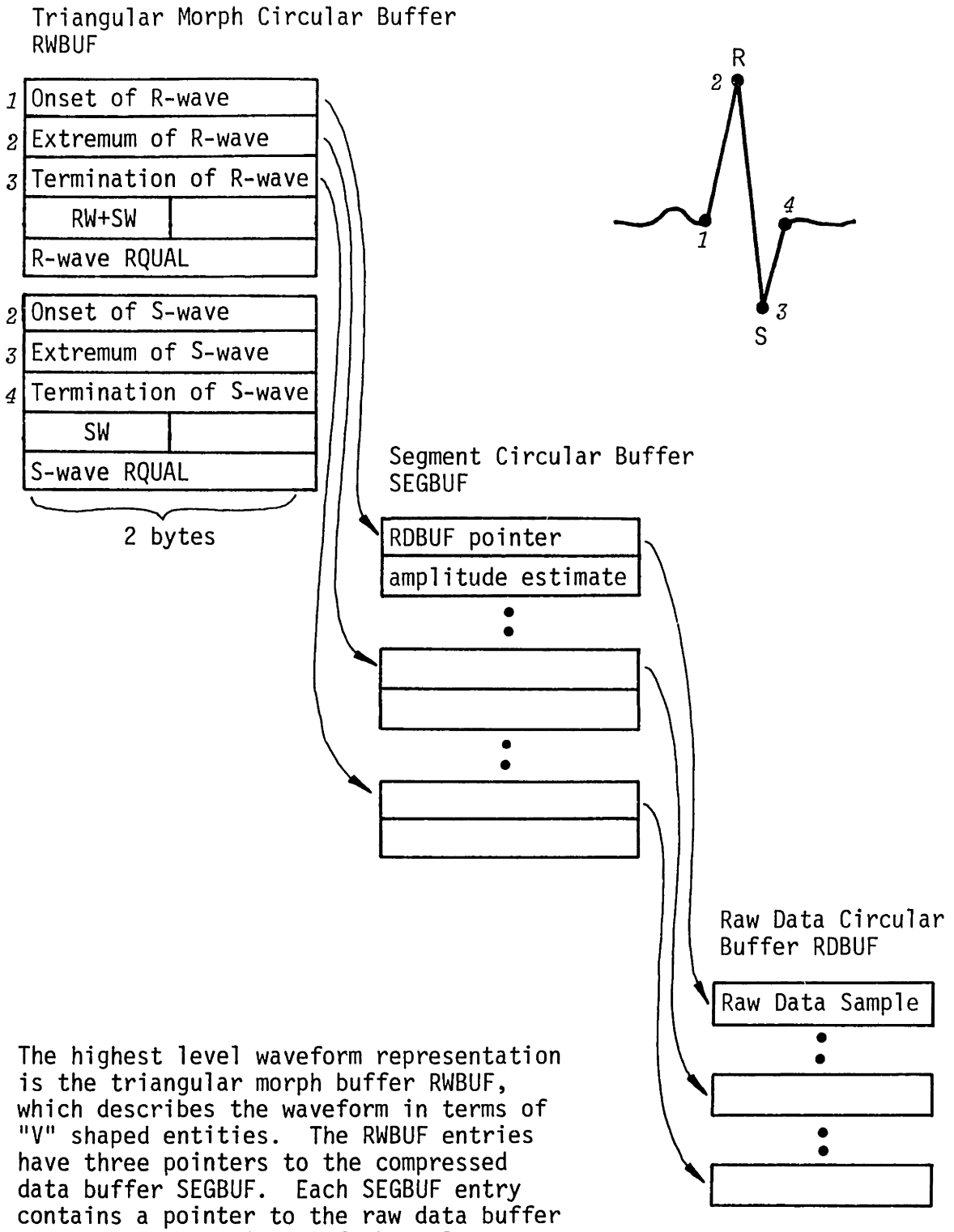


Figure 5-10: ECG Processing Stages




The highest level waveform representation is the triangular morph buffer RWBUF, which describes the waveform in terms of "V" shaped entities. The RWBUF entries have three pointers to the compressed data buffer SEGBUF. Each SEGBUF entry contains a pointer to the raw data buffer RDBUF, and an estimate of the voltage amplitude at that point in time.

Figure 5-11: Major Data Structures

# 6. MORPHOLOGY CLASSIFICATION

After the QRS complex has been detected and delineated, its morphology and timing must be characterized for final classification. Morphologic information, combined with timing information, is used to determine the final classification of the QRS complex.

Morphological characterization can improve algorithm performance in several areas. First, it provides a mechanism for classifying beats in the context of other beats. This is usually done by comparing the candidate QRS morphology with examples of morphologies seen earlier. By using average features based on beats of similar morphology, it is possible to make a more accurate determination of the most probable physiologic origin of the beat. For example, the bedside monitor uses average prematurity and width of all complexes in a family in determining the final beat label, rather than relying on measurements based on individual beats. Second, the set of ectopic beats can be subdivided into classes (according to the presumed site of ectopic focus) in monitoring situations where multiform VPBs are present. Third, morphologic analysis can be used to reduce false detection of T-waves as PVCs, and to reject artifacts that are non-physiological in nature.

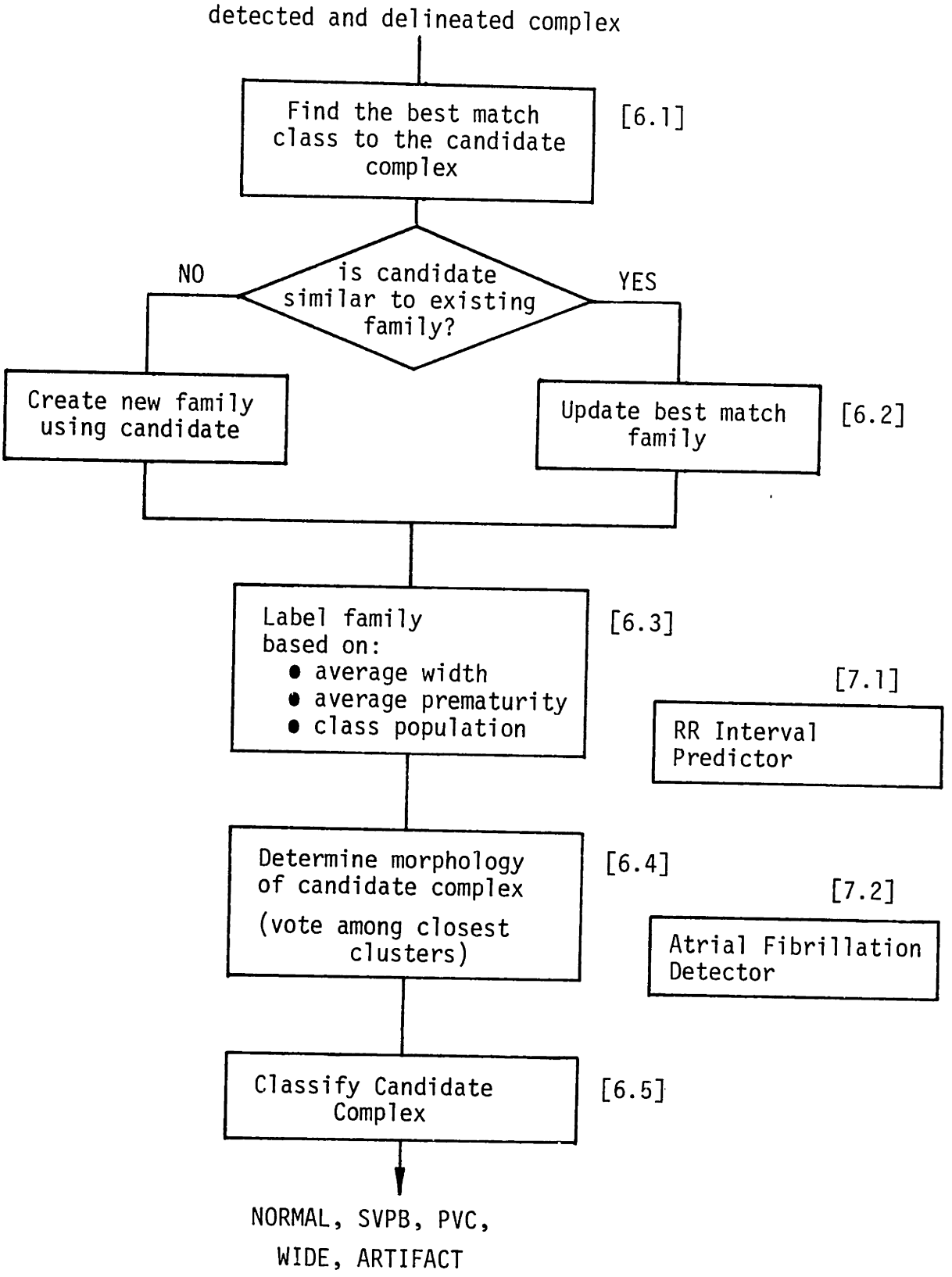



Figure 6-1: QRS Classification Procedure

The overall procedure for beat classification is shown in Figure 6-1. The first step is to compare the morphology of the candidate QRS complex with a catalogue of morphologies that were seen earlier. This requires the computation of a distance measure that indicates the difference between two morphologies. If the candidate complex is similar to an existing family, then the candidate complex is incorporated into the family. If the candidate is not similar to any of the families, then it becomes the basis for starting a new family.

The average prematurity and width of all complexes that belong to the family, and the population of the family, are used to assign its final label. The set of possible labels includes:

| | |
|---|---|
| NORMAL | not premature or wide (presumably supraventricular) |
| PVC | premature and wide (presumably ventricular ectopic) |
| SVPB | premature, but not wide |
| WIDE | not premature, but wider than normals |
| ARTIFACT | very narrow, likely to be artifact |

The prematurity of the candidate beat is based on a predicted RR interval computed elsewhere in the algorithm. Width thresholds are based on running average histograms of QRS width, and thus can adapt to individual patients.

The next step is to determine the morphology of the candidate complex, which generally will be identical to the family that it best matches. There are two exceptions to this rule, however. The first exception is that the candidate will be labelled NORMAL if it is sufficiently similar to a NORMAL family, even though it may be more similar to another family with a different label (e.g. PVC). The second exception handles cases

where the algorithm is uncertain of the label of the best match family, and instead looks at less similar families that are more likely to have accurate labels.

The final classification of the complex is based on its morphology and rhythmic context. If a beat has normal morphology, it is declared NORMAL or SVPB, depending on its prematurity. If a beat matches a ventricular ectopic morphology, it is declared a PVC, regardless of its timing.

The prematurity and rhythmic context of the candidate beat are determined by program stages described in the next chapter. The major functions of the timing stages are (1) to determine the prematurity of the candidate complex by comparing it with a predicted R-R interval, and (2) to determine whether the beat is interpolated. Another procedure detects whether atrial fibrillation is present, and modifies several morphology classifier thresholds to deal with the increased likelihood of aberrated supraventricular beats during atrial fibrillation.

## 6.1 Feature Extraction and Distance Measure

A suitable measure of similarity is required for morphologic comparison. A similarity measure is based on (1) the feature set used to characterize QRS morphology, and (2) the "distance" metric used to compare feature vectors. Although many feature sets and distance measures have been applied to ECG analysis, two techniques have enjoyed especially widespread use.

One similarity measure, commonly termed "template matching" or "correlation", uses data samples centered about the QRS detection point as the feature vector. The sampled data vectors are compared using a linear or squared error norm, or the correlation coefficient. Correlation has a number of potential advantages:

(1) it is an integrative measure, and thus is relatively insensitive to noise,

(2) it uses sample values, and not noise sensitive measurements, and

(3) it does not require accurate delineation of QRS onset and termination.

On the other hand,

(1) correlation measures are sensitive to waveform alignment, and

(2) are computationally intensive, especially if waveforms are time-shifted to maximize the degree of match.

Another popular methodology, termed "heuristic feature extraction", describes QRS morphology in terms of its height, width, area, and other features. The principal advantages of feature extraction are that

(1) it significantly reduces the dimensionality of the waveform, permitting the use of more sophisticated distance measures and classifiers, and

(2) time alignment of waveforms is not required, since most "heuristic" features are independent of a time origin.

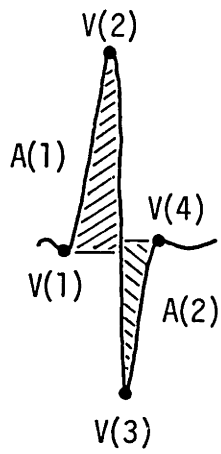
For these reasons, feature extraction is less computationally demanding than correlation. However, feature extraction has several drawbacks:

(1) features such as QRS width and height are not "robust", since they are sensitive to noise and delineator error, and

(2) grossly different morphologies can (with some feature sets) be represented with the same vector features.

The Feature Set used by the Bedside Monitor

The bedside monitor uses a QRS morphology feature set that benefits from some of the advantages of both correlation and feature extraction. Each QRS complex is described in terms of its triangular morphs (which roughly correspond to classical Q, R, and S waves). For example, a monophasic QRS complex is represented as a single triangular morph, whereas a biphasic complex (shown to the right) is represented by two triangular

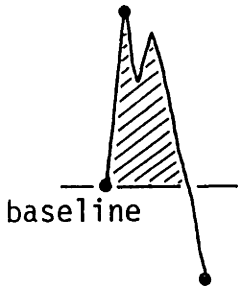


morphs. One feature vector consists of the time-ordered sequence of onset, extrema, and termination voltage amplitudes V(i), and the other feature vector consists of the time-ordered sequence of triangular morph areas A(i).

This representation arises quite naturally from the triangular morph description of the QRS complex. The principal advantages of this representation are that the voltage and area sequences preserve the temporal character of the QRS complex that would otherwise be lost if only "envelope" features such as height, area, and width were used. This representation still retains the principal benefits of feature extraction, which include a substantial reduction in the dimensionality of the waveform, and a time-independent feature set that does not require the determination of a fiducial point, nor alignment of waveforms for peak degree-of-match.

There are other advantages as well. The triangular morph representation allows the classifier to "delete" morphs (suspected noise or artifact) in order to improve the degree of similarity. Also, to the extent that inversibility is desirable, one can roughly reconstruct the original waveform from its feature vectors. The principal disadvantage of this representation is that the onset and termination voltages and areas are not robust measures, and are sensitive to delineator error.

The SEGMENT voltage amplitude estimates, rather than the raw data values, are used for the voltage amplitude array V(i). Since the compressor returns a voltage estimate that is the average of the upper and lower bounds of the ZOI aperture, the voltage estimate is fairly insensitive to the level of low-amplitude, high-frequency noise. The area values are computed by integrating the raw data values with respect to a baseline set at the onset or termination voltage of the triangular morph that is closest to the extremum (see adjacent figure). The area measure gives a rough indication of the width of the triangular morph. The raw data values, rather than the segment approximation, are integrated to compute the area since it is computationally faster, and minimizes the effects of high frequency artifact.




The voltage and area sequences are computed for every candidate complex, and are saved as a temporary entry in the morphology catalogue. If a complex does not have a Q and/or S-wave, the voltage amplitudes V(i) for the missing morph(s) are set to the corresponding onset and/or termination voltage of the R-wave, and the missing morph area(s) are set to zero.

QRS Complex Alignment

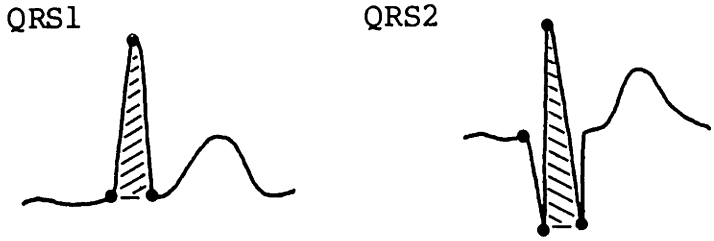
The distance measure computation requires the alignment of the voltage and area sequences for the best degree of match. Each QRS complex is described in terms of an R-wave (with a positive or negative peak), and possible Q and S-waves (with opposite polarity to the R-wave). The alignment procedure first time-shifts the complexes to match the polarity of the triangular morphs.

There are two cases to consider. In the first case, the two R-waves have the same polarity and thus the voltage and area sequences are properly aligned. The next step is to determine whether the Q and S-waves should be included in the distance computation. Two choices are possible: use the union (logical-OR) of the Q, R, and S-waves of the two complexes, or use only the morphs in common (logical-AND). This is illustrated below:

| | | |
|---|---|---|
| | +/- - R S | RS complex |
| | +/- Q R - | QR complex |
| | --------- | |
| logical-OR | +/- Q R S | the union of morphs |
| logical-AND | +/- - R - | only the morphs in common |

The logical-OR of the morph bits is the most "complete" waveform comparison, and tends to increase the distance value by including the greatest number of morphs. The logical-AND of the morph bits favors the smallest distance measure, since it removes triangular morphs that are not in common. Although the logical-AND alignment tends to remove triangular morphs due to

artifact and noise, it can also remove morphs that are physiologically significant. This is illustrated in the figure below, where the complexes would be considered similar with the logical-AND of the morph bits (since the Q and S-waves of QRS2 are "removed"), but are substantially different waveforms. For this reason, the logical-OR alignment is used.




The two morph alignments (AND and OR) were compared using selected data base tapes that represent a relatively difficult cross section of ECG waveforms. The results are shown below, where:
fn (false negatives) number of PVCs labeled NORMAL,
fp (false positives) number of NORMALs labeled PVC,
tp (true positives) number of correctly identified PVCs,
tn (true negatives) number of correctly identified NORMALs.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Date: | 12/6/80 | | | | | | |
| mod: | 'Logical-OR' alignment | | | | | | |
| Tape 106 | fn | 118 | fp | 2 | tp 398 | tn 1456 | + |
| Tape 203 | fn | 133 | fp | 7 | tp 297 | tn 2459 | + |
| Tape 207 | fn | 22 | fp | 5 | tp 53 | tn 1622 | + |
| Tape 214 | fn | 109 | fp | 1 | tp 143 | tn 1952 | + |
| Tape 223 | fn | 247 | fp | 10 | tp 225 | tn 2057 | + |
| Total | fn | 629 | fp | 25 | tp 1116 | tn 9546 | |
| mod: | 'Logical-AND' alignment | | | | | | |
| Tape 106 | fn | 136 | fp | 2 | tp 382 | tn 1456 | + |
| Tape 203 | fn | 122 | fp | 13 | tp 309 | tn 2455 | + |
| Tape 207 | fn | 22 | fp | 6 | tp 53 | tn 1620 | + |
| Tape 214 | fn | 101 | fp | 2 | tp 151 | tn 1951 | + |
| Tape 223 | fn | 222 | fp | 4 | tp 250 | tn 2063 | + |
| Total | fn | 603 | fp | 27 | tp 1145 | tn 9545 | |

The logical-AND performed somewhat better, but not significantly so. The lower overall false negative rate is probably due to the smaller number of families that are created using the logical-AND alignment. However, the logical-OR alignment is used in the bedside monitor, since it retains the most "complete" morphology description of the complexes.

The second case occurs when the two QRS complexes have R-waves with different polarity. To align them, the candidate is shifted to the left and right. If matches are possible with both alignments, a preliminary distance calculation is performed which compares the area sequences for both alignments, and the alignment with the minimum distance is selected.

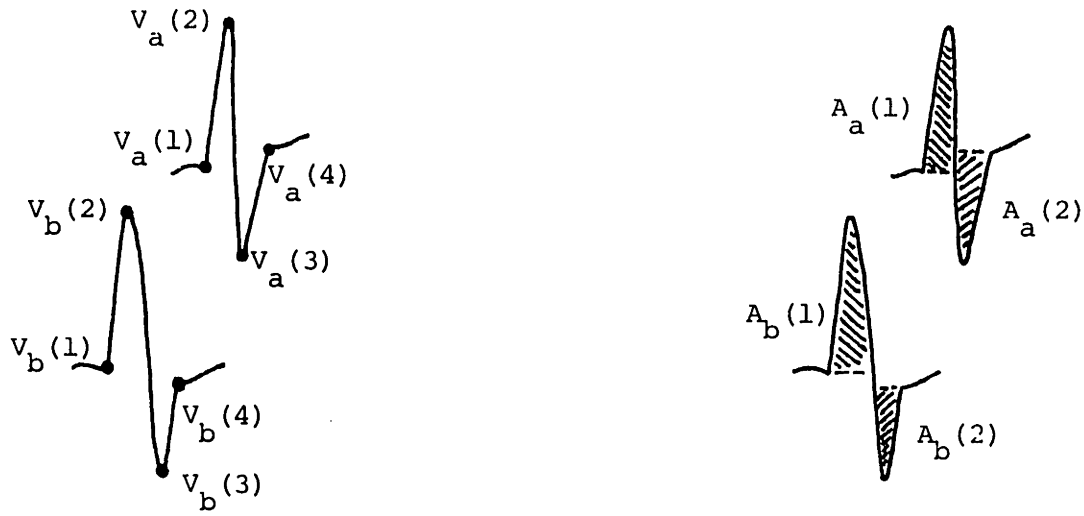
| | | | |
|---|---|---|---|
| Complex A: | +/- | Q R S | shift complex to right, |
| Complex B: | -/+ | Q R S | compute preliminary distance, |
| Complex A: | +/- | Q R S | shift complex to left, |
| Complex B: | -/+ | Q R S | compute preliminary distance, and pick the best alignment. |
| Complex A: | +/- | - R - | In some cases, |
| Complex B: | -/+ | - R - | no match is possible. |

The end result is the best possible alignment of the triangular morphs for the distance calculation described in the next section.

QRS Morphology Distance Measure

QRS morphology is compared by calculating the difference between the voltage V(i) and area A(i) sequences of the aligned complexes. The distance measure is based on the "normalized vertex error" and "normalized area error", illustrated in Figure 6-2.

The normalized vertex error is the sum of the absolute values of the differences between the onset, extrema, and termination voltages of the two complexes. The value "B" in the numerator is the median voltage difference (which minimizes numerator) and is designed to remove any DC offset between the




Extrema Amplitudes  Triangular Morph Areas

Figure 6-2: QRS Morphology Feature Set and Distance Measure

QRS morphology is described by
(1) the onset, extrema, and offset voltage amplitudes, and
(2) the areas of the triangular morphs of the complex.

Morphologies are compared by computing the two normalized errors:

Normalized Vertex Error =

$$\frac{\sum_{i=1}^{\Delta+2} | V_a(i) - V_b(i) + B |}{\frac{1}{2}\{ \sum_{i=1}^{\Delta+1} |V_a(i+1)-V_a(i)| + \sum_{i=1}^{\Delta+1} |V_b(i+1)-V_b(i)| \} + V_o}$$

B is computed to minimize the numerator

total 'perimeter'

Normalized Area Error =

$$\frac{\sum_{i=1}^{\Delta} | A_a(i) - A_b(i) |}{\frac{1}{2}\{ \sum_{i=1}^{\Delta} A_a(i) + \sum_{i=1}^{\Delta} A_b(i) \} + A_o}$$

area error

total area

where

$\Delta$ is the number of triangular morphs ( $\Delta$ = 2, above)
V is voltage amplitude
A is area

two complexes. This is similar to subracting the mean (arithmetic average) difference from two arrays that are compared with a squared error metric. The error term is normalized by the average total perimeter of the reference and candidate complexes.

The normalized area error is the sum of the absolute differences of the areas of the corresponding morphs, and is normalized by the total average area of the two complexes.

The additive terms $V_o$ and $A_o$ in the denominator are added to make the distance measure less sensitive to morphologic variations of extremely low amplitude complexes. The current value for $V_o$ is 3 (vertices) * 50 uV, and $A_o$ is 50 msec * 50 uV. For normal complex amplitudes, $V_o$ and $A_o$ can be considered to be zero.

The final distance measure is the sum of the squares of the normalized vertex and area errors. If an overflow occurs while computing the normalized errors, or adding their squares, the distance is limited to less than one. The fractional part of the distance is expressed as a 16-bit unsigned integer. Thus, the distance between two roughly similar complexes (25% feature variation) would be

(1/4)**2 + (1/4)**2 = 1/8 = 8192./65536.

The fractional part is 8192 (decimal), or 2000H (hexadecimal).

The principal reason for selecting this distance measure was its computational efficiency. If there are N morphs to be compared, the distance calculation requires

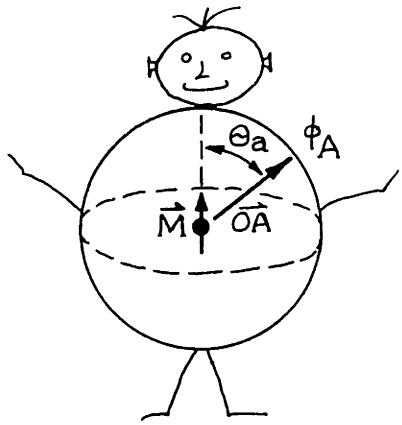
7N+4 adds,
4N+4 subtracts,
4N+4 absolute values,
2 divides,
2 multiplies, and a
sort to find the median entry of an N+2 integer array.

The add, subtract, and absolute value operators can be computed rapidly on the 8080 microprocessor, whereas the divide and multiply are substantially slower software routines. Using a linear, rather than squared, error norm greatly reduces the number of time-consuming multiplies.

Another important property of the distance measure is that it is normalized with respect to the overall amplitude of the QRS complexes, rather than (1) the statistical variance of the features, or (2) fixed normalization values (of a "typical" QRS of arbitrary amplitude and duration). The principal reason for avoiding a statistically normalized distance measure (e.g. Mahalanobis) were the issues of cluster overlap and drift. For example, an ECG record with fusion beats will often exhibit morphologies intermediate to normals and PVCs. An unsupervised adaptive clustering algorithm could "accidently" include some fusion beats in a normal cluster. The observed variance would increase, making it more likely that additional fusion beats would be included. If no absolute limit is set on the variance, the normal cluster could easily "grow" to include normals, fusions, and even PVCs.

A fixed normalization (with respect to an arbitrary size QRS complex) probably would not be the most effective normalization for physiological variations of QRS size. When a patient moves around in bed, for example, the heart can move with respect to the monitoring electrodes. If we model the heart as a current dipole embedded in a spherically shaped, linear, isotropic homogeneous conductor, the potential measured on the surface of the sphere (chest) is in the form:

$$\phi_A = |\vec{M}| \cdot |\vec{OA}| \cdot \cos(\theta_a)$$




for a small perturbation of angle $d\theta$, the change in potential $d\phi_A$ is:

$$d\phi_A = -|\vec{M}| \cdot |\vec{OA}| \cdot \sin(\theta_a) d\theta.$$

This demonstrates that the change in magnitude of the QRS complex due to positional changes is roughly proportional to its amplitude. Thus, as a rough estimate of the expected variance of QRS features, the distance metric was normalized to the overall size of the QRS complex.

## 6.2 Morphology Clustering

The bedside monitor uses adaptive clustering for morphology classification. The classifier "learns" the normal and abnormal morphologies for the patient by seeking naturally occuring modes in feature space, on the assumption that beats with similar morphology will agglomerate in tightly packed clusters. The classifier uses a "nearest neighbor" algorithm for morphology clustering. If the distance between the candidate complex and the nearest cluster is small, the candidate complex is considered to be a member of the cluster, and the cluster description is updated to include the features of the new member. If the candidate complex is not similar to any cluster, then it defines the beginning of a new cluster.

The clustering procedure used by the bedside monitor is outlined in Figure 6-3, and is discussed in detail in the following sections. The topics are:

(1) the characterization of each cluster,

(2) the determination of an adaptive similarity threshold,

(3) the incorporation of the candidate into the cluster it matches, and

(4) the management of dormant clusters.

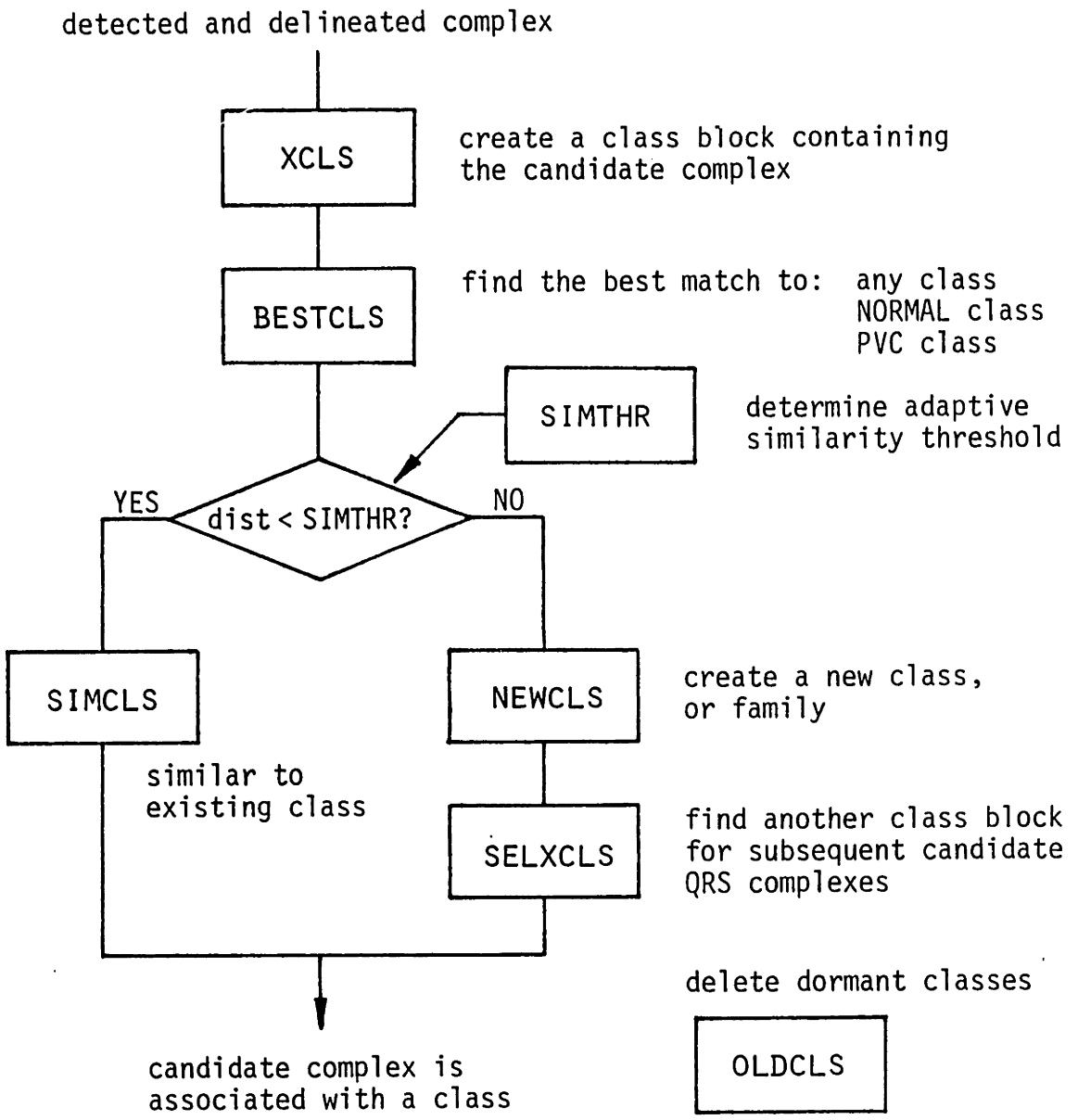



Figure 6-3: Morphology Clustering Procedure

Creating a Candidate Morphology Class Block -- XCLS

One of the MXNCLS [16] class blocks is always reserved for the candidate complex. At the time the candidate class block is created, the triangular morph vertex and area values are computed, and the QRS triangular morph structure is determined. Each class block has a 20 byte header followed by a three 8-byte triangular morph descriptors shown below:

Morphology Class Block

CLSLBL Class label <1 byte>
  NULL [0] the class is inactive
  NORMAL normal (supraventricular) complexes
  PVC ventricular ectopic (premature and wide) complexes
  SVPB premature, narrow complexes
  WIDE probable abberrant morphology (on-time and wide)
  ARTIFACT probable artifact (very narrow)
  UNKNOWN unclassifiable morphology
CLSNUM Class number <1 byte>
  Increments (mod 256) for each new class
CLSZOI ZOI aperture value when the class was created <1 byte>
QRSTM QRS triangular morph structure <1 byte>
  Includes the polarity (+/-) of the R-wave,
  and indicates presence of Q and S morphs
CTIME Time of class creation <4 bytes>
MTIME Time class was most recently matched <4 bytes>
POP Class population <2 bytes>
  Incremented whenever a candidate complex matches the class.
  Initially 1.
WIDTH Width (running average) <2 bytes>
PREM Prematurity-Ratio (running average) <2 bytes>
XDST Candidate's distance to this class <2 bytes, temporary>

Each class block has a triplet (Q, R, and S) of triangular morphs, each with the following descriptors:

ONVAL Onset voltage value <2 bytes>
PKVAL Peak (extremum) voltage value <2 bytes>
OFVAL Offset (termination) voltage value <2 bytes>
AREA Morph area <2 bytes>

Searching for the best match class -- BESTCLS

After the candidate class block has been initialized, it is then compared with all other active class blocks in the catalogue. During the search, inactive class blocks are skipped, and all other class blocks are compared using the distance measure described in the previous section. The following values are determined at the conclusion of the search:

(BSTDST) Best (smallest) distance to any class
(BSTNDST) Best distance to the nearest NORMAL class (with population greater than 5)
(BSTVDST) Best distance to the nearest PVC class (if there are no PVC classes, return largest distance).

Similarity distance threshold -- SIMTHR

After the distance to the nearest class (of any type) has been determined, the next step is to determine whether or not the candidate is sufficiently similar to the class to be included in it. This decision is made by comparing the distance to a threshold: if the distance is smaller, the candidate is considered similar; otherwise, the candidate is considered to be a new morphology.

The similarity threshold can have a significant impact on classifier performance. A small distance threshold finely "samples" morphology space, but at the cost of requiring many small clusters to characterize it. If the distance threshold is too large, there is the danger that different morphologies will be included in the same cluster, and thus labeled incorrectly.

Early versions of the bedside monitor clustering algorithm used a fixed similarity threshold that was set at a small value (1800H) so that the monitor could distinguish fine variations of morphology. This worked fine for ECGs with relatively stable normal and ectopic morphologies, but was totally inadequate for records with moderate levels of background noise, or significant variations in QRS morphology. In some instances, well over 200 morphology classes (for a catalogue size of 16 classes) were created during a half hour period. The morphology classes were "recycled" so frequently that they never persisted long enough to aquire a sufficient population to permit them to be labeled accurately. Increasing the similarity threshold improved performance on noisy ECGs, but reduced the algorithm's ability to make fine distinctions required by ECGs that had fusion beats or PVCs with morphology similar to normals.

Thus, an adaptive similarity threshold was devised to accomodate different levels of background noise and morphology variability. The similarity threshold is a function of (1) the distance of the candidate complex to the nearest normal class, and (2) the number of active morphology classes. As illustrated in Figure 6-4a, the similarity threshold increases as the number of active classes increases (indicating substantial morphologic variability). The increased threshold reduces the probability that new classes will be created, allowing the classes already in the catalogue to attain a large population.

The similarity threshold is also increased if the distance of the candidate complex to the nearest normal class is large. Since the major goal of morphologic clustering is to accurately identify the boundary (in morphology space) between normal and ectopic morphologies, its precise definition is critical only in the region near normal complexes. Therefore, the similarity threshold is increased for classes with clearly abnormal morphologies. This strategy doesn't tend to fill the class catalogue with normal classes, since there is usually less morphology variation for normals than ectopics. The effect on cluster distribution is shown in Figure 6-4b, which illustrates the finely "sampled" morphology space near the NORMAL-PVC boundary, and the coarser sampling farther away from it.

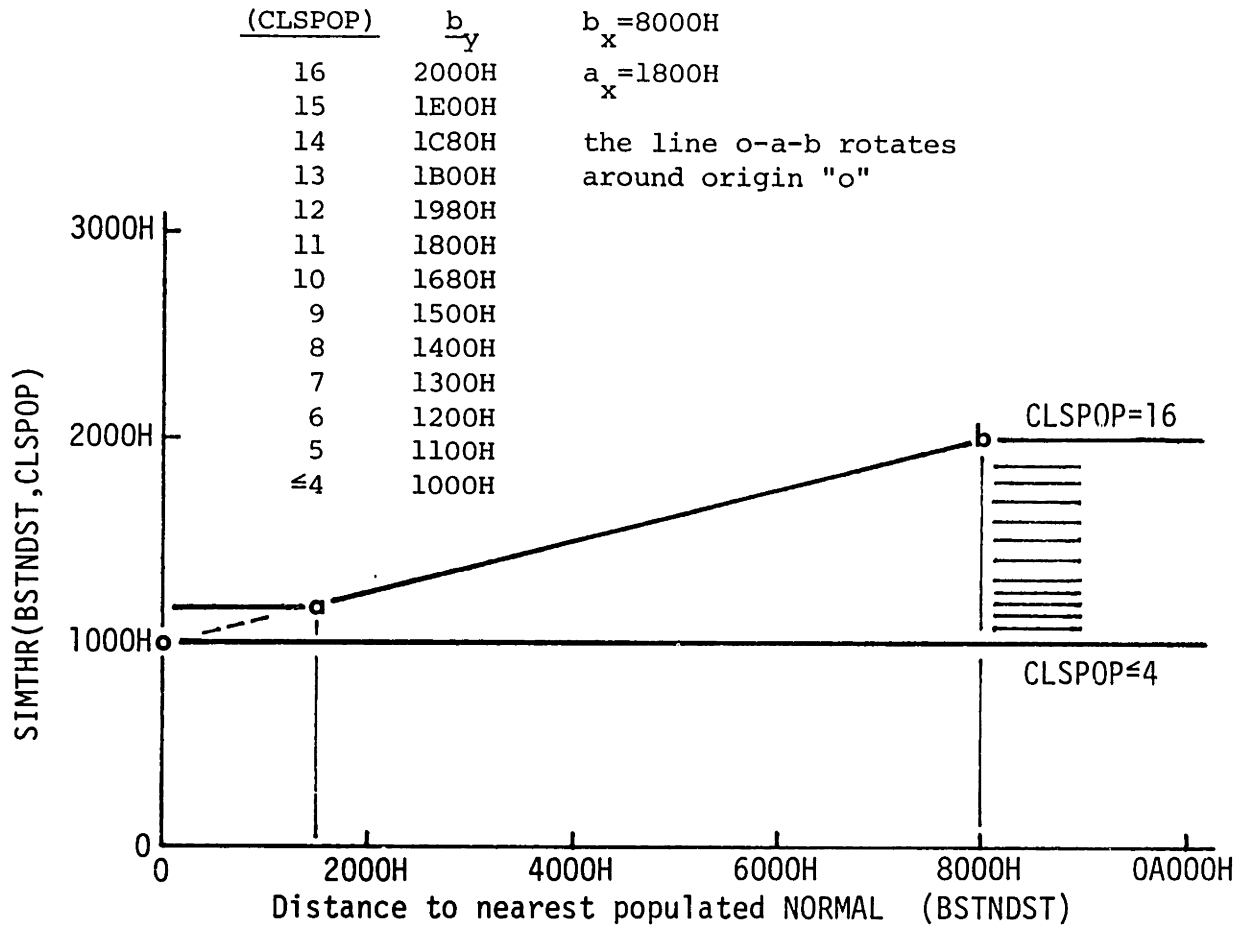



Figure 6-4a: SIMTHR similarity distance threshold

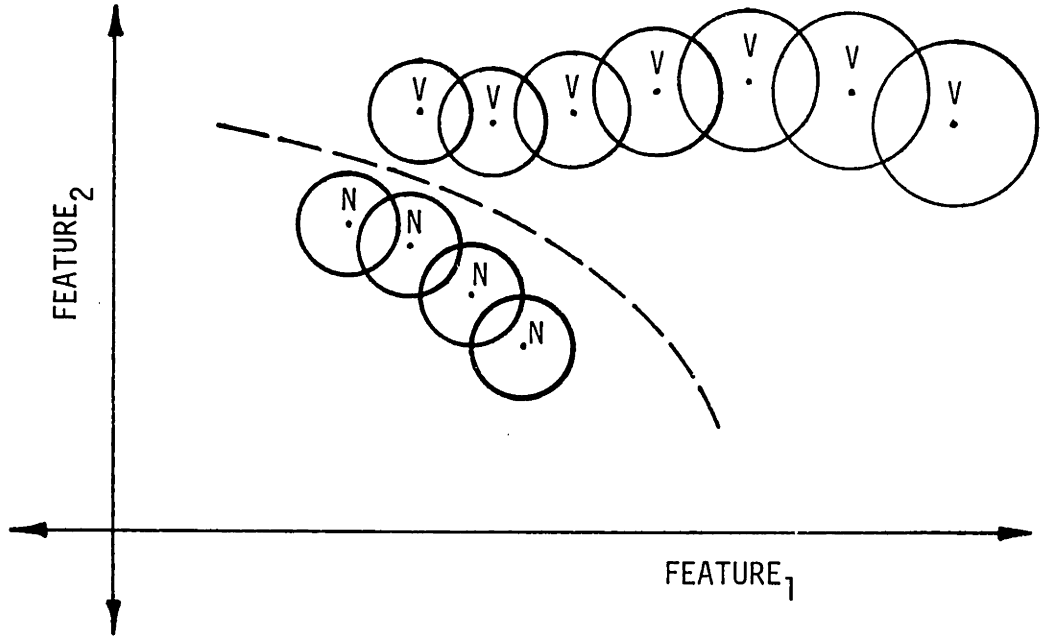



Figure 6-4b: Cluster Distribution

Several experiments with the annotated ECG data base were performed to develop the adaptive similarity threshold. The false negative rate was significantly improved for tape 106, primarily because the algorithm successfully recognized a PVC morphology that was previously considered NORMAL. This is a consequence of the finer sampling of morphology space. The increase in false positives for tape 203 is within the trial-to-trial variability for this very difficult tape.

```
Date:     12/6/80
mod:      'Logical-AND' alignment.
Tape 106    fn  136  fp    2  tp  382  tn 1456  +
Tape 203    fn  122  fp   13  tp  309  tn 2455  +
Tape 207    fn   22  fp    6  tp   53  tn 1620  +
Tape 214    fn  101  fp    2  tp  151  tn 1951  +
Tape 223    fn  222  fp    4  tp  250  tn 2063  +
Total       fn  603  fp   27  tp 1145  tn 9545

Date:     12/7/80
mod:      Step Ds change:
mod:       Ds=1400H for Dn<6000H,  Ds=1600H for Dn>6000H,
mod:       NORMAL classes need pop .ge. 1 to be considered.
Tape 203    fn  136  fp    5  tp  295  tn 2467
Tape 214    fn  114  fp    1  tp    0  tn 1952

Date:     12/7/80
mod:      Linear Ds change with Dn:
mod:       Ds=1400H for Dn=2000H, Ds=1C00H for Dn=8000H
mod:       Also depends on number of active classes
mod:       Normal population must be greater than 8
Tape 106    fn  109  fp    0  tp  407  tn 1458
Tape 203    fn  138  fp   13  tp  292  tn 2451
com:      Helped somewhat on tape 106, not much (can help) on tape 203.

Date:     12/9/80
mod:      Linear Ds change with Dn:
mod:       Ds=1200H for Dn=2000H, Ds=2000H for Dn=8000H.
mod:      SELXCLS weightings modified to preserve PVC classes.
Tape 106    fn   94  fp    2  tp  422  tn 1456  +
Tape 203    fn  125  fp   23  tp  305  tn 2443  +
Tape 207    fn   21  fp    9  tp   54  tn 1617  +
Tape 214    fn  100  fp    1  tp  151  tn 1952  +
Tape 223    fn  237  fp    3  tp  234  tn 2064  +
Total       fn  577  fp   38  tp 1166  tn 9532
com:      Sensitivity improved for all tapes, but FP increased on 203.
com:      Both modifications were retained.
```

Candidate matches existing class -- SIMCLS

If the distance between the candidate complex and the nearest class is less than the similarity threshold, the candidate is considered similar to the class, and is incorporated into it. The time of last match (MTIME) is set to the current time, and the class population (POP) is incremented. The average prematurity-ratio (PREM) and width (WIDTH) are updated at a later time.

One crucial design decision was whether to update the morphology description (voltage and area sequences) of the cluster. Updating the morphology description would permit the cluster(s) to track slow morphology changes due to axis shifts and positional changes of the patient. This strategy had a number of disadvantages, however. The major drawback was class (template) drift, which was particularly acute if normal and ectopic morphologies were similar. If ectopic beats were incorrectly included in a normal cluster, the cluster template would drift somewhere in between the true normal and ectopic template, and cause classifier error. Some mechanism was necessary to "tie down" the cluster template so that it wouldn't drift too far from the first beat that established it. Thus it was decided to use the first QRS complex that created the cluster as its template, and not update the template with subsequent similar complexes. Positional changes and axis shifts would be handled instead with the creation (and deletion) of multiple clusters that would track the morphologic changes.

Candidate complex does not match an existing class -- NEWCLS

If the distance between the candidate complex and the nearest class is larger than the similarity threshold, the candidate is considered to be a new morphology. The temporary class block that was used to contain the description of the candidate morphology now becomes an "active" class, and several of the entries in the class block are updated to reflect its new status:

(1) The class label (CLSLBL) is changed from INACTIVE [0] to UNKNOWN.

(2) A class ID number (CLSNUM) is assigned to the block.

(3) The value of the ZOI aperture is saved in (CLSZOI), and serves as a measure of the signal quality at the time the class was created.

(4) The time of class creation (CTIME), and the time of last match (MTIME) are set to the current time.

(5) The class population (POP) is set to one.

Making room for new classes -- OLDCLS and SELXCLS

Since there are a finite number of class blocks available, a mechanism for removing dormant classes is required. Two procedures are used to remove class blocks. One procedure (OLDCLS) is invoked once a minute, and deletes "dormant" morphology classes. The primary reason for removing dormant classes is to minimize the number of classes that have to be searched for the best morphology match. The presence of dormant classes does not seem to have an adverse effect on classifier performance, only the computational load that it represents. A

class is considered dormant if it satisfies one of the criteria listed below:

(1) If the class is NORMAL, and has not been matched for 5 minutes. This aggressively deletes extraneous NORMAL classes due to axis shifts or positional changes.

(2) If the population (of any type of class) is one, and more than 10 minutes have elapsed. This deletes classes that are likely to be examples of artifact (which rarely match other instances of artifact).

(3) If a class of any type, and with any population, has not been matched for more than 20 minutes.

The other mechanism used to remove classes is invoked when the morphology catalogue is full, and room must be made for a new class. This is done by subroutine SELXCLS, which is called whenever a new morphology class has been created. SELXCLS first searches for any free class blocks, and if one is available, it is selected to be the new class block for candidate complexes.

If there are no free class blocks available, SELXCLS searches for the two most populated normal classes, and flags them "not to be deleted". This preserves some knowledge about the normal morphology for the patient during prolonged episodes of noise or ventricular flutter or fibrillation. During such periods of intense morphologic activity, classes are constantly being created and removed, and could easily obliterate all previous knowledge about the patient's morphology.

The benefits of keeping the two most populated classes in the morphology catalogue are illustrated with data base tape 207. This tape has several long runs of ventricular tachycardia, and several minutes of ventricular flutter. These episodes exhibited substantial morphologic variability, and rapidly exhausted the catalogue. The 10/4/80 version of the algorithm did not retain

the most populated classes, and was unable to recover properly after the run of ventricular flutter. The 10/5/80 version, however, still "remembered" what the NORMAL complex looked like for the patient, and thus had a lower false positive rate.

Date: 10/4/80
Tape 207 fn 31 fp 53 tp 44 tn 1574
obs: NORMAL classes are flushed during VT runs.

Date: 10/5/80
mod: Two most populated classes are preserved by SELXCLS.
Tape 207 fn 32 fp 11 tp 43 tn 1616

After deciding which classes should be preserved, SELXCLS then selects the most dormant class, and deletes it from the catalogue. This is done by computing a "dormant class score" for each class, and selecting the class with the highest score. The score is based on family classification, age (elapsed time since a candidate complex matched the class), and the noise level (ZOI aperture) when it was first created. The dormant class score is the product of three terms:

(class weighting coefficient) *
(class ZOI aperture) *
(class age + 1 minute).

The class weighting coefficients are:

20 for PVC classes (tends to preserve PVC classes)
90 for ARTIFACT classes (tends to discard ARTIFACT classes)
and
40 for all other classes (NORMAL, SVPB, and WIDE).

Although the equation and weighting coefficient are somewhat ad-hoc, they have several desirable properties. First, the score tends to preserve PVC classes, and remove classes suspected of being ARTIFACT. Second, if the incoming data was noisy when the class was first created, it's quite likely that the class has a template corrupted by noise, and thus should be discarded over

one that is relatively noise free. The primary indication of class dormancy, however, is its age. The one minute bias has been added so that classes created during periods of intense noise can be deleted immediately, thus preserving noise-free morphology classes.

Discussion

This concludes the description of the morphology clustering algorithm used by the bedside monitor. Again, the overall goal of morphology clustering is to group together QRS complexes with similar morphology. This is shown in Figure 6-5, which illustrates the creation and removal of clusters during a relatively noise-free 30 minute ECG record. The figure indicates, however, that the performance of a feature-extraction based clustering algorithm can be sensitive to delineation error. For example, classes 2 and 3 really represent the same ectopic focus, but were considered different by the classifier because they were delineated in slightly different ways. Classes 5 and 9 also represent the same ectopic focus, but since the "T-wave" of the PVC in class 9 was retained, it was considered a different morphology class.

One important clustering parameter is the number of morphology class blocks in the catalogue. A large number of class blocks allows the algorithm to retain more morphologic information about previous QRS complexes, and permits finer sampling of morphology space. On the other hand, the computational load and memory requirements for additional class blocks can become prohibitive. The number of class blocks used by the bedside monitor is MXNCLS [16], which works well for relatively noise free ECGs. From time to time, however, the MXNCLS was increased in an effort to improve performance, as is

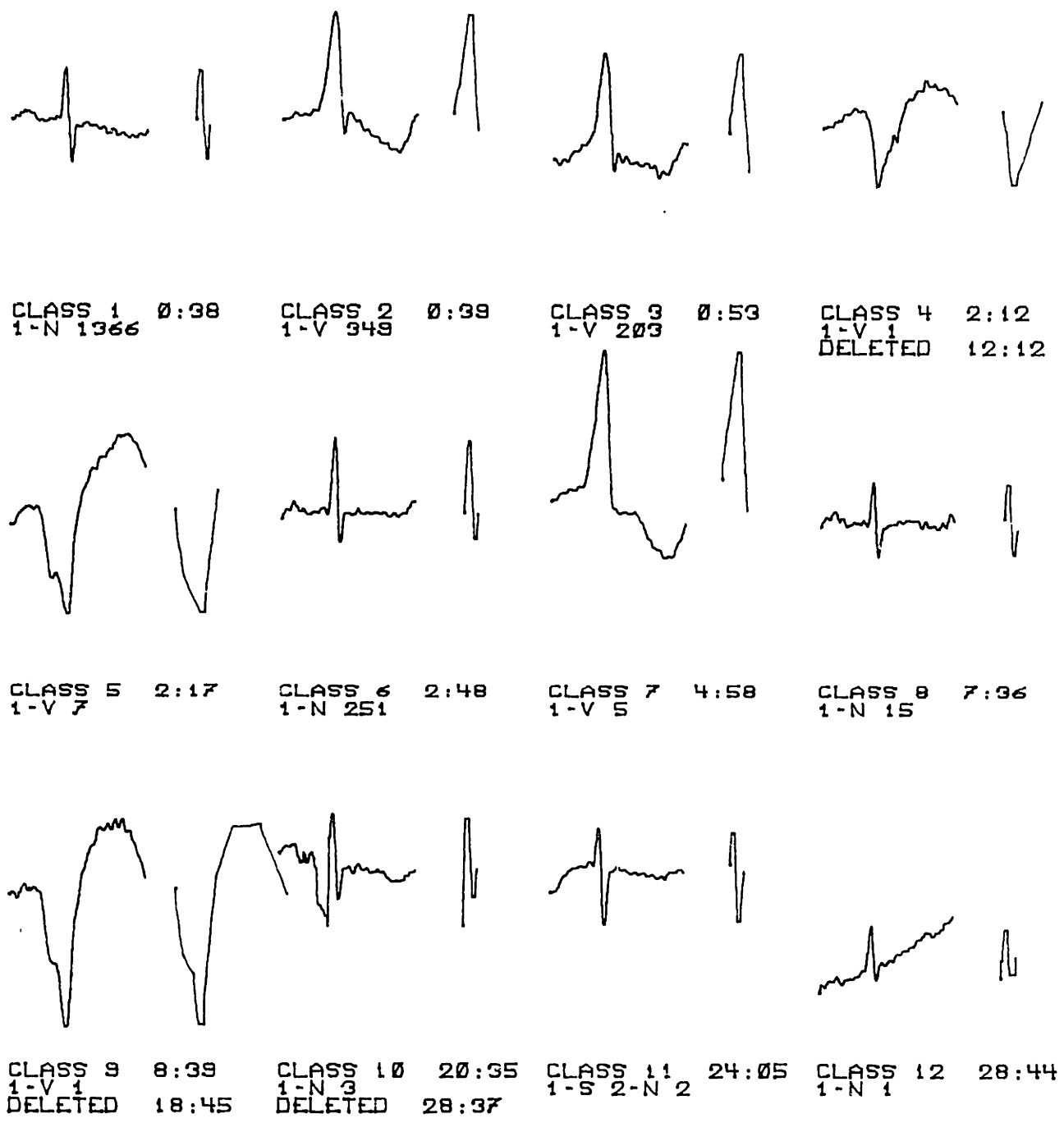



Figure 6-5: QRS Morphology Classes

This figure depicts the different QRS morphology classes that were found in a 30 minute ECG segment. For each class plot, the raw ECG and delineated QRS complex are shown. The first line lists the class number and the time it was created; the second line shows the class classification (N-normal, V-PVC) and total population at the end of 30 minutes. Classes that were deleted are also identified.

shown in the data below:

```
Date:     1/1-B/81
mod:      16 class blocks.
Tape 106    fn   52  fp    7  tp  462  tn 1457  +
Tape 200    fn   87  fp    7  tp  722  tn 1730  +
Tape 203    fn  126  fp   17  tp  304  tn 2443  +
Tape 207    fn   23  fp    8  tp   52  tn 1618  +
Tape 214    fn   92  fp    1  tp  160  tn 1952  +
Tape 223    fn  204  fp    0  tp  258  tn 2067  +
Total       fn  584  fp   40  tp 1958  tn 11267

Date:     1/1-C/81
mod:      24 class blocks.
Tape 106    fn   62  fp    8  tp  453  tn 1456  +
Tape 200    fn   94  fp    0  tp  715  tn 1737  +
Tape 203    fn  118  fp   28  tp  312  tn 2437  +
Tape 207    fn   34  fp    9  tp   40  tn 1618  +
Tape 214    fn   95  fp    3  tp  157  tn 1950  +
Tape 223    fn  215  fp    4  tp  257  tn 2063  +
Total       fn  618  fp   52  tp 1934  tn 11261
obs:      Performance is somewhat worse.
```

This data indicates that increasing the number of class blocks from 16 to 24 increased both false negatives and false positives! One possible explanation is that the rest of the algorithm has been "tuned" for 16 class blocks, and that other parameters must be adjusted to obtain improved performance with more class blocks.

## 6.3 QRS Morphology Family Identification

Once a family has been updated to include the candidate complex, the family must be identified with a label which is meaningful to the physician. Labels can be based on standard ECG terminology such as "PVC", "SVPB", "NORMAL", etc. Such labels can mislead the physician, however, since most monitoring systems only perform an incomplete analysis of the ECG (for example, most systems do not detect P-waves). Another approach is to identify complexes with "non-physiologic" labels (numbers, letters, or other codes) that are precisely defined, but at the disadvantage of being much less meaningful to the clinician.

Cluster identification can be quite difficult for an unsupervised classifier, since there are no absolute rules that can be used for labeling. Of the relatively few measures that can be used, three have enjoyed the most widespread use:

(1) The duration of the QRS complex, since the spread of depolarization from a ventricular ectopic focus is generally slower than for a normally conducted beat.

(2) The relative timing of the QRS complex, since PVCs are usually premature. Late ectopic beats are usually inhibited by an on-time, normal complex.

(3) Relative population, since the most common complex is usually the "normal" complex.

A number of conditions preclude any of the above as being totally reliable measures for identifying the physiologic origin of a family. The apparent duration of the QRS complex can be erroneously lengthened by noise that mimics Q and S waves. The presence of bundle-branch block or other intraventricular conduction defects can also result in abnormally "wide" beats. In other instances, the QRS of a VPB may appear to be narrow becauses its initial or terminal portions are isoelectric. In fact, if the normal rhythm is bundle branch block, the PVCs may actually be narrower than normal complexes. Finally, aberrated atrial premature beats may be wider than normal complexes, but are not of ventricular origin.

Prematurity can also be an unreliable indicator of physiologic origin. Supraventricular premature beats may be present, and must be compared on the basis of morphology to determine their origin. Atrial fibrillation and flutter or sinus arrhythmia can also be present, and preclude the sole use of timing measurements on individual beats.

Relative population counts of different morphologies can often be a useful indicator of the "normal" morphology of the patient, since normal beats are generally the most common. However, prolonged episodes of ventricular rhythms such as bigeminy and flutter can fool program decisions based solely on class population. Also, it is often difficult to obtain accurate population counts with classifiers that have a finite number of morphology class blocks, since classes can be deleted at any time

to make room for new ones.

Bedside monitor family label "feature set"

The bedside monitor identifies morphology clusters with one of five labels: NORMAL, PVC, WIDE, SVPB, or ARTIFACT. These classifications are based on the average prematurity-ratio and width of complexes that belong to the cluster, and the population count of the cluster. The average prematurity-ratio and width are computed as running averages in order to track the most recent beats in the cluster.

The first step is to compute (CLSPOP), which is the class population, but with an upper bound that establishes the weighting of the candidate's prematurity and width into the average.

(CLSPOP) = MIN (candidate class population count) and
20 if the class was previously NORMAL, 10 otherwise.

The average class prematurity-ratio ($CLSPR_i$) and width ($CLSDUR_i$) are updated with the candidate prematurity-ratio ($PREM_c$) and width ($WIDTH_c$) with the two equations:

$$(CLSPR_i) = ((CLSPOP)-1)*(CLSPR_i) + (PREM_c)) / (CLSPOP)$$

$$(CLSDUR_i) = ((CLSPOP)-1)*(CLSDUR_i) + (WIDTH_c)) / (CLSPOP),$$

and the family classifier variables are then updated:

$$(CLSPR) = (CLSPR_i)$$

$$(CLSDUR) = (CLSDUR_i)$$

Thus the prematurity-ratio (CLSPR) and width (CLSDUR) are updated with a uniform average until (CLSPOP) attains its upper limit. At that point, the average becomes a running average where the current beat is incorporated with a weight of 1/(CLSPOP). The averages (CLSPR) and (CLSDUR), and the class population (CLSPOP) now become the parameters for determining the morphology label for $CLASS_i$.

Determining the width range for normal QRS complexes

Since the family morphology classification depends in part on QRS width, a threshold must be computed to determine if the family consists of wide or narrow (normal) beats. There is no absolute threshold that works well for all patients, especially when bundle branch block or other conduction defects can prolong the duration of the "normal" QRS complexes. A complex is usually considered "wide" if its duration is greater than 110 to 120 milliseconds.

The QRS width is the duration of the largest (RQUAL) subcomplex of the entire QRS complex. This was used (rather than the width of the entire QRS complex) to avoid multimodal distributions that would arise if the delineation of the QRS complex was inconsistent. For example, the complex could have a Rs configuration where the S-wave was occassionally included in the description of the QRS complex. This would produce a double-mode histogram that would make the determination of the normal-abnormal width threshold more difficult. Using a single

morph for the width measurement eliminates this difficulty, but at the cost of not identifying the full width of a multiphasic QRS complex.

The bedside monitor maintains several histograms of QRS duration which are used to determine the width range for normal QRS complexes. The three histograms are:

RWHIST -- width histogram for all QRS complexes,
OTWHIST -- width histogram for all "on-time" and "late" QRS complexes that are less than 5% premature (10% if atrial fibrillation is present), and
PWHIST -- width histogram for all "premature" beats that are more than 15% premature.

All three histograms represent a "running average" of the width distribution, since each bin value is multiplied by 7/8 once every 30 beats. Thus the histograms reflect the width distribution of the most recent beats, with a decay time on the order of a few minutes. The bin width is one sample duration, or 1/SFREQ [1/360] of a second. An example of the RWHIST width histogram is shown in Figure 6-6a.

The modes of the RWHIST width histogram (all QRS complexes) are delineated according to the procedure outlined below. Each mode is subsequently labeled "normal" or "premature" based on the relative population of on-time and premature beats in the corresponding bin in the OTWHIST and PWHIST histograms. The mode

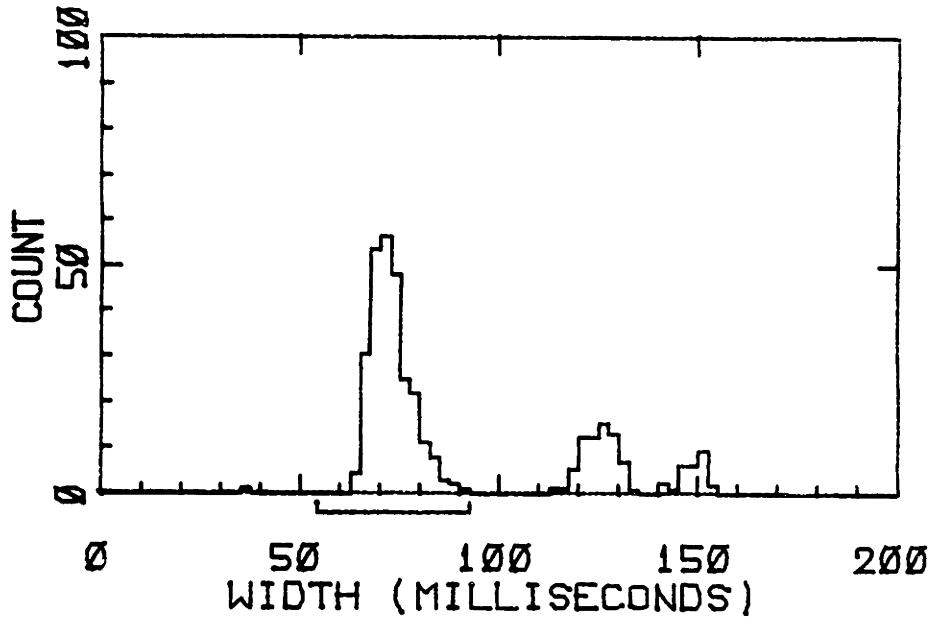



Figure 6-6a: QRS Width Histogram

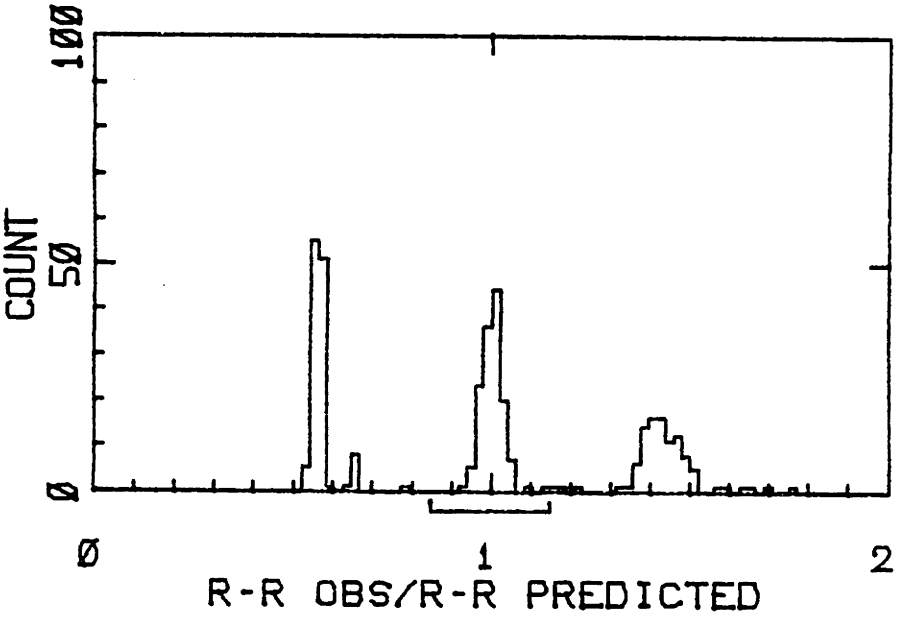



Figure 6-6b: QRS Prematurity-Ratio Histogram

The R-wave width histogram and the prematurity-ratio histogram are used to set the decision regions for morphology family classification. Histogram bins are reduced by 7/8 every 30 beats.

search begins with the highest mode, and as each mode is delineated and labeled, it is masked so that the next highest mode can be found.

(1) find the highest bin in RWHIST that hasn't been found before (by ignoring all bins with a non-zero MASK array entry),

(2) search left and right for a bin value less than than one-third the highest bin in the mode, and then

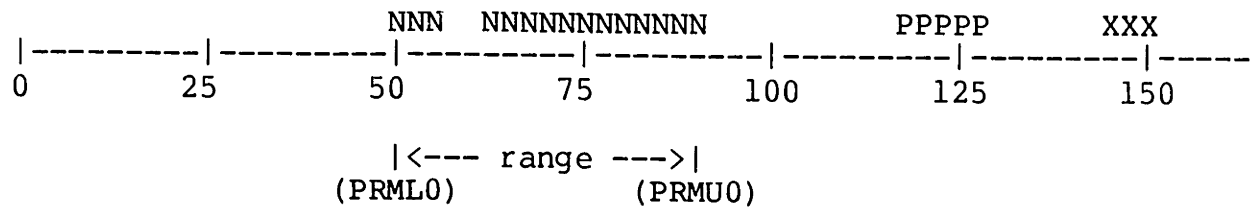
(3) search left and right for a local minimum.

This delineates the largest unlabeled histogram mode in RWHIST, which is then labeled according to its total population, and the relative population of on-time and premature beats. The labeling is done by storing an "N", "P", or "X" in the search MASK array (which was initially cleared).

(4) if the histogram mode has less than 15% of the total number of counts in RWHIST, label it "X" (to discard it).

(5) label the histogram mode "N" for normal and "P" for premature. Let #N be the number of on-time beats in OTWHIST, and let #P be the number of premature beats in PWHIST that are in the bins corresponding to the delineated bins in RWHIST. Label the mode "N":
  if #N > #P with no atrial fibrillation present
  if 9/8 #N > #P with atrial fibrillation present
Otherwise, label the mode "P".

The steps (starting at 1) are repeated for the four largest histogram modes. The bins corresponding to complexes 50 to 60 milliseconds wide are forcibly labeled "N" to guarantee the presence of a "normal" mode.




The mask array is then searched starting from the right and left-most margins until a "N" label is encountered, thereby delineating multi-modal distributions properly. The values that are determined are:

(PRML0) the lower width limit for normal complexes, and

(PRMU0) the upper width limit for normal complexes.

Why go to all the trouble to use histograms to determine the width threshold? Simpler methods have been used by others, which typically use a running average of the previous normal QRS complexes. Such schemes are linked to the classifier's accuracy, however, which in turn is linked to the accuracy of the width threshold determination, and so on. For the bedside monitor, it was decided to decouple the width threshold determination from the classifier, and make it a stand-alone module.

Determining the prematurity-ratio range

A running average histogram of prematurity-ratio (PRHIST) is also maintained (see Figure 6-6b). The histogram is used to measure the degree of sinus arrhythmia that is present, and set the prematurity thresholds used to label morphologic families. Each bin represents a 2% change in prematurity-ratio.

Only the central mode (within 20% of "on-time") of the prematurity-ratio histogram is delineated. A search is conducted to the left and right of the histogram peak until the bin value is less than one-third the peak bin value, followed by a search for a local minimum. The minimum to the left of the central peak is used to establish the threshold that separates premature beats from on-time beats. The lower threshold (PRMLO) is limited to be not less than 15%, and can only be extended if central histogram mode is broadened by the presence of sinus arrhythmia or atrial fibrillation.

Morphology Class Labeling

Now that the range for normal QRS width, and level of sinus arrhythmia have been established, it's now time to label the cluster. Three features are used:

(CLSPR) average class prematurity-ratio,

(CLSDUR) average class width, and

(CLSPOP) class population.

The decision regions used for family labeling are shown in Figure 6-7, and is depicted as a two dimensional decision space with prematurity and width as the primary features. The set of family labels includes:

NORMAL not premature or wide (presumably supraventricular);

PVC premature and wide (presumably ventricular ectopic);

SVPB premature, but not wide;

WIDE not premature, but wide; and

ARTIFACT very narrow, likely to be artifact.

For illustration purposes, the third dimension (CLSPOP) can be viewed as a perturbation of the of the (CLSPR) and (CLSDUR) decision boundaries. The width range for normal QRS complexes (NRWU0) and (NRWL0), the level of sinus arrhythmia (PRML0), and the presence of atrial fibrillation (AFDFLG) determine the basic thresholds for the classifier. The decision vertices are set according to the rules listed in Table 6-1.

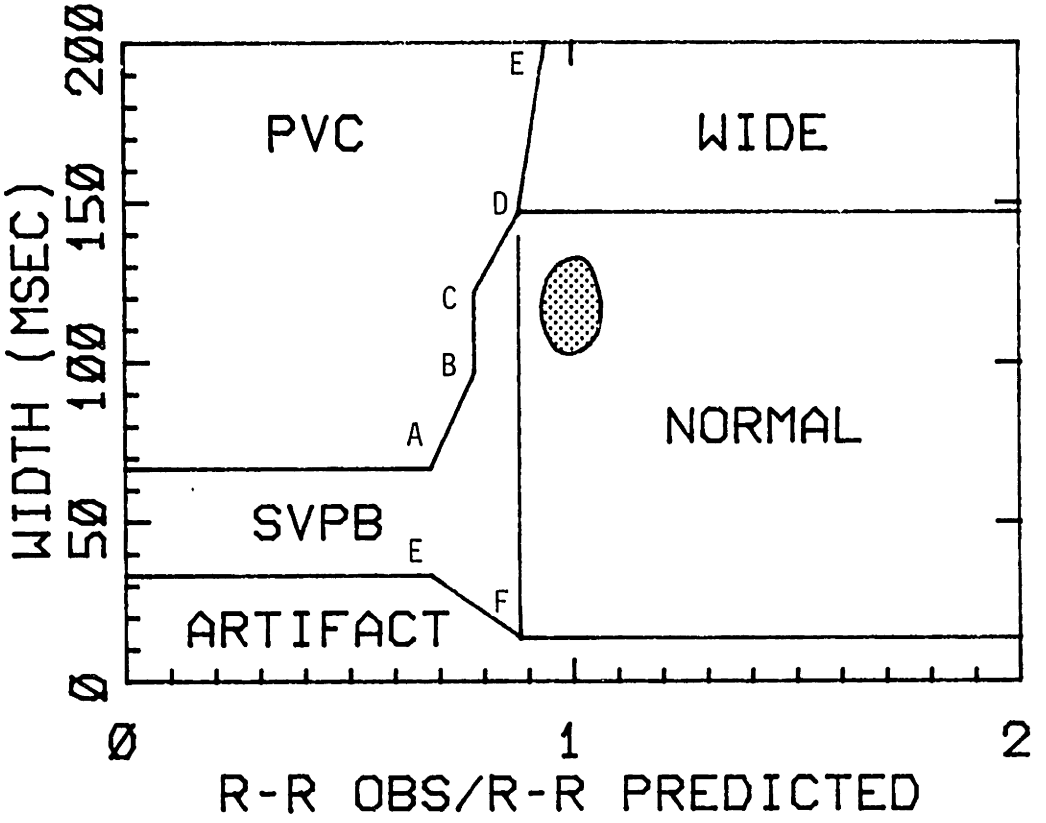



Figure 6-7: Morphology Labeling Decision Regions

The figure above shows the decision regions used to label morphology classes. The values of the decision region vertices are listed in Table 6-1, and are based on the class population count (CLSPOP) and the thresholds derived from the QRS width and prematurity-ratio histograms. A class is labeled according to the region that contains the average class width (CLSDUR) and prematurity-ratio (CLSPR).

Boundaries are shown for:

(CLSPOP) = 4
(NRWU0) = 133 ms
(NRWL0) = 33 ms
(PRML0) = 0.8
(AFDFLG) = NOT-AF

Table 6-1: Decision Region Vertices

The decision region vertices used by the family classifier are listed in the table below:

Input arguments:

| | |
|---|---|
| (CLSPOP) | class population |
| (NRWU0) | upper width limit for normal QRS complexes |
| (NRWL0) | lower width limit for normal QRS complexes |
| (PRML0) | lower prematurity-ratio threshold |
| (AFDFLG) | NOT-AF or AF |

Decision Region Vertices:

| | | |
|---|---|---|
| A: | (PRA) | ((PRBC)+40%)/2 |
| | (WA) | POP( (NRWU0)+10ms , 35ms ) |
| B: | (PRBC) | NOT-AF: POP( 40% , 12% )<br>AF: POP( 40% , 16% ) |
| | (WB) | POP( (NRWU0)+10ms , MIN(80ms,MAX((NRWU0)-20ms,50ms)) ) |
| C: | (PRBC) | same as above |
| | (WC) | POP( (NRWU0)+10ms , MAX((NRWU0)-20ms,50ms) ) |
| D: | (PRD) | NOT-AF: POP( MIN(40%,(PRML0)) , 8% )<br>AF: POP( 40%+4% , 12% ) |
| | (WD) | POP( (NRWU0)+25ms , (NRWU0)+15ms ) |
| E: | 6%<br>200ms | |
| F: | (PRF) | ((PRBC)+40%)/2 |
| | (WF) | MIN((MAX((NRWL0),30ms)),(WA)) |
| G: | (PRG) | same as (PRD) |
| | (WG) | MAX((NRWL0,30ms)-20ms |

Functions:

MAX(A,B) returns maximum value of A or B
MIN(A,B) returns minimum value of A or B

POP(A,B) scales from A to B as (CLSPOP) goes from 1 to 8, and depends whether AFIB is present:

| (CLSPOP) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | , then |
|---|---|---|---|---|---|---|---|---|---|---|
| factor = | 0 | 5 | 8 | 11 | 14 | 16 | 16 | 16 | 16 | if NOT-AF, |
| factor = | 0 | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | if AF. |

POP(A,B) = [ A*16 + (B-A)*factor ] / 16

The decision region vertices are based on numerous experiments with the annotated ECG data base, and on clinical trials of the bedside monitor. The detailed results from some of the data base experiments are discussed at the end of this section, and only the broad principles that underly their design are discussed here.

The most critical decision boundary separates PVCs from NORMALs (vertices B, C, and D). This boundary separates what the algorithm considers to be ventricular ectopic morphologies (because they are, on the average, premature and wide) from normal morphologies (which are usually on-time, and not wide). The other decision boundaries are somewhat less important, since the primary goal of the bedside monitor is to monitor the level of ventricular ectopic activity. In order to illustrate the operation of the family classifier, three cases are discussed in the following sections, and are illustrated in Figures 6-8a, b, and c.

Figure 6-8a shows the decision boundaries for a new cluster (whose population is one). Since this is the first time the algorithm has "seen" the complex, both the prematurity-ratio and width thresholds are extended well beyond the bounds for normal width and prematurity. The basic strategy here is to minimize false positives by imposing very strict prematurity and width requirements on new complexes that might have been (or were distorted by) artifact.

After a particular morphology has been seen several times, it is quite unlikely that it is artifact or a noise corrupted QRS complex. Also, the average prematurity-ratio and width will be more accurate, since they now are the average for several complexes. Thus the decision boundaries can be drawn closer towards the region of normal prematurity and width (see Figure 6-8b). By using average prematurity, errors due to sinus arrhythmia and atrial fibrillation are reduced, and the prematurity boundaries can be brought in closer to "on-time". By using an average of several width measurements, errors due to individual width mis-measurement are also reduced.

In fact, to classify a populated class, we can just about discard width as a criterion. If a particular morphology is always seen early, and has a distinctly different morphology from normal complexes, then it probably is an ectopic beat (or an aberrated supraventricular beat). By relying on average prematurity rather than width, the bedside monitor can accurately classify PVCs that are as narrow, or narrower than normal complexes. This helps in situations where the ectopic complexes are isoelectric in the monitored lead, or with patients who have conduction defects that prolong their normal QRS complexes (see Figure 6-8c).

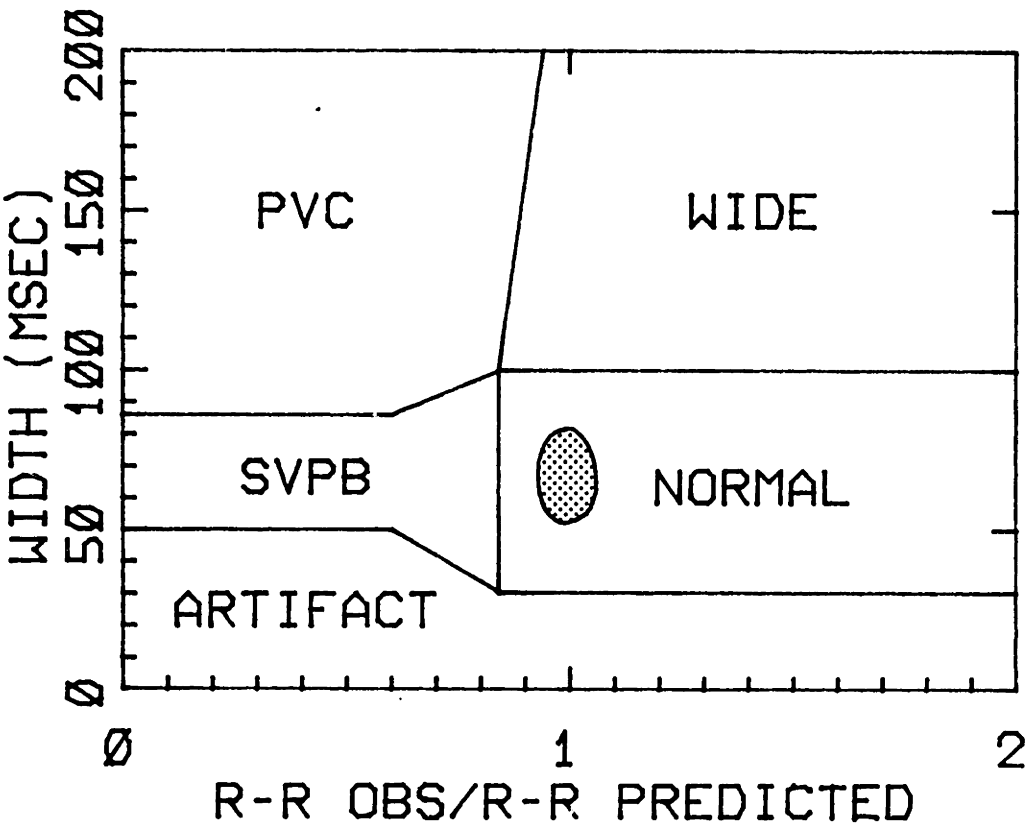



Figure 6-8a: Decision Regions for New Morphology Classes

The decision regions for a new class (whose population is 1) are shown above. Since the complex has been seen only once, the prematurity and width thresholds are extended well beyond the bounds for normal width and prematurity to minimize the number of false positives.

The decision boundaries are shown for:

(CLSPOP) = 1
(NRWU0) = 80 ms
(NRWL0) = 50 ms
(PRML0) = 0.85
(AFDFLG) = NOT-AF

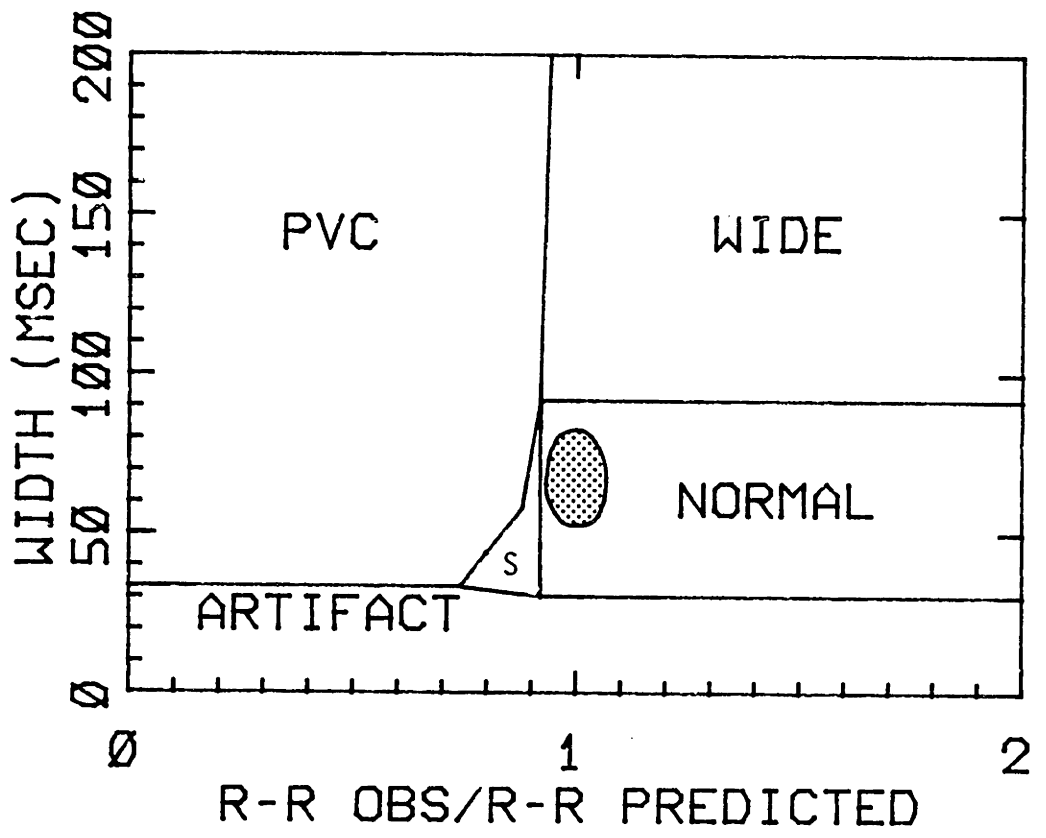



Figure 6-8b: Decision Regions for Populated Classes

The decision regions for well populated classes are shown above. After a particular morphology has been seen several times, it is unlikely to be artifact or a noise corrupted QRS complex. The average prematurity-ratio and width will be more accurate, and thus the decision regions are drawn inwards towards normal prematurity and width. Note that a PVC need not be wider than normals to be classified as a PVC, provided that it always is premature, and has a morphology different than normals.

The decision boundaries are shown for:

(CLSPOP) = 9
(NRWU0) = 80 ms
(NRWL0) = 50 ms
(PRML0) = 0.85
(AFDFLG) = NOT-AF

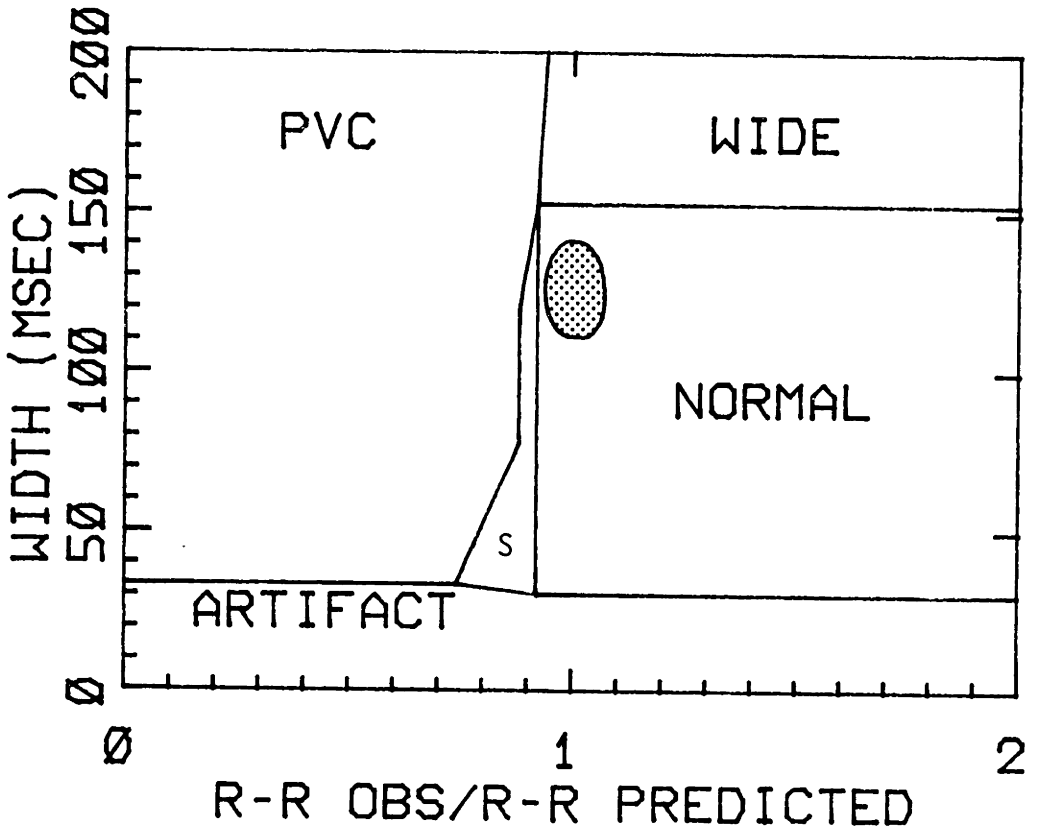



Figure 6-8c: Decision Regions for Wide Normal Beats

This figure illustrates how the decision boundaries adapt to a patient with wide normal beats (most likely due to conduction defects that prolong the normal QRS complex). Again, PVCs need not be wider than normals to be accurately classified, provided that they are (on the average) premature, and have a different morphology from normal QRS complexes.

The decision boundaries are shown for:

(CLSPOP) = 9
(NRWUO) = 140 ms
(NRWLO) = 50 ms
(PRMLO) = 0.85
(AFDFLG) = NOT-AF

The decision region vertices are based on numerous experiments with the annotated digital data base, and on clinical trials of the bedside monitor. The overall goal was to reduce the number of false negatives while keeping the false positive rate very low (below 0.2% for the entire data base). Unfortunately, there was no "magic handle" that would reduce both, and usually reducing one would inevitably increase the other.

The population scaling function POP (see Table 6-1) illustrates how difficult the optimization task can be. The POPulation function interpolates between thresholds based on the population of the candidate's family. Originally the interpolation was linear (for a population change from one to nine). It was clear, however, that the need to significantly extend the decision vertices away from the normal was really only necessary when the class population was one. Once the morphology was seen again, the decision vertices could be moved back towards normals, but a lesser increments as the class population increased. This was tested on 1/5/81, and improved the PVC sensitivity for all tapes (except 203), without a significant increase in the number of false positives. The underlying rhythm for tape 203 is atrial fibrillation, and so the linear (more gradual) interpolation was re-installed, and is used if atrial fibrillation is detected. This was retested on tape 203, and significantly reduced the number of false positives.

com: Linear POPulation interpolation.
Date: 1/4/81

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Tape 106 | fn | 56 | fp | 6 | tp | 459 | tn 1457 | + |
| Tape 200 | fn | 89 | fp | 1 | tp | 720 | tn 1736 | + |
| Tape 203 | fn | 131 | fp | 26 | tp | 297 | tn 2432 | + |
| Tape 207 | fn | 24 | fp | 11 | tp | 51 | tn 1615 | + |
| Tape 214 | fn | 94 | fp | 3 | tp | 157 | tn 1950 | + |
| Tape 223 | fn | 223 | fp | 5 | tp | 249 | tn 2062 | + |
| Total | fn | 617 | fp | 52 | tp | 1933 | tn 11252 | |

Date: 1/5/81
mod: Non-linear POPulation interpolation.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Tape 106 | fn | 51 | fp | 8 | tp | 464 | tn 1456 | + |
| Tape 200 | fn | 70 | fp | 1 | tp | 739 | tn 1736 | + |
| Tape 203 | fn | 123 | fp | 70 | tp | 306 | tn 2396 | + |
| Tape 207 | fn | 24 | fp | 12 | tp | 51 | tn 1615 | + |
| Tape 214 | fn | 83 | fp | 5 | tp | 169 | tn 1948 | + |
| Tape 223 | fn | 210 | fp | 5 | tp | 262 | tn 2062 | + |
| Total | fn | 561 | fp | 101 | tp | 1991 | tn 11213 | |

obs: SENS improved somewhat for tapes 106, 200, 214, 223.
obs: Increased FP for tape 203.

Date: 1/5/81
mod: Interpolation for AFIB is linear.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Tape 203 | fn | 133 | fp | 39 | tp | 296 | tn 2426 | |

obs: Reduced FP for tape 203.

Summary of morphology labeling algorithm

In summary, the bedside monitor uses a morphology labeling algorithm that uses adaptive thresholds based on running average histograms of prematurity and width. The decision boundaries adjust to the population of the classes by requiring very strict prematurity and width requirements for new morphology classes, but are relaxed as the population of a class increases.

It is important to note that the label assigned to a morphology class is independent of the labels of other classes with similar morphology. For example, a class is not labeled NORMAL if it is similar to other NORMAL classes, but instead is labeled to its own prematurity, width, and population count. This approach prevents possibly incorrect labels from propagating from one class to the next.

Another advantage of the classifier is that it readily lends itself to a very readable and interpretable display. This has been invaluable in developing and debugging the algorithm, since it can easily pinpoint the major reasons why a QRS complex is classified incorrectly. For example, all the plots shown in Figure 6-8 are available as real-time displays on the clinical prototype of the bedside monitor. In addition to drawing the decision boundaries, the individual prematurity and width for QRS complexes are also drawn. The two dimensional display also makes it much easier to explain to users how the algorithms makes its decisions.

## 6.4 QRS Complex Morphology Classification

Now that the label of the best-match morphology family has been updated, the next step is to determine the morphology of the candidate QRS complex. Early versions of the algorithm labeled the candidate complex with the label of the family to which it was most similar. This strategy is optimum if the morphology clusters are accurately labeled (as would be the case with a supervised classifier).

The bedside monitor, however, uses an unsupervised classifier that is not infallible. The family classifier is in fact designed to err on the side of false negatives, rather than false positives, when the family has low population. This suggests the strategy of labeling the candidate complex based on highly populated classes that are somewhat less similar to it, rather than the best match class if it has a low population.

Atrial fibrillation and flutter are two other conditions that are likely to cause family classification error. The difficulty arises from the irregular rhythm and the increased morphologic variability of QRS complexes. This indicates the need for a "similarity to normals" threshold such that the candidate would be labeled normal if it was sufficiently similar to a well-populated normal class, even though it might be more similar to a PVC class.

Candidate QRS morphology labeling algorithm

Thus the strategy of labeling the QRS with the best match family label was replaced with one that accounts for likely family classifier errors. There are three basic properties of the new algorithm:

(1) If the candidate complex is sufficiently similar to a normal family, label the QRS "normal", even though the candidate may be more similar to a PVC class. The distance threshold should be increased during atrial fibrillation to account for the greater likelihood of aberrated normal beats.

(2) If the best match family has a low population (and thus more likely to be labeled incorrectly), then label the candidate using less similar, but more populated classes.

(3) If the best match family has a large population, always base the classification on it. (Rule (1) has precedence, however.)

Every morphology class that is similar to the candidate complex is considered by the QRS complex labeling algorithm. The candidate QRS classifier maintains a list of entries organized by classification label (NORMAL, PVC, SVPB, WIDE, and ARTIFACT). Each table entry contains the total population and a cumulative score for all families that have the same morphology label. The score is computed for each family based on its distance to the candidate complex and its population, and is accumulated according to the family's morphologic label. If the cumulative score for NORMAL classes is sufficiently high, then the QRS complex is always considered to have NORMAL morphology.

Otherwise, the candidate is identified with the label (NORMAL, PVC, SVPB, WIDE, and ARTIFACT) that has the highest cumulative score. The exact procedure is detailed on the following page and in Figure 6-9.

Candidate QRS Morphology Labeling Algorithm

In the discussion that follows,
"i" refers to a particular morphology family, and
"j" refers to one of the family labels
NORMAL, SVPB, PVC, WIDE, or ARTIFACT.

First, the cumulative population $CPOP_j$ and cumulative score $CSCORE_j$ are initialized:

Set $CSCORE_j = 0$ for j = 1(1)5
$CPOP_j = 0$ for j = 1(1)5

Then, repeat steps (1) through (3) until the closest family is not similar enough to the candidate complex to contribute to the cumulative score; then exit to (4).

(1) Find the next closest family to the candidate complex. Assume it has $CLSPOP_i$ and label j.

(2) $CSCORE_j =$

$$CSCORE_j + [\ LWDST(distance_i) * \sum_{n=CPOP_j}^{CPOP_j+CLSPOP_i} LWPOP(n)\ ]$$

(3) $CPOP_j = CPOP_j + CLSPOP_i$
GOTO (1)

The functions $LWDST(distance_i)$ and LWPOP(n) are specified in Figure 6-9a and 6-9b respectively.

(4) Compare $CSCORE_{j=NORMAL}$ with one of the thresholds:
198*170 if no atrial fibrillation present
[LWDIST(1300H) at full population]
198*140 if atrial fibrillation is present
[LWDIST(1600H) at full population]

If the CSCORE is larger than the appropriate threshold, then the candidate complex has NORMAL morphology.

Otherwise,
label the candidate with the label that has the highest cumulative score.

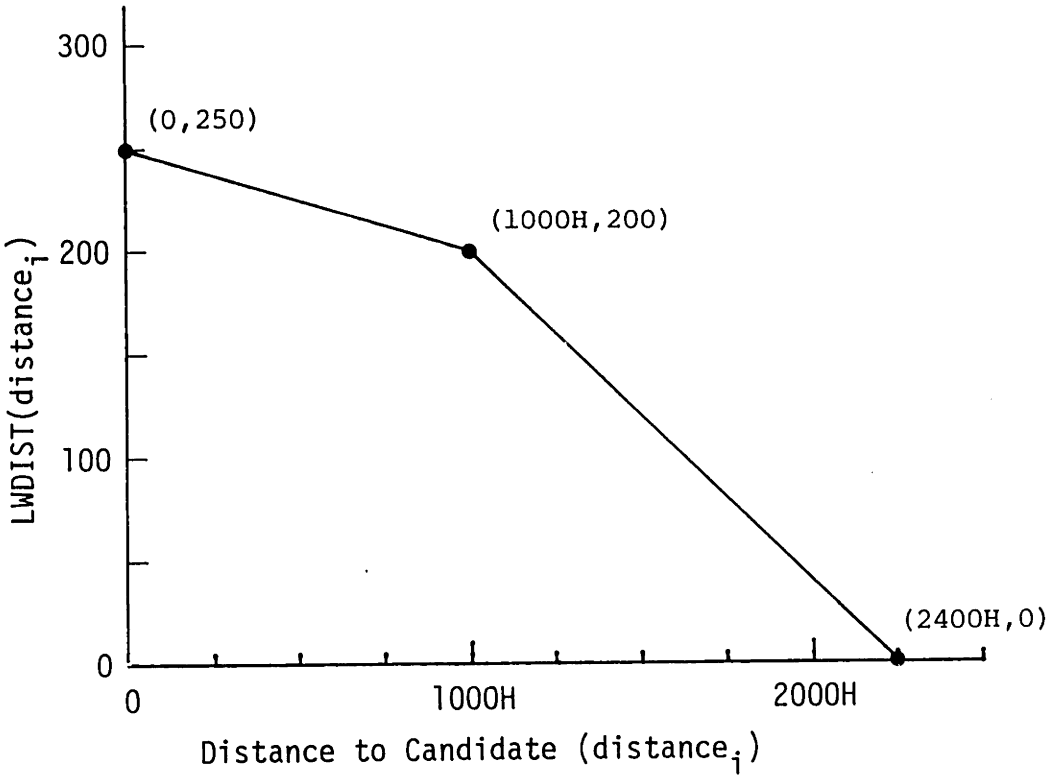



Figure 6-9a: Distance Weighting Function LWDIST

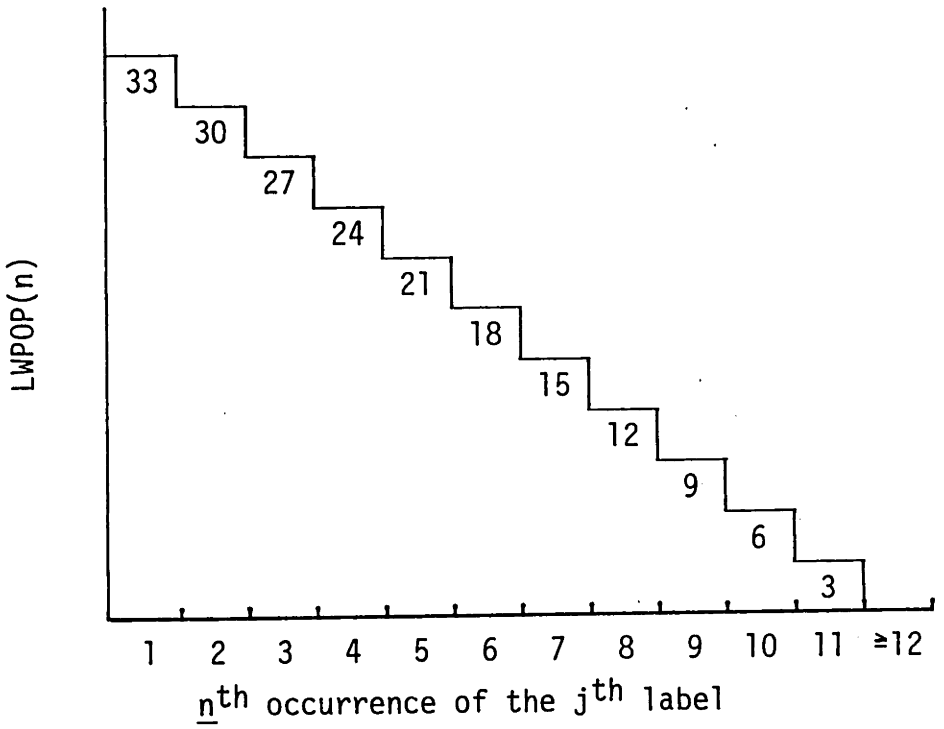


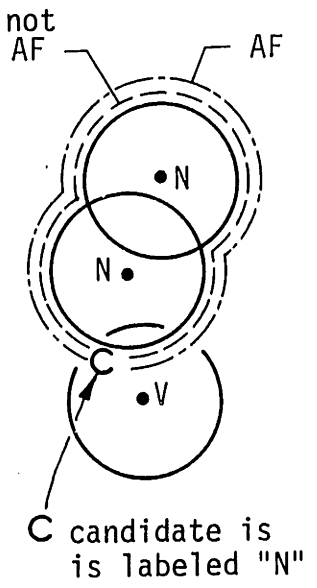

Figure 6-9b: Population Weighting Function LWPOP

The function LWDIST(distance) is a monotone decreasing function of morphologic distance. LWDIST returns a maximum value for similar complexes [distance between 0 and 1000H], and drops off rapidly to zero for complexes that are not similar [distance greater than 2400H]. Other than satisfying the above criteria, the function is somewhat arbitrary.

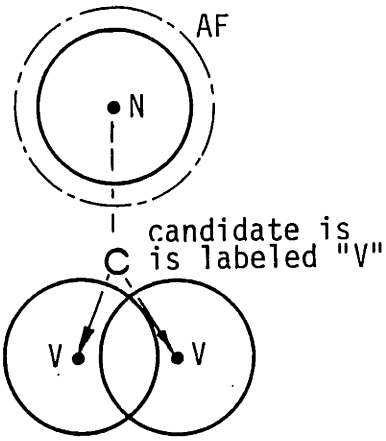
The use of the function LWPOP(n) is somewhat more complicated. LWPOP(n) returns a weight based on the class population (for a particular class label), starting with classes that are most similar to the candidate. For example, say that the closest NORMAL class has a population of 2. LWPOP returns the first two values LWPOP(1) and LWPOP(2) [33+30], which is multiplied by LWDIST(distance) and accumulated in the cumulative score CSCORE. If the next closest NORMAL class has a population of 4, LWPOP will then return LWPOP(3) + LWPOP(4) + LWPOP(5) + LWPOP(6) [27+24+21+18].

The primary goal of LWPOP is to heavily weight the families that are closest to the candidate complex. Once the cumulative population for a given class label is more than eight or so, no other NORMAL classes need to be considered. Basically, a family with a population of one thousand should not have much more weight than one with ten, since we are only trying to establish the accuracy of a family label in the region of morphology space near the candidate complex.

The figure to the right illustrates how a candidate complex (c) can be labeled "N", even though it is more similar to a class with a "V" label. This is especially important during atrial fibrillation, since it is likely that aberrated families will be erroneously labeled "V". For this reason, the threshold surrounding the normal classes is increased during atrial fibrillation.




Let's consider another fairly common situation where a true PVC class has a slightly different morphology than PVCs seen earlier. If it is not sufficiently wide and premature, it will likely be labeled a SVPB instead of a PVC until its population increases. In this case, the classifier relies more heavily on populated "V" classes, and labels the candidate "V". When the candidate's own class attains a sufficiently high population, subsequent complexes near it will be labeled "V".



An important parameter is the distance at which a candidate is always labelled NORMAL, even though it may be more similar to another class with a different label. The final thresholds that were selected were
CSCORE=198*170 [LWDIST(1300H)] when AF is not present, and
CSCORE=198*140 [LWDIST(1600H)] when AF is present.
A series of evaluations with different values of Dn were tried (on a very difficult set of tapes) to see the effects (if any) of different values of Dn. Relatively little change was observed in classifier performance.

```
Date:     1/5/81
com:      CSCORE=198*180 no AF  [LWDIST(1200H)]
com:      CSCORE=198*160 with AF  [LWDIST(1400H)]
Tape 106    fn   51  fp    8  tp  464  tn 1456  +
Tape 200    fn   70  fp    1  tp  739  tn 1736  +
Tape 203    fn  133  fp   39  tp  296  tn 2426  +
Tape 207    fn   24  fp   12  tp   51  tn 1615  +
Tape 214    fn   83  fp    5  tp  169  tn 1948  +
Tape 223    fn  210  fp    5  tp  262  tn 2062  +
Total       fn  571  fp   70  tp 1891  tn 11243

Date:     1/9/81
mod:      CSCORE=198*160 no AF  [LWDIST(1400H)]
mod:      CSCORE=198*140 with AF  [LWDIST(1600H)]
Tape 106    fn   56  fp    6  tp  459  tn 1458  +
Tape 200    fn   84  fp    1  tp  725  tn 1736  +
Tape 203    fn  127  fp   35  tp  302  tn 2431  +
Tape 207    fn   26  fp   21  tp   49  tn 1605  +
Tape 214    fn   90  fp    2  tp  162  tn 1951  +
Tape 223    fn  209  fp    4  tp  253  tn 2063  +
Total       fn  592  fp   69  tp 1950  tn 11244
obs:      FP reduced slightly on all tapes, except 207.
obs:      Slight increase in FN on all tapes.

Date:     1/13/81
mod:      CSCORE=198*170 no AF [LWDIST(1300H)]
com:      CSCORE=198*140 with AF  [LWDIST(1600H)]
Tape 106    fn   53  fp    6  tp  463  tn 1458  +
Tape 200    fn   85  fp    3  tp  724  tn 1734  +
Tape 203    fn  143  fp   14  tp  287  tn 2449  +
Tape 207    fn    8  fp    9  tp   54  tn 1618  +
Tape 214    fn   90  fp    3  tp  162  tn 1950  +
Tape 223    fn  210  fp    6  tp  262  tn 2061  +
Total       fn  589  fp   41  tp 1952  tn 11270
obs:      PVC sensitivity unaffected.
obs:      False positives reduced for 203.
```

## 6.5 Final QRS classification

At this point, the individual QRS complex morphology has been labeled as NORMAL, SVPB, PVC, WIDE, or ARTIFACT. The local and global timing of the QRS complex must also be considered to determine the final classification of the complex. The detailed issues involved in computing the prematurity of a beat and determining atrial rhythm are discussed in the next chapter. The interface to the QRS classifier is very simple, however: local timing is based solely on the prematurity ratio, and the only two global rhythm categories are whether or not atrial fibrillation is present.

An individual QRS complex retains the PVC, WIDE, or ARTIFACT classification regardless of its local timing or global rhythm. If the QRS morphology label is NORMAL or SVPB, then the complex is considered to have supraventricular morphology, and is labeled NORMAL or SVPB based on its individual prematurity, and whether atrial fibrillation is present. If atrial fibrillation is present, supraventricular complexes are always labelled NORMAL. If atrial fibrillation is not present, supraventricular complexes are labeled according to their prematurity. If the prematurity-ratio is less than (PRML0) [usually at least 15% premature], the complex is labeled a SVPB; otherwise it is labeled NORMAL.

## 7. TIMING CLASSIFICATION

Both timing and morphologic information are required to classify QRS complexes. Two types of timing information may be derived from the ECG:

(1) "local" timing, which deals with the individual QRS complex and its rhythmic context within the beats that surround it, and

(2) "global" timing, which deals with the long term statistical properties of the RR interval sequence.

The local timing of a QRS complex is used to determine its prematurity, which is subsequently used to update the label of the complex and the morphologic family to which it belongs. Other rhythm questions such as whether or not the complex is "interpolated", or follows a "compensatory pause", are also suitable for local timing analysis. All of the above timing classifications are relative to a "predicted RR interval", which is an estimate of the expected interval between two normal complexes. The bedside monitor, like most other ECG analysis algorithms, uses a running average of the preceeding normal-normal intervals to predict the duration of the next normal-normal RR interval.

The long term statistical properties of the RR interval sequence (spanning tens of seconds to several minutes) can also be used to extract information about the ECG. Global timing information can be used to indirectly identify several atrial arrhythmias without detecting P-waves. These include normal

sinus rhythm, sinus arrhythmia (due to respiratory effects), and atrial fibrillation and flutter. Global rhythm analysis performed by the bedside monitor is currently limited to identifying atrial fibrillation and flutter. Several techniques were developed and evaluated, and the best (based on tests with the annotated ECG data base) was incorporated into the final algorithm. The principal use of the atrial fibrillation detector is to modify several morphology classifier thresholds to account for the increased likelihood of aberrated beats and irregular timing during atrial fibrillation.

## 7.1 RR Interval Prediction

The bedside monitor performs local rhythm analysis by comparing the measured RR interval with the predicted RR interval expected for the next normal complex. The predicted RR interval (PRRINT) is used by a number of algorithm modules, which are listed below:

(1) The predicted interval is used to compute the prematurity ratio for individual QRS complexes. The prematurity ratio is the ratio of the observed RR interval (ORRINT) divided by the predicted interval (PRRINT). The individual prematurity ratio is used to update the average prematurity ratio of the family that best matches the candidate QRS complex. It is also used to distinguish between supraventricular premature beats (SVPBs) and "on-time" NORMAL beats.

(2) The prematurity ratio of individual complexes is used to update the prematurity ratio histogram. The width of the on-time central mode of the histogram is used to determine the degree of sinus arrhythmia, and sets several classifier thresholds.

(3) The R-wave detector uses the predicted RR interval to determine whether it is likely that a low amplitude beat was "missed" if the interval between two detected R-waves is prolonged. If it is prolonged, the R-wave detection thresholds are lowered to recognize low amplitude R-waves.

(4) The bedside monitor alarm processor uses the predicted interval to determine whether a beat was skipped or missed.

RR interval predictor design

Most RR interval predictors reported in the literature use a running average of the most recent RR intervals to predict the duration of the next normal RR interval. More sophisticated predictors assign less weight (or even discard) intervals that are bounded by an ectopic beat, but still must provide a failsafe mechanism in case there are no recent normal-normal RR intervals. An ideal predictor would have mechanisms to recognize and deal with sinus arrhythmia (which is a periodic modulation of the RR interval sequence due to respiration), and would make optimum use of morphologic knowledge about each beat in the sequence.

One predictor described in the literature is used by the ARGUS monitoring system (Nolle,1972). It uses two running averages of QQ intervals. One of the averages is derived from QQ intervals between consecutive pairs of normal QRS complexes, and is used if at least one normal QQ doublet occurred in the last twenty QQ intervals. Otherwise, the average QQ interval for all complexes (of any type) is used. The QQ interval (instead of the RR interval) is used, since it is a more consistent timing fiducial for patients with wide QRS complexes. Each running average had a time constant of eight intervals.

One (supposedly) important design issue is the number of previous RR intervals included in the running average. This number depends on the application to which the predictor is applied. If the primary use of the predictor is to identify isolated SVPBs, then an average of only one or two intervals is the best, since an average with a short time constant would be able to track sinus arrhythmia and other cyclic changes of instantaneous heart rate. An average of six to ten intervals may be required to detect runs of SVTA, since it won't instantly adapt to the change in heart rate. Another advantage of using more intervals in the average is that it decreases its sensitivity to detector and classifier errors. Missed beats (due to detector error) and falsely detected beats (T-waves or artifact) will not seriously disturb the predicted interval for an average based on many intervals.

The bedside monitor RR interval predictor

Early versions of the bedside monitor used a uniformly weighted running average of the last eight normal-normal intervals to predicted the duration of the next normal sinus interval. If no normal-normal intervals were seen in the last 16 complexes, a uniformly weighted average of the last ten intervals (of any beat type) was used. The onset of the R-wave (onset of the major deflection of the QRS complex) was used as the timing fiducial, since it is the most reliable indicator of atrial timing, and the least sensitive to morphology variations of the QRS complex.

This predictor worked quite well for most ECG waveforms and rhythm patterns. However, ECG records with irregular rhythm and variable QRS morphology provoked undesirable interaction between the morphology classifier and the predictor described above. If a particular morphology happened to be slightly premature (on the average), it would be labeled PVC by the morphology classifier. If the morphology variation was dependent on the coupling interval to the preceeding QRS complex, then morphologies with a shorter (than the average) coupling interval would be labeled PVC, and would not be included in the running average of normal-normal intervals. The running average, which now includes only longer than average intervals, would tend to predict a longer than average interval. This would make it even more likely that an average duration interval was considered premature, and cause more families to be labeled PVC. This

failure mode usually corrected itself once the predictor switched to the average of all beat types, but not until after a long run of false positives PVCs was declared. This certainly represented an undesirable feedback path in the software, where the predictor relied on the accuracy of the classifier, which in turn relied on the accuracy of the predictor.

Thus, a new RR interval predictor was developed that based its prediction on RR intervals weighted according to the complexes' morphologic distance to the nearest normal, rather than their final classification. This reduced the software coupling between the classifier and predictor, and practically eliminated the problem of false PVC run detection. Even though a complex was incorrectly labeled a PVC by the classifier, the interval that it bounded would be given full weight if the complex was sufficiently similar to a normal family. If the complex had a different morphology than the normal, then its interval was weighted much less. The new predictor also uses the relative and absolute duration of the RR interval to discard outliers, and to prevent abnormally prolonged RR intervals from corrupting the predicted interval value.

The RR interval predictor used by the bedside monitor consists of two parts. The first part predicts the interval based on a weighted average of the preceeding intervals. It also determines whether the complex preceeding the current complex was interpolated, and if it is, computes the corrected observed RR interval (ORRINT) and predicted RR interval (PRRINT) for the

current complex. The second part of the predictor is invoked after the classifier has classified the QRS complex, and uses the final classification and morphologic distance to normals to determine the weighting coefficient for the interval.

RR interval predictor -- part 1

The first part of the predictor (subroutine RRPR1) computes the weighted average of the preceeding RR intervals, and stores the predicted interval in (PRRINT). The observed RR interval is computed by subtracting the time between the current and preceeding QRS complexes. The next step is to determine whether or not the preceeding complex was interpolated, according to the following rules:

If the current complex has NORMAL morhphology, and
the preceeding complex is a PVC, and
the complex before it has NORMAL morphology,
then the interval between the two normal complexes surrounding the PVC is tested to see if it is close to the predicted RR interval. If the RR interval between the two normal complexes is within +30% and -15% of the predicted interval (PRRINT), then the preceeding beat is declared interpolated, and the observed RR interval (ORRINT) is set to the duration of the interval between the two normal complexes surrounding the PVC. The assymetrical thresholds are used since it is likely that the AV junctional conduction time (PR interval) is increased by the interpolated beat.

The corrected value for the observed RR interval (ORRINT) and the predicted interval (PRRINT) are then used by the morphology classifier to determine the prematurity of the current complex, and also to update the average prematurity ratio of the class to which the complex belongs.

RR interval predictor -- part 2

The second part of the RR interval predictor (subroutine RRPR2) is invoked after the candidate complex has been classified. The final morphology classification and the distance to the nearest normal family is used to compute the weighting coefficient for each RR interval. The form of the weighted average is shown below:

$$(\text{PRRINT}) = \frac{\sum_{i=1}^{N} W(i) * \text{ORRINT}(i)}{\sum_{i=1}^{N} W(i)}$$

The number of averaged intervals N is variable, but is limited to the number of previous intervals saved in the averaging array [up to 14]. The quantity that actually determines when the summing is complete is the cumulative weight $\sum W(i)$.

The weighting factor W(i) for individual RR intervals is based on

(1) the morphologic labels of the complexes that bound the interval,

(2) their morphologic distance to the nearest normal family

(BSTNDST), and

(3) the relative and absolute duration of the interval.

The weighting coefficients and functions that were used are discussed in the following paragraphs, and shown in Figure 7-1.

First, a weight based on the morphologic label of the current complex is determined. The maximum weight [8/8] is assigned to a NORMAL complex, and a partial weight [2/8] is assigned to a complex labeled ARTIFACT. If the complex is neither NORMAL nor ARTIFACT, then its morphologic weight is based on its distance to the nearest normal family according to the function shown in Figure 7-1a. A complex similar to normals is assigned a maximum weight [of 8/8], whereas complexes very dissimilar are assigned a partial weight [as low as 1/8]. Note that a weight of zero cannot be assigned to an complex, even if it is classified ARTIFACT or PVC. This prevents the predictor from being completely insensitive to the incoming RR interval sequence. The final weight for the complex is saved, and the smallest of the morphologic weights of the two complexes that bound the interval is selected.

The other two weighting factors depend on the duration of the RR interval, rather than the morphology of the two complexes that define it. If the RR interval is prolonged (compared to (RRPRED)), the interval is weighted less, as shown in Figure 7-1b. This is designed to reduce the effect of outliers due to complexes missed by the event detector, or prolonged pauses. Also, intervals whose absolute duration is excessively long are

also weighted less, according to the function shown in Figure 7-1c. During normal operation, however, both interval based weights are usually at their full value [8/8], and have very little effect on the operation of the predictor. The final weight used for the current interval is the smallest of the morphologic weight and the relative and absolute duration weights. The interval value (ORRINT) and its weight are saved in a circular buffer [14 entries deep].

The predicted interval is the average of the previous weighted intervals. Rather than using a fixed number of intervals, however, the summation is performed over the most recent intervals until the cumulative weight exceeds a threshold. In this way, the average smoothly adjusts to include more intervals if there is a mix of normal and ectopic beats. If all the complexes are normal (maximum weight), fewer intervals are incorporated into the predicted value. The predictor terminates the summation when the cumulative weight is 6*8/8 (6 normal-normal intervals) during normal sinus rhythm, and 10*8/8 (10 normal-normal intervals) during atrial fibrillation.

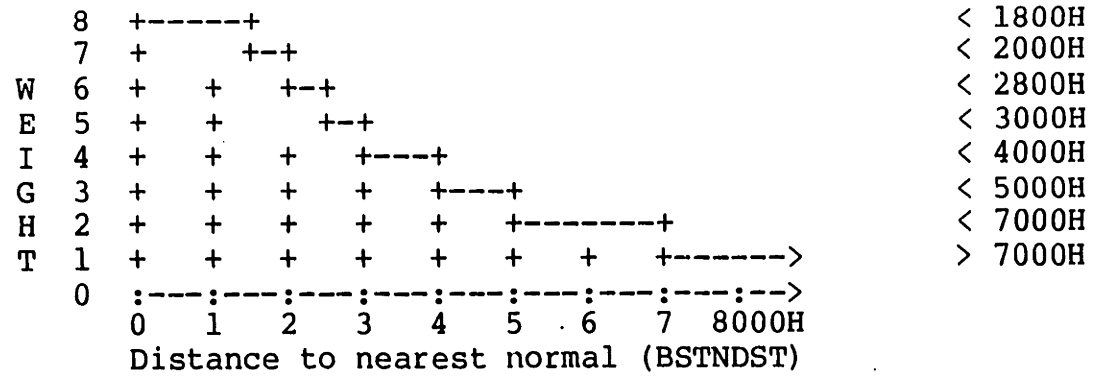



Figure 7-1a: Beat Weighting Based on Morphology

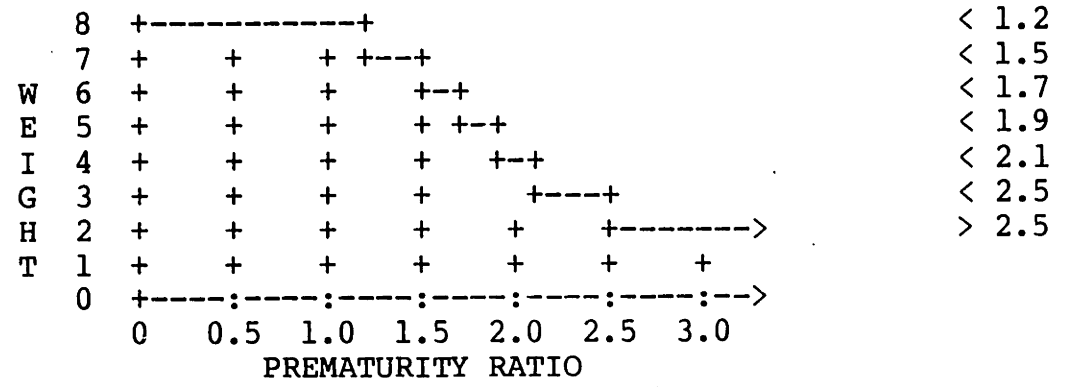



Figure 7-1b: Interval Weighting Based on Prematurity-Ratio

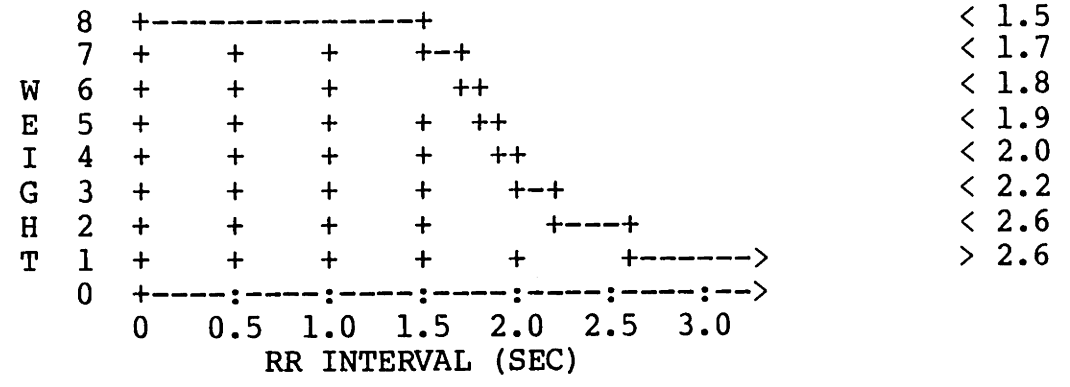



Figure 7-1c: Interval Weighting Based on Duration

Experimental results for the predictor

A series of experiments was performed to determine the optimum number of intervals averaged by the predictor. The performance measure was the detection accuracy for supraventricular premature beats (SVPBs) using two subsets of the annotated ECG data base that contain frequent SVPBs and runs of SVTA.

A "condensed" version of the ECG data base was developed to test parts of the algorithm that required only beat and rhythm labels, and their time of occurrence. The more compact format allowed the entire set of 48 half-hour "truth" annotation channels to be put on a single mass storage unit, making experiments such as the one described here easy to perform. The label and time of occurrence for each QRS complex was sequentially read from the condensed data base, and used as input to different versions of the interval predictor. Since the predictor weights the intervals according to morphology, the following weights were assigned to the data base beat labels shown below:

supraventricular beats
NORMAL, NPC, APC, SVPB, AEB, NEB [8/8]

abnormal morphology beats
LBBB, RBBB, ABERR, FUSION, NESC [6/8]

ventricular beats
PVC, FLWAV, VESC, PACE [1/8]

and artifact
UNKNOWN, SPIKE [1/8]

The prematurity ratio was used to update the prematurity ratio histogram for either SVPBs or NORMALs, depending on the classification of the complex.

Figure 7-2 shows SVPB detection performance for a subset of the 100 series annotated data base tapes. The subset includes tapes 100, 101, 108, 112, 114, 117, 118, 121, and 124, which have SVPBs, but relatively few runs of SVTA. The background rhythm for this subset is predominately normal sinus rhythm.

Figure 7-3 shows SVPB detection performance for a subset of the 200 series annotated data base tapes. This subset includes tapes 200, 201, 202, 205, 207, 209, 213, 215, 219, 220, 222, 223, 228, 231, and 233, which contain SVPBs and runs of SVTA. The background rhythm for this subset includes normal sinus rhythm, sinus arrhythmia, and other episodes of irregular rhythm. Regions of atrial flutter fibrillation, and runs of ventricular flutter, and tapes with paced rhythm were excluded from the study.

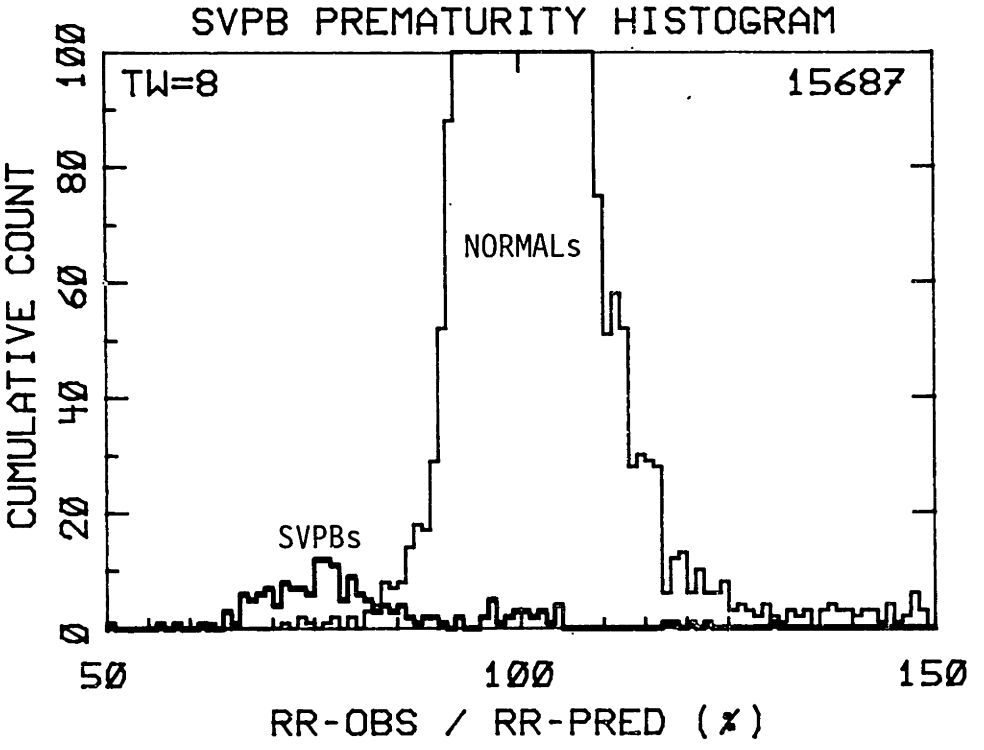
The SVPB prematurity histogram (heavy line) depicts the prematurity ratio histogram for SVPBs, and the NORMAL prematurity histogram (thin line) shows the prematurity ratio histogram for all NORMAL complexes. The cumulative error plot shows SVPB classification error as a function of prematurity ratio threshold. The cumulative error curve starting from the upper left corner is the fraction of false negative SVPBs (FNS); the error curve starting from the lower left corner is the fraction of false positive SVPBs (FPS). Both cumulative error curves are normalized to the number of true SVPBs in the subset.

The performance measure is the total error (FPS+FNS) using a prematurity ratio threshold that minimizes the sum of the two cumulative errors. The prematurity threshold can be obtained by noting where the SVPB and non-SVPB histograms intersect; at any other prematurity threshold, the total error will be larger.

The RR interval predictor was evaluated using a total cumulative weight thresholds of TW = 1*8, 2*8, 6*8, and 10*8 on the two data subsets. The cumulative weights correspond to a predictor that uses the most recent 1, 2, 6, or 10 normal-normal intervals. The error rates for the different weights and data subsets are shown in the following table:

| Trial | WT | P-R threshold | FPS | FNS | FPS+FNS |
|---|---|---|---|---|---|
| 100's, | 1*8 | 83% | 10% | 30% | 40% of # of SVPBs |
| | 2*8 | 85% | 6% | 27% | 33% |
| | 6*8 | 85% | 4% | 25% | 29% |
| | 10*8 | 84% | 3% | 25% | 28% |
| 200's, | 1*8 | 80% | 12% | 60% | 72% |
| | 2*8 | 82% | 16% | 50% | 66% |
| | 6*8 | 83% | 15% | 45% | 60% |
| | 10*8 | 83% | 12% | 40% | 52% |

It was relatively surprising to see that the number of averaged intervals did not substantially affect the performance on the 100 series tapes. The performance on the 200 series tapes was significantly worse, due to the irregular background rhythms for these tapes, and the greater number of long runs of SVTA. As the number of averaged intervals is increased, the height of the "on-time" SVPB histogram mode decreases. This illustrates how a longer time-constant average adapts more slowly to the change in heart rate due to the SVTA, permitting more SVPBs in the run to be properly classified.



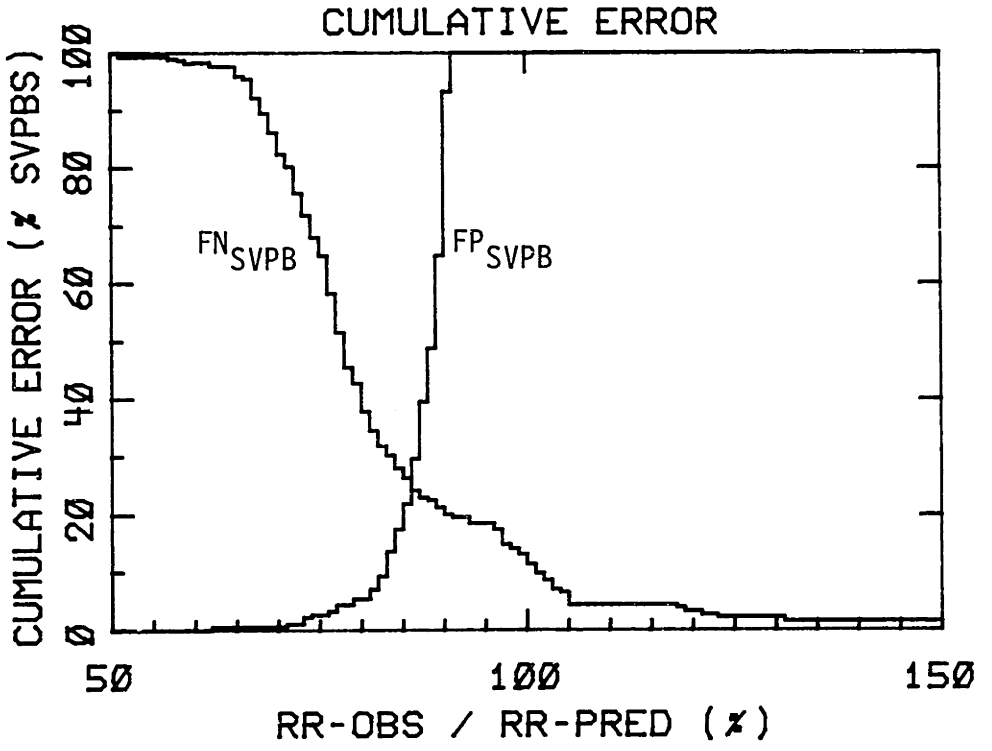
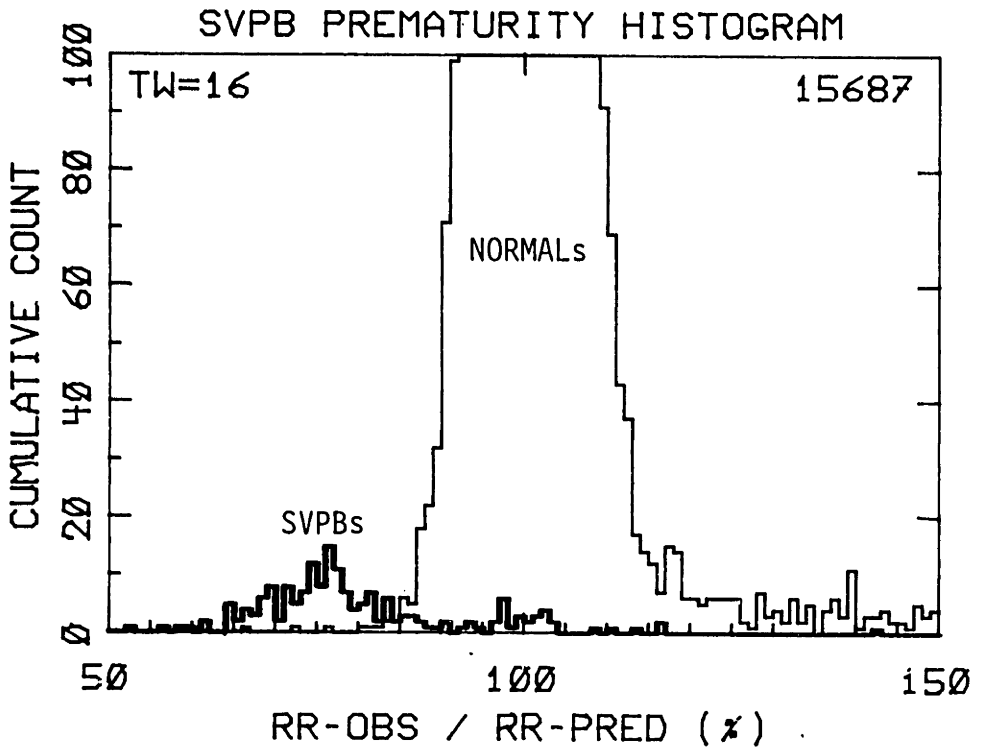




Figure 7-2a: SVPB Detection Performance
100 series with TW = 1*8



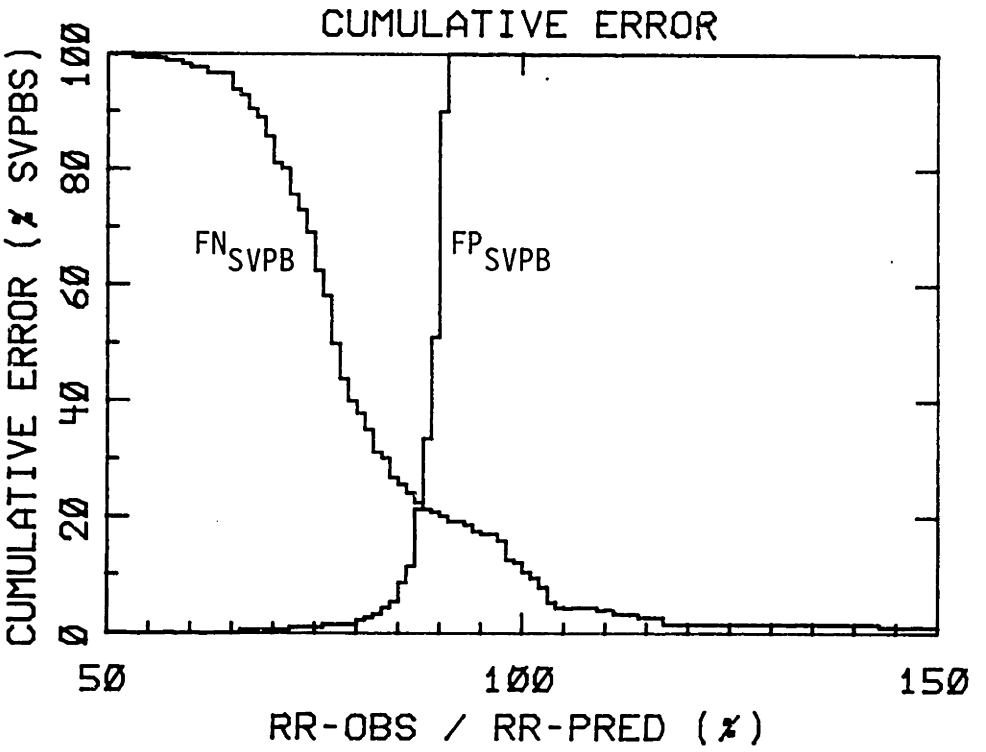
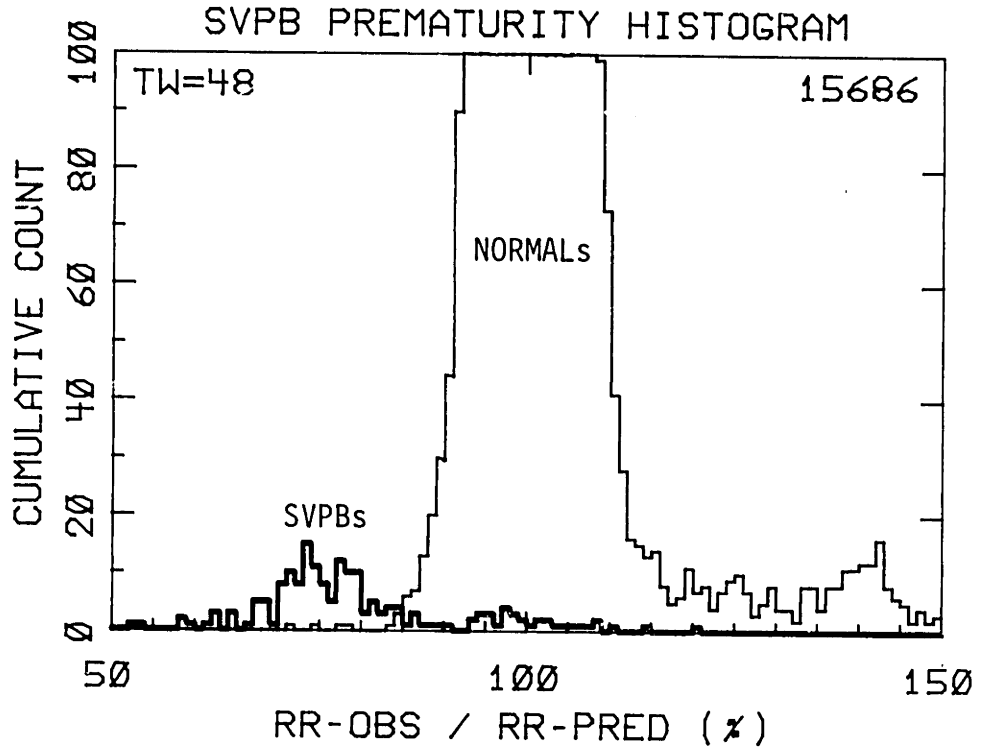




Figure 7-2b: SVPB Detection Performance
100 series with TW = 2*8



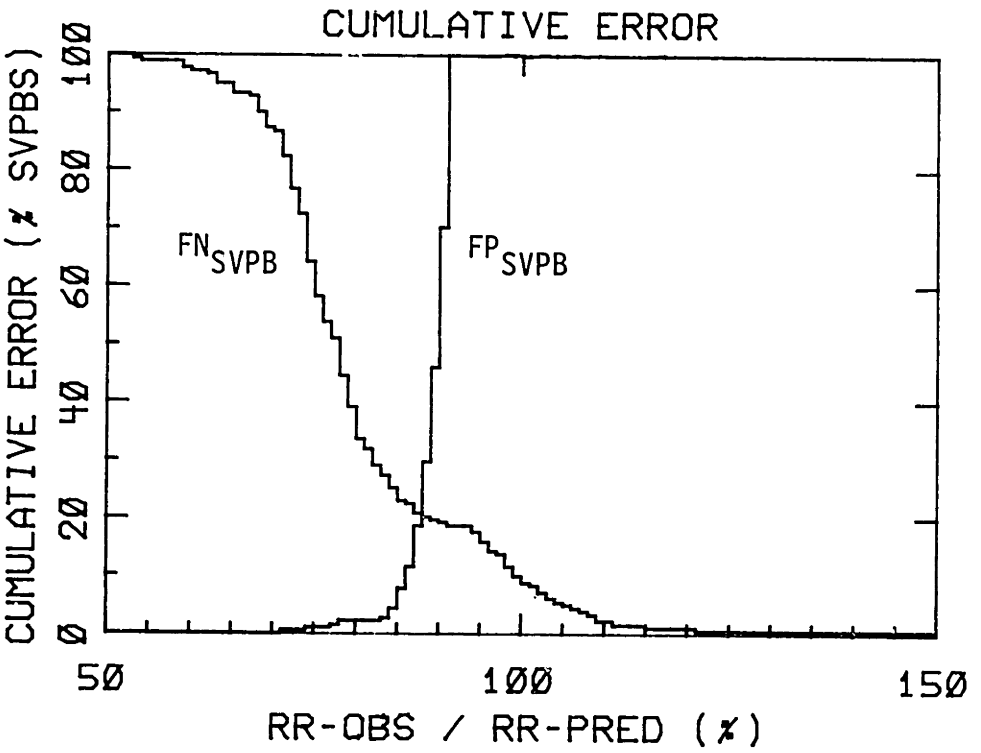
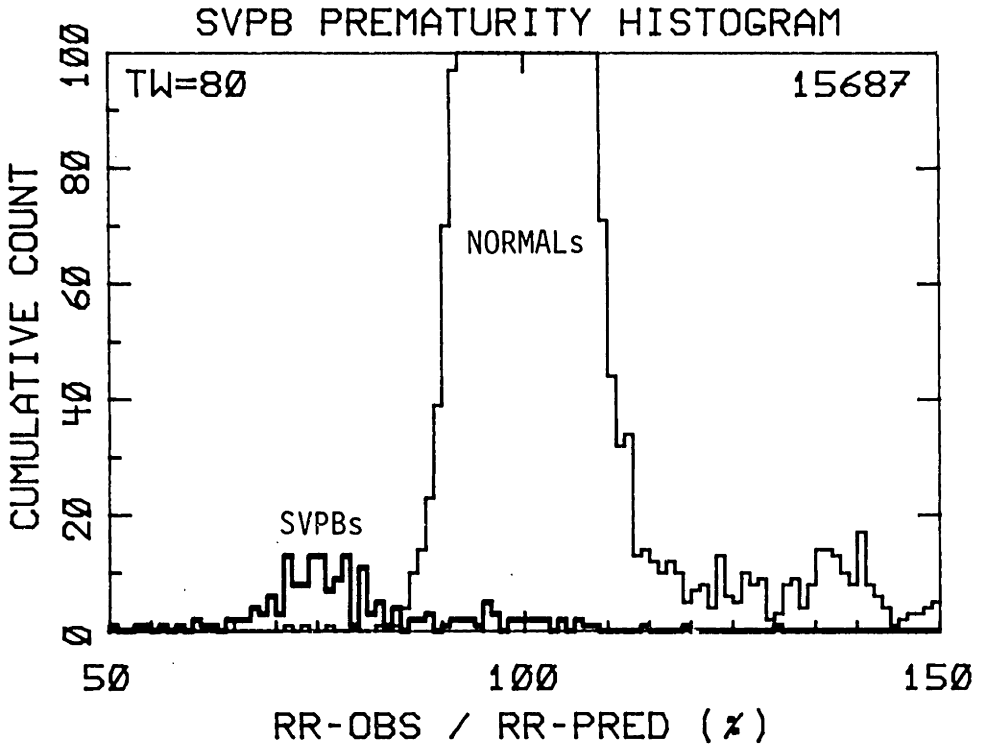




Figure 7-2c: SVPB Detection Performance
100 series with TW = 6*8



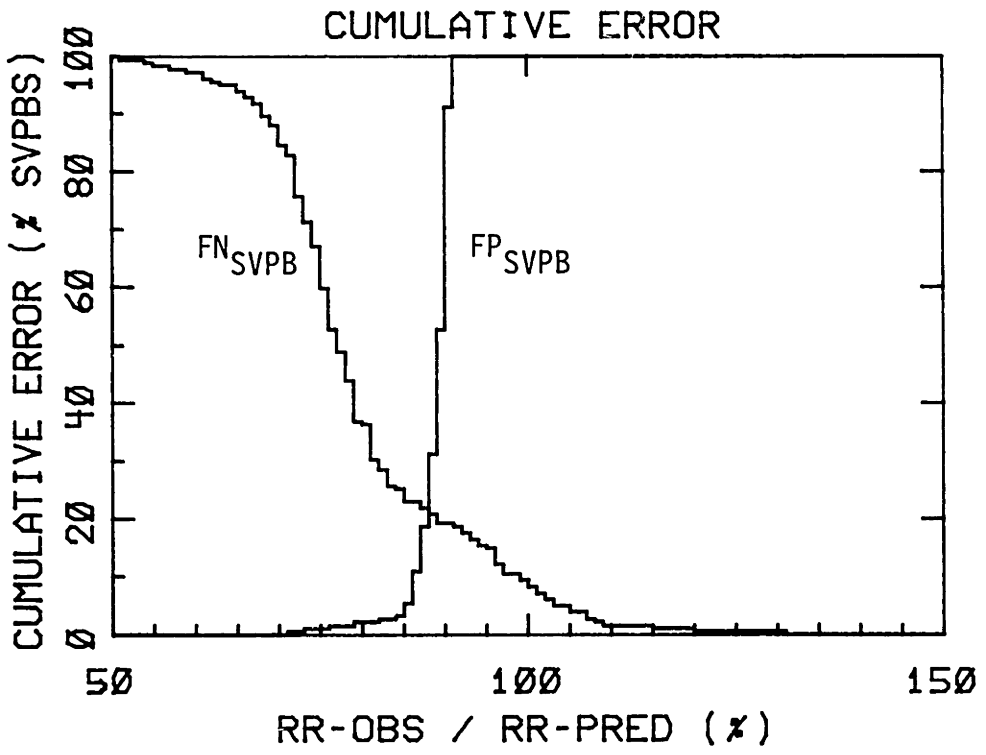
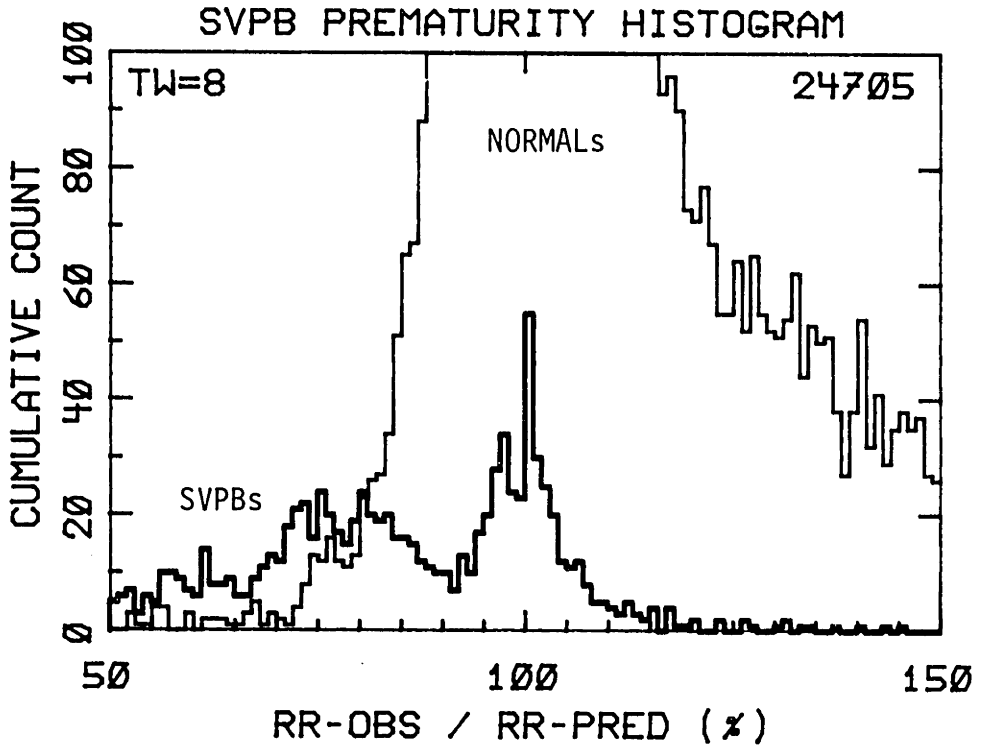




Figure 7-2d: SVPB Detection Performance
100 series with TW = 10*8



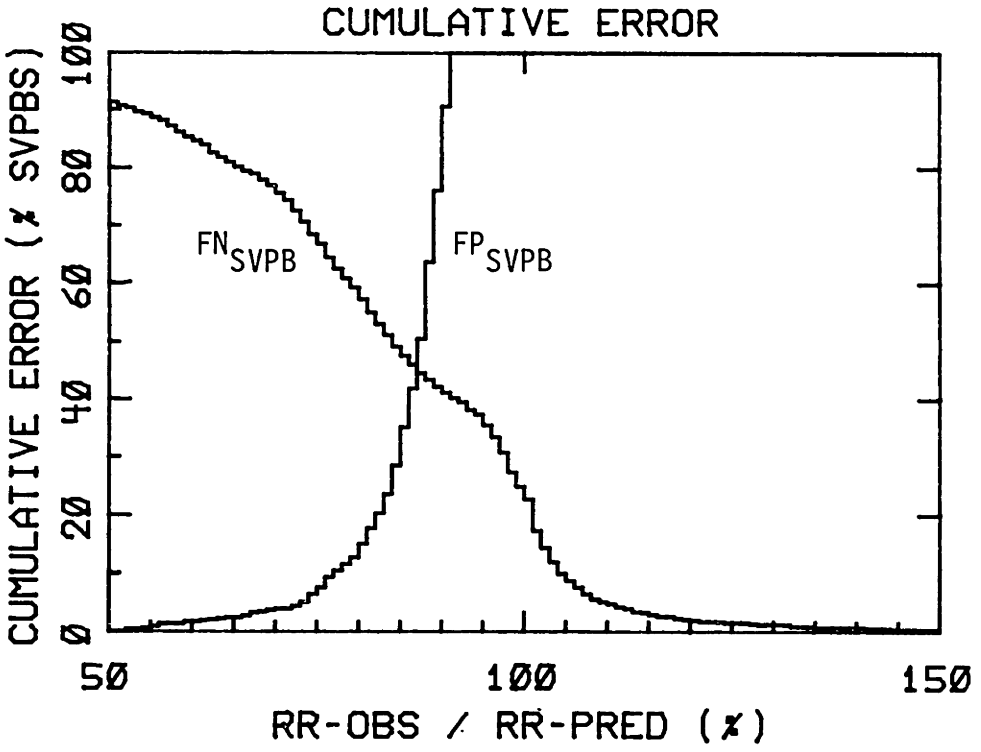
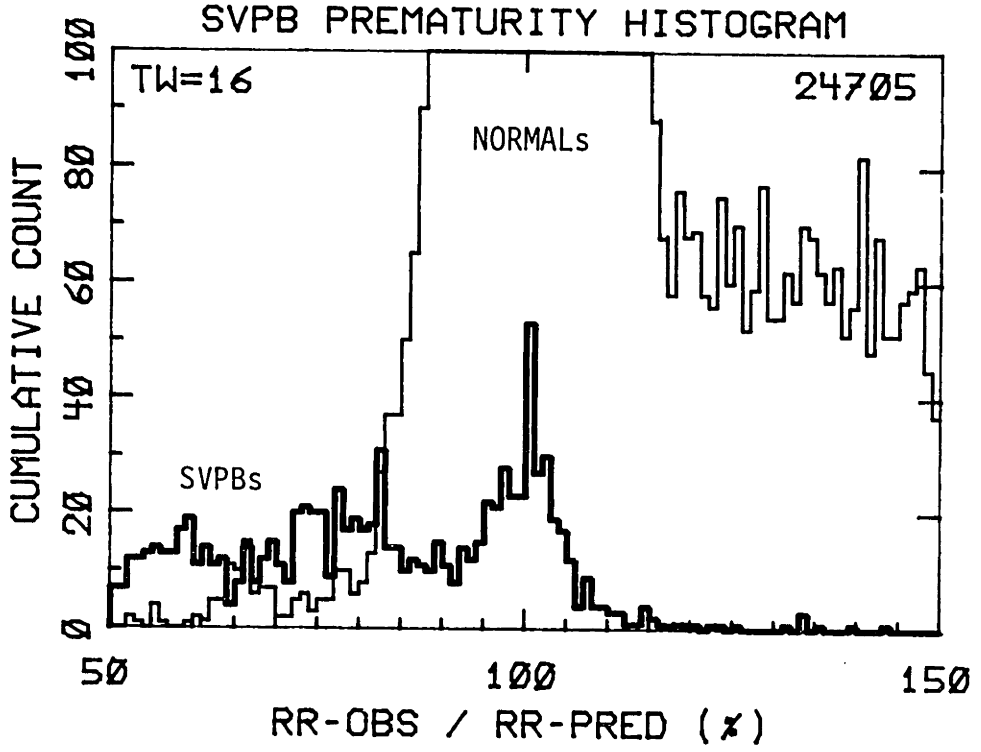




Figure 7-3a: SVPB Detection Performance
200 series with TW = 1*8



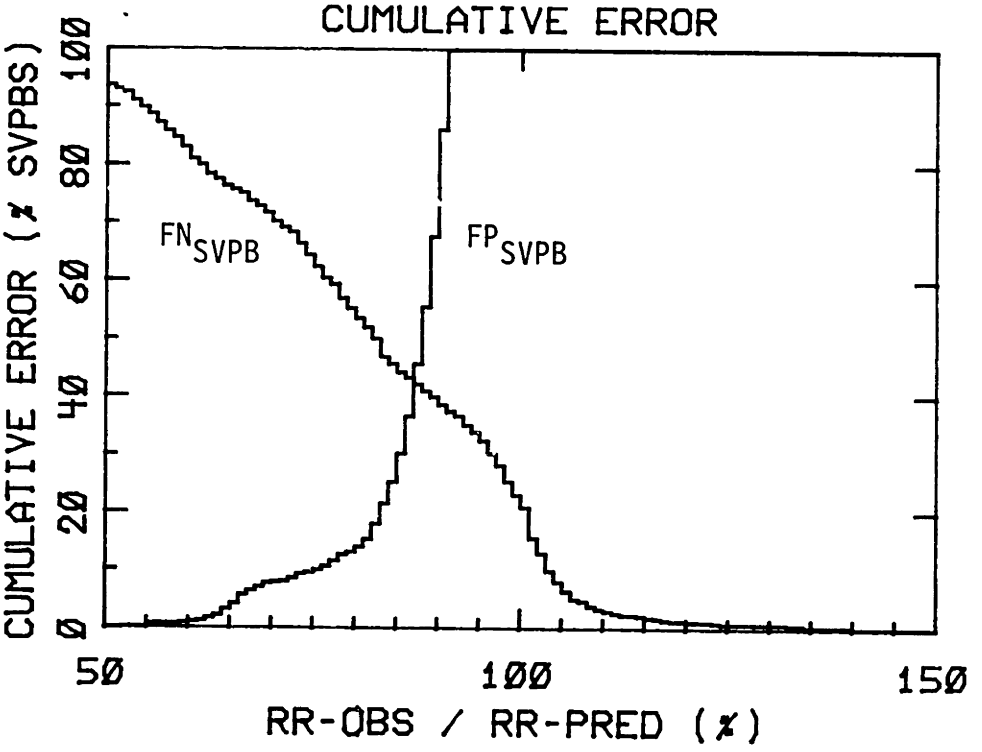




Figure 7-3b: SVPB Detection Performance
200 series with TW = 2*8

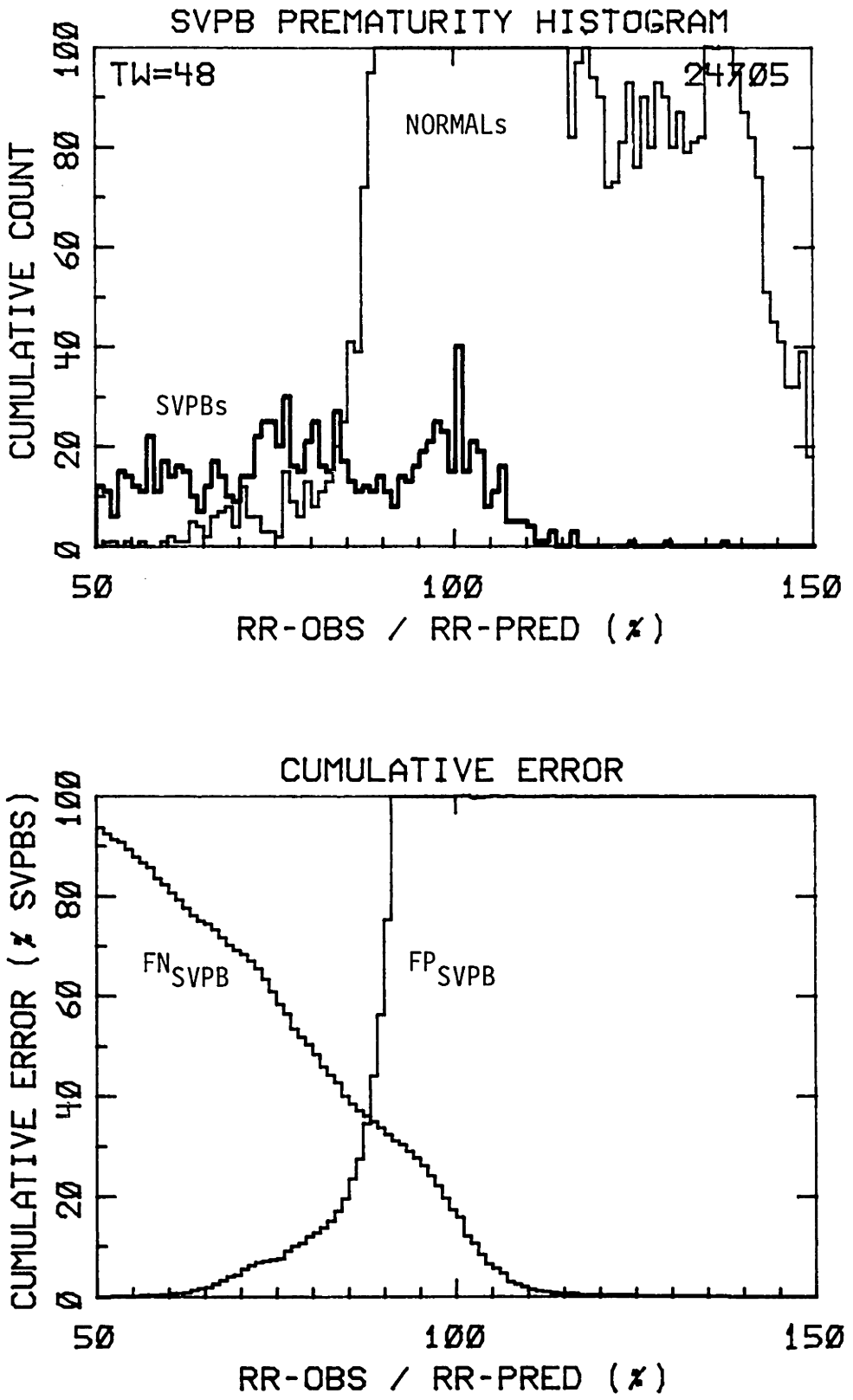
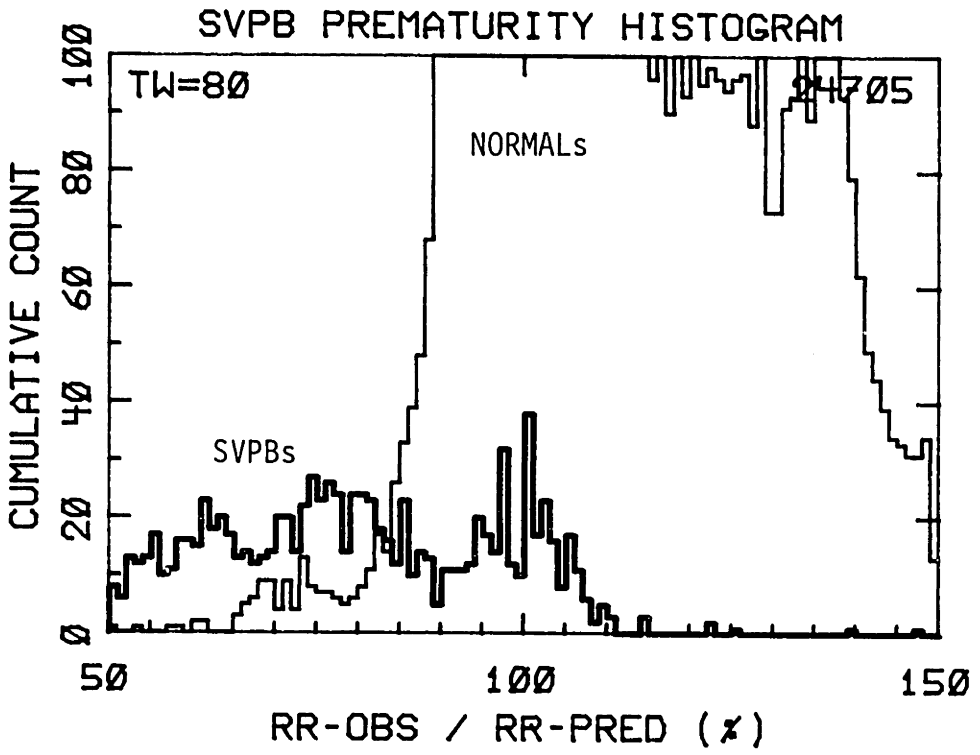



Figure 7-3c: SVPB Detection Performance
200 series with TW = 6*8



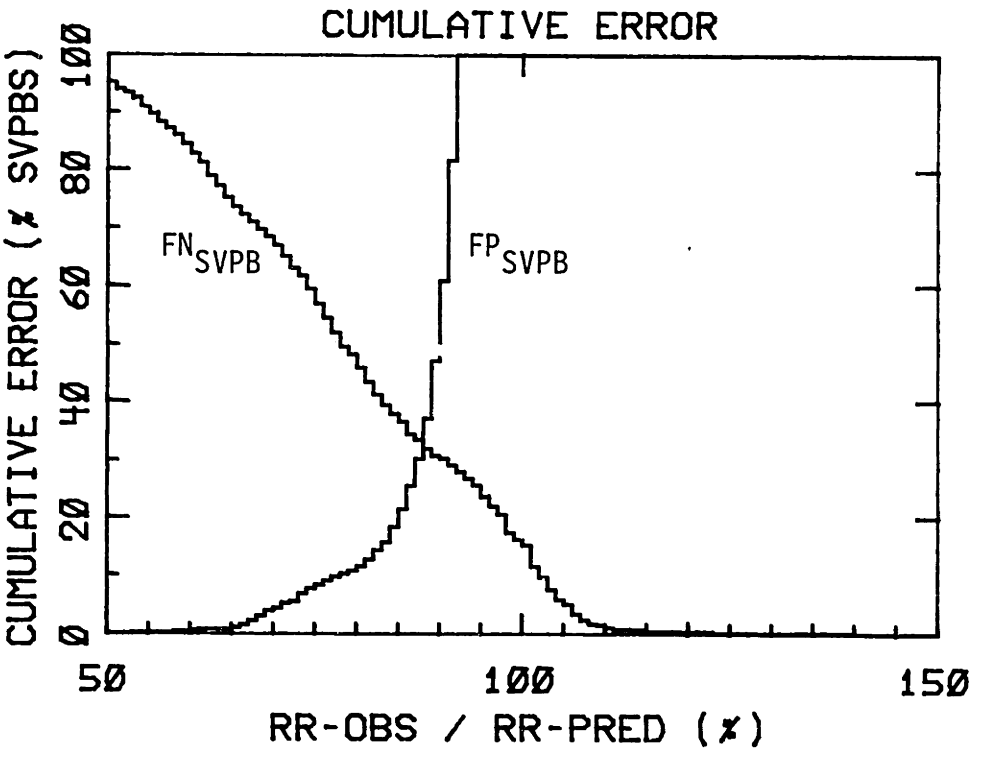




Figure 7-3d: SVPB Detection Performance
200 series with TW = 10*8

PVC detection performance

Another important question is how the interval predictor affects PVC detection performance. The PVC detection accuracy of the entire algorithm was evaluated using a small (TW=12) and large (TW=6*8,10*8) cumulative weight threshold. The data base subset used for this test is similar to that used for developmental evaluations described in the preceeding chapters. Except for tape 203, the performance for the two weight thresholds is nearly the same, indicating that PVC detection accuracy is fairly insensitive to the number of beats incorporated in the RR interval predictor. This is not surprising, since the morphology classifier maintains its own average prematurity ratio for each morphology class, and this is relatively insensitive to the type of predictor that is used. The larger cumulative weight threshold (TW=6*8 during NSR, and 10*8 during AF) was retained because of the better SVPB and SVTA detection accuracy with the longer term average.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Date: | 1/17/81 | | | | | | | |
| com: | TW=12 | | | | | | | |
| Tape 106 | fn | 62 | fp | 6 | tp | 453 | tn 1458 | + |
| Tape 200 | fn | 76 | fp | 0 | tp | 733 | tn 1737 | + |
| Tape 203 | fn | 139 | fp | 16 | tp | 287 | tn 2434 | + |
| Tape 207 | fn | 24 | fp | 6 | tp | 51 | tn 1621 | + |
| Tape 214 | fn | 96 | fp | 2 | tp | 0 | tn 1950 | + |
| Tape 223 | fn | 211 | fp | 1 | tp | 261 | tn 2066 | + |
| Total | fn | 608 | fp | 31 | tp | 1785 | tn 11266 | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Date: | 5/25/81 | | | | | | | |
| mod: | TW=6*8 if NSR, 10*8 if AF | | | | | | | |
| Tape 106 | fn | 68 | fp | 6 | tp | 447 | tn 1457 | + |
| Tape 200 | fn | 74 | fp | 0 | tp | 736 | tn 1736 | + |
| Tape 203 | fn | 135 | fp | 49 | tp | 293 | tn 2406 | + |
| Tape 207 | fn | 29 | fp | 7 | tp | 46 | tn 1620 | + |
| Tape 214 | fn | 97 | fp | 2 | tp | 155 | tn 1951 | + |

Tape 223 fn 192 fp 1 tp 280 tn 2066 +
Total fn 595 fp 65 tp 1957 tn 11236
obs: no substantial change (except FP increase for 203).

## 7.2 Atrial Fibrillation Detection

The long term statistical properties of the RR interval sequence can also analyzed to determine the underlying atrial rhythm. Instead of attempting to identify P-waves or atrial flutter waves directly, the RR interval sequence can be examined for patterns indicative of atrial fibrillation and flutter. During periods of atrial fibrillation and flutter, the atria exhibit a chaotic, non-synchronous pattern of electrical activity. This is accompanied by intermittent transmission through the AV node, leading to an irregular ventricular response at a rate between 50 to 200 bpm. In addition, the morphologic features of the QRS complex are apt to vary due to the irregular baseline during atrial fibrillation and flutter.

The bedside monitor makes a binary decision about whether or not atrial fibrillation is present and then adjusts several classifier thresholds to account for the increased variability of QRS morphology features. Specifically, the minimum prematurity thresholds for declaring a family to be a PVC family are increased, and the minimum distance that a complex must be from the nearest normal family is enlarged to account for the increased morphologic variability. When atrial fibrillation is detected, all supraventricular beats are labeled NORMAL.

Several techniques to detect atrial fibrillation were developed and evaluated. Two of them are discussed, along with a comparative evaluation. The first technique that was tried (but not used in the bedside monitor) treats the RR interval sequence as a Markov process, and attempts to determine whether or not atrial fibrillation could produce the observed RR interval sequence. The second method is based on how well previous RR intervals predict the value of the current RR interval. It offers somewhat better performance than the Markov chain approach, and thus was incorporated in the bedside monitor.

## 7.3 Markov Atrial Fibrillation Detector

Statistical approaches using Markov chain models can be used to indirectly determine the underlying rhythm. The approach is based on work by Gersh, who used Markov chain models to distinguish between atrial fibrillation and atrial fibrillation with occasional PVCs (Gersh,1975), and to recognize other rhythm patterns (Gersh,1970). In the work described here, Markov chain models were used to distinguish between normal sinus rhythm (NSR) and atrial fibrillation and flutter (AFIB).

### Discussion

The sequence of RR intervals is first transformed into a three symbol Markov chain sequence where each interval is classified as short (S), regular (R), or long (L). The transformation uses 85% and 115% prematurity ratio thresholds to

classify the intervals. The sequence of symbols (S, R, and L) describes the time pattern of RR intervals, and is used to determine the underlying rhythm.

The a posteriori probablility of a given sequence of intervals

X = [ X(1), X(2), X(3), ... X(N) ]

with initial state X(0) arising from a given rhythm model r is, by the Markov hypothesis,

P(X|r) = P(X(1)|X(0),r) * P(X(2)|X(1),r) * ... * P(X(N)|X(N-1),r)

where P(X(i)|X(i-1),r) is the probability for the transition from state X(i-1) to state X(i) for the rhythm model r.

The maximum likelihood procedure for choosing the most probable rhythm model is to choose r such that P(X|r) is maximum. Since only a binary rhythm decision (r = AF or NSR) must be made, a transition matrix can be defined in terms of the log of the transition probability ratios:

TRMAT(S1,S2) = K * log [P(S2|S1,AF)/P(S2|S1,NSR)]

where K is an arbitrary constant, and S1 and S2 are one of the possible states S, R, or L. Thus a transition matrix entry whose value is greater than zero represents a transition that is more likely to occur during atrial fibrillation than during normal sinus rhythm.

A pair of interval classifications defines a transition. For each transition, the corresponding matrix entry is incorporated into a running average that, if greater than zero, indicates that AFIB is present, and if less than zero, indicates that AFIB is not present.

Markov AF detector performance

The Markov atrial fibrillation detector was evaluated using a subset of the annotated ECG data base. The subset was chosen to provide a wide variety of rhythms to exercise the detector, and includes the following tapes: 210 and 221, which have only atrial fibrillation as the underlying rhythm; 201, 202, 203, 219, and 222, which contain episodes of atrial fibrillation and normal sinus rhythm; and 207, 209, 213, 220, and 223, which have other non-AFIB rhythms such as bigeminy, trigeminy and NSR.

The transition matrix values for the Markov AFIB detector were obtained by scanning the data subset, and accumulating transition counts for pairs of RR intervals that did not contain ventricular ectopic beats. The transition counts for non-AFIB and AFIB rhythms are shown below, as well as the log transition probability ratio (log base 10), and the final transition matrix values used by the detector.

| Trans-ition | Transition Counts AF | NOT-AF | Log Ratio | TRMAT Entry |
|---|---|---|---|---|
| S to S | 649 | 232 | +.4467 | +45 |
| S to R | 907 | 192 | +.6743 | +67 |
| S to L | 495 | 505 | -.0087 | -01 |
| R to S | 929 | 388 | +.3792 | +38 |
| R to R | 3268 | 12413 | -.5796 | -58 |
| R to L | 1096 | 231 | +.6762 | +68 |
| L to S | 405 | 309 | +.1175 | +12 |
| L to R | 1170 | 402 | +.4693 | +47 |
| L to L | 792 | 519 | +.1836 | +18 |

The data subset was then used as test data for the Markov atrial fibrillation detector, using the transition matrix coefficients listed above. The transition matrix entry for each observed RR interval pair was incorporated in a running average (over the last 16 intervals). The value of the running average specified which bin to increment in one of two histograms, depending on the underlying rhythm obtained from the data base rhythm annotations. One histogram was updated only during periods of atrial fibrillation, and the other when atrial fibrillation was not present.

Detector performance histograms and cumulative error plots were obtained, and are shown in Figures 7-4a and 7-4b. Figure 7-4a shows the Markov detector performance when intervals containing PVCs are excluded. Figure 7-4b shows detector performance when intervals containing PVCs are included, and indicates the detector's robustness when morphologic information is not available. The detection threshold for lowest total number of errors, and the corresponding error rates are shown in the table below. The results here are compared with another technique described in the next chapter.

| Trial | AF threshold | FPAF | FNAF | FPAF+FNAF |
|---|---|---|---|---|
| PVCs excluded | -23 | 27% | 5% | 32% of beats in AF |
| PVCs included | -18 | 30% | 4% | 34% |

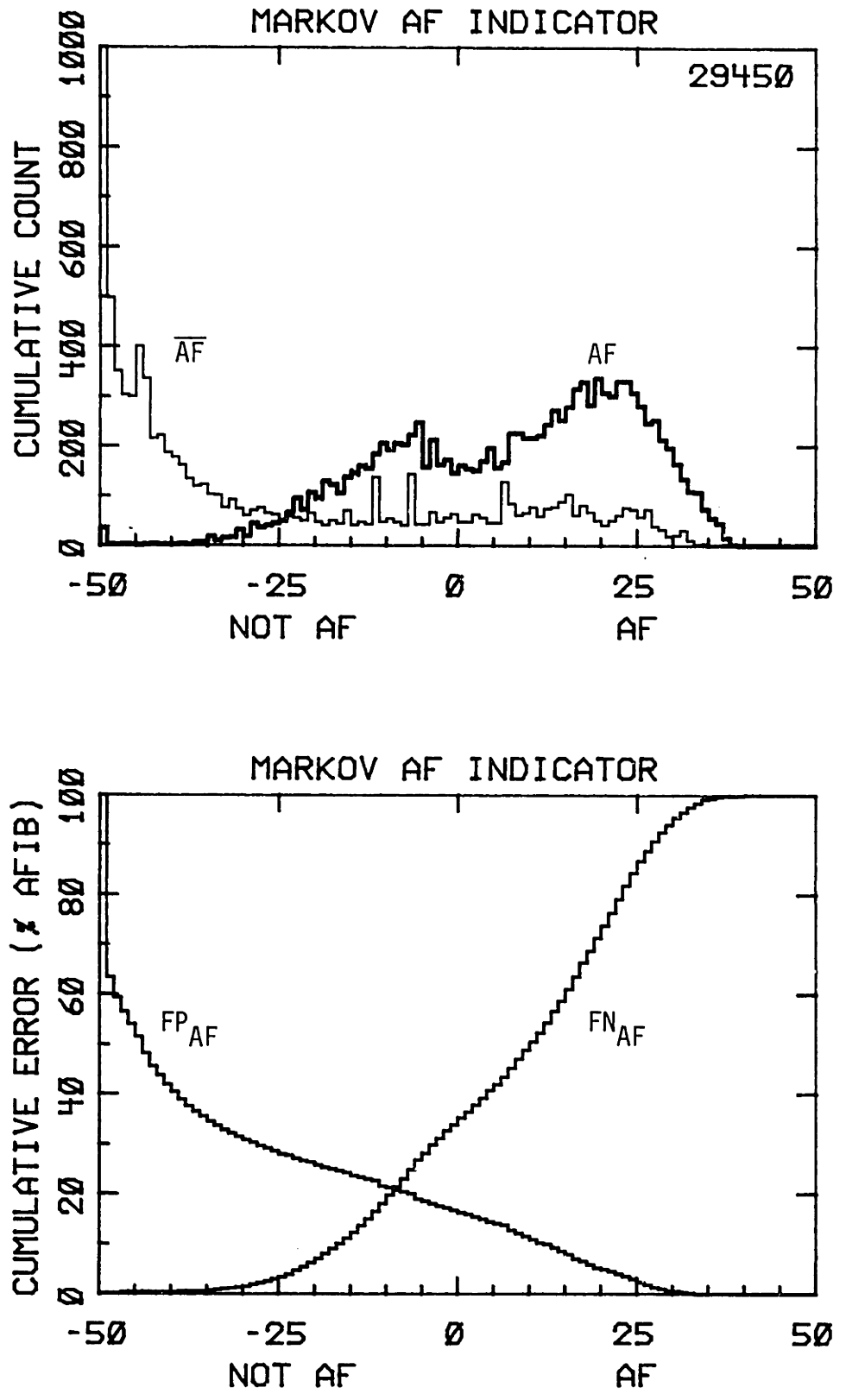



Figure 7-4a: Markov AF Detector Performance Excluding PVCs

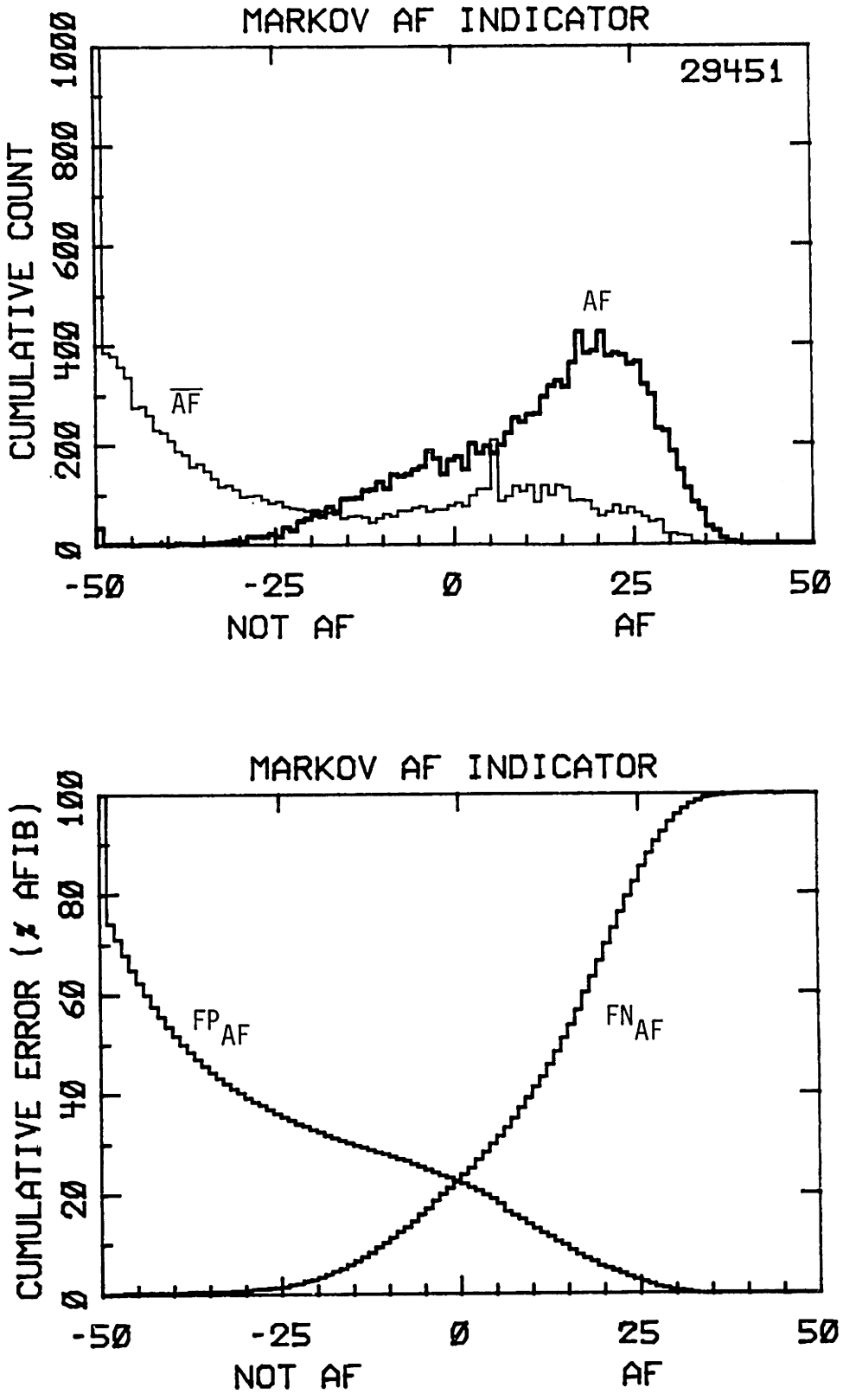



Figure 7-4b: Markov AF Detector Performance Including PVCs

## 7.4 Predictive Atrial Fibrillation Detector

Another property of atrial fibrillation is that there is no long-term temporal regularity to the RR interval sequence. The mean atrial rate of untreated atrial fibrillation is in the range of 400 to 600 impulses per minute, while the ventricular response for humans is in the range of 100 to 160 beats per minute. Only a fraction of the incoming atrial impulses succeed in initiating ventricular depolarizations.

This suggested another way of detecting atrial fibrillation, based on how well previous RR intervals predict the current RR interval. During atrial fibrillation, none of the preceeding intervals will consistently predict the value of the current interval. During normal sinus rhythm or sinus arrhythmia, however, the first preceeding interval is usually a good predictor of the current interval. During bigeminy, the second preceeding interval, and during trigeminy, the third preceeding interval are good predictors for the current interval. Thus, one may use the predictive error of the best predicting interval as an indicator for atrial fibrillation. During atrial fibrillation the predictive error will be large, whereas for most other rhythms the predictive error will be small.

The predictive AF detector

The predictive atrial fibrillation detector maintains a circular buffer of the previous M [M=10] RR intervals. Each interval in the buffer is flagged if it is a normal-normal interval. The difference between the current RR interval and each of the intervals in the buffer is computed. The difference is averaged into another array of M running averages if both intervals are normal-normal intervals, according to the formula:

ERROR(I) = (|ORRINT-ORRINT(I)| - ERROR(I))/16 + 15*(ERROR(I))/16

where I is the Ith preceeding interval.

Thus, each of the M running averages contains the average predictive error for the Ith preceeding interval.

The response of the predictive error array is shown in Figure 7-5 for two different rhythms. Figure 7-5a shows the response for sinus arrhythmia, which can be seen as a sinusoidal modulation of the RR interval sequence. The error array shows that the preceeding interval, and the 5th preceeding interval are good predictors for the current interval. In fact, this technique provides a computationally efficient method to determine whether sinus arrhythmia is present, and if it is, also allows a rough estimate of its period. Also, an RR interval predictor could use the error array to adjust the relative weights of previous beats into the predicted RR interval (provided that the frequency of sinus variation remains constant). For the rhythm in Figure 7-5a, a predictor that used the first, fourth, fifth, and sixth preceeding intervals would be

able to track the periodic variations of the RR interval sequence.

The response of the predictor array to atrial fibrillation is shown in Figure 7-5b. In this case, none of the preceeding intervals is a good predictor of the current interval, and the predictive error is quite large (150 msec, on the average). There is no particular interval that consistently predicts better than the others. This suggests a technique for recognizing atrial fibrillation: select the best predicting interval, and use its average predictive error as an indicator for atrial fibrillation. If the minimum predictive error is low, no AF is present; if the error is large, then AF is present.

Predictive AF detector performance

The predictive atrial fibrillation detector was evaluated using the same data base subset that was used to test the Markov AF detector. Four variations of the predictive detector were evaluated:

BIN Best Interval Normalized
The best predicting interval error normalized by (PRRINT).

BIU Best Interval Unnormalized
The best predicting interval error.

PIN Preceeding Interval Normalized
The preceeding interval error normalized by (PRRINT).

PIU Preceeding Interval Unnormalized
The preceeding interval error.

Detector performance histograms and cumulative error plots are shown in Figures 7-6 and 7-7. Two histograms were accumulated

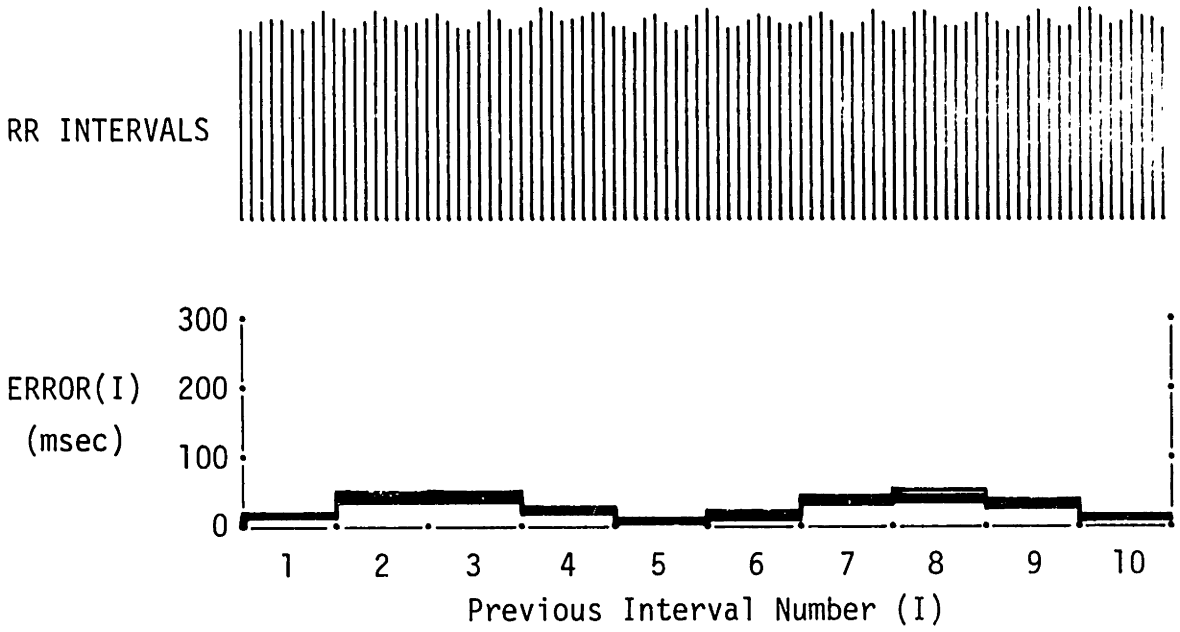



Figure 7-5a: Predictive Error for Sinus Arrhythmia

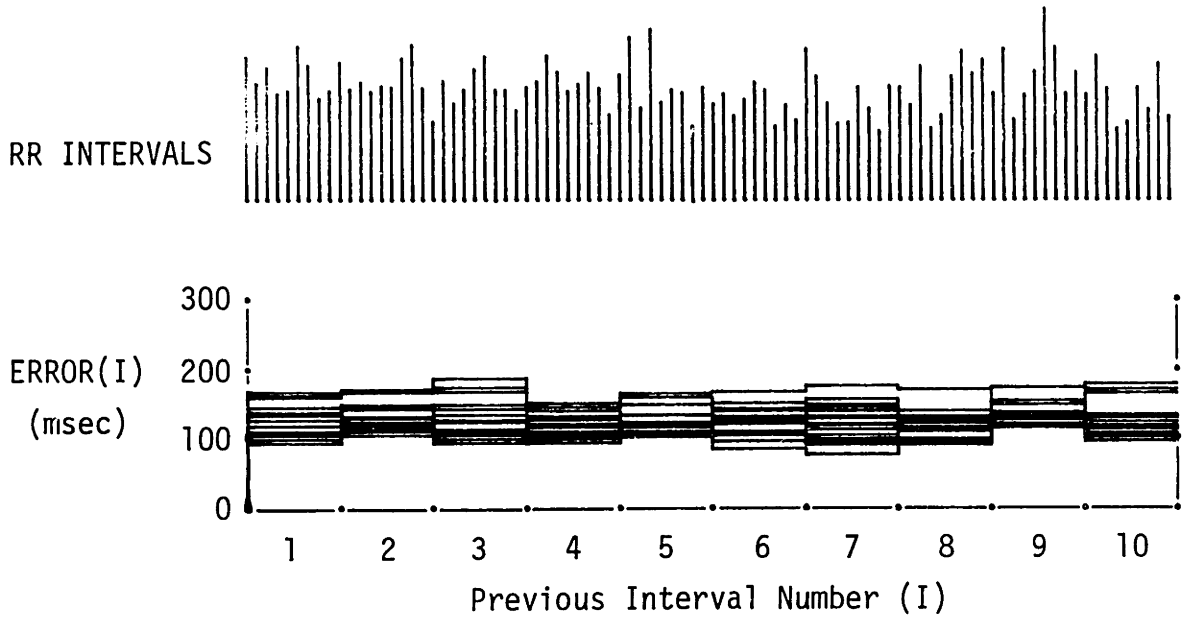



Figure 7-5b: Predictive Error for Atrial Fibrillation

For both figures, the sequence of RR intervals is indicated by the series of vertical lines. The error array is drawn every 5th interval.

for each trial: one represents the predictive error during AF, the second for all other rhythms. The corresponding cumulative error plot (as a function of detector threshold) are expressed as a percent of the total number of beats during atrial fibrillation (based on data base rhythm annotations). Figure 7-6 shows the predictive AF detector when intervals containing PVCs are excluded. Figure 7-7 shows detector performance when intervals containing PVCs are included, in indicates the detector's robustness when morphologic information is not available. The overall trial results are shown in the table below, along with the results for the Markov atrial fibrillation detector.

| Trial | AF threshold | FPAF | FNAF | FPAF+FNAF |
|---|---|---|---|---|
| PVCs excluded | | | | |
| BIN | 7% | 23% | 5% | 28% of beats in AF |
| BIU | 45ms | 28% | 3% | 31% |
| PIN | 9.5% | 24% | 10% | 34% |
| PIU | 60ms | 29% | 8% | 37% |
| Markov | -23 | 27% | 5% | 32% |
| PVCs included | | | | |
| BIN | 9% | 23% | 7% | 30% |
| BIU | 60ms | 27% | 5% | 32% |
| PIN | 10% | 35% | 4% | 39% |
| PIU | 70ms | 37% | 5% | 42% |
| Markov | -18 | 30% | 4% | 34% |

The Best Interval Normalized predictive AF detector had the best overall performance, but not by a significant amount (it had a 4% lower error rate than the Markov AF detector). It was selected as the AF detector for the bedside monitor. Another reason for its selection was that the predictive detector could be used to identify other rhythm patterns. Athough not currently implemented in the bedside monitor, the predictive error array

could be used to improve the performance of the RR interval predictor by recognizing sinus arrhythmia, and estimating its period.

Although the evaluation was used to select the best technique for recognizing atrial fibrillation, the actual detection thresholds were selected based on performance on individual data base tapes, and "human factors" considerations. It was felt that a lower false-positive rate (less than 23%) was desirable so that SVPBs and runs of SVTA would not be mysteriously "missed" (since all supraventricular beats are labeled NORMAL when AF is detected). Because of the non-stationary characteristics of atrial fibrillation, hysterisis was also incorporated. Two thresholds are used: if the normalized best predictive error exceeds 15%, AF is declared; if the predictive error is less than 8%, not-AF is declared; and for values between the two thresholds, the previous state (AF,not-AF) is declared.

The final classifier performance is shown in the two tables below: The table on the left shows the performance for the threshold (7%) that minimized the total error, the table on the right shows the performance for the pair of hysteritic thresholds. Although the total error rate is worse for the latter, the false positive rate for declaring AF was considered more acceptable. Another benefit of the pair of hysteretic thresholds is that the AF detector won't oscillate between the AF and not-AF states.

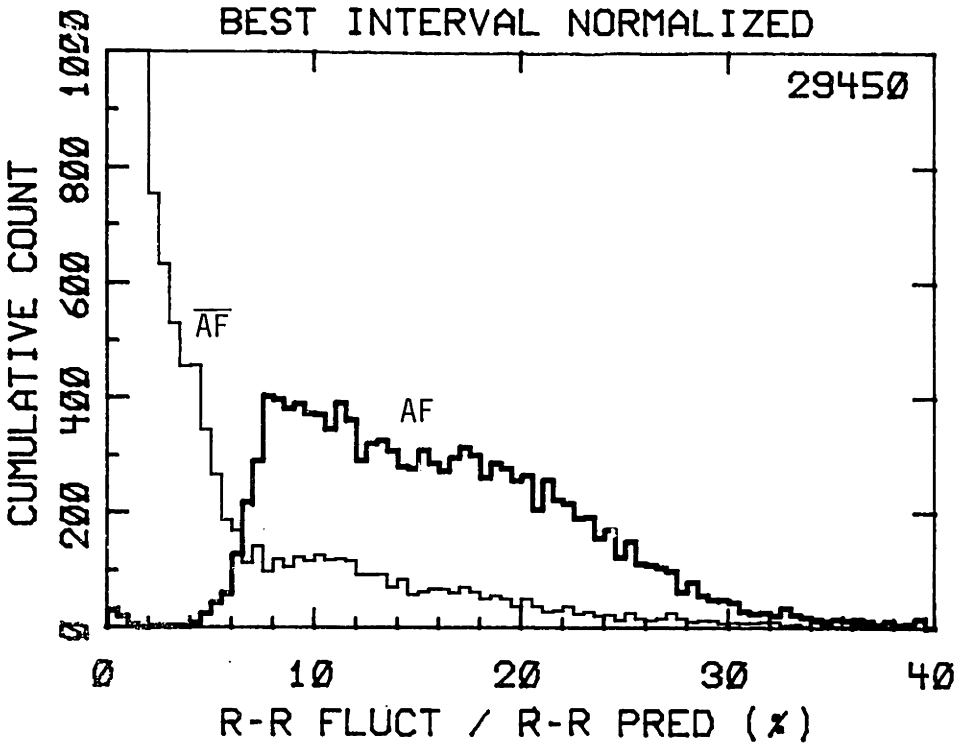
BIN AF detector (7%)

| truth | | $\overline{AF}$ | AF |
|---|---|---|---|
| | $\overline{AF}$ | 14489 | 2806 |
| | AF | 554 | 11602 |

BIN AF detector (8%-15%)

| truth | | $\overline{AF}$ | AF |
|---|---|---|---|
| | $\overline{AF}$ | 15383 | 1912 |
| | AF | 3551 | 8605 |

The data base subset used to evaluate the atrial fibrillation detectors was a particularly difficult one. The rhythm switched frequently between atrial fibrillation and and regular rhythm on many of the tapes. The frequent transitions were responsible for many of the errors, since the running average (over the previous 16 beats) could not rapidly respond to transitions between atrial fibrillation and normal sinus rhythm. Also, one of the tapes (tape 210) has a very regular ventricular response during atrial fibrillation, and substantially contributes to the total overall error.



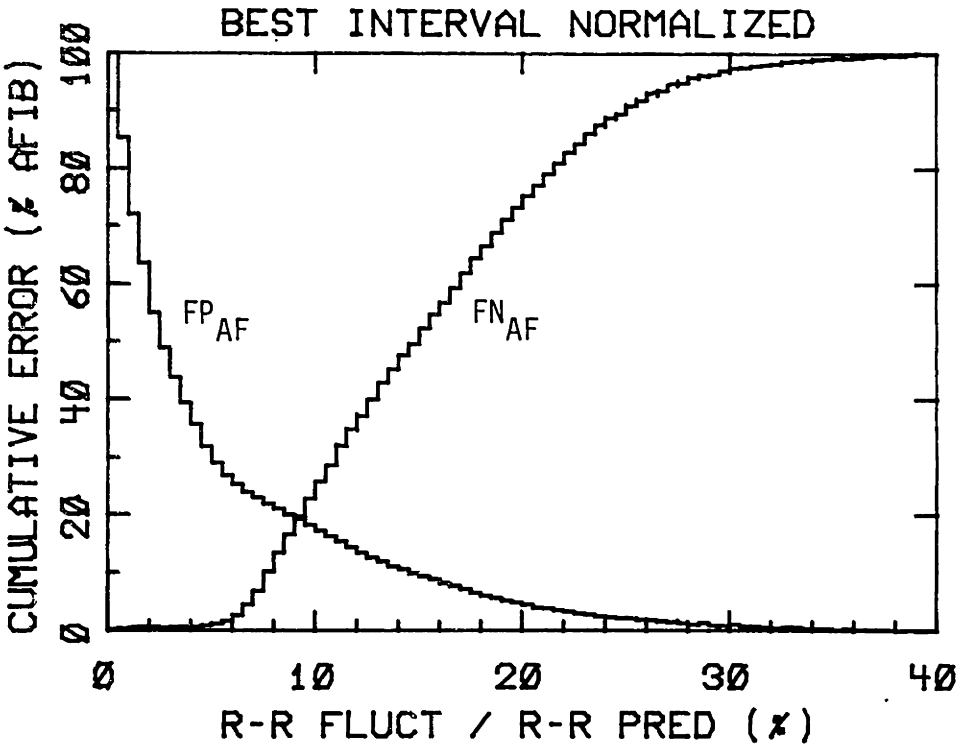
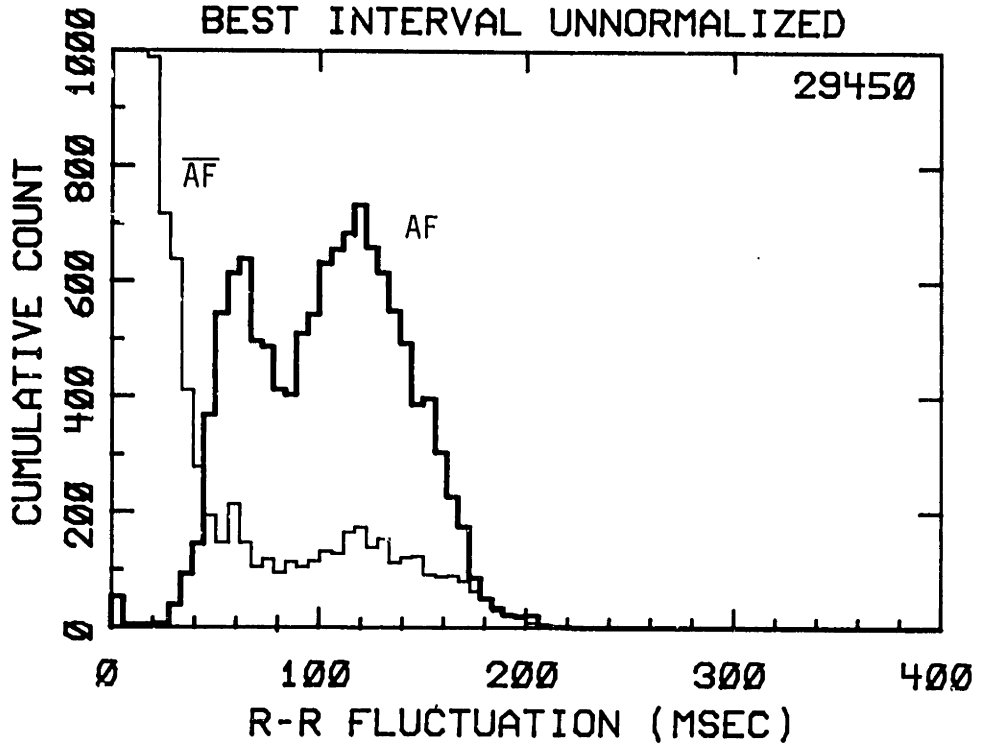




Figure 7-6a: Best Interval Normalized AF Detector Performance Excluding PVCs



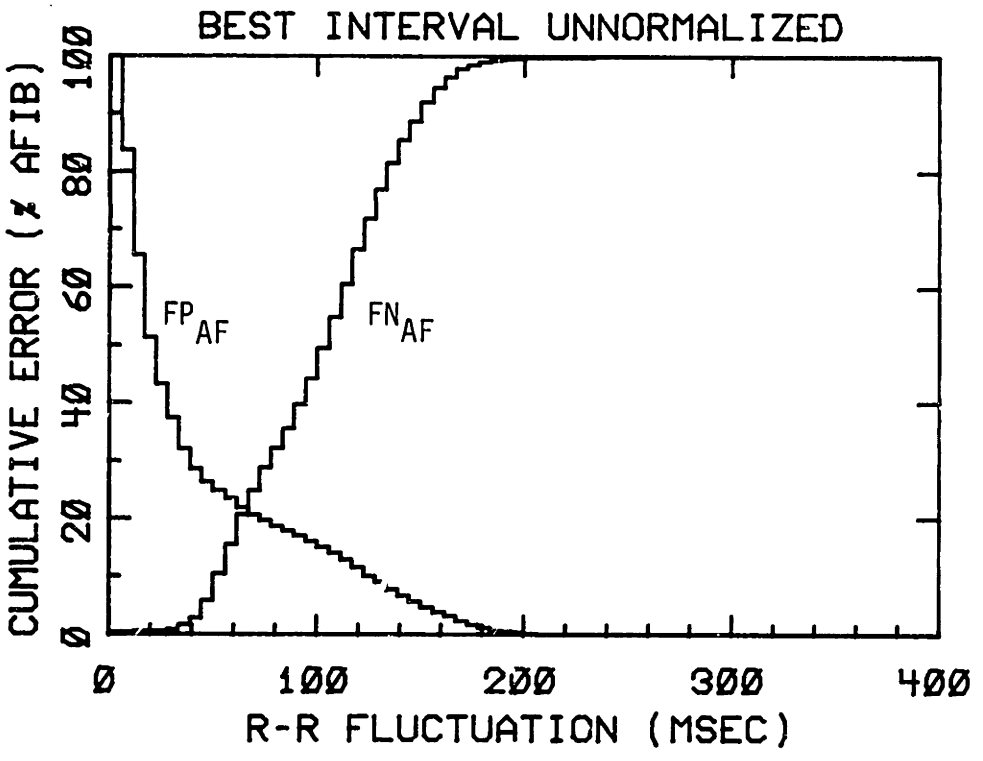
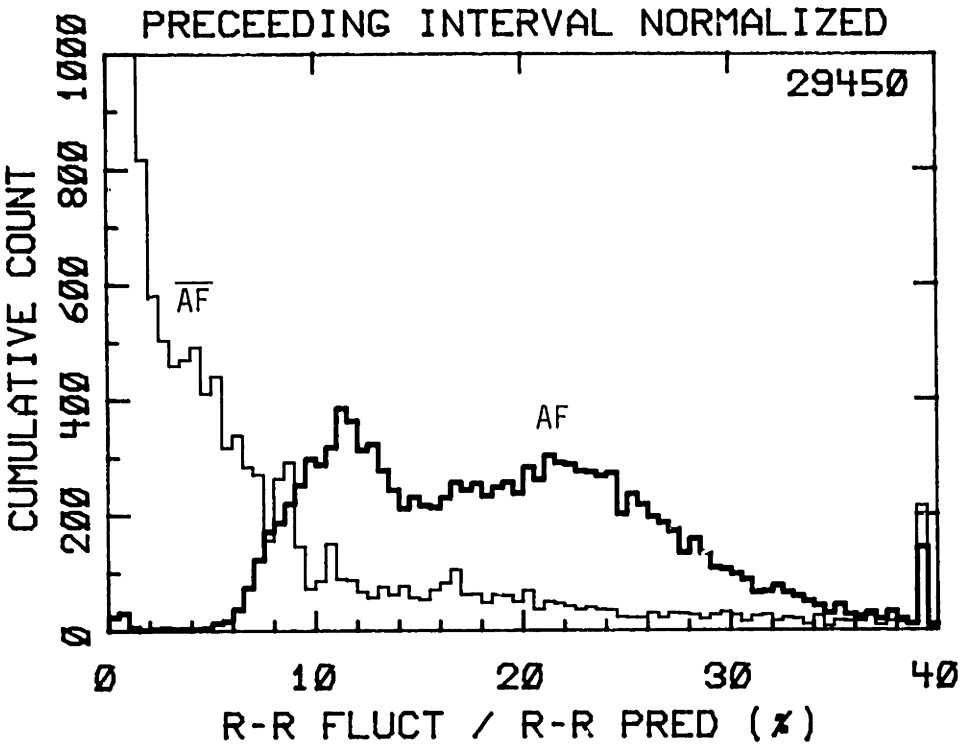




Figure 7-6b: Best Interval Unnormalized AF Detector Performance Excluding PVCs



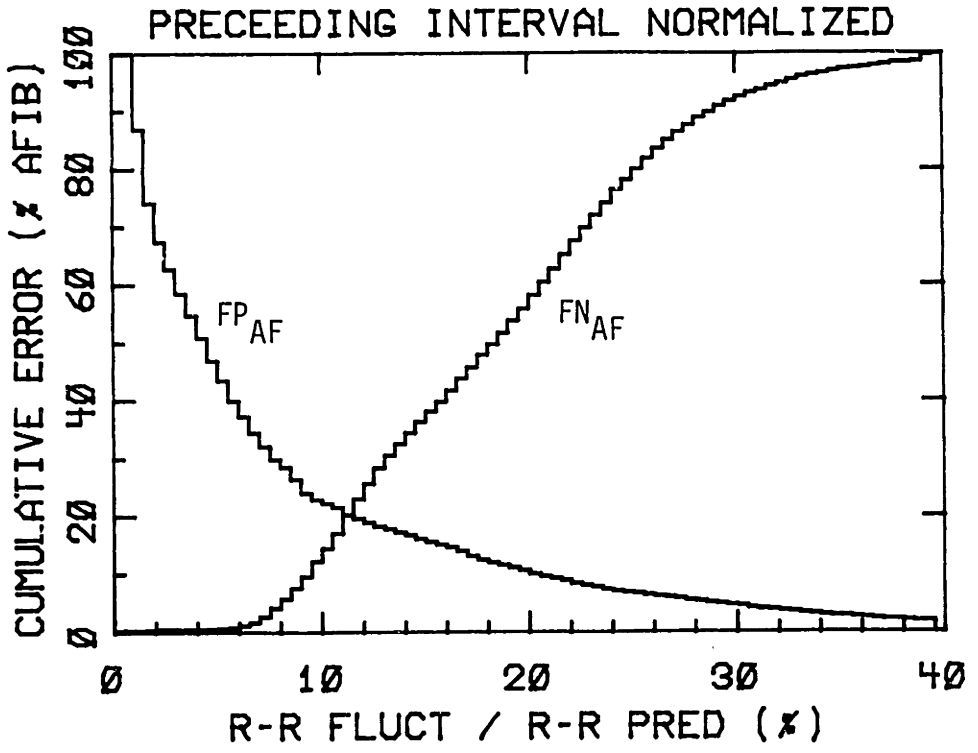
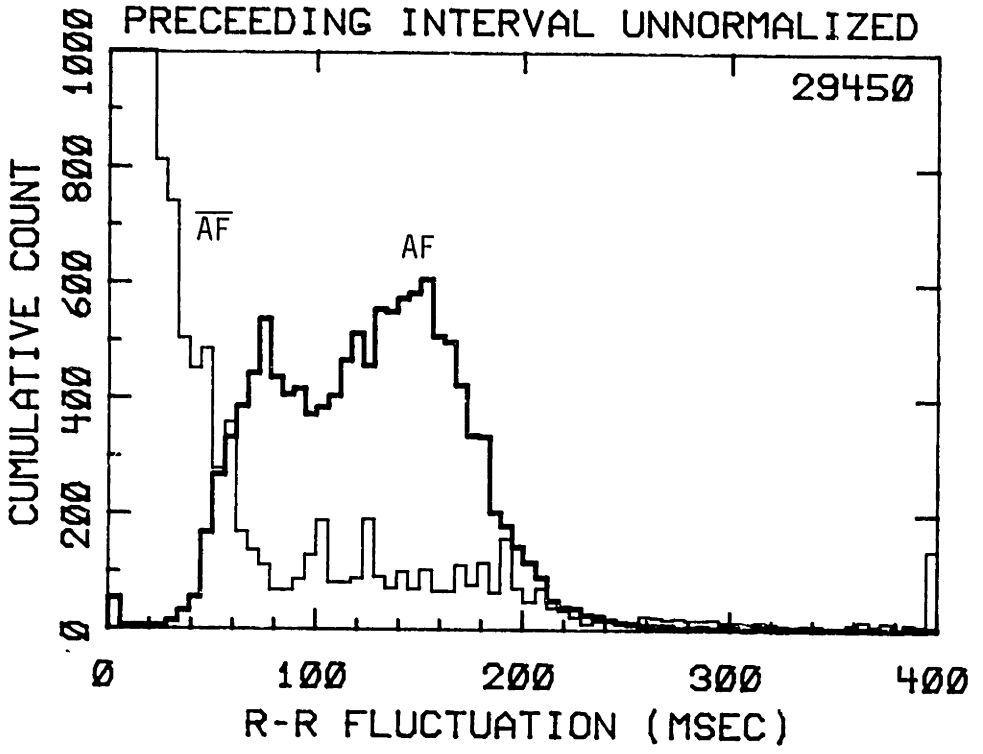




Figure 7-6c: Preceeding Interval Normalized AF Detector Performance Excluding PVCs



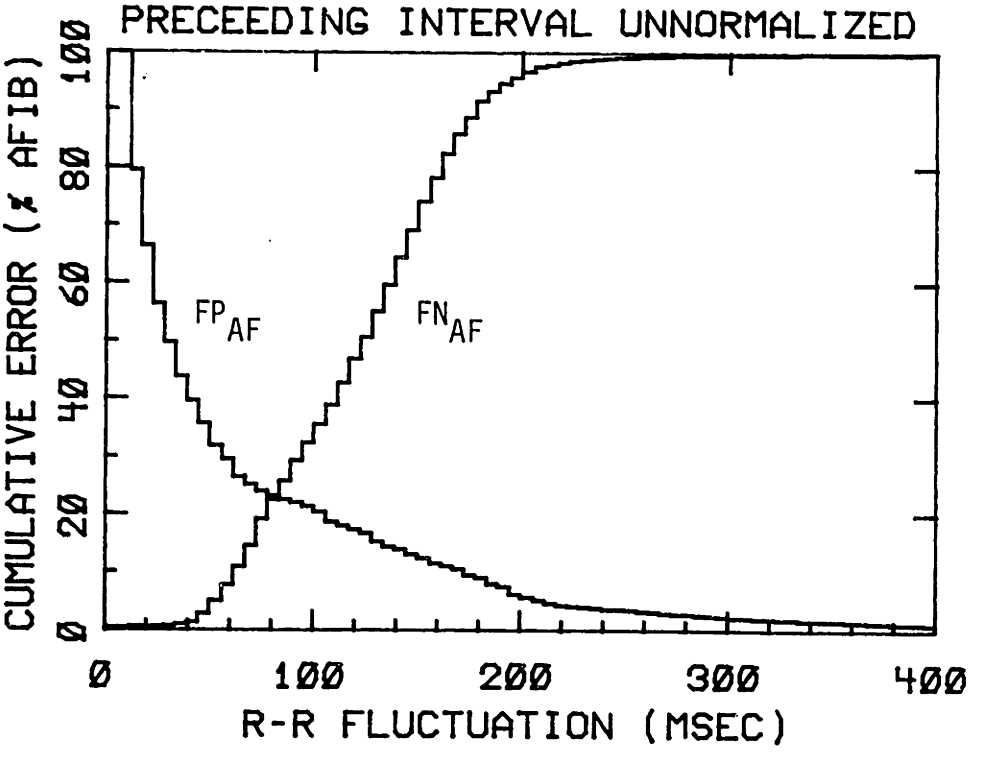




Figure 7-6d: Preceeding Interval Unnormalized AF Detector Performance Excluding PVCs

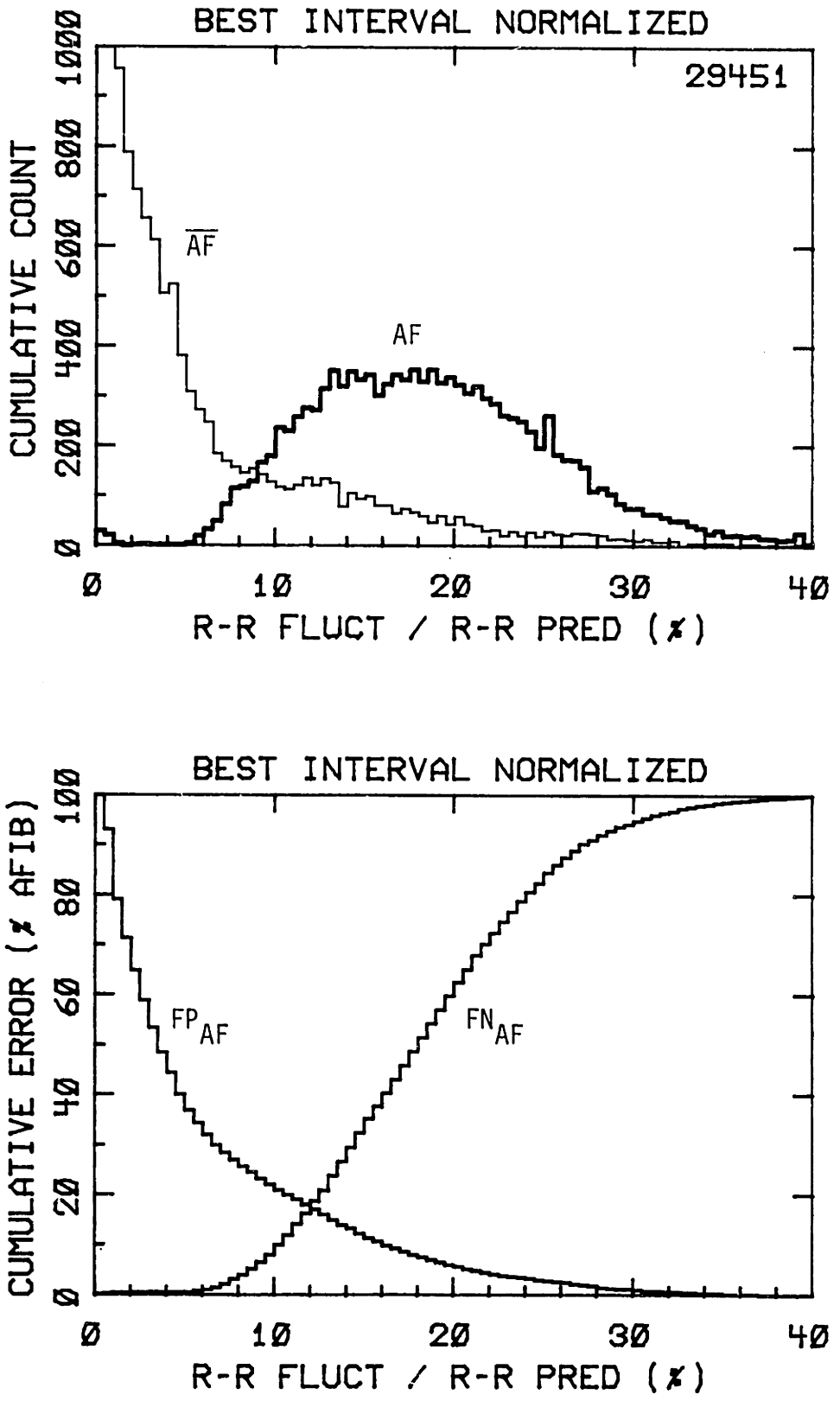
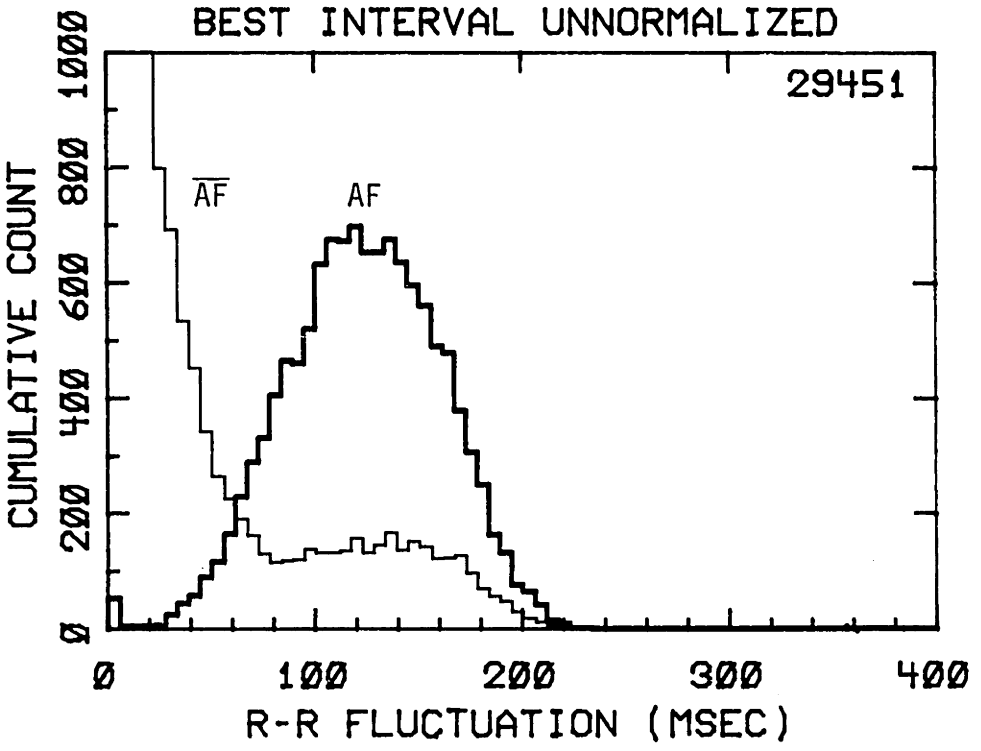



Figure 7-7a: Best Interval Normalized AF Detector Performance Including PVCs



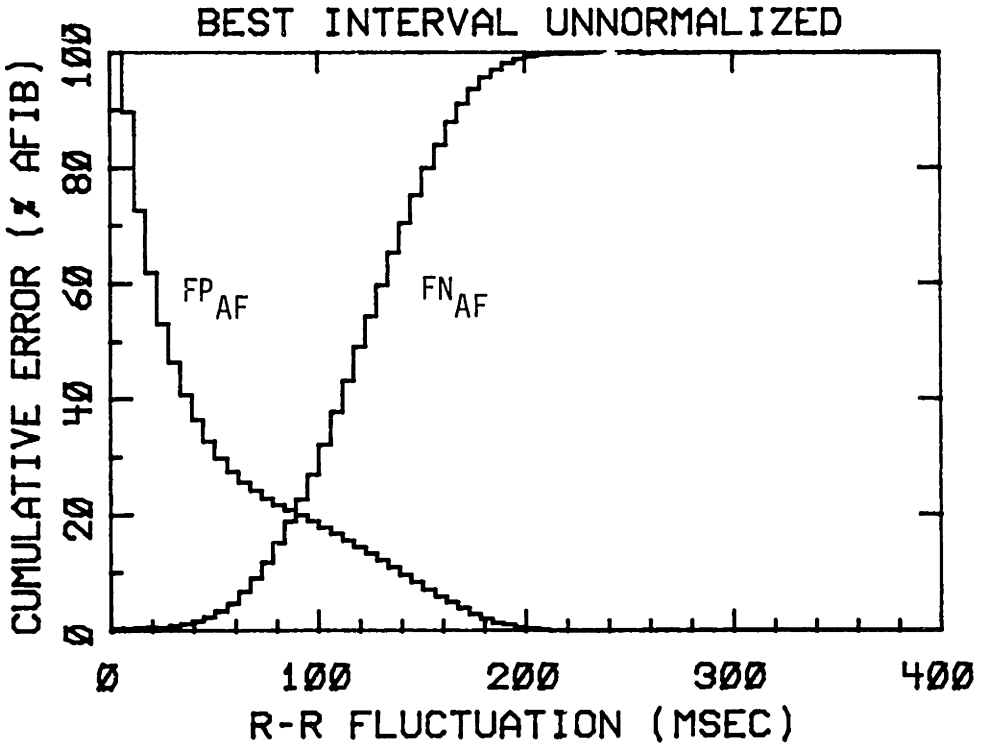
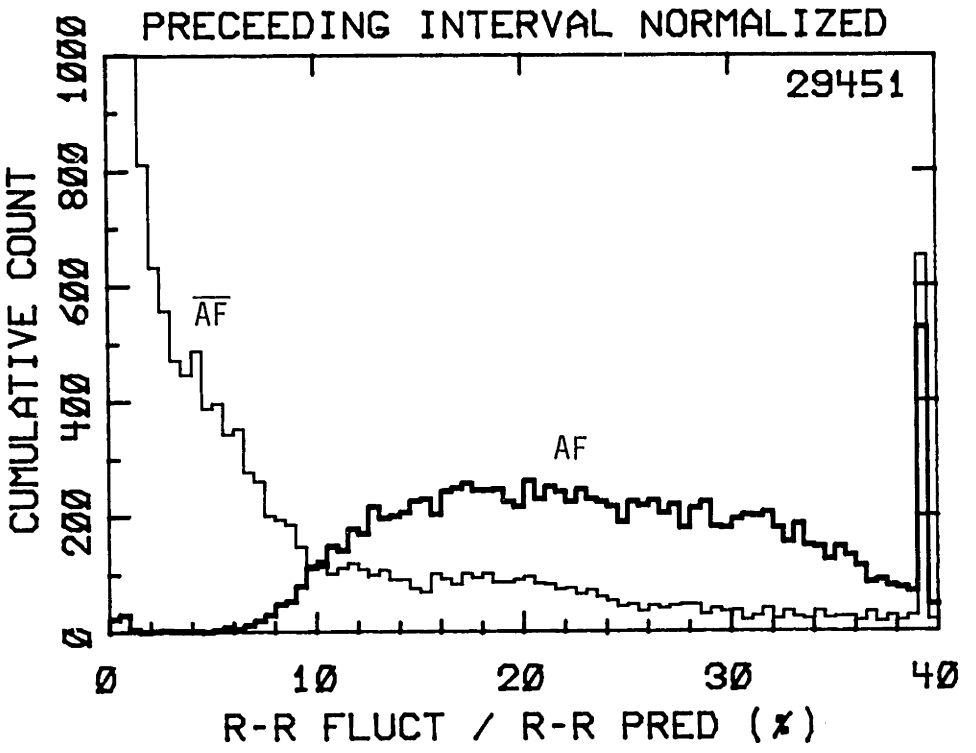




Figure 7-7b: Best Interval Unnormalized AF Detector Performance Including PVCs



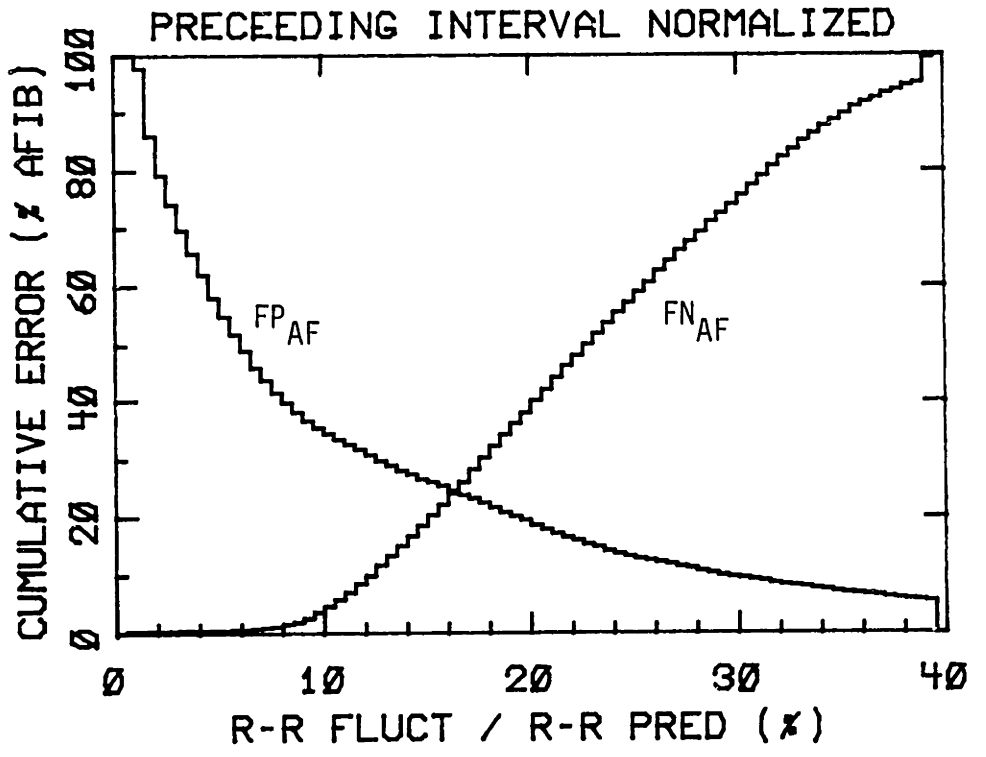
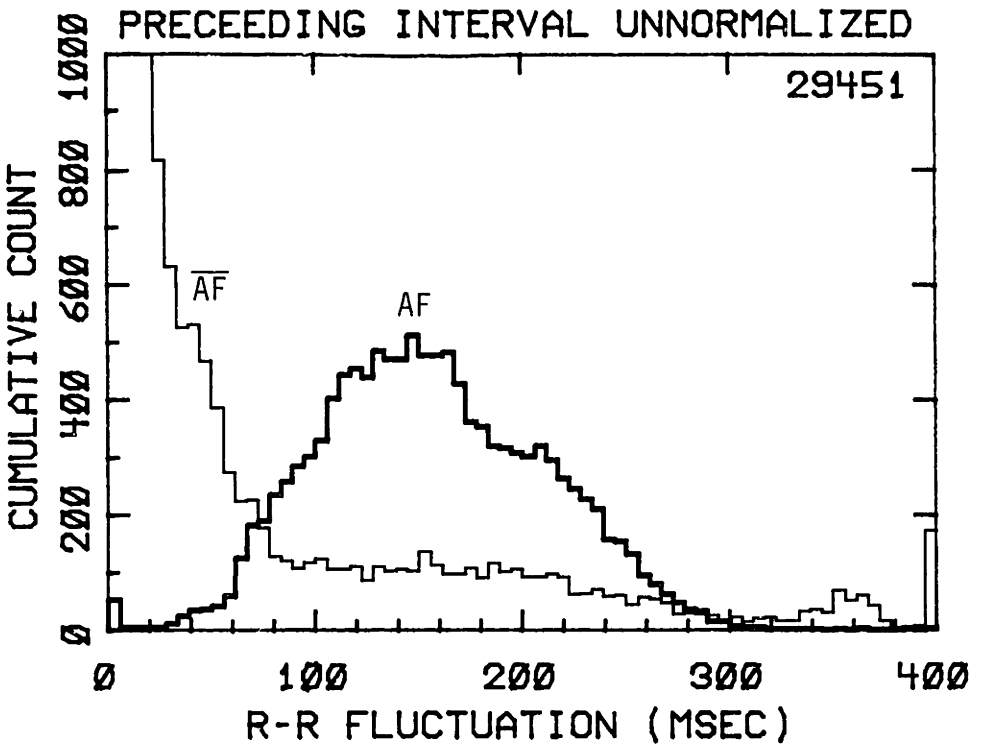




Figure 7-7c: Preceeding Interval Normalized AF Detector Performance Including PVCs



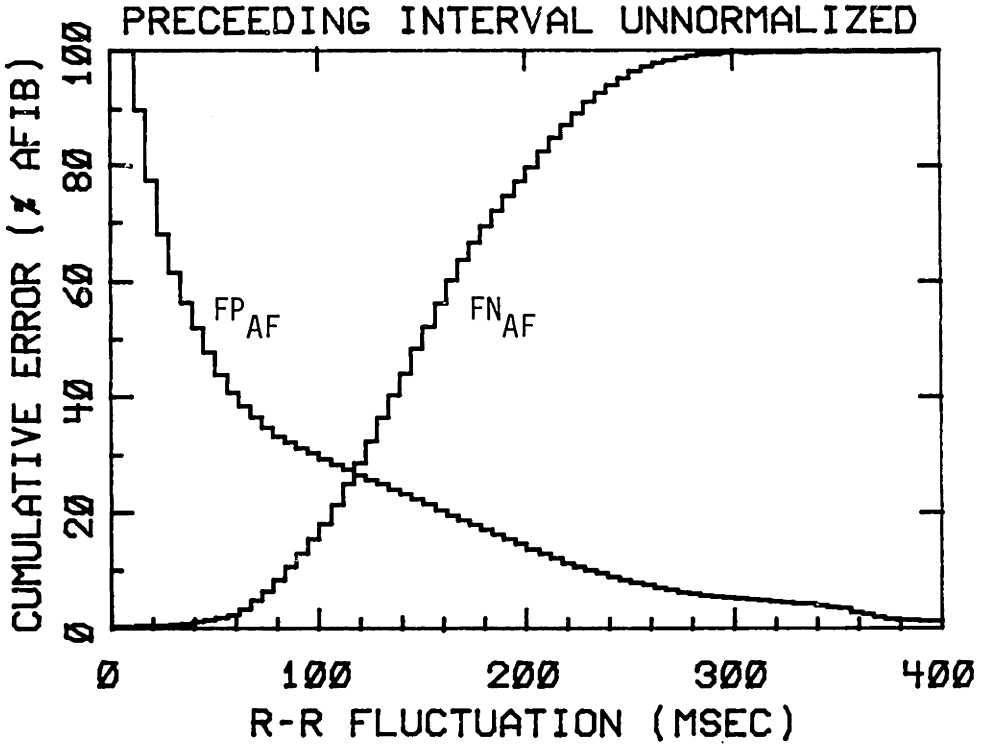




Figure 7-7d: Preceeding Interval Unnormalized AF Detector Performance Including PVCs

Summary

Timing and rhythm analysis play an important role in classifying QRS complexes. Local timing analysis is based on a predicted RR interval, and is used to update the morphologic label of the family to which the candidate complex belongs, and to classify premature supraventricular beats. The predicted RR interval is based on a weighted running average of the preceeding RR intervals, and is designed to partially exclude ventricular ectopic beats from the average. Global rhythm analysis can be used to indirectly determine the underlying atrial rhythm based on the long term statistical properties of the RR interval sequence, rather than attempting to detect P-waves. Currently, the bedside monitor detects the presence of atrial fibrillation based on the ability of previous RR intervals to consistently predict the current RR interval.

The major application of timing and rhythm analysis in this work was for the detection and classification of ectopic beats. However, there are other potential clinical applications for RR interval analysis as well. In ongoing research at our lab (Cohen,1978, Akselrod,1981), random process theory is being applied to the analysis of beat-to-beat fluctuations of electrocardiographic parameters, and how they relate to (a) autonomic nervous system regulation of cardiac activity, and (b) intracardiac conduction processes. In this work, power spectral analysis of variations in the RR interval sequence may provide a quantitative non-invasive means of assessing the level of

sympathetic and parasympathetic tone to the heart. In a related area, the RR interval sequence can be used to estimate the parameters for a physiologic model of the AV node during atrial fibrillation, and may provide a means for optimizing pharmacologic regulation of this dysrhythmia.

# 8. ARTIFACT REJECTION

The preceeding chapters have the described the processing steps that are used to acquire, detect, and classify the QRS complex. One must also, however, deal with real world signals that often contain artifact and noise. Prolonged episodes of noise can overload a processor by exceeding its ability to compress the data to a manageable level. In other instances, artifact can distort the morphology of QRS complexes, or can be falsely detected as cardiac events.

The first section of this chapter deals with the issue of processor overload, and techniques that can be used to minimize this problem. The second section of this chapter discusses several techniques that can be used to cope with artifact and noise. Several of these issues have been discussed in previous chapters, and are only briefly reviewed here.

## 8.1 Handling Processor Overload

Several conditions can cause a processor to fall behind the incoming data. The most common of these are:

(1) Large amplitude, moderate to high frequency noise that contains many "possible" QRS complexes that require careful screening by the algorithm. In our case, such artifact can rapidly fill the segment buffer (SEGBUF) and the triangular morph buffer (RWBUF).

(2) Extremely high heart rates.

(3) Additional processor tasks such as continuous and frequent requests for data displays, hardcopy plots, and strip chart recordings. In the case of the bedside monitor, the strip chart recorder and plotter operate concurrently with the analysis algorithm, but each process requires about 10 to 15 percent of the processor.

There are several mechanisms available to reduce the computational load on the processor.

Increase compression ratio

When the processor begins to fall behind the incoming data, the processor workload can be reduced by increasing the compression ratio. Although some degradation in system performance will occur with an increased compression ratio, it is not as catastrophic as actually discarding the incoming data. This is accomplished by allowing the zero order interpolator (ZOI) aperture to automatically adjust to produce a relatively constant compression ratio. In the presence of high-frequency, large amplitude artifact, the aperture is increased to reduce the number of segments produced per second. However, if the aperture exceeds a fixed threshold, the data is declared noisy, and processing is suspended.

Shed non-essential tasks

Second, non-essential tasks can be skipped. For example, our monitor normally provides an annotated real-time display of the incoming ECG, and overdraws the QRS complex with the segment representation of the data to highlight it. This display can take a large fraction of the processor's time when the heart rate and QRS amplitude are high. If the algorithm starts falling behind, the overdrawing of the delineated complex is terminated until the algorithm can catch up with the incoming data (they are still labeled, however).

Detecting and handling buffer overflow

If none of the above procedures help, the algorithm will eventually fall behind the incoming data to the point where any of the three circular buffers will overflow. In this condition, incoming data overwrites old data that is still being processed by the algorithm. To prevent this, all three buffers are periodically examined (once every 128 samples) to see if all triangular morph buffer pointer and segment buffer point to valid regions of the raw data buffer RDBUF. If the earliest unprocessed morph in the triangular morph buffer references data within 128 samples of the end of the raw data buffer, all triangular morphs in RWBUF are discarded, and the analysis is restarted at the current incoming data. The fact that the previous eight seconds of data are lost is reported to the summary trend and statistical display routines. This "failsafe" mechanism is rarely executed, however, since both the adaptive

ZOI aperture and the display shutdown are adequate to deal with most processor overload conditions.

## 8.2 Artifact and noise rejection

Signal artifact and noise can arise from many sources, and create problems that plague all real-time monitoring systems. Baseline shift due to poor electrode contact, muscle artifact, and powerline interference frequently corrupt the ECG. Artifact can distort the morphology of QRS complexes and cause them to be missclassified. In other instances, artifact can be falsely detected as QRS complexes.

Some of the techniques used by the bedside monitor to cope with artifact and noise are discussed in the following paragraphs. A number of these techniques were presented in previous chapters, and are listed here for completeness.

### Analog and digital filtering

Filtering the raw data is an important step in improving the signal to noise ratio of the ECG. The bedside monitor uses a single-pole high pass filter (with a two second time constant) to remove DC electrode offset potentials, and a non-linear diode network for rapid recovery from defibrillator pulses. A linear-phase, two pole low-pass filter with a cutoff frequency of 72 Hertz is used to attenuate high-frequency muscle artifact, and also serves as an anti-aliasing filter. A 60 Hertz notch filter is used to suppress powerline interference.

Adaptive zero order interpolator aperture

In algorithms which incorporate data compression, the compression ratio may be dynamically adjusted to limit the data output rate. Thus, during noisy periods, higher compression ratios are used. In the case of the bedside monitor, the value of the ZOI aperture is dynamically adjusted to produce a maximum of 20 to 30 segments per second. If high frequency, moderate amplitude noise is present, the aperture increases to suppress it. Thus, the value of the ZOI aperture can serve as an amplitude estimate for high frequency noise. If the aperture exceeds a fixed threshold [144 sample counts, or 20% of nominal QRS amplitude], the incoming data is declared noisy, and further processing is suspended.

Analog to digital converter saturation

Baseline artifact due to loose electrodes will often saturate the A/D converter, and can be used as an indicator of poor signal quality. The bedside monitor tests the voltage amplitude of every extremum in the segment buffer (SEGBUF) for possible A/D converter saturation. If the voltage is within 5% of the upper or lower A/D converter range, the incoming data is declared noisy. If the incoming data is declared noisy (because the ZOI aperture is too large, or the A/D is saturated), processing is suspended until the signal is "noise free" for at least three seconds. Processing then resumes one second after the last instance of saturation or excessive ZOI aperture.

Artifact rejection based on width

Some solitary artifacts may be rejected on the basis of their non-physiologic morphology. In our system, the "R-wave quality" of a triangular morph is computed based on the morph's width and amplitude. The width function is designed to emphasize morph widths that are within the normal physiologic range for QRS complexes. Events that are 30 to 120 msec wide are favored, whereas much wider events (T-waves or large baseline excursions) are weighted less. Events less than 30 msec wide (that are likely to be artifact or 60 Hz powerline interference) are rejected by setting the width weight to zero.

Artifact rejection by the event detector

The accuracy of the event detector can have a major impact on the overall performance of the analysis algorithm. Although subsequent processing stages can be designed to account for possible detector errors, they still are at the mercy of detector errors such as missed beats. One notable feature of the bedside monitor event detector is its R-wave quality ordered search in which the morphs with the highest signal to noise ratio are detected and labeled first. Thus, lower signal to noise ratio morphs can be labeled in the context of previously detected and labeled morphs. For example, if there is a prolonged interval between two detected R-waves, detection thresholds are lowered to recognize possible low amplitude beats. If the detected R-waves exhibit a regular RR interval, the detection threshold is not lowered, thus rendering the event detector more immune to noise.

Artifact rejection based on family classification

If an instance of artifact is falsely detected as an event, it can still be rejected on the basis of its morphology. There are several mechanisms provided by the morphology classifier to minimize the number of false positives due to artifact and noise. First, a family is labeled ARTIFACT if it is too narrow (compared with absolute limits, or with the width thresholds based on the QRS width histogram RWHIST). Second, the morphology classifier decision boundaries depend on the number of events in a class, and are designed to require very strict prematurity and width thresholds to label a complex/family "PVC" when it is seen for the first time. However, once a complex has been seen several times, the prematurity and width thresholds are relaxed.

The preceeding paragraphs have discussed how the bedside monitor copes with noise and artifact up to the point where the QRS complex has been classified. However, there is one more artifact detection stage before the complex is finally transmitted to the summary trend and statistical displays.

Artifact rejection based on morphology variability

The variability of QRS morphology can also be used to recognize the presence of artifact. This technique relies on the observation that it is unlikely that one instance of artifact will be similar to another instance of artifact, or other QRS complexes. If a series of falsely detected events occur during an episode of severe artifact, the average distance between the events and the nearest clusters will be large. It is unlikely

that a run of "new" ectopic beats will exhibit the same morphology variability as would an episode of artifact.

Thus, there is one last software filter that labeled QRS complexes must pass through before they are transmitted to the summary trend and statistical displays. The filter (subroutine REJECT) analyzes the morphology of groups of beats to determine if they're more likely to be artifact or real QRS complexes. If a group of consecutive beats has a high average distance to the nearest families, all the beats in the group are relabeled UNKNOWN, and are labeled with a "?" on all annotated displays.

Subroutine REJECT uses a running average of a transformed distance measure. The transformed distance is used so that very dissimilar complexes won't grossly affect the running average. The distance transform is shown in Figure 8-1a, which illustrates how the transformed distance "saturates" for large distances. The running average (over the last 1 to 3 beats) is then compared to two thresholds. If the average transformed distance exceeds the lower threshold DACCEPT, REJECT "holds off" transmitting the labeled QRS complexes in order to test subsequent complexes for artifact. If the average transformed distance falls below DACCEPT, the section of data is not considered noisy, and the original event labels are transmitted to the summary trend routines. If the average transformed distance exceeds a higher threshold DREJECT, all complexes in the region are relabeled UNKNOWN until the average transformed distance drops below DACCEPT. If no decision can be made (the average remains between

DACCEPT and DREJECT), the beats are relabeled UNKNOWN only if (in the same epoch of data) the DREJECT threshold was exceeded.

Using a pair of hysteretical thresholds improves performance by labeling the set of complexes UNKNOWN only if there was clear evidence of artifact. Also, hysterisis reduces the frequency of transitions between the two modes. The detailed operation of REJECT is described on the next page, and an example of hysteretical operation is shown in Figure 8-1b.

The value of the transformed distance thresholds DACCEPT and DREJECT, the number of beats in the average distance, and the distance transform RJDIST all affect the performance of this stage. The exact transform function was somewhat arbitrary: the major requirement was that it "saturated" for large distances so that bona-fide new morphologies would not grossly affect the running average distance. The thresholds were selected empirically, but satisfy the requirement that a single isolated beat with a different morphology surrounded by other beats that were seen before would not be labeled UNKNOWN. Thus, new PVC forms would be correctly classified the first time they were seen, provided that the surrounding data was relatively noise free.

```
Synopsis of artifact rejection algorithm REJECT.
This routine is executed once for each complex.

IF (REJCNT) < 2

THEN
     IF (((RJDIST(BSTDST) < DACCEPT) OR (Complex is NORMAL)
     THEN
          (REJDST) = RJDIST(BSTDST)
          (REJCNT) = 1
          (REJFLG) = FALSE
          All events up to and including the current complex are
          transmitted to the summary trend routines.
     ELSE
          (REJDST) = ((REJDST)+RJDIST(BSTDST))/2
          (REJCNT) = 2
          Do not transmit events --
          not sure if artifact or not.

ELSE
     (REJCNT) = MIN((REJCNT)+1,3)
     (REJDST) = [(REJDST)*((REJCNT)-1)+RJDIST(BSTDST)]/(REJCNT)

     IF ((REJDST) > DREJECT)
     THEN
          (REJFLG) = TRUE
          Relabel all pending events as UNKNOWN,
          and transmit to summary trend routines.
     ELSE

     IF ((REJDST) < DACCEPT)
     THEN
          IF ((REJFLG) .EQ. TRUE)
          THEN
               Search backwards through pending events
               until first dissimilar beat is found, and
               relabel preceeding events as UNKNOWN.
          ELSE
     (REJFLG) = FALSE
     (REJCNT) = 1
     (REJDST) = RJDIST(BSTDST)
     ELSE

     IF (DACCEPT < (REJDST) < DREJECT)
     THEN
          Not sure whether artifact or not.
          If too far behind incoming data,
          relabel all complexes UNKNOWN
          only if (REJFLG) = TRUE.
     ELSE

RETURN
```

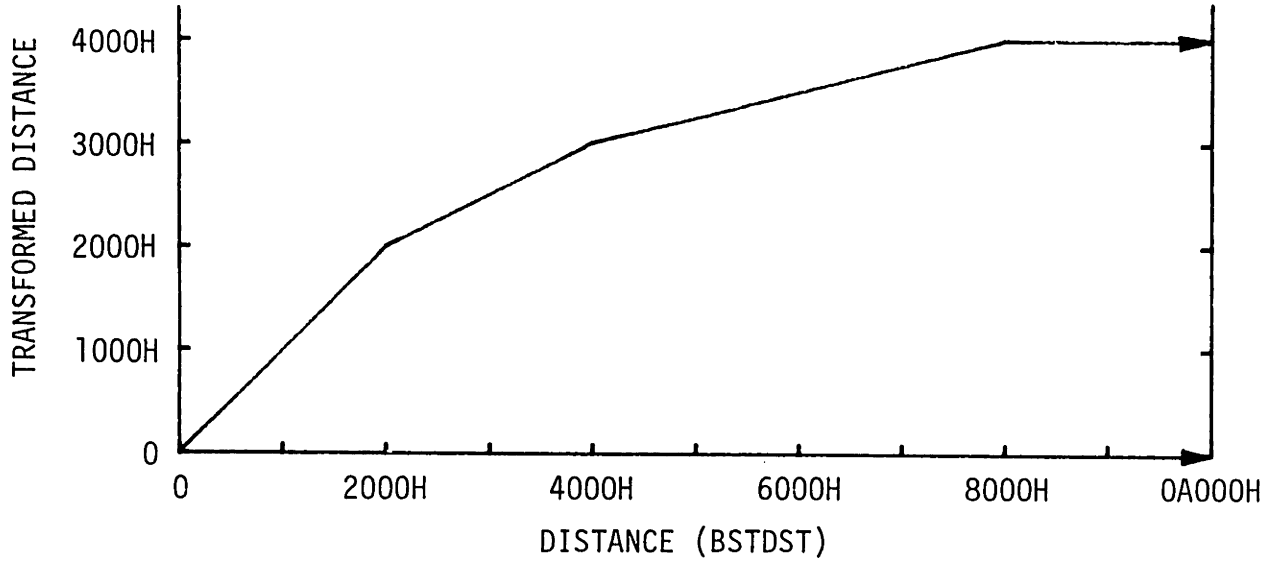



Figure 8-1a: Distance Transformation RJDIST(BSTDST)

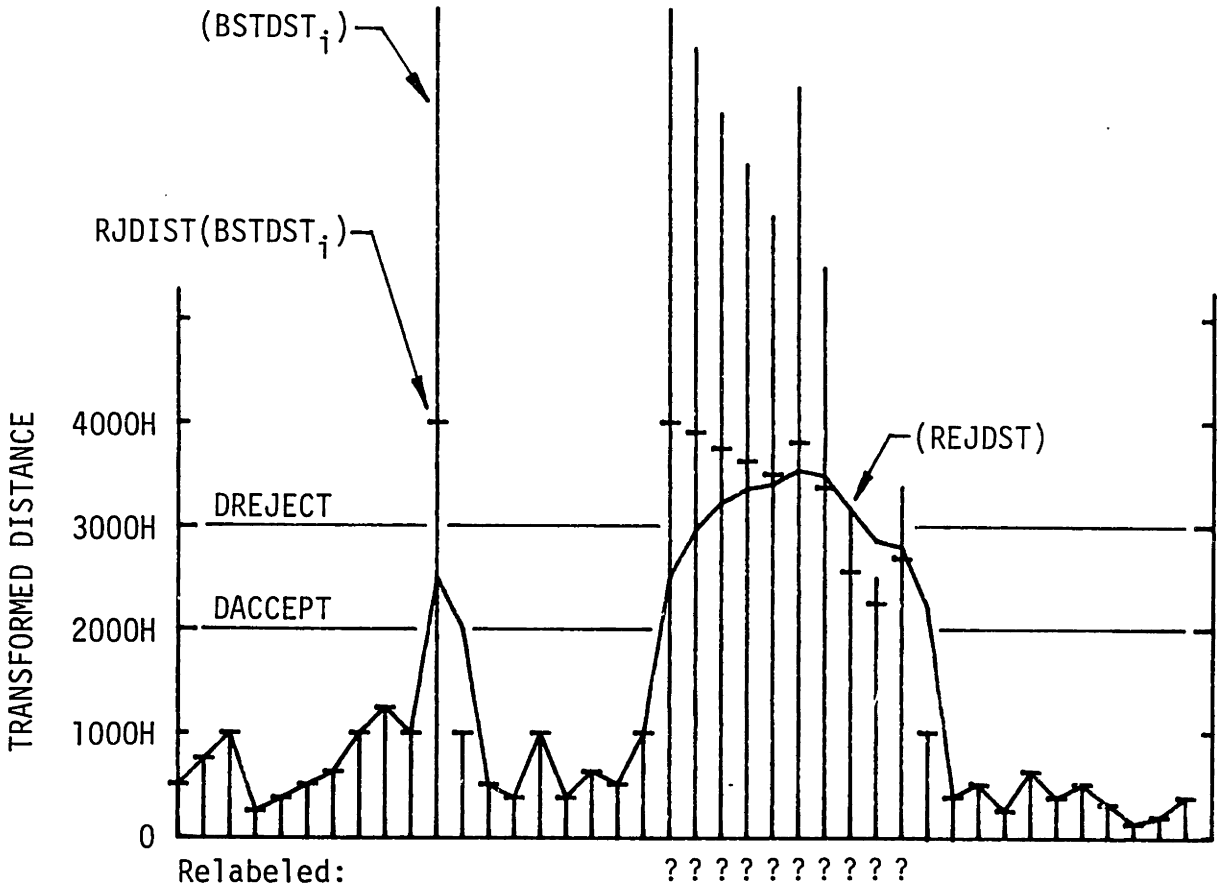



Figure 8-1b: Dynamic Operation of REJECT

One criticism of this technique is that a run of ventricular tachycardia (which can be considered a sequence of complexes with somewhat variable morphology) could be incorrectly identified as artifact, and relabeled UNKNOWN. This rarely occurs in actual practice, since once within the run, ventricular morphologies tend to be similar to other morphologies in the same run. If the run is recognized by VTCHECK (which looks for a contiguous sequence of overlapping triangular morphs), subroutine REJECT is bypassed.

Several examples of artifact and their detection are shown in Figure 8-2. Figure 8-2a depicts an episode of high-frequency noise (typically due to muscle artifact) that was detected because the ZOI aperture increased. The other two episodes are examples of baseline noise, often due to poorly applied or loose electrodes. In Figure 8-2b, the baseline noise was successfully rejected based on the morphologic variabilty of the detected "events". Figure 8-2c shows a more difficult case where an episode of baseline artifact that was falsely labeled a salvo of PVCs. The salvo was so labeled because it is comprised of overlapping triangular morphs. On closer examination, however, two normal complexes can be seen within the "salvo". This also illustrates that it is often not possible to reliably distinguish between some types of noise and ventricular tachycardia, flutter, and fibrillation. In this case, it is better to generate a false positive strip and have it reviewed by the clinician.

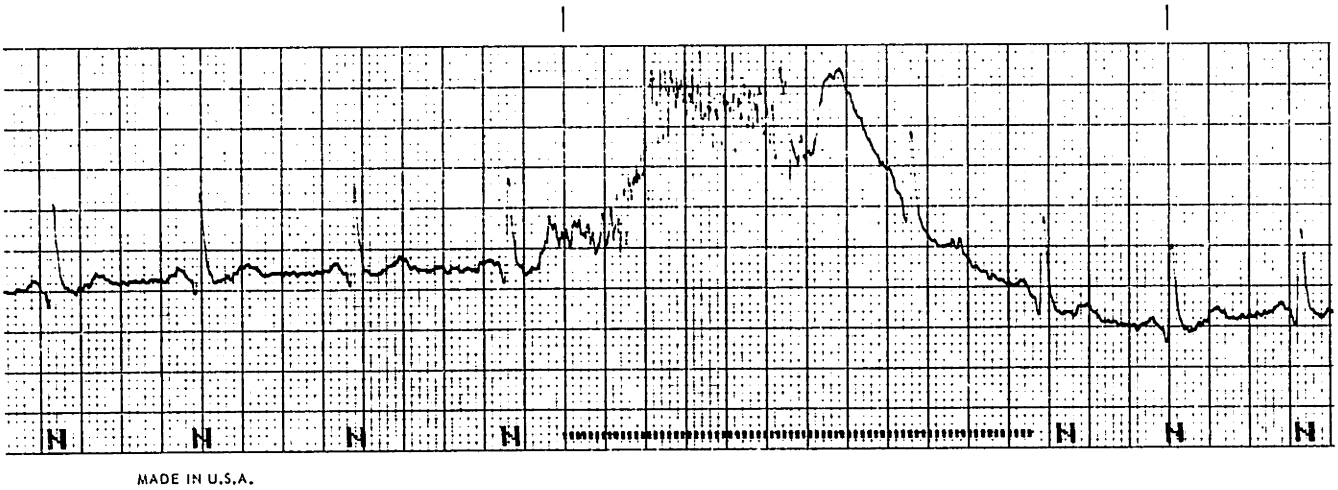



Figure 8-2a: Episode of High Frequency Artifact

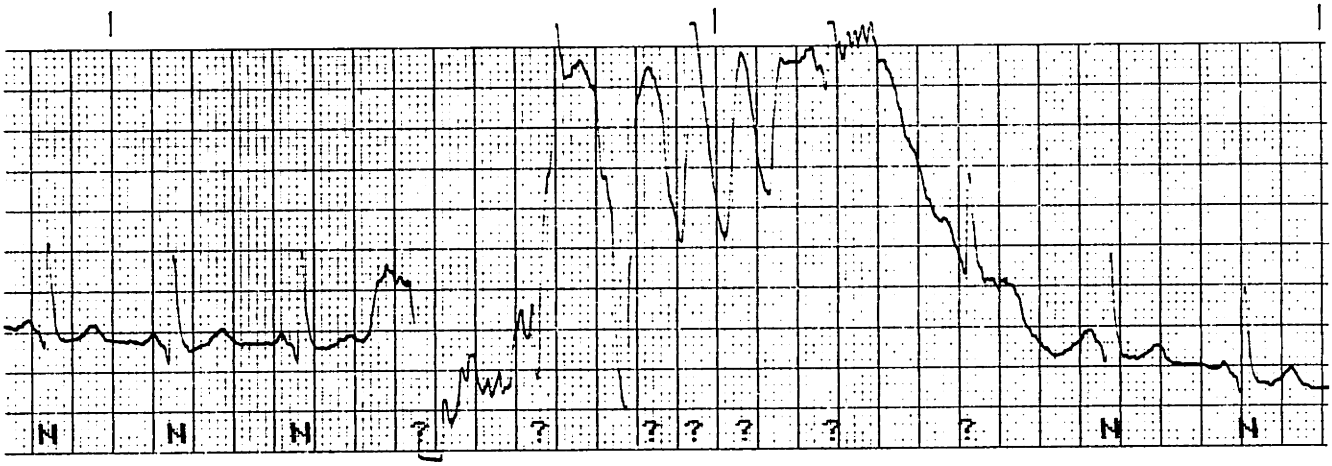



Figure 8-2b: Baseline Artifact Relabeled UNKNOWN

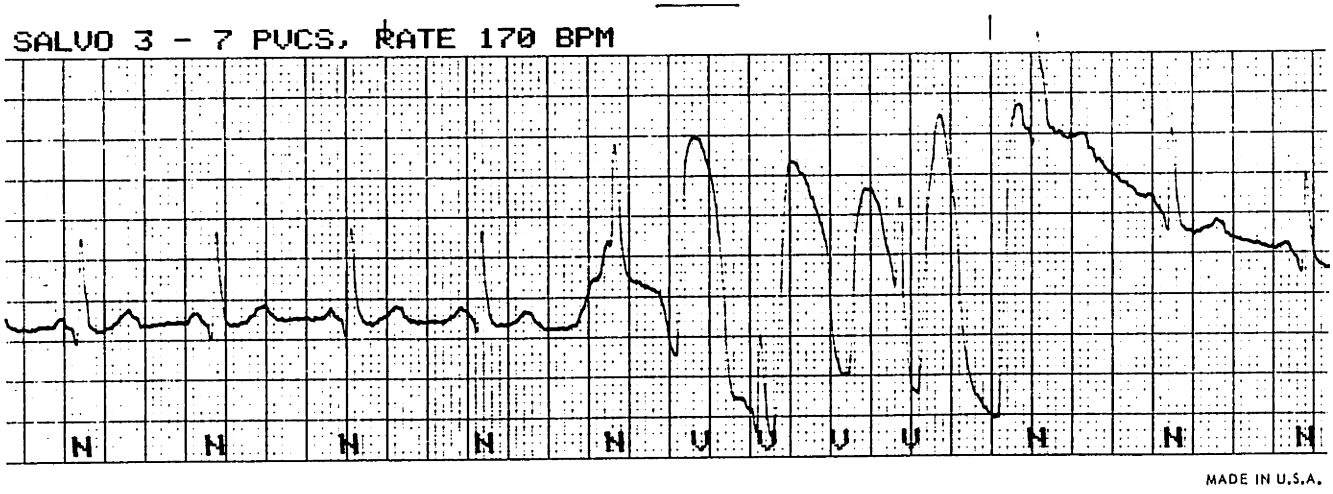



Figure 8-2c: Baseline Artifact Falsely Labeled Salvo PVCs

# 9. SUMMARY AND EPISODE REPORTS

The final sequence of QRS complexes must now be summarized and documented in a clinically usable form. The documentation strategy that is used can have a significant impact on the clinical usefulness of a monitoring system, since it has the responsibility for condensing and documenting thousands of heartbeats over long periods of time. It should also document important changes in patient status, while keeping false and redundant alarms to a minimum.

In a closely related issue, a monitoring system should be easy to use, and require a minimum of user interaction to operate. Any system that is inconvenient to use or disrupts the the activities of the nurse or doctor is unlikely to succeed in the clinical environment. This is especially important for an instrument like the bedside monitor, which will see most of its use in general surgical and medical wards, rather than in the ICU or CCU which have specially trained staff.

The design of a monitoring system's front console can significantly affect the user acceptability of the system. The bedside monitor console illustrates one approach to this important man - machine interface, as shown in Figure 9-1. Each key is labeled with its function, and provides feedback to the user by lighting when the requested function is performed. The labeled keys provide a command menu that allows the clinician to deduce its operation almost "by inspection". This was considered

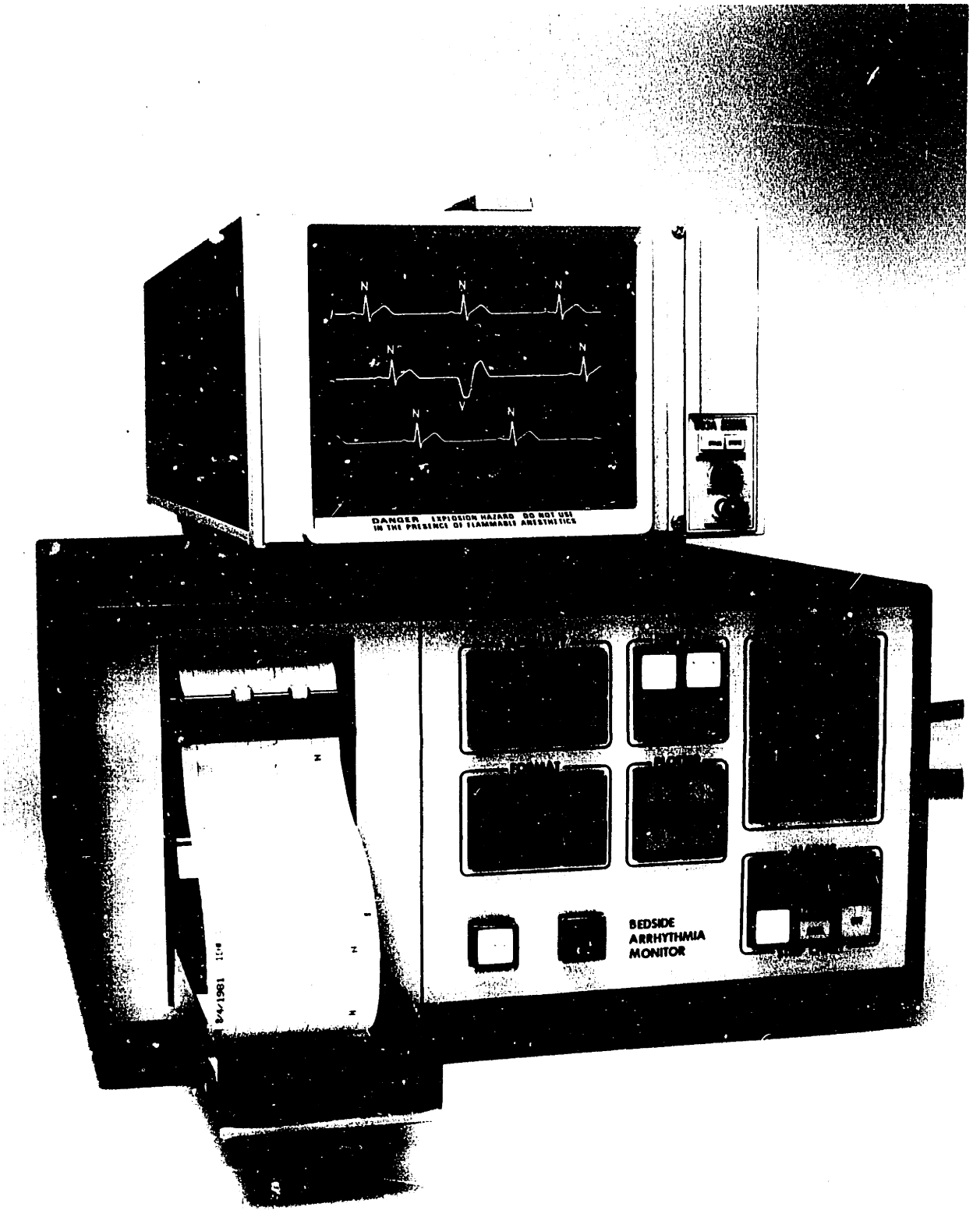



Figure 9-1: The Bedside Arrhythmia Monitor - Model II

superior to a typewriter-like keyboard that doesn't reveal the intent of the system. The bedside monitor was designed to look like an instrument, rather than a computer whose appearance might be more intimidating to the clinical staff.

Starting the bedside monitor

The bedside monitor also has a numeric keypad that is used to enter patient information, and to tailor the operation of the monitor to an individual patient. When the monitor is turned on, it requests the patient's identification number, the date, and the time (in "military" format). This is shown in Figure 9-2. After the information has been entered, the analysis of the ECG begins.

Real-time annotated ECG display

Another important aspect of human engineered design is to provide feedback to the user about how well the system is performing. It is often not possible to guarantee a certain level of accuracy from patient to patient, since an algorithm's performance can be sensitive to the morphology and rhythm of the ECG. The bedside monitor provides this feedback with a real-time annotated ECG display (Figure 9-3) where each beat is labeled with its classification:

"N" for a normal QRS complex,
"V" for a PVC,
"S" for a supraventricular premature beat,
"W" for a wide, on-time beat,
"?" for complexes likely to be distorted by noise,
"↑" for a very narrow complex likely to be artifact, and
"L" for a learning beat (only the first 50 complexes).

The delineated complex is highlighted by overdrawing it with its

straight-line approximation, and its classification is drawn near the extremum of the R-wave. Three traces are drawn, maintained for a few seconds, and then the screen is redrawn with new incoming data.

Labeling the beats provides important feedback about the monitor's accuracy, and helps establish the proper degree of confidence in the system. The clinician can get a reasonable "feel" for the monitor's accuracy by casual observation. Although such an estimate is crude, it is probably much better than no real-time feedback at all.

The bedside monitor also provides displays of trend information and special displays that illustrate the algorithm's decision making process. The system also detects and automatically documents clinically important episodes on the strip chart recorder.

```
PLEASE ENTER PATIENT ID# --
NUMBER <ENTER>
ID#? 3039

TODAY'S DATE --
MONTH <ENTER> DAY <ENTER> YEAR <ENTER>
DATE? 7/11/81

TIME --
(0-23) HRS <ENTER> MIN <ENTER>
TIME? 13:54
```

Figure 9-2: Entering Patient Information

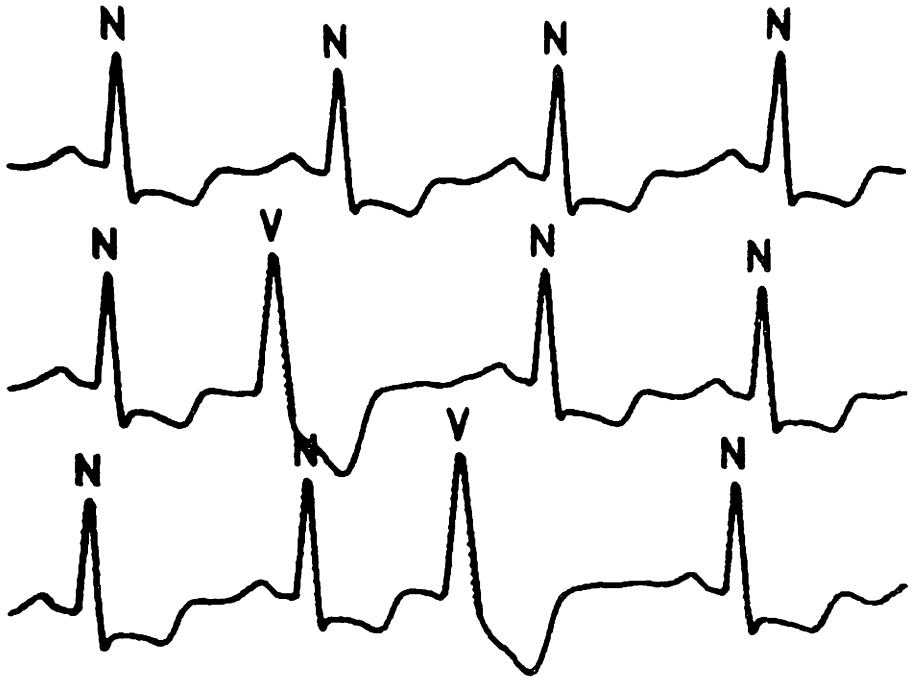



Figure 9-3: Annotated Real Time ECG Display

## 9.1 Bedside Monitor Trend Displays

The bedside monitor maintains several trends that are immediately available upon demand by the clinician. The trends cover the most recent 15 minutes, 1 hour, 3 hours, and 12 hours for the following parameters:

| | |
|---|---|
| Heart Rate | (in beats per minute) |
| SVPB Rate | (in SVPBs per minute) |
| PVC Rate | (in PVCs per mintue), and |
| PVC Run Length | (longest run of consecutive PVCs) |

The different time spans allow the clinician to select the one most appropriate for the application at hand. The 15 minute trend is best suited for observing transient changes (e.g. an exercise stress test), whereas the longer trends are more useful for documenting long term changes due to drug therapy or diurnal variations. The vertical axis scale is fixed for each parameter (rather than adjusting to the peak value of the trend) so that the "visual impact" of the trend is always the same. Also, the time axis shifts at discrete time intervals (rather than showing the greatest amount of data) so that trends for different parameters can be temporally aligned on the hardcopy plotter.

Examples of the heart rate and SVPB rate trends are shown in Figure 9-4a and 9-4b, and PVC rate and PVC run-length trends are shown in Figure 9-4c and 9-4d, respectively. These plots were made with a hardcopy plotter attached to the bedside monitor, which can record the contents of the current display in any one of the four quadrants of a sheet of paper.

The PVC Run Length trend is a novel display which shows the longest run of consecutive PVCs in each update interval. The longest run of ventricular tachycardia (VT: rate greater than 100 bpm) and the longest run of accelerated idioventricular rhythm (AIVR: rate less than 100 bpm) are noted in each update interval. The run lengths are drawn as a solid vertical line with height proportional to the length of the longest run of VT, and as a dashed line with height proportional to the length of the longest run of AIVR. If the run of VT is longer than the run of AIVR, the solid line completely obscures the shorter one for AIVR, which is appropriate considering the relative severity of the two ventricular rhythms.

The trend plot characteristics are summarized in Table 9-1. Each trend spans a "trend duration", the time axis shifts every "time axis shift", and a new data point is added every "update interval". The PVC and SVPB rate trends accumulate and average their event counts over longer periods than the heart rate since their event rates are usually lower. If more than a certain amount of artifact is present in a trend update interval, the interval is blanked on the display.

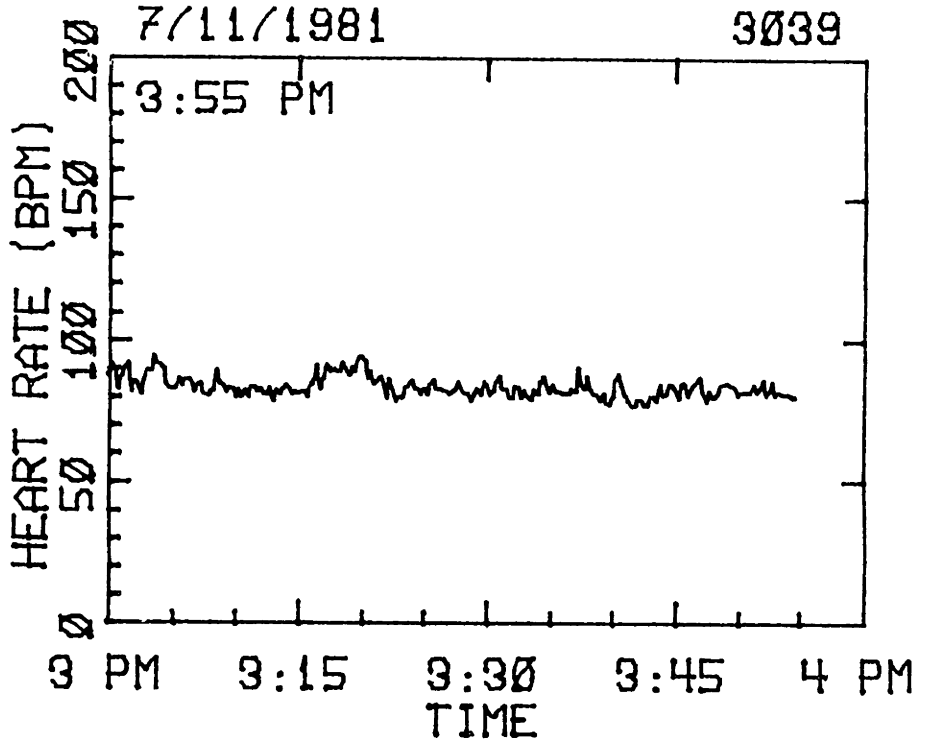



Figure 9-4a: One Hour Heart Rate Trend

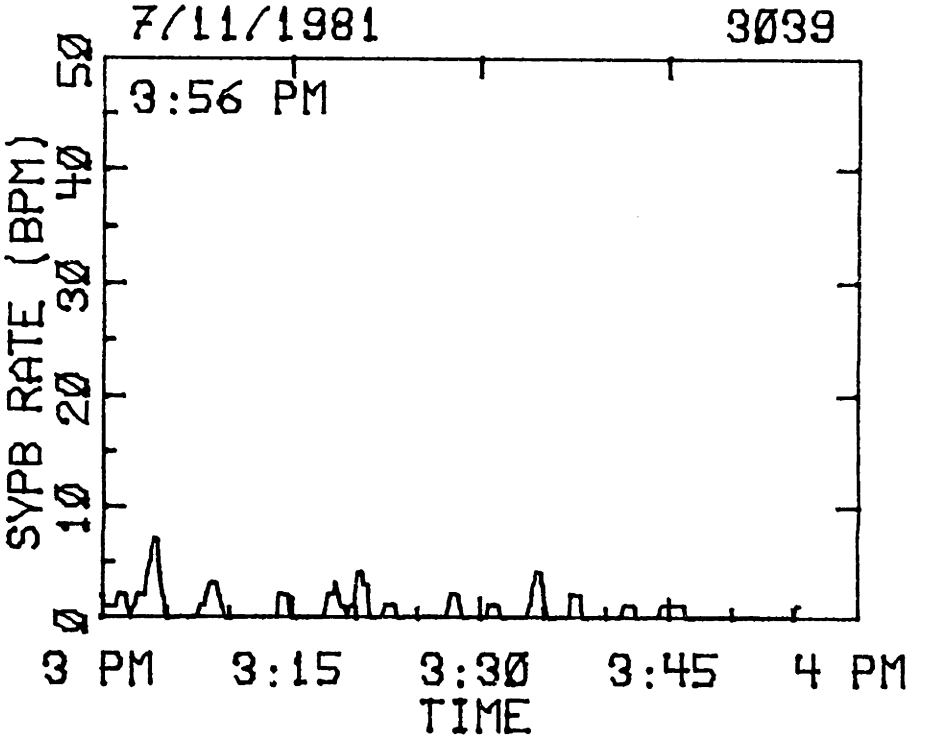



Figure 9-4b: One Hour SVPB Rate Trend

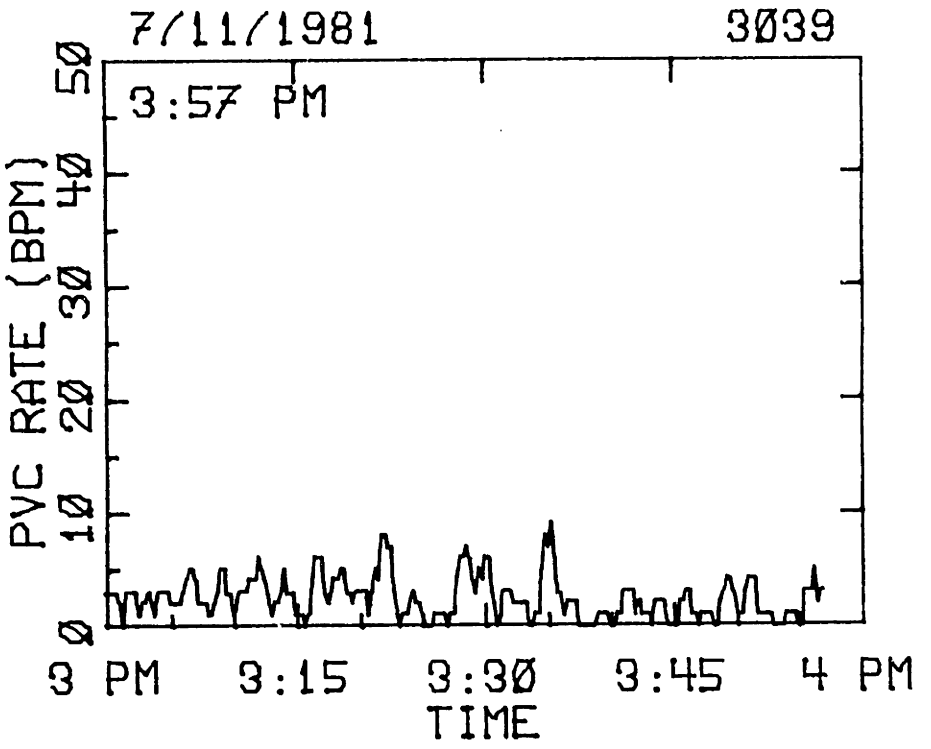



Figure 9-4c: One Hour PVC Rate Trend

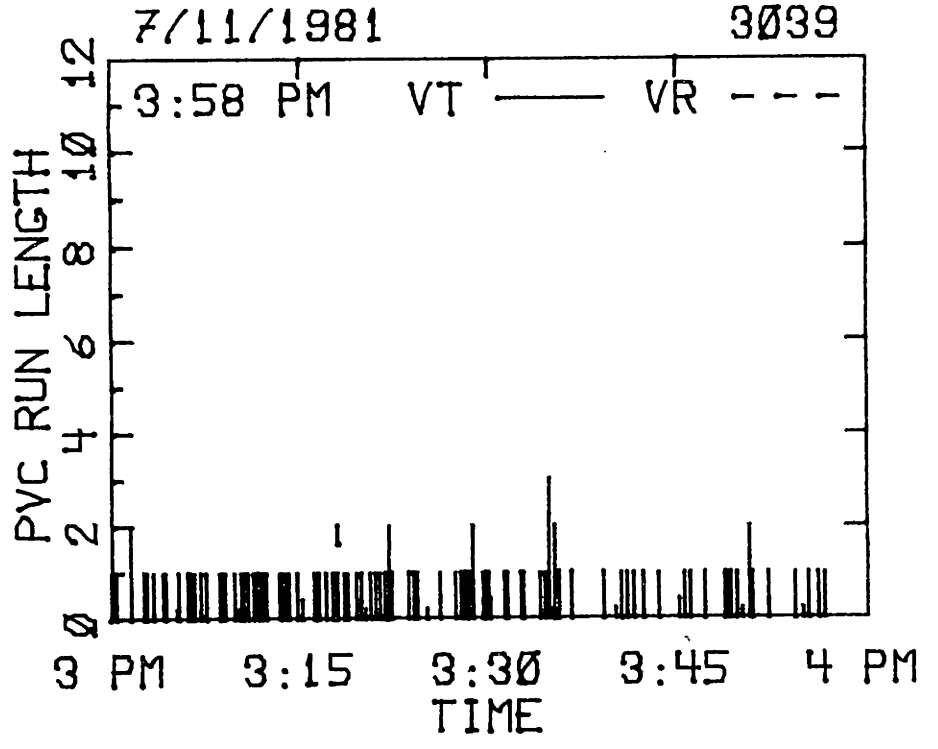



Figure 9-4d: One Hour PVC Run Length Trend

Table 9-1: Bedside Monitor Trend Plot Parameters

For all trends:

| | | | | |
|---|---|---|---|---|
| Trend Duration: | 15 min | 1 hour | 3 hours | 12 hours |
| Time Axis Shift: | 5 min | 15 min | 1 hour | 1 hour |
| Update Interval: | 5 sec | 15 sec | 1 min | 5 min |
| Displayed Points: | 180 | 240 | 180 | 144 |

HEART RATE trends:

| | | | | |
|---|---|---|---|---|
| Averaging interval: | 5 sec | 15 sec | 1 min | 5 min |
| Artifact limit: | 1/2 sec | 1/2 sec | 5 sec | 30 sec |
| Max Heart Rate: | 254 bpm | | | |

When the heart rate is computed for the 15 min and 1 hour trends, the epoch duration is aligned to the nearest QRS complex. Periods of artifact are removed from the epoch duration. The heart rate used for the alarm processor is based on the average heart rate over the last 15 seconds.

PVC and SVPB RATE trends:

| | | | | |
|---|---|---|---|---|
| Averaging interval: | 30 sec | 1 min | 2 min | 5 min |
| Maximum rate: | 254 bpm | | | |

A circular buffer of the number of PVCs in 5 second epochs is maintained, and each 5 second epoch is flagged as "noisy" if there is more than 1/2 second of artifact. The entire averaging interval is declared "noisy" if more than one-quarter of the 5 second epochs is "noisy". The PVC and SVPB rate used for the alarm processor is based on the average rate over the last 1 minute.

PVC RUN LENGTH trend:

| | | | | |
|---|---|---|---|---|
| Epoch Interval: | 5 sec | 15 sec | 1 min | 5 min |
| Maximum Run Length: | 15 beats | | | |

Ventricular runs with rate less than 100 bpm are AIVR. Ventricular runs with rate greater than 100 bpm are VT. The longest run for both AIVR and VT terminating in the epoch are saved. There can be no artifact within a run.

## 9.2 Statistical Displays

The bedside monitor also provides several statistical summary displays. These displays have limited clinical usefulness, and are more useful as diagnostic aids for the algorithm.

The RR interval histogram (Figure 9-5a) is the histogram of all RR intervals since the beginning of the monitoring period. This histogram can be used to compare the relative populations of normal and premature beats. However, gradual rate changes can broaden the modes of the histogram, making it impossible to identify modes due to ectopic beats. The PVC coupling interval histogram (Figure 9-5b) is the histogram of the coupling (RR interval) for PVCs that follow a normal complex. The modes of the PVC Coupling histogram can sometimes be used to estimate the relative population of different ectopic foci. Again, this display can be misleading, since multiple-mode histograms can also arise from artifact or delineator error.

A number of other displays provide a detailed look at the algorithm's decision making process. Running average histograms (each bin is reduced by 7/8 every 30 beats) for QRS width (Figure 9-6a) and prematurity-ratio (Figure 9-6b) show the current distribution for these two parameters. The width range for NORMAL QRS complexes and the "on-time" mode of the prematurity-ratio histogram are underlined to indicate their delineation by the algorithm.

The thresholds derived from the two histograms are used to set the family classifier decision boundaries, which are depicted in Figure 9-7a (the boundaries are drawn for a class population of one). When a QRS is detected and delineated, a "+" is drawn to indicate its width and prematurity ratio. Although the actual QRS classification is based primarily on morphology, this real-time display is often very informative. Another display (Figure 9-7b) shows the distance between individual QRS complexes (indicated with a "+") and the nearest NORMAL and PVC classes. This particular figure indicates that there is a PVC class that is similar to the normal QRS complex, which is indeed the case for the ECG waveform shown in Figure 9-3. The principal use for this display is to determine whether classifier error is due to very similar NORMAL and PVC morphologies.

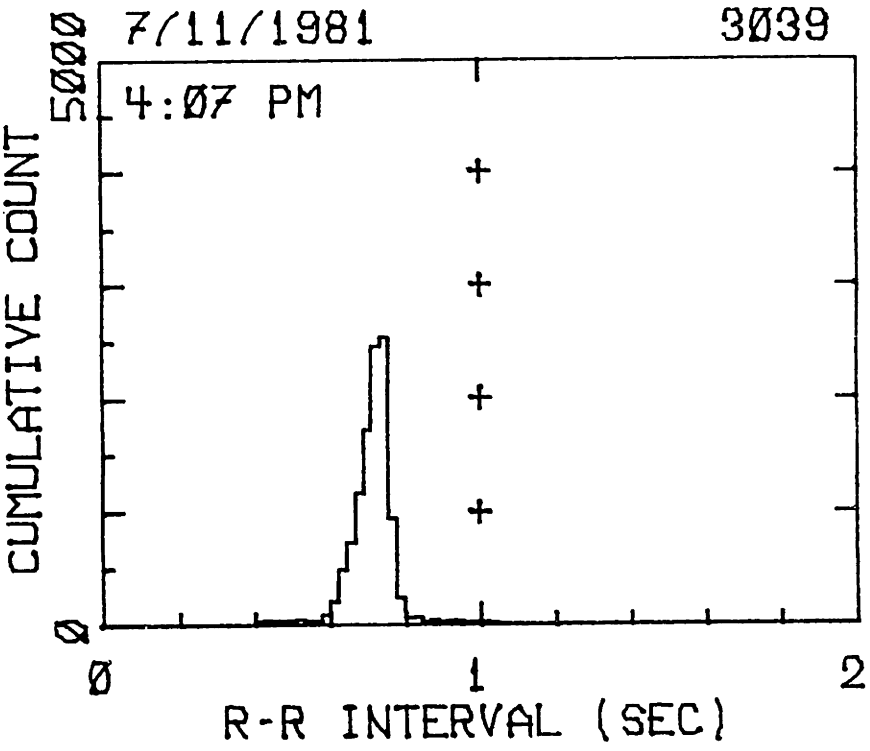



Figure 9-5a: RR Interval Histogram

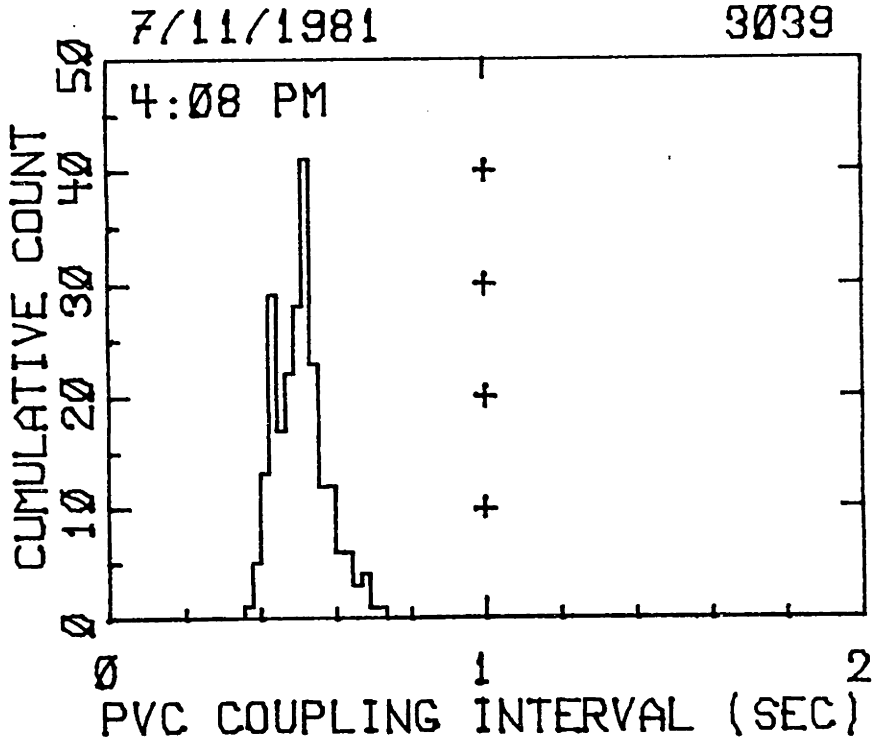
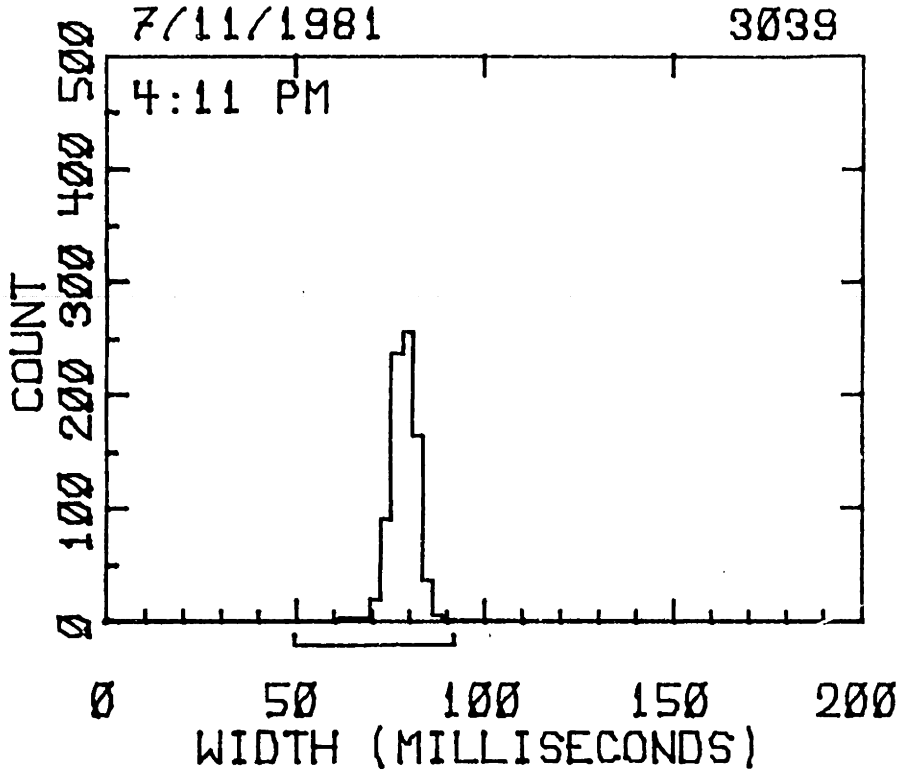



Figure 9-5b: PVC Coupling Interval Histogram



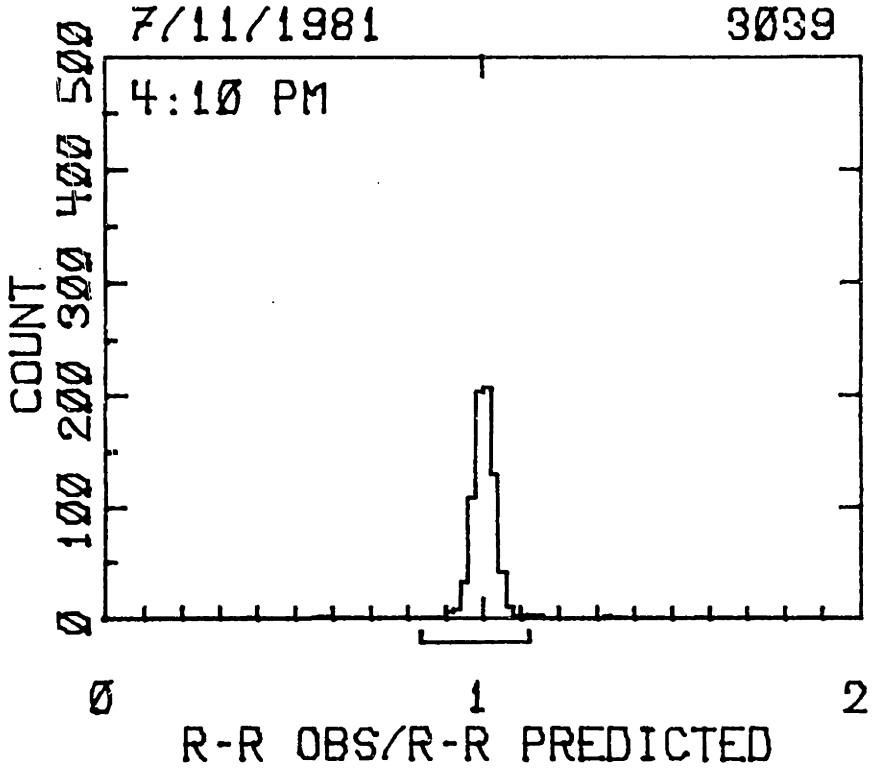




Figure 9-6a: Running Average QRS Width Histogram

Figure 9-6b: Running Average Prematurity Ratio Histogram

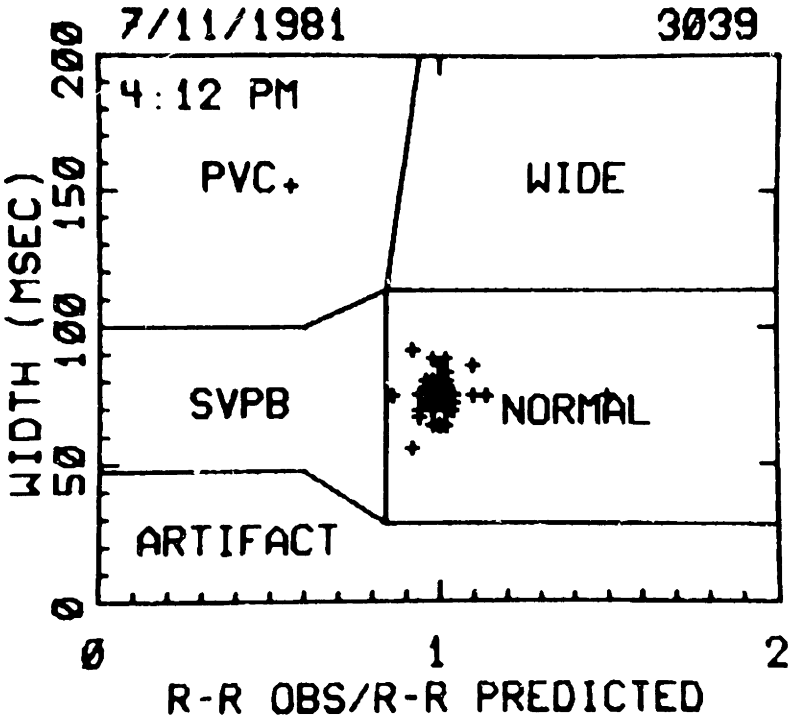



Figure 9-7a: Classifier Decision Regions

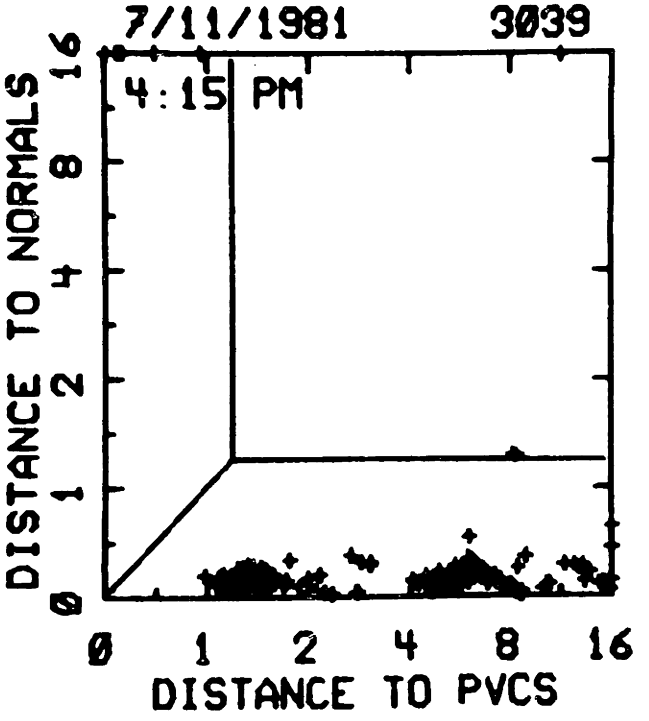



Figure 9-7b: Candidate Distance to NORMALs and PVCs

## 9.3 Detection of Alarm Conditions

One of the major functions of any automated arrhythmia monitoring system is to detect and document clinically significant rhythms. This aspect of monitoring intrudes the most on the clinical routine, and thus requires careful consideration.

The major emphasis of this discussion will be on episode detection and documentation strategies used by the bedside monitor, with less attention devoted to the issue of alarms. During clinical trials, the bedside monitor was used primarily as a real-time replacement for a Holter recorder, and was always used with a conventional monitor with rate alarms. An alarm capability is currently being added to the monitor, but was not installed at the time this thesis was written.

The episode detection and documentation strategy can be viewed as two separate processing stages. The first is concerned with determining the patient's overall status, and detecting clinically significant trends and episodes. Here the system can determine that the patient's heart rate is in "normal" range, that the patient has "frequent PVCs", and has had several episodes of "ventricular bigeminy". The task of the second stage is to decide which of these conditions is important enough to bring to the attention of the clinician. It should note and report significant changes in patient status while minimizing redundant alarms.

Several of the concepts embodied in the bedside monitor episode documentation procedure are based on the alarm processor for the Stanford Arrhythmia Monitoring System (Sanders,1975). Their approach to alarm processing has proven to be quite successful, and has been incorporated in several commerical multipatient monitors.

Determination of patient status

The episode detection stage recognizes a variety of rate and rhythm conditions, which are summarized in Table 9-2. Three of the categories ("heart rate", "PVC rate", and "ventricular rhythm") consist of mutually exclusive "states" that the patient can be in. For example, the patient's ventricular rhythm state can be only "no PVCs", "isolated PVC", "trigeminy", etc. Other conditions, such as "SVTA" or "asystole" are reported as isolated events. The patient's state is continuously reported to the episode and alarm processor, which makes the final decision to report the condition to the clinician. The interface between the two parts is done with only two routines:

ALSET(i), which indicates the onset of the ith alarm, and

ALCLR(i), which indicates the termination of the ith alarm.

Heart Rate status

Heart rate, and especially changes in heart rate, is a fundamental monitoring parameter. However, the onset of a heart rate change is often not well defined. Heart rates can vary considerably from patient to patient, and it is difficult to establish rate standards that will work well with all patients.

This could be circumvented by allowing the user to adjust the upper and lower heart rate limits, but this would make the operation of a monitoring system less automatic.

Another way to detect heart rate changes is to use the patient's current heart rate as a standard, and report a status change whenever the rate changes by a certain amount. This technique, however, is unsatisfactory due to the difficulty in defining what the "current" heart rate is. If the current heart rate is based on the average over the last few minutes of monitoring, then gradual changes will not be detected. If the "current" heart rate is based on the rate at the beginning of the monitoring session, then the system will be unable to track gradual changes in heart rate.

The bedside monitor uses seven overlapping heart rate ranges to describe the patient's heart rate status, and are listed in Table 9-2. For example, if a patient's heart rate increases from 100 to 130 bpm, it is noted as a change from "normal heart rate" (60 to 120 bpm) to "mild tachycardia" (110 to 150 bpm). The heart rate ranges overlap to provide hysteresis so that heart rates near a rate threshold will not cause the patient's status to alternate between two adjacent ranges. Currently the heart rate limits are fixed, but they do provide sufficient discrimination to follow moderate heart rate changes. This technique does not have the disadvantages mentioned earlier: it can automatically adapt to the patient's nominal heart rate, and detect significant changes, no matter how gradual their onset.

The heart rate status is based on the one hour trend rate which is updated every 15 seconds. Transient rate changes such as a pause or asystole are detected as separate conditions. For each transition from one rate range to another (e.g. code n to m), ALCLR(m) is called, followed by a call to ALSET(n).

PVC Rate status

A similar procedure is used to determine changes in PVC rate. Frequent PVCs are divided into three categories, with (ascending) thresholds at 6, 12, and 24 PVCs per minute. Although these limits are arbitrary, they represent meaningful distinctions to cardiologists. Subsequent thresholds are twice the previous, and represent a change that may indicate the need for more agressive treatment. The PVC rate is updated every 15 seconds, and the rate is the average rate over the preceeding 60 seconds.

Ventricular Rhythm

Ventricular rhythms ranging from isolated PVCs and trigeminy to long runs of ventricular tachycardia are also recognized (see Table 9-2). Ventricular rhythms are recognized using a software shift register in which a "1" is stored for a PVC, and a "0" for a beat with any other label. A rate test is also used to discriminate between "ventricular tachycardia" with rate greater than 100 bpm and "accelerated idioventricular rhythm", with rate less than 100 bpm.

Supraventricular Rhythm

Runs of supraventricular tachycardia (SVTA) are declared when four or more consecutive SVPBs are detected, and have an average rate over 100 bpm.

Pauses and Asystole

Other clinically important episodes deal with the absence of heart beats. A "missed beat", for example, is a prolonged RR interval (relative to the predicted RR interval). "Asystole" is declared if there is more than four seconds of artifact-free ECG. If monitoring has been inhibited for more than ten seconds by excessive artifact, an alarm condition is declared since the patient's status would have been unknown for too long.

Conditions not detected

The bedside monitor does not explicity check for R-on-T (early cycle) PVCs, nor does it document multiform ectopic activity as a separate alarm category. Although the monitor is capable of recognizing changes in morphology, it is often difficult to tell when a new morphology is due to artifact or axis shift. By documenting enough examples of PVCs per hour, one should get a reasonable sampling of different morphologies. Another approach (not currently implemented) would be to output examples that the algorithm recognized as multiform, but without the label "multiform", and let the clinician make his own determination.

Table 9-2: Table of Alarm Conditions
Detected by the Bedside Monitor

| | | |
|---|---|---|
| | **Heart Rate** | |
| HRL3 | Severe Bradycardia | 0 to 45 bpm |
| HRL2 | Moderate Bradycardia | 40 to 55 bpm |
| HRL1 | Mild Bradycardia | 50 to 65 bpm |
| HRNM | Normal Heart Rate | 60 to 120 bpm |
| HRH1 | Mild Tachycardia | 110 to 150 bpm |
| HRH2 | Moderate Tachycardia | 140 to 180 bpm |
| HRH3 | Severe Tachycardia | 170 to 255 bpm |
| | **PVC Rate** | |
| VR00 | Low PVC rate | 0 to 6 PVCs/min |
| VR06 | Moderate PVC rate | 4 to 12 PVCs/min |
| VR12 | High PVC rate | 10 to 24 PVCs/min |
| VR24 | Very high PVC rate | 21 to 255 PVCs/min |
| | **Ventricular Rhythm** | **Minimum Beat Sequence** |
| PVC0 | No PVCs | |
| PVC1 | Isolated PVC | x V x |
| PVCT | Trigeminy | x V x x V x x V x |
| PVCB | Bigeminy | x V x V x V x V x |
| PVC2 | Couplet | x V V x |
| AVR3 | Slow 3 - 7 beat VT, rate < 100 | x V V V x |
| RVT3 | Salvo 3 - 7 beat VT, rate > 100 | x V V V x |
| AVR8 | Slow 8+ beat VT, rate < 100 | x V V V V V V V V V - |
| RVT8 | Salvo 8+ beat VT, rate > 100 | x V V V V V V V V V - |
| | **Supraventricular Rhythm** | |
| SVTA | Run of SVTA,<br>> 4 beats, rate > 100 bpm | x S S S S - |
| | **Pauses and Asystole** | |
| MISS | Missed Beat | |
| | if previous beat was NORMAL: | RRI > 1.5*(PRRINT) |
| | if previous beat not NORMAL: | RRI > 2.0*(PRRINT) |
| ASYS | Asystole | RRI > 4 seconds |
| NSIG | Poor Signal | Artifact > 10 seconds |

## 9.4 Alarm and Episode Documentation Strategy

After the status of the patient has been determined, it is now necessary to decide whether a clinically significant change has occurred. If so, it should be brought to the attention of the clinician in a manner that indicates the severity of the episode. For example, a minor rate decrease need only be documented in a trend, whereas an episode of asystole should be documented on a strip chart, and reported to the user with an audible alarm. Providing the same visual and audible cue for alarms of different severity would inevitably diminish the user's reaction to situations that require immediate attention. The alarm processor should also suppress repetitive minor conditions, and report only clinically significant changes in patient status.

Many of these principles are embodied in the bedside monitor's episode and alarm processor. The major characteristics of the alarm processor are:

(1) Alarms are ranked by severity, so that only the most severe patient condition is brought to the attention of the clinician.

(2) Alarm conditions are associated in groups of the same type so that repetitive alarms for the same condition are suppressed.

(3) The operation of the alarm/episode documentation system is automatic, but can be easily tailored by the clinician.

(4) Alarm episodes are automatically documented and brought to the attention of the clinician according to their relative severity.

Alarms are ranked by severity

The alarm conditions detected by the bedside monitor (listed in Table 9-2 in the previous section) are listed in Table 9-3 according to their relative severity. The ranking is somewhat arbitrary, but in general the more life threatening conditions are listed further down the list. The ranking provides a basis for selecting the severest condition if two or more alarm conditions exist at the same time. For example, if a "couplet" and "moderate PVC rate" occur at the same time, the couplet will be documented on the strip chart. If an episode of "asystole" is detected, all other alarms are suppressed, a strip chart recording is made, and its presence is made apparent to the clinician.

Similar alarms are grouped together

In addition to being ranked by severity, alarms are also grouped to the type of alarm. The primary goal of grouping similar alarms is to minimize the number of redundant alarms for conditions that are more or less severe manifestations of the same phenomena. There are five alarm groups:

Type "h": High heart rate group.
Type "l": Low heart rate group.
Type "v": High PVC rate group.
Type "V": PVC sequence group.
Type "Q": Signal quality group.

The alarm groupings are indicated in Table 9-3. A particular alarm code such as "8-bt VT" can belong to several groups: high heart rate (h), high PVC rate (v), and PVC sequence (V).

Table 9-3: Prioritized List of Alarm and Episode Conditions

| Code | | Type | #/hr | strip? | alarm? |
|---|---|---|---|---|---|
| HRNM | rate 60-120 | ---1h | 0 | - | - |
| VR00 | 0-6 PVCs/min | --v-- | 0 | - | - |
| PVC0 | no PVCs | -V--- | 0 | - | - |
| HRH1 | rate 110-150 | ----h | 3 | - | - |
| HRL1 | rate 50-65 | ---1- | 3 | - | - |
| PSIG | artifact | Q---- | 2 | - | - |
| HRH2 | 140-180 bpm | ----h | 3 | - | - |
| HRL2 | 40-55 bpm | ---1- | 3 | - | - |
| VR06 | 4-12 PVCs/min | --v-- | 4 | - | G |
| VR12 | 10-24 PVCs/min | --v-- | 4 | - | G |
| VR24 | 21- PVCs/min | --v-- | 5 | - | Y |
| PVC1 | Isolated PVC | -V--- | 4 | S | G |
| PVCT | Trigeminy | -V--- | 4 | S | G |
| PVCB | Bigeminy | -V--- | 4 | S | Y |
| MISS | Missed Beat | ---1- | 4 | S | G |
| SVTA | Run SVTA | ----h | 4 | S | G |
| FORM | Multiform PVC | -V--- | 5 | - | - |
| RONT | R-on-T PVC | -V--- | 5 | - | - |
| PVC2 | Couplet | -V--- | 5 | S | Y |
| AVR3 | 3-7 bt slow VT | -V--- | ALL | S | Y |
| RVT3 | 3-7 bt VT | -Vv-- | ALL | S | Y |
| AVR8 | 8- bt slow VT | -V--- | ALL | L | Y |
| RVT8 | 8- bt VT | -Vv-h | ALL | L | R |
| HRH3 | Rate >170 | ----h | 10 | S | R |
| HRL3 | Rate <45 | ---1- | 10 | S | R |
| NSIG | No signal | Q--1- | 10 | L | Y |
| ASYS | Asystole | ---1- | 10 | L | R |

Alarm priority is determined by order on this list.
Lowest priority alarm is listed first.

Short strip: 8 second strip.
Long strip: Strip for duration of episode, but not longer than 30 seconds.

Green alarm: "Green Alarm" light blinks until acknowledged, or condition abates.
Yellow alarm: "Yellow Alarm" light blinks until acknowledged, or 15 minutes after condition abates.
Red alarm: "Red Alarm" light blinks until acknowledged, or 30 minutes after condition abates. Chimes continuously until acknowledged, or after condition abates.

h High heart-rate group
l Low heart-rate group
v High PVC-rate group
V PVC sequence group
Q Poor signal quality group

If a specific alarm condition occurs too frequently, all related alarms with a lower priority are inhibited. For example, if several couplets have been documented, lesser PVC sequence alarms (such as "isolated PVC") are inhibited. If a patient has a range of ventricular ectopic activity including isolated PVCs, couplets, and runs, then the bedside monitor will document most of the runs, fewer examples of couplets, and even fewer examples of isolated PVCs. The principal benefits of this approach are: (1) it saves strip chart paper, and reduces the amount of strip chart documentation that has to be reviewed by the clinician, and (2) it minimizes the number of multiple redundant alarms for the same condition, and thus reduces the intrusion of the monitor on the clinical routine. Only the most severe manifestation of each class of alarms is brought to the attention of the clinician.

Adjustable alarm rates

The number of strips that are documented per hour can be tailored to the requirements of an individual patient. The nominal number of strips produced per hour are listed in Table 9-3. For example, the bedside monitor will document up to 4 PVCs per hour, 5 couplets per hour, but will document all occurrences of ventricular tachycardia. The number of strips produced per hour for each alarm can be changed by the clinician from 0 strips (no strips) to 60 strips per hour. A number greater than 60 will cause all episodes of a given alarm to be documented and reported by the bedside monitor. Figure 9-8 illustrates the procedure for modifying the alarm rate table.

As long as the number of occurrences for each alarm is less than the specified rate, every episode will be documented and reported. If the rate exceeds the specified limit for the particular alarm, all related lower-priority alarms will be inhibited for a "refractory interval" that is the reciprocal of the alarm rate. After the refractory time elapses, new strips can be made for the alarm.

The alarm occurrence rate is computed as a running average. For every instance of a particular alarm condition, its alarm rate count is incremented by one. For every minute of time, the alarm rate is decreased by 1/64 of its current value. Thus, over an hour's time, the alarm rate reflects the average rate for a given alarm code. This was considered more desirable than using a fixed refractory period for each alarm because it would allow the bedside monitor to document a transient burst of one type of alarm until more than the specified alarm rate of that type of alarm occurred. For example, if five couplets occurred within one minute's time, all of them would be documented provided none had occurred earlier. Ectopic activity often does not occur at a uniform rate, so allowing the transient rate to exceed the specified alarm rate allows the monitor to capture transient bursts of ectopic activity.

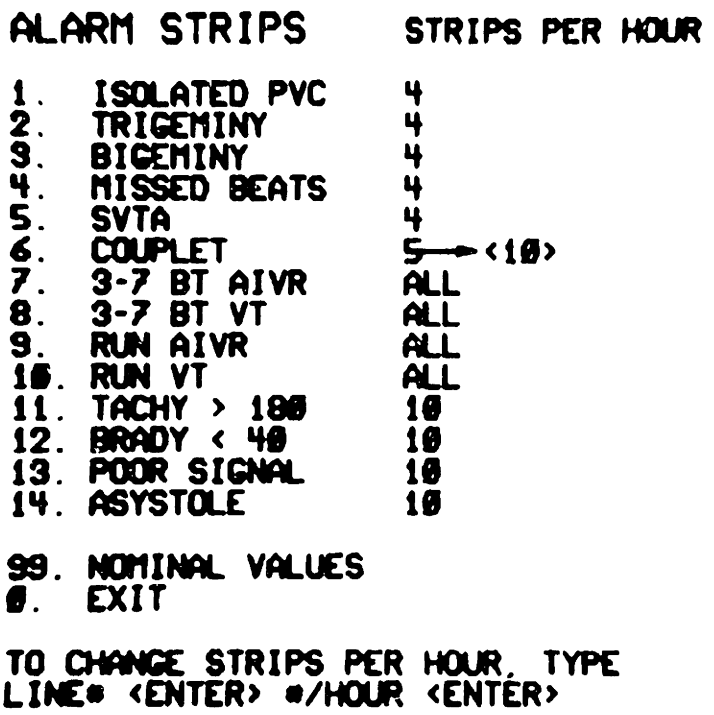



Figure 9-8: Modifying the Episode Documentation Rate

The rate at which alarm episodes are documented by the strip chart recorder can be adjusted by the user by specifying (by line number) the alarm rate to be modified, and then entering the desired rate. A rate of zero inhibits all strips for the specified alarm; a rate greater than 60 (per hour) will cause all episodes to be recorded.

What is done when an alarm occurs

If the current alarm

(1) has the highest priority,

(2) occurs at least 8 seconds after a higher priority alarm, and

(3) has a refractory timeout of zero,

then a strip chart recording is made, and a visual and audible indication of the alarm is generated.

The type of action taken by the bedside monitor for each alarm is listed in Table 9-3. For lesser alarms such as an isolated PVC, a short (8 second) strip chart recording is made. For more severe alarms, a continuous strip chart recording is made until the alarm condition stops, is acknowledged by the clinician, or if more than 30 seconds has elapsed since the onset of the alarm.

In addition to making a strip chart recording, the bedside monitor can summon aid with one of three (red, yellow, or green) flashing pushbuttons, and a chime. Each category can be individually enabled and disabled by the clinician. The three categories are based on the severity of the alarm:

RED Red alarms are issued for life-threatening arrhythmias that require immediate attention. This category includes asystole, prolonged runs of ventricular tachycardia, or extremely high or low heart rate. For a red alarm:

a strip chart recording is made,

a red light flashes on the monitor until it is acknowledged by the clinician, or 30 minutes after the alarm condition abates, and

a chime is generated until acknowledged by the clinician, or the condition abates.

YELLOW Yellow alarms are reserved for less severe arrhythmias that do not require immediate attention. This category includes rhythms such as short runs of VT and couplets. For a yellow alarm:

a strip chart recording is made, and

a yellow light flashes on the monitor until it is acknowledged by the clinician, or 15 minutes after the condition abates.

GREEN Green alarms are relatively minor such as an isolated PVC, or high PVC rate. For a green alarm:

a strip chart recording is made, and

a green light flashes on the monitor until it is acknowledged by the clinician, or the condition abates.

Synopsis of strip chart and alarm processor

The operation of the strip chart and alarm processor is summarized in Table 9-3, and is stated algorithmically on the following page. The onset and termination of each alarm condition is reported by calling ALSET(i) at the onset of the condition, and ALCLR(i) when the condition abates. When ALSET(i) is called, a series of tests are performed to see whether the alarm condition should be brought to the attention of the clinician. If so, the strip chart recording, audible tone, and flashing light is activated according to the rightmost columns of Table 9-3.

If the production rate for alarm(i) is too high, that alarm, and all lower priority alarms are inhibited by setting the corresponding refractory timers in ALREF(n<i) to the reciprocal of the desired alarm production rate ALRATE(n<i). Once a minute, ALTIMER decrements non-zero refractory timers in ALREF, and diminishes the ALRATE entries for inactive alarms by a fixed ratio so that they reflect the average alarm rate.

Episode and Alarm Processor Parameters

Constants:

| | |
|---|---|
| NALARM | Number of alarms |

Constant arrays:

| | |
|---|---|
| ALTYPE(NALARM) | alarm type bits [Q,V,v,l,h] |
| ALUSER(NALARM) | desired # strips per hour for alarm(i) (can be set by the user) |

Variable arrays:

| | |
|---|---|
| ALRATE(NALARM) | alarm rate (per hour) for alarm(i) |
| ALREF(NALARM) | refractory timeout array |
| ALACT(NALARM) | alarm ACTIVE or INACTIVE |

Description of Alarm and Episode Documentation Processor

SUBROUTINE ALSET for alarm code (i)

```
     ALRATE(i) = ALRATE(i) + 1
     IF (ALACT(i) .EQ. TRUE)      RETURN
     ALACT(i) = TRUE
     IF (ALREF(i) .NE. 0)         RETURN
     IF (alarm(i) does not have highest priority
         of all active alarms)    RETURN
     IF (a higher priority alarm occurred
         in the last 8 seconds)   RETURN
     IF (ALUSER(i) .EQ. 0)        RETURN
; If alarm(i) rate is too high, set refractory timeout
; for alarm(i) and related lower priority alarms.
     IF (ALRATE(i) .GT. ALUSER(i))
     THEN
          DO for n = 1 to i
             IF ((ALTYPE(i).AND.(ALTYPE(n)) .NE. 0)
             THEN
                  ALREF(i) = 60/ALUSER(i)
             ELSE
          LOOP
     ELSE
     EXECUTE alarm-onset-routine(i)
     RETURN
```

SUBROUTINE ALCLR for alarm code (i)

```
     IF (ALACT(i) .EQ. TRUE))
     THEN
          ALACT(i) = FALSE
          EXECUTE alarm-off-routine(i)
     ELSE
     RETURN
```

ALTIMER is invoked once a minute

```
     DO for n = 1, NALARM
        IF (((ALACT(i).EQ.FALSE) .AND. (ALREF(i).NE.0))
        THEN
             ALREF(i) = ALREF(i) - 1
        ELSE
     LOOP
     DO for n = 1, NALARM
        ALRATE(i) = ALRATE(i) - ALRATE(i)/64
     LOOP
     RETURN
```

Strip chart recorder output

The primary output of the alarm processor are annotated strip chart recordings of the ECG waveform. Each strip includes the patient identification, date, and time, and the reason for the strip. Each beat is labeled with the algorithm's classification, and thus provides vital feedback to the user about how the algorithm made its decision. Several alarm strips are shown in Figures 9-9 a, b, and c, which were produced during the hour trended in Figure 9-4.

An hourly summary report is also provided by the bedside monitor. The hourly summary includes the:

minimum, average, and maximum heart rate,

minimum, average, and maximum PVC rate,

number of isolated PVCs,

number of couplets, and

number of ventricular runs and longest run length.

The hourly summary strips provide hardcopy backup in case the algorithm crashes for some reason, or the power to the monitor is interrupted.

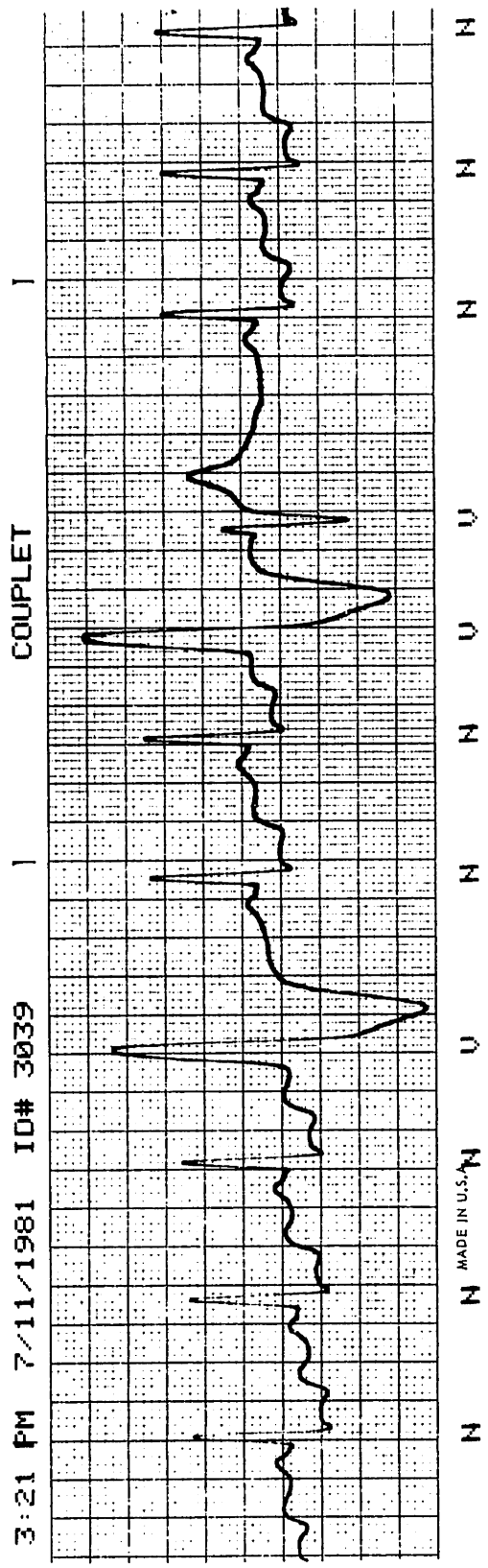



Figure 9-9a: Couplet

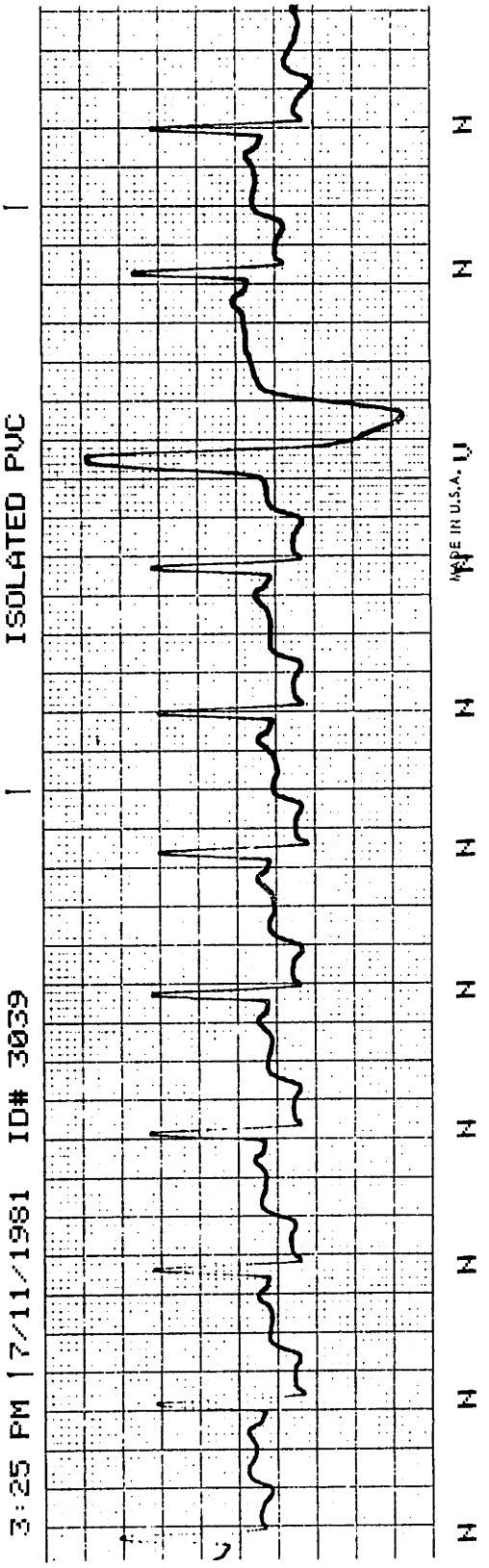



Figure 9-9b: Isolated PVC

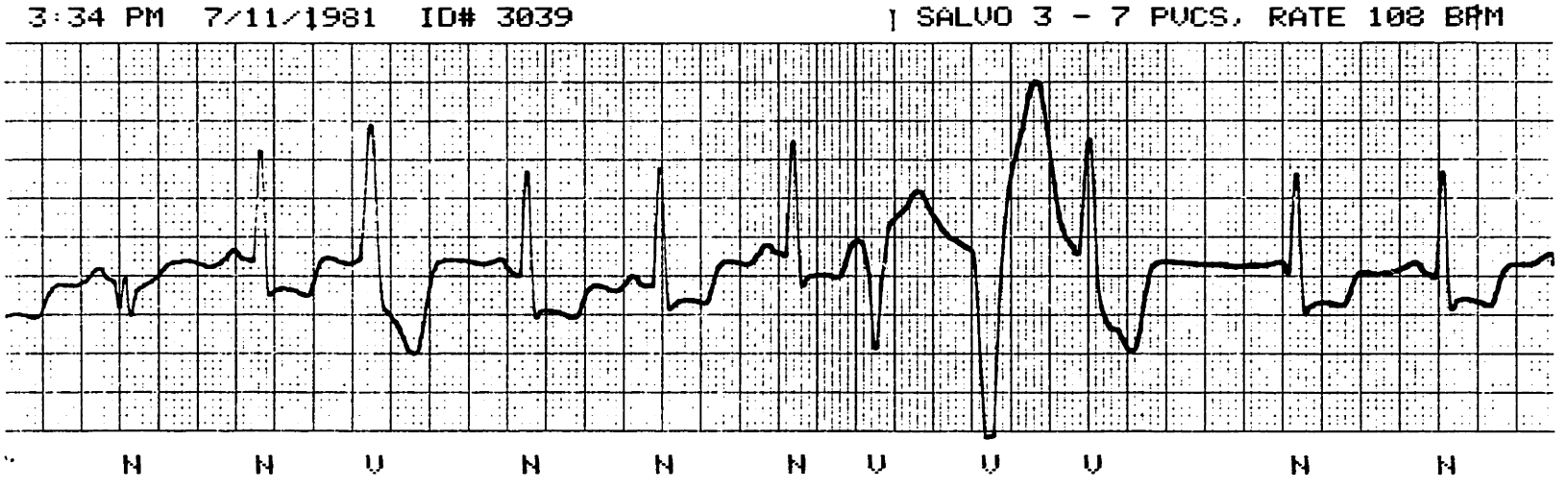



Figure 9-9c: Salvo of 3 PVCs

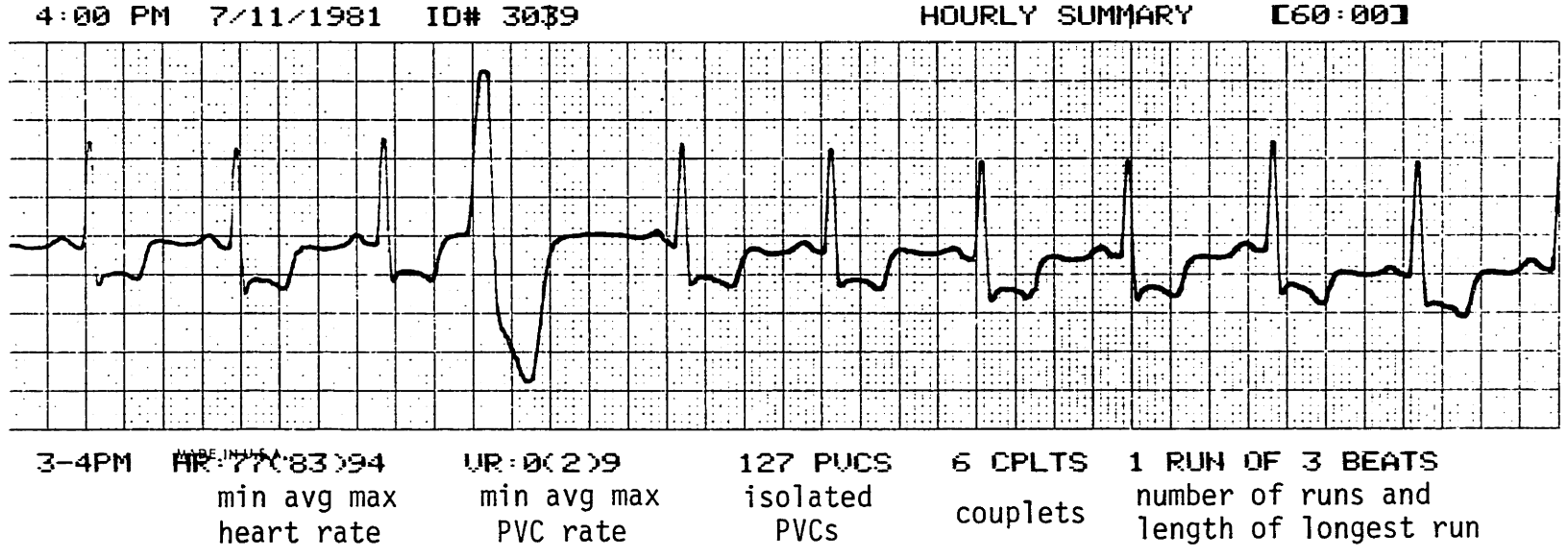



Figure 9-9d: Hourly Summary Strip

## 9.5 Bedside Monitor Log Reports

In addition to providing reports directly to the clinician, the bedside monitor can also produce machine readable reports. The present form of the report is quite simple, and was used primarily to document the output of the bedside monitor, and to see how the clinical staff made use of the system.

The monitor's log is an ASCII file that contains the following information:

(1) patient ID number, and starting date and time,

(2) all user requests for displays and plots

(3) all strip chart titles (abbreviated), and

(4) hourly summary data.

An example of a log report is shown in Figure 9-10. The log file format allows it to be directly "compiled" by STOIC, a stack oriented language developed at the Biomedical Engineering Center for Clinical Instrumentation at MIT (Sachs,1978). Since the log is an ASCII text file, "corrections" can be easily incorporated with a text editor. The corrected files were used to tally the number and types of errors that the bedside monitor made during long term Holter tape tests and clinical evaluation.

The log file can be transmitted at any time to a larger computer or portable cassette recorder. A special "log buffer status" display (Figure 9-11) is provided by the monitor and allows the user to send the entire log, or only the data since the previous transmission.

Figure 9-10: Example of Bedside Monitor Log Report

```
"3039" ID#                          patient ID
"BAM VERSION 7/10/1981" SFTID       software version
13:54 7/11/1981                     time and date

13:54 USER                          user requested strip
  /
  /                                 [13:54 to 15:00 removed]
  /
15:01 CPLT                          couplet
15:06 MISS                          missed beat
15:08 PVC                           isolated PVC
15:17 CPLT
15:21 CPLT
15:25 PVC
15:28 CPLT
15:34 3 BT 108 BPM VT               run of VT
15:35 CPLT
15:49 PVC
15:50 CPLT
HR-1H                               1 hour heart rate trend
PLOT                                  and hardcopy plot
SR-1H                               1 hour SVPB rate trend
PLOT
VR-1H                               1 hour PVC rate trend
PLOT
VS-1H                               1 hour PVC run length trend
PLOT
RAWDSP                              incoming ECG display
FREEZE                              and "freeze" it (for photograph)

16:00 HREPORT  3600 DUR             hourly summary
77 83 94 HR                         min avg max heart rate
0 2 9 VR                            min avg max PVC rate
5016 BEATS                          number of beats
127 PVCS                            number of isolated PVCs
6 CPLTS                             number of couplets
1 VRUNS 3 VMXBT                     1 3-beat run VT

16:04 MISS                          "missed" beat
RR-HST                              RR interval histogram
PLOT
16:07 PVC
VC-HST                              PVC coupling histogram
PLOT
PR-HST                              Prematurity-ratio histogram
PLOT
RW-HST                              R-width histogram
PLOT
CLS-DSP                             Classifier display
DST-DSP                             NORMAL-PVC distance display
ALMENU                              alarm rate modification
```

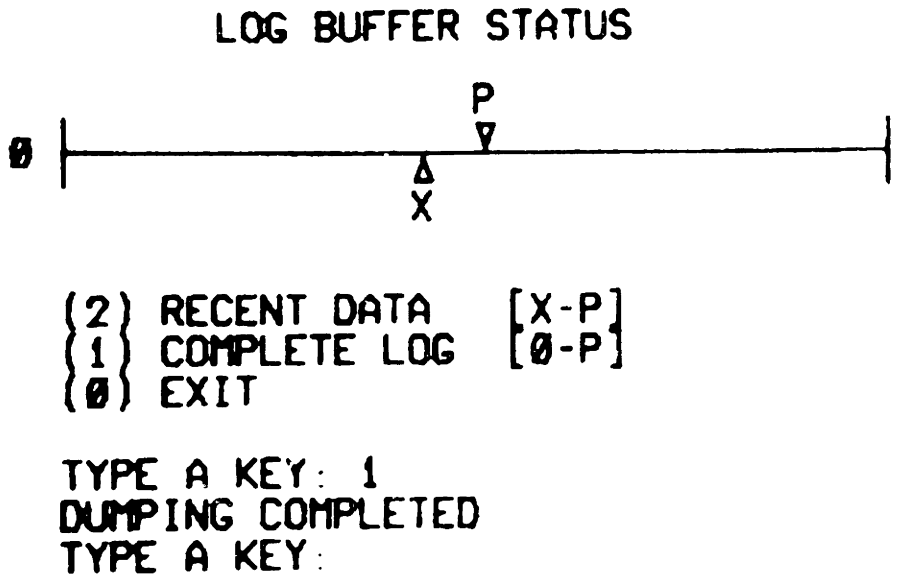



Figure 9-11: Log Buffer Status

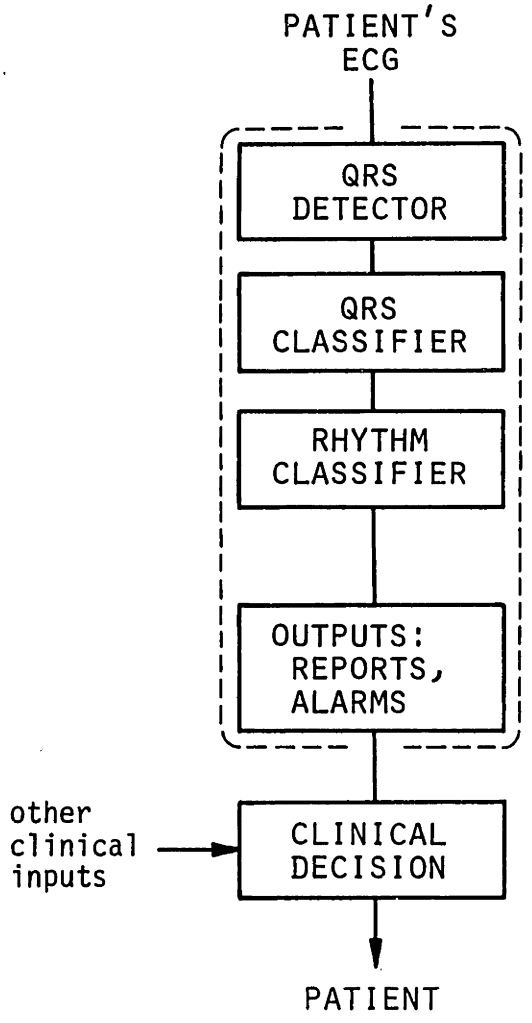
The status of the log buffer is available as a display. The "P" indicates how full it is, and the "X" indicates the point of last transmission. A "Ø" at the left of the bar display indicates that the beginning of the log is still available; if not, the buffer has "wrapped around". The entire contents [Ø-P] or the contents since the last transmission [X-P] can be sent over the serial communications port.

# 10. DATA BASE DEVELOPMENT AND EVALUATION

Evaluation of arrhythmia detectors is a difficult task that concerns both system designers and prospective users, yet lacks a generally accepted methodology. Evaluation questions can vary from the microscopic ("How well does the QRS detector work?") to the macroscopic ("What is the chance that the device will mislead the physician?"). Answering such varied questions requires a multifaceted evaluation as illustrated in Figure 10-1. A "beat-by-beat" evaluation is the only method that adequately tests the first stages of the system: the QRS detector, classifier, and overall rhythm classifier. The use of long-term analog (Holter) tape recordings which have been carefully scanned can also play an important role in evaluating the accuracy of the detection algorithm and its alarm and documentation strategies. Finally, the real-world interface between the system and the clinician must be studied to assess the importance of the device in clinical decision making and the confidence of the users in its performance. This level of evaluation is often more subjective and qualitative than data base trials, but is of critical importance.

## Annotated ECG Data Base Evaluation

The principal quantitative evaluation of the arrhythmia detection algorithm was performed using an annotated digital ECG data base. An annotated data base for beat-by-beat evaluation offers several advantages for algorithm evaluation. It permits




EVALUATION QUESTION(S) and
*-Evaluation Technique(s)*

QRS detection accuracy?
*-Beat-by-beat analysis*

PVC and SVPB detection accuracy?
*-Beat-by-beat analysis*

Accuracy of ectopic run detection
Accuracy of rhythm classification
(Pauses, atrial fibrillation, etc.)
*-Beat-by-beat analysis*
*-Long term analog recordings*

Effectiveness of alarm and
documentation strategy?
*-Beat-by-beat analysis*
*-Comparison with other techniques*
*-Observe user interaction*

Importance of instrument in
clinical decision making?
Level of user confidence?
*-Document real-world demand*
*-Assess impact of withdrawal of new technology*
*-Interviews with users*

Figure 10-1: Evaluation Questions and Techniques

automatic measurement of detector sensitivity and specificity for QRS complexes, PVC's, SVPB's, couplets, runs, etc. Reproducable experiments are possible which document the effects of algorithm changes during the development process; and the availability of a standard data base for evaluations (as opposed to development) would permit meaningful comparison of different algorithms.

Long Term Analog Tape Evaluation

On the other hand, annotated digital data bases are limited by their relatively small size. It is not at all clear, for example, that 50 to 100 half-hour ECG excerpts are an adequate representation of the "universe", nor that system performance measured with a data base will accurately predict its performance in the clinical environment. Thus, the second phase of the evaluation uses 8 hour segments of long-term Holter recordings as test data. The reports produced by the bedside monitor are compared hour-by-hour with the reports prepared by Holter tape scanning technicians. Although this procedure is not as precise as a beat-by-beat evaluation, it expands the scope of the evaluation in several areas:

1) it exposes the algorithm to far more data than is available with the annotated data base,
2) it compares the system with a clinically accepted technique (Holter scanning), and
3) it exercises the entire arrhythmia detection system - the detector, classifier, and episode documentation strategy.

Clinical Evaluation

Assessing the clinical usefulness and cost effectiveness of technology in the hosptal is very subjective and and qualitative. Some questions that can be asked include:

1) Does the bedside monitor increase the speed of patient drug trials because of its immediate results? Does the instrument reduce the number of days that a patient has to remain in the hospital?

2) Does the monitor eliminate the need for Holter monitoring of inpatients? Is the instrument cost effective when compared against inpatient Holter recording or other arrhythmia monitoring techniques?

3) Does the monitor increase the safety of drug trials because of the immediate availability of trend and statistical summaries?

4) Do doctors and nurses prefer it over Holter recording?

5) Does it work reliabily enough in a clinical setting?

A "formal" clinical trial was not attempted as part of this study. Although such a trial would have been quite valuable, it was considered beyond the scope of this thesis. It was felt that another measure of the monitor's clinical value would be its acceptance by doctors and nurses. Its acceptability could be measured by noting the number of requests for the monitor, and by conducting interviews with the personnel who use it. In an attempt to make this measurement more quantitative, a data logging feature was installed that recorded all user interaction with the front console and also maintained a log of all ECG sample strip episodes created during the monitoring period. The log was useful for determining the production rate of false positive strips, and the frequency of user requests for the monitor's trend and statistical displays.

## 10.1 ECG Data Base: Development

This section describes the design and development of the annotated digital ECG data base (Schluter,1978, MIT-BIH,1980). The creation of this resource was a major effort, and was funded, in part, by government, industry, and university resources. The author would like to acknowledge the many hours of work that others have contributed to its development and dissemination:

Scott Peterson, who supervised the detailed selection, digitization, annotation, and editing at Beth Israel Hospital. He also contributed substantially to the development of software needed for using the data base for algorithm evaluation.

George Moody, who converted the data base to the AHA data base format, and who contributed in a major way to the tape directory (MIT-BIH,1980).

Ted Baker, who wrote the latest version of the annotation comparator which was used to obtain the data base results shown in this thesis.

Larry Siegel, who wrote the displaying waveform editor.

Cheryl Jackson, who was responsible for most of the detailed transcription of cardiologist annotations, the comparison and quality control functions, and who assembled the final manuscript of the tape directory.

Diane Perry, who, as chief technician in the Arrhythmia Laboratory, helped to identify suitable data, and helped in the annotation process.

The physicians who helped with the task of beat-by-beat annotation of the ECGs, and finally,

Dr. Roger Mark, final arbiter of all disputes, and his colleague, Dr. Walt Olson.

Data Selection

The data base currently consists of 48 half-hour 2-channel ECG excerpts containing a wide variety of ventricular and supraventricular arrhythmias, conduction abnormalities, and artifact. The excerpts were selected from a library of over 4000 24-hour Holter tape recordings obtained by the Arrhythmia Laboratory of the Beth Israel Hospital in Boston, Massachusetts, from both inpatients (60%) and outpatients (40%). Recordings were made using Avionics model 445 2-channel electrocardiocorders, and typically employed modified lead II and modified lead V1 electrode configurations.

Twenty-three of the excerpts (the "100 series") were selected randomly to serve as a representative cross section of ECG waveforms and artifact that an arrhythmia detector might encounter. The particular Holter tape and half-hour segment were chosen using a table of random numbers, and the segment was incorporated into the data base if at least one of the two ECG channels was interpretable.

The other 25 of the data base tapes (the "200 series") were selected to adequately represent more rarely occurring but clinically important events. Episodes of ventricular tachycardia, ventricular flutter, and other ventricular rhythms were included in this set. Several of the tapes were selected on the basis that they would be very difficult to analyze due to background rhythm, morphology variation, and artifact.

ECG data aquisition

The two-channel analog tape epochs were digitized at a sampling rate of 360 Hz with 11-bit resolution over a 5 mV range. The samples were stored as 8-bit signed differences, permitting one-half hour of two channel data to be recorded on a single 3M (DC300A) 1/4 inch data cartridge. The overall bandwidth was 0.1 Hz to 100 Hz at -3 dB down, which was considered adequate for testing arrhythmia analysis algorithms. A more significant limitation, however, was the skew between the two analog channels of the Holter recorder The skew was as large as 40 milliseconds for some recorders, and may cause some difficulties for two channel analysis algorithms.

The digitized tape was subsequently annotated by a simple QRS detector algorithm which arbitrarily labeled each beat as "normal". This preliminary label was used only as an aid in the cardiologist annotation protocol.

Annotation protocol

Each tape was then edited, beat-by-beat, by two independent experts. The editing was done using 30 minute, two channel strip chart recordings of the data, where each beat recognized by the simple QRS detector had been labeled "N". All beats that were falsely detected, missed, or not NORMAL were relabeled by the human annotators. The underlying rhythm and episodes of artifact were also noted on the strip chart recording.

The annotated strip chart recordings were then incorporated onto the digitized tape using a "waveform editor". The editor supports a refreshed display of 8 seconds of two channel ECG data, and provides a joystick controlled cursor and keyboard for labeling. The new annotations were saved on tape and simultaneously displayed, allowing abnormal beats to be correctly identified, and noise and rhythm episodes to be noted. The final output of an editing session was a tape containing the new information, as well the original ECG data and earlier annotations of the old tape. The annotation format allows up to eight annotation channels to be saved on the tape.

Reconciliation of beat labels

After the tape was edited by two independent observers, the annotations were compared by a program that noted all disagreements, and a third strip chart recording was made to document them. Often the discrepancies are the result of editor entry error, or mental "eye blink" in which an annotator failed to note an event. On other occassions there was disagreement about debatable definitions regarding fusion or aberrated beats. In any case, the discrepancies were resolved by a single expert (usually Dr. Roger Mark) in consultation with others. The corrections were then added to the tape using the editor, producing a final tape containing the sampled 2-channel ECG and consensus annotations.

Auditing the data base

The consensus annotation channel was then checked by an "auditor" program that verified the logical relationship between QRS complex labels and rhythm labels. For example, all complexes within a run of ventricular tachycardia were examined to see if they had a "PVC" label. Omitted beats were detected by listing the longest RR intervals and their time of occurrence. After the consensus annotation channel was audited and corrected, it became the final "truth" annotation channel for the tape.

Dissemination of the data base

For internal use at MIT and Beth Israel Hospital, each data base record was stored on a 3M (DC300A) 1/4 inch tape cartridge. The data was stored in a format (see Appendix A-1) in which the original ECG data and associated annotations were merged in sequential "epochs" (rather than as separate files). This made it possible to have real-time access to the data and annotations that otherwise would have been impossible on a sequential mass storage device.

The data base is also available on 9-track magnetic tape in a format compatible with the American Heart Association (AHA) data base for Ventricular Arrhythmia Detectors (Hermes,1980a). However, the additional annotation codes provided by the MIT-BIH data base have been added as a superset that specify atrial arrhythmias, conduction defects, changes in predominant rhythm, and episodes of noisy data. The additional annotation codes are described in the tape directory and format specification

(MIT-BIH,1980) for the data base.

Recently we have distributed the MIT-BIH data base on analog FM tape. The primary reason for this format is economic, since the cost of equipment to play back digital tapes is often prohibitive. The data format (see Appendix A-2) is certainly universal: most, if not all, ECG monitoring systems can accept analog data. The analog data is recorded on two of the four FM channels, and annotation data and 5-second time ticks are encoded as EIA RS-232 bit-serial voltage levels compatible with most computers.

A "compressed" annotation format was also developed which saves only the truth annotations in a very compact format (Appendix A-3). This format was used for experiments that required only beat and rhythm labels, and their time of occurrence. For example, the RR interval predictor and atrial fibrillation detectors described in chapter 8 could be evaluated on the entire 48 half-hour data base in under 30 minutes.

## 10.2 ECG Data Base: Description

The data base consists of 48 records, each containing 30 minutes of two-channel ECG with beat, rhythm, and artifact annotations. A comprehensive set of beat and rhythm annotations was used to make the data base as useful as possible for developmental purposes (see Tables 10-1a and 10-1b). For example, supraventricular beats are divided into 11 subcategories, and thus allow one to determine if a false positive PVC was really a SVPB (a bad mistake) or an aberrated SVPB (somewhat more forgivable).

Rhythm annotations make it possible to evaluate algorithm performance with different underlying rhythms. For example, regions of atrial fibrillation and flutter are indicated on the data base, permitting automatic evaluation of atrial fibrillation detectors. Detection algorithms for ventricular flutter and fibrillation could also be tested in a similar fashion. Rhythm annotations are mutually exclusive, although they may overlap simulataneous episodes of artifact.

Table 10-1a: MIT-BIH Data Base Annotation Codes

The MIT-BIH annotation codes are listed below. Also listed are a suggested mnemonic, the corresponding AHA annotation code, and description. The AHA annotation code is an ASCII character from the set { N V E F R P Q U [ ] }. The R (R-on-T) code is currently not used in the MIT-BIH data base.

| Mnemonic | MITBIH Code | AHA | Description |
|---|---|---|---|
| NORMAL | 01 | N | normal QRS |
| LBBB | 02 | N | left bundle branch block |
| RBBB | 03 | N | right bundle branch block |
| ABERR | 04 | N | aberrantly conducted beat |
| PVC | 05 | V | ventricular Premature Beat |
| FUSION | 06 | F | fusion beat **** |
| NPC | 07 | N | nodal premature beat |
| APC | 08 | N | atrial premature beat |
| SVPB | 09 | N | nodal or Atrial premature beat |
| VESC | 10 | E | ventricular escape beat |
| NESC | 11 | N | nodal escape beat |
| PACE | 12 | P | paced beat |
| UNKNOWN | 13 | Q | cannot identify, QRS like event. |
| NOISE | 14 | O,U | beginning of NOISE * |
| QUIET | 15 | O | end of NOISE * |
| SPIKE | 16 | O | single QRS-like artifact |
| P | 17 | O | P-wave ** |
| Q | 18 | O | Q-wave ** |
| R | 19 | O | R-wave ** |
| S | 20 | O | S-wave ** |
| T | 21 | O | T-wave ** |
| COMMENT | 22 | O | comment (text) annotation *** |
| FIRST | 23 | O | first annotation (optional) ** |
| LAST | 24 | O | last annotation (optional) ** |
| BBB | 25 | N | left or right bundle branch block beat |
| PACESP | 26 | O | pacemaker spike without capture |
| AXIS | 27 | O | axis shift |
| ONSET | 28 | O | rhythm onset (text) annotation *** |
| OFFSET | 29 | O | offset comment ** |
| LEARN | 30 | O | learning ** |
| FLWAV | 31 | O | ventricular flutter wave |
| VFON | 32 | [ | onset of ventricular flutter/fibrillation |
| VFOFF | 33 | ] | end of ventricular flutter/fibrillation |
| AEB | 34 | N | atrial ectopic beat |
| NEB | 35 | N | nodal ectopic beat |
| MISSB | 36 | O | missed beat |
| BLAPB | 37 | O | blocked APB |
| AXLMX | 38 | O | legal codes are < AXLMX ** |

Table 10-1b: Rhythm Onset Annotations

Rhythm onset annotation (MIT-BIH annotation code 28) include an ASCII string which begins with a "(":

| | |
|---|---|
| (AB | atrial bigeminy |
| (AFIB | atrial fibrillation |
| (AFL | atrial flutter |
| (AT | atrial tachycardia |
| (B | ventricular bigeminy |
| (BI | first degree heart block |
| (BII | second degree heart block |
| (BIII | third degree heart block |
| (IVR | idioventricular rhythm |
| (N | normal sinus rhythm |
| (NOD | nodal (A-V junctional) rhythm |
| (PAT | paroxsysmal atrial tachycardia |
| (PREX | pre-excitation (WPW) |
| (SBR | sinus bradycardia |
| (SVTA | supraventricular tachyarrhythmia |
| (T | ventricular trigeminy |
| (VFIB | ventricular fibrillation |
| (VFL | ventricular flutter |
| (VT | ventricular tachycardia |

Notes:

Sparse comment annotations exist on a few records. They are:

| | |
|---|---|
| PSE | pause |
| TS | tape slippage |

* Annotation codes 14 and 15 are used in pairs.
If the subtype is <1>, no beats are labeled until the next code 15.

** These codes are reserved for future use, and should be ignored.

*** This annotation has a text string, terminated by a null <0>.

**** In the context of paced rhythm (tapes 102, 104, 107, and 217), annotation code 6 is used for pacemaker fusion beats.

The predominant rhythm categories for the 100 series (randomly chosen) and 200 series (selected) data base tapes are shown in Tables 10-2a and 10-2b, respectively. The 100 series tapes primarily exhibited normal sinus rhythm; the 200 series tapes, on the other hand, offer a wider (and more difficult) variety of background and ectopic rhythms.

A rough breakdown of the number of different types of QRS complexes is shown in Table 10-3a (100 series) and Table 10-3b (200 series). Three of the tapes have pacemaker fusion beats, and seven of the tapes exhibit bundle branch blocks that offer significant challenges to algorithms that rely heavily on QRS width for beat classification.

Several of the tapes can offer significant challenges to any algorithm. One tape in particular (203) has a predominant rhythm of atrial flutter and fibrillation (so there goes timing!), with frequent multifocal PVCs, numerous aberrantly conducted beats, plus QRS morphology changes due to axis shift (and there goes morphology!). It also has considerable muscle artifact and baseline shifts on both ECG channels. It truly is "a very difficult tape, even for humans!". A more complete description of the data base is available in the tape directory (MIT-BIH,1980).

Table 10-2a: Predominant Rhythms for 100 Series Tapes

| Tape | NSR | VT | SVTA | AFIB | Other |
|---|---|---|---|---|---|
| 100 | x | | | | |
| 101 | x | | | | |
| 102 | x | | | | Paced |
| 103 | x | | | | |
| 104 | x | | | | Paced |
| 105 | x | | | | |
| 106 | x | x | | | |
| 107 | x | | | | Paced |
| 108 | x | | | | |
| 109 | x | | | | |
| 111 | x | | | | |
| 112 | x | | | | |
| 113 | x | | | | |
| 114 | x | | x | | |
| 115 | x | | | | |
| 116 | x | | | | |
| 117 | x | | | | Sinus bradycardia |
| 118 | x | | | | |
| 119 | x | | | | Bigeminy, trigeminy |
| 121 | x | | | | |
| 122 | x | | | | |
| 123 | x | | | | Sinus bradycardia |
| 124 | x | x | | | Nodal rhythm, idioventricular rhythm |

Table 10-2b: Predominant Rhythms for 200 Series Tapes

| Tape | NSR | VT | SVTA | AFIB | Other |
|---|---|---|---|---|---|
| 200 | x | x | | | |
| 201 | x | | | x | Nodal rhythm |
| 202 | x | | | x | |
| 203 | x | x | | x | Atrial flutter |
| 205 | x | x | | | |
| 207 | x | x | x | | V flutter, idioventricular rhythm |
| 208 | x | | | | |
| 209 | x | | x | | |
| 210 | | x | | x | |
| 212 | x | | | | |
| 213 | x | x | | | Bigeminy, trigeminy |
| 214 | x | x | | | |
| 215 | x | x | | | |
| 217 | x | x | | | Paced |
| 219 | x | | | x | Pauses |
| 220 | x | | x | | |
| 221 | | x | | x | |
| 222 | x | | x | x | Atrial flutter and bigeminy, nodal rhythm |
| 223 | x | x | | | Bigeminy, trigeminy |
| 228 | x | | | | |
| 230 | x | | | | Pre-excitation (WPW) |
| 231 | x | | | | Second degree heart block |
| 232 | | | | | Sinus bradycardia |
| 233 | x | x | | | |
| 234 | x | | x | | |

Table 10-3a: QRS Types for 100 Series Tapes

| Tape | N | N SVPB | N BBB | N Aberr | N Other | V PVC | E VESC | F FUS | P PACED | Q UNKNOWN |
|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 2239 | 33 | | | | 1 | | | | |
| 101 | 1860 | 3 | | | | | | | | 2 |
| 102 | 99 | | | | | 4 | | 56 | 2028 | |
| 103 | 2082 | | 2 | | | | | | | |
| 104 | 163 | | | | | 2 | | 666 | 1381 | 18 |
| 105 | 2526 | | | | | 41 | | | | 5 |
| 106 | 1507 | | | | | 520 | | | | |
| 107 | | | | | | 59 | | | 2078 | |
| 108 | 1740 | 4 | | | 1 NESC | 16 | | 2 | | |
| 109 | | | 2492 | | | 38 | | 2 | | |
| 111 | | | 2123 | | | 1 | | | | |
| 112 | 2537 | 2 | | | | | | | | |
| 113 | 1789 | | | 6 | | | | | | |
| 114 | 1821 | 11 | | | | 43 | | 4 | | |
| 115 | 1953 | | | | | | | | | |
| 116 | 2302 | 1 | | | | 109 | | | | |
| 117 | 1534 | 1 | | | | | | | | |
| 118 | | 96 | 2166 | | | 16 | | | | |
| 119 | 1543 | | | | | 444 | | | | |
| 121 | 1861 | 1 | | | | 1 | | | | |
| 122 | 2476 | | | | | | | | | |
| 123 | 1515 | | | | | 3 | | | | |
| 124 | 1531 | 2 | | | 29 NPC<br>5 NESC | 47 | | 5 | | |

Table 10-3b: QRS Types for 200 Series Tapes

| Tape | N | N APC | N BBB | N Aberr | N Other | SV | V PVC | E VESC | F FUS | P PACED | Q UNKNOWN |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 200 | 1743 | 30 | | | | | 826 | | 2 | | |
| 201 | 1625 | 30 | | 97 | 1 NPC<br>10 NESC | | 198 | | 2 | | |
| 202 | 2061 | 36 | | 19 | | | 19 | | 1 | | |
| 203 | 2531 | | | 2 | | | 444 | | 1 | | 4 |
| 205 | 2571 | 3 | | | | | 71 | | 11 | | |
| 207 | | 107 | 1545 | | | | 105 | 105 | | | |
| 208 | 1587 | | | | 2 | SVPB | 992 | | 373 | | 2 |
| 209 | 2621 | 382 | | | | | 1 | | | | |
| 210 | 2423 | | | 22 | | | 191 | 1 | 10 | | |
| 212 | 923 | | 1825 | | | | | | | | |
| 213 | 2641 | 25 | | 3 | | | 220 | | 362 | | |
| 214 | 2002 | | | | | | 257 | | 1 | | 2 |
| 215 | 3196 | 2 | | | | | 164 | | 1 | | |
| 217 | 244 | | | | | | 162 | | 260 | 1542 | |
| 219 | 2082 | 7 | | | | | 64 | | 1 | | |
| 220 | 1954 | 94 | | | | | | | | | |
| 221 | 2031 | | | | | | 396 | | | | |
| 222 | 2062 | 208 | | | 1 NPC<br>212 NESC | | | | | | 1 |
| 223 | 2029 | 72 | | 1 | 16 | AEB | 473 | | 14 | | |
| 228 | 1688 | 3 | | | | | 362 | | | | |
| 230 | 2255 | | | | | | 1 | | | | |
| 231 | 314 | 1 | 1254 | | | | 2 | | | | |
| 232 | | 1382 | 397 | | 1 | NESC | | | | | |
| 233 | 2230 | 7 | | | | | 831 | | 11 | | |
| 234 | 2700 | | | | 50 | NPC | 3 | | | | |

## 10.3 Beat by Beat Comparison

The digitized tapes were used as input to a version of the ECG analysis program which can mark the tape with its own beat-by-beat annotations. During the evaluation, the same copy of the algorithm was used for all ECG excerpts, and no adjustment of parameters was permitted. Since algorithm performance may be dependent on how much "idle time" there is, the tapes were played back in real-time.

The program's annotations were compared automatically with the tape's truth annotations on a beat-by-beat basis. Up to 150 msec of QRS timing error was allowed to determine if a beat was a true or false detection, or missed. If two or more algorithm detections occurred within the window, they were tabulated and reviewed manually, and later incorporated into the final results.

Two matrices were produced by the comparison of annotations in each excerpt. One was a comparison of truth vs. algorithm beat labels, where each matrix element B(i,j,k,l) contained a beat-pair count, and i,j,k,l are:

i: truth type code (0 if false QRS detection)
j: algorithm type code (0 if missed)
k: current truth rhythm code (NSR, AFIB, VT, ...)
l: current truth noise level (noisy/not noisy)

The other matrix contained truth vs. algorithm ventricular run lengths, where each matrix element VR(i,j,k,l) contained paired run lengths for consecutive PVC or ventricular escape beats, and i,j,k,l were:

i: truth V-run length
j: algorithm V-run length
k: current truth rhythm
l: current truth noise level

The two matrices could be printed on the line printer, and also saved on mass storage for subsequent analysis. The results for one tape (tape 200) are shown in Table 10-4a on the next page. The upper confusion matrix shows beat-by-beat performance where the truth annotation code (i) is listed row by row, and the algorithm annotation code (j) is listed column by column. For example, 46 true PVCs were incorrectly labeled NORMAL by the algorithm. The lower confusion matrix (Table 10-4b) compares the ventricular run lengths on the truth channel and those detected by the algorithm. This matrix shows that although 14 "truth" couplets were correctly identified, 24 were identified only as single PVCs, and 5 were missed entirely.

As a result of the automatic comparison process, errors in beat identification can be documented on a strip chart recorder or displaying waveform editor for later review. Incorrect classifications are noted on the strips with a COMMENT annotation that lists both the truth and algorithm annotation codes. An example is shown in Figure 10-2.

Table 10-4a: Beat-by-Beat Confusion Matrix for Tape 200

TOTAL RESULTS FOR TAPE # 200
WINDOW = 150 MSEC
REFERENCE = TRUTH; TEST = PSS2

| | | Algorithm Beat Label | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | MISS | N | AB | V | FUS | APC | SVPB | LEARN |
| | NOTQRS | - | 7 | - | - | - | - | 2 | - |
| Data Base | N | - | 1686 | 2 | - | - | - | 20 | 33 |
| | AB | - | - | - | - | - | - | - | - |
| | V | 1 | 46 | 11 | 736 | - | - | 16 | 16 |
| | FUS | 1 | - | - | 1 | - | - | - | - |
| | APC | - | 23 | - | - | - | - | 7 | - |
| | SVPB | - | - | - | - | - | - | - | - |
| | LEARN | - | - | - | - | - | - | - | - |

Table 10-4b: PVC Run Length Confusion Matrix for Tape 200

V RUN RESULTS FOR TAPE # 200 --
WINDOW = 150 MSEC
REFERENCE = ATRUTH; TEST = PSS2

| | | Algorithm Run Length | | | | |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 |
| Data Base | 0 | - | - | - | - | - |
| | 1 | 52 | 668 | - | - | - |
| | 2 | 5 | 24 | 14 | - | - |
| | 3 | - | 1 | 1 | 2 | - |
| | 4 | - | - | 1 | - | 1 |

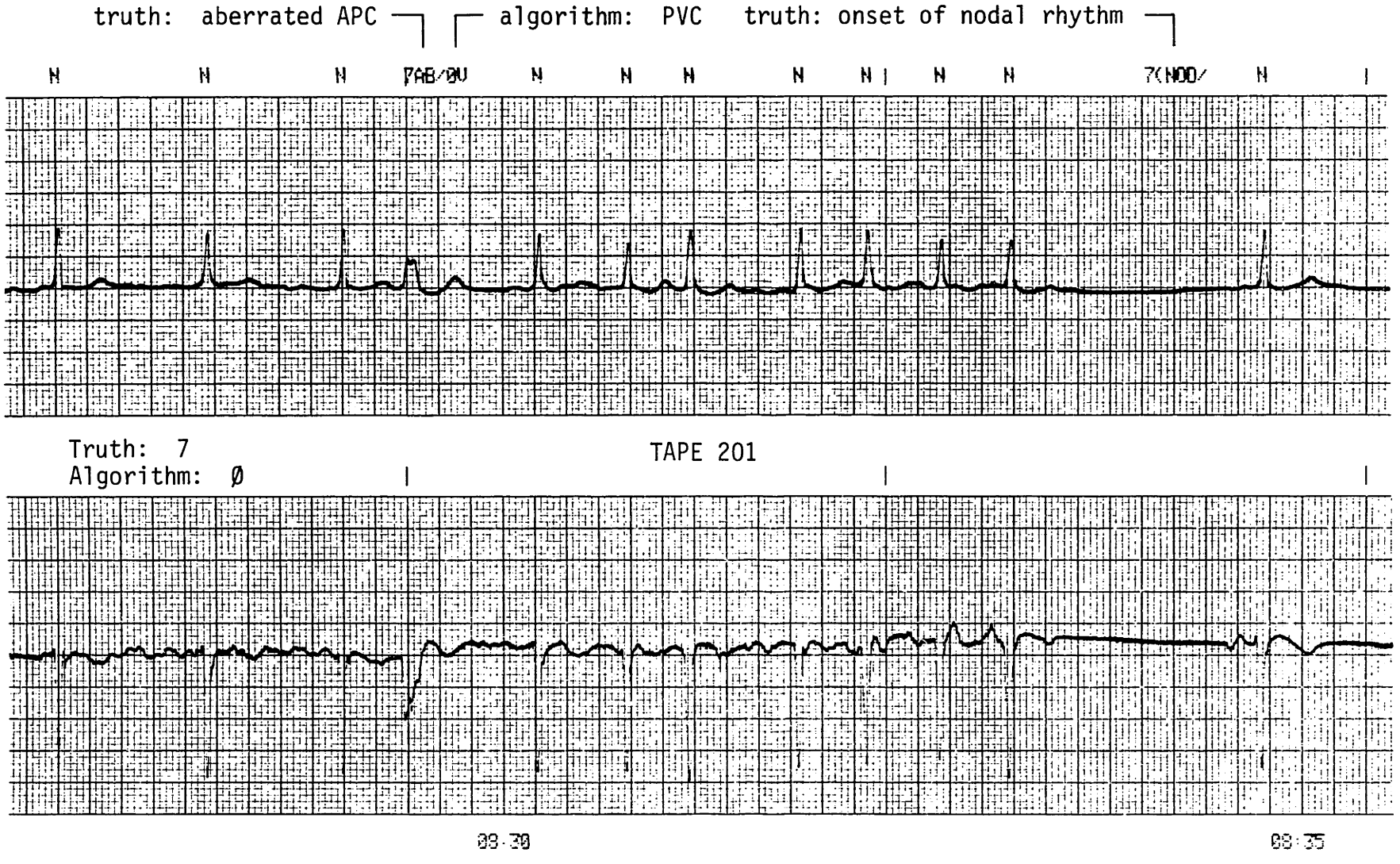



Figure 10-2: Strip Chart Recording of Comparator Output

## 10.4 Beat-by-Beat Evaluation Results

An evaluation of the bedside arrhythmia monitor was performed using the entire 48 tape MIT-BIH database, and illustrates the power of the technique. The following policies were adopted regarding beat and rhythm labels:

1) Learning beats (the first fifty non-ventricular-flutter beats) were excluded.

2) Regions of ventricular flutter and fibrillation were excluded (and are present only on tape 207). All other background rhythms and episodes containing artifact were included.

3) Beat labels were re-mapped into a smaller number of different categories:

   NORMAL: NORMAL, LBBB, RBBB, NESC, PB, UNKNOWN, BBB, AEB, NEB

   SVPB: NPC, APC, SVPB, ABERR

   PVC: PVC, VESC, FL

   FUSION: Fusion beats were discarded when computing PVC sensitivity and positive predictivity, since many of the fusion beats exhibited near-normal morphology and timing. However, pacemaker fusion beats (FUSION beats on paced tapes 102, 104, and 217) were considered NORMAL.

   $\overline{\text{PVC}}$: This group included NORMAL and SVPB group beats.

   $\overline{\text{QRS}}$: Falsely-detected and missed beats.

Aggregate Results

A number of commonly used algorithm performance measures can be derived from the overall results shown in Table 10-5a, where the beat label categories have been merged into five subcategories:

NORMAL, PVC, FUSION, SVPB, and $\overline{QRS}$.

The results can be further condensed into four categories, which will serve as the basis for subsequent evaluation of the algorithm as a PVC detector:

$\overline{PVC}$, PVC, FSN, and $\overline{QRS}$.

The condensed results for the entire data base and subsets of the data base are shown in Tables 10-5b to 10-5e as a 4 x 3 confusion matrix:

| | | Algorithm | | |
|---|---|---|---|---|
| | | $\overline{PVC}$ | PVC | $\overline{QRS}$ |
| Truth | $\overline{PVC}$ | TN | FP | MN |
| | PVC | FN | TP | MV |
| | FSN | FUSN | FUSP | MFUS |
| | $\overline{QRS}$ | EN | EP | -- |

with the following entries:

TN = True Negatives = true NORMALs labeled NORMAL
FN = False Negatives = true PVCs labeled NORMAL
FUSN = true FUSION beats labeled NORMAL
EN = Extra Negatives = false detections labeled NORMAL
FP = False Positives = true NORMALs labeled PVC
TP = True Positives = true PVCs labeled PVC
FUSP = true FUSION beats labeled PVC
EP = Extra Positives = false detections labeled PVC
MN = Missed Negatives = missed NORMALs
MV = Missed Positives = missed PVCs
MFUS = missed FUSION beats.

Table 10-5a
Evaluation results for entire data base
Bedside Monitor, August 1981

| DATA BASE | | ALGORITHM NORMAL | PVC | FSN | SVPB | $\overline{QRS}$ |
|---|---|---|---|---|---|---|
| | NORMAL | 95613 | 86 | - | 215 | 502 |
| | PVC | 678 | 5973 | - | 368 | 87 |
| | FSN | 413 | 374 | - | 1 | 5 |
| | SVPB | 2107 | 84 | - | 523 | 21 |
| | $\overline{QRS}$ | 150 | 116 | - | 37 | |

Table 10-5b
Entire Data Base

| | $\overline{PVC}$ | PVC | $\overline{QRS}$ |
|---|---|---|---|
| $\overline{PVC}$ | 98458<br>99.30% | 170<br>0.17% | 523<br>0.53% |
| PVC | 1046<br>14.7% | 5973<br>84.1% | 87<br>1.2% |
| FSN | 414 | 374 | 5 |
| $\overline{QRS}$ | 187 | 116 | |

Table 10-5c
Atrial fib/flutter only

| | $\overline{PVC}$ | PVC | $\overline{QRS}$ |
|---|---|---|---|
| $\overline{PVC}$ | 11095<br>99.19% | 44<br>0.39% | 47<br>0.42% |
| PVC | 175<br>18.5% | 759<br>80.0% | 14<br>1.5% |
| FSN | 9 | - | 2 |
| $\overline{QRS}$ | 46 | 8 | |

Table 10-5d
100 Series only

| | $\overline{PVC}$ | PVC | $\overline{QRS}$ |
|---|---|---|---|
| $\overline{PVC}$ | 44684<br>98.97% | 34<br>0.08% | 430<br>0.95% |
| PVC | 197<br>14.9% | 1114<br>84.0% | 15<br>1.1% |
| FSN | 7 | 4 | - |
| $\overline{QRS}$ | 92 | 53 | |

Table 10-5e
200 Series only

| | $\overline{PVC}$ | PVC | $\overline{QRS}$ |
|---|---|---|---|
| $\overline{PVC}$ | 53774<br>99.58% | 136<br>0.25% | 93<br>0.17% |
| PVC | 849<br>14.7% | 4859<br>84.1% | 72<br>1.2% |
| FSN | 407 | 370 | 5 |
| $\overline{QRS}$ | 95 | 63 | |

QRS detector accuracy

Two performance measures that can be used to describe QRS detector accuracy are "QRS sensitivity" and "QRS positive predicitivity", defined and computed (based on Table 10-5b) below:

QRS Sensitivity = QRS Se =

$$= \frac{\text{number of true QRS complexes detected}}{\text{number of true QRS complexes}}$$

$$= \frac{TN+FN+FUSN+FP+TP+FUSP}{TN+FN+FUSN+FP+TP+FUSP + MN+MV+MFUS}$$

$$= \frac{106435}{106435+615} = 99.43\%$$

QRS Positive Predictivity = QRS P+ =

$$= \frac{\text{number of true QRS complexes detected}}{\text{number of detected QRS complexes}}$$

$$= \frac{TN+FN+FUSN+FP+TP+FUSP}{TN+FN+FUSN+FP+TP+FUSP + EN+EP}$$

$$= \frac{106435}{106435+303} = 99.72\%$$

Of the 615 beats (MN+MV+MFUS) missed by the QRS detector, 372 (60%) were pacemaker fusion beats where only the pacemaker spike was delineated, and rejected as artifact because it was too narrow. Other missed beats were either low amplitude, or were present in artifact that was not processed by the algorithm. Most false positive detections were due to artifact, and approximately 30 false detections were due to T-waves.

PVC detection accuracy

Two commonly used performance measures for PVC detection accuracy can be derived from the results shown in Table 10-5b:

PVC Sensitivity = PVC Se =

$$= \frac{\text{number of true PVCs labeled PVC}}{\text{number of true PVCs}} = \frac{TP}{TP+FN+MV} = 84.1\%$$

PVC Positive Predictivity = PVC P+ =

$$= \frac{\text{number of true PVCs labeled PVC}}{\text{number of algorithm PVCs}} = \frac{TP}{TP+FP+EV} = 95.4\%$$

The overall results shown in Table 10-5b indicate an average 84% sensitivity and 95% positive predictivity for PVC detection. The results for the entire data base may be subclassified using the rhythm and noise annotations to test the algorithm in specific areas. For example, the presence of atrial fibrillation and flutter is an important subcategory, since many algorithms rely on timing for event detection and classification. The results for atrial fibrillation and flutter categories are shown in Figure 10-5c, and indicates a PVC sensitivity of 80.0% and PVC positive predicitivity of 93.5%. The algorithm's performance was somewhat degraded for this rhythm category since timing information cannot be relied upon to classify complexes.

The overall results can also be broken down according to whether the tape was randomly chosen (100 series tapes) or selected (200 series tapes). These results are shown in Tables 10-5d and 10-5e, and indicates that PVC sensitivity was about the

same for the two sets, and the PVC positive predictivity was somewhat higher (96.1%) for the 200 series.

The aggregate results, however, do not reflect the substantial tape-to-tape variability in performance which may be clinically important. Results for individual tapes are listed in Table 10-6 according to the TN, FN, FUSN, EN, FP, TP, FUSP, EP, MN, MV, and MFUS counts. Table 10-7 lists QRS and PVC sensitivity and positive predictivity for individual tapes. The results are also shown graphically in Figure 10-3, on which is plotted the PVC detection and classification counts MP, FN, TP, FP, and EP. These figures indicate a wide variability in performance between individual tapes. PVC sensitivity and positive predictivity ranged from zero to 100%, an observation not conveyed in the overall totals.

Table 10-6

Tabulated Results for Individual Trials, August 1981
(VFL and Learning Phase Beats are excluded)

| TAPE | TN | FN | FUSN | EN | FP | TP | FUSP | EP | MN | MV | MFUS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 2220 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 101 | 1806 | 0 | 0 | 3 | 6 | 0 | 0 | 0 | 1 | 0 | 0 |
| 102 | 1885 | 3 | 0 | 0 | 0 | 1 | 0 | 0 | 245 | 0 | 0 |
| 103 | 2033 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 104 | 2048 | 0 | 0 | 11 | 0 | 1 | 0 | 0 | 127 | 1 | 0 |
| 105 | 2453 | 26 | 0 | 56 | 16 | 12 | 0 | 17 | 14 | 0 | 0 |
| 106 | 1457 | 67 | 0 | 1 | 6 | 450 | 0 | 11 | 0 | 3 | 0 |
| 107 | 2022 | 37 | 0 | 0 | 1 | 17 | 0 | 2 | 0 | 5 | 0 |
| 108 | 1687 | 8 | 2 | 17 | 2 | 6 | 0 | 4 | 10 | 0 | 0 |
| 109 | 2442 | 15 | 0 | 0 | 0 | 23 | 0 | 0 | 0 | 0 | 0 |
| 111 | 2070 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| 112 | 2487 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 113 | 1760 | 0 | 0 | 0 | 1 | 0 | 0 | 14 | 0 | 0 | 0 |
| 114 | 1776 | 0 | 0 | 3 | 1 | 43 | 4 | 0 | 4 | 0 | 0 |
| 115 | 1901 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 116 | 2226 | 1 | 0 | 0 | 0 | 105 | 0 | 2 | 26 | 3 | 0 |
| 117 | 1484 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 118 | 2210 | 12 | 0 | 0 | 1 | 3 | 0 | 0 | 0 | 0 | 0 |
| 119 | 1506 | 0 | 0 | 0 | 0 | 431 | 0 | 2 | 0 | 0 | 0 |
| 121 | 1808 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |
| 122 | 2424 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 123 | 1463 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 |
| 124 | 1516 | 26 | 5 | 0 | 0 | 21 | 0 | 0 | 0 | 0 | 0 |
| 200 | 1738 | 73 | 0 | 9 | 0 | 736 | 1 | 0 | 0 | 1 | 1 |
| 201 | 1633 | 42 | 2 | 0 | 63 | 156 | 0 | 2 | 14 | 0 | 0 |
| 202 | 2059 | 10 | 1 | 1 | 5 | 7 | 0 | 0 | 2 | 1 | 0 |
| 203 | 2438 | 140 | 1 | 37 | 24 | 283 | 0 | 12 | 27 | 17 | 0 |
| 205 | 2522 | 3 | 11 | 0 | 0 | 66 | 0 | 0 | 1 | 2 | 0 |
| 207 | 1623 | 121 | 0 | 4 | 6 | 45 | 0 | 15 | 1 | 22 | 0 |
| 208 | 1556 | 22 | 101 | 5 | 0 | 949 | 263 | 5 | 8 | 2 | 2 |
| 209 | 2951 | 0 | 0 | 3 | 0 | 1 | 0 | 2 | 0 | 0 | 0 |
| 210 | 2390 | 60 | 7 | 7 | 0 | 128 | 0 | 0 | 7 | 3 | 2 |
| 212 | 2697 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 213 | 2617 | 6 | 263 | 0 | 0 | 213 | 99 | 1 | 0 | 1 | 0 |
| 214 | 1951 | 93 | 1 | 1 | 2 | 159 | 0 | 2 | 2 | 1 | 0 |
| 215 | 3149 | 7 | 1 | 1 | 0 | 154 | 0 | 0 | 0 | 0 | 0 |
| 217 | 1991 | 21 | 0 | 2 | 4 | 123 | 0 | 13 | 1 | 17 | 0 |
| 219 | 2039 | 9 | 1 | 0 | 0 | 54 | 0 | 0 | 0 | 0 | 0 |
| 220 | 1996 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 221 | 1988 | 5 | 0 | 0 | 0 | 379 | 0 | 1 | 0 | 3 | 0 |
| 222 | 2373 | 0 | 0 | 6 | 31 | 0 | 0 | 1 | 27 | 0 | 0 |
| 223 | 2066 | 196 | 9 | 0 | 1 | 276 | 5 | 0 | 0 | 0 | 0 |
| 228 | 1649 | 7 | 0 | 14 | 0 | 343 | 0 | 5 | 3 | 0 | 0 |
| 230 | 2203 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 231 | 1517 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| 232 | 1731 | 0 | 0 | 3 | 0 | 0 | 0 | 2 | 0 | 0 | 0 |
| 233 | 2199 | 34 | 9 | 2 | 0 | 781 | 2 | 2 | 0 | 2 | 0 |
| 234 | 2698 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 |

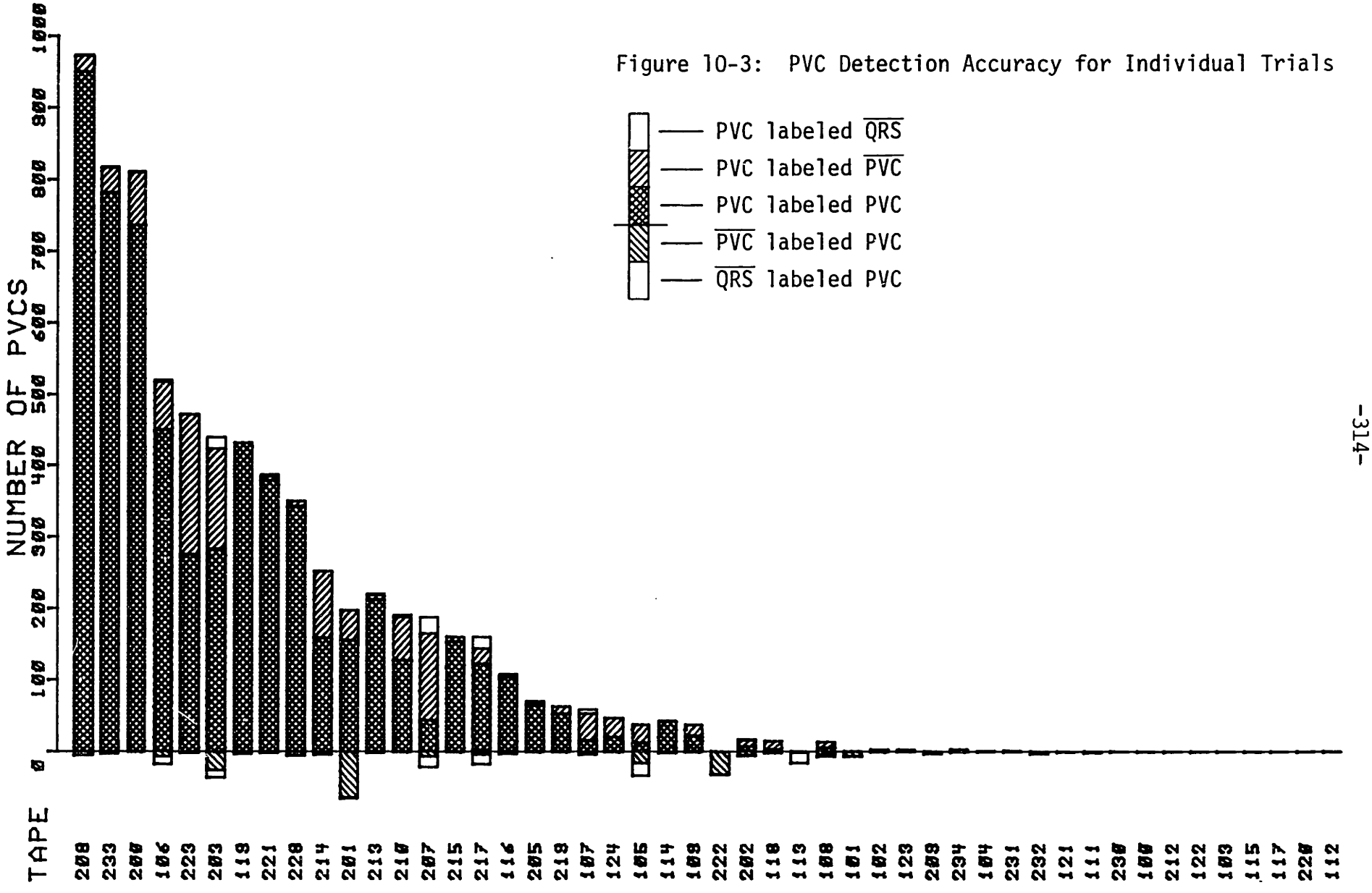
Table 10-7

QRS and PVC Detection Accuracy for Individual Trials

| TAPE | QRS Se | QRS P+ | PVC Se | PVC P+ | TP |
|---|---|---|---|---|---|
| 100 | 100.00% | 100.00% | 100.00% | 100.00% | 1 |
| 101 | 99.94% | 99.83% | -- | 0.00% | 0 |
| 102 | 88.52% | 100.00% | 25.00% | 100.00% | 1 |
| 103 | 100.00% | 100.00% | -- | -- | 0 |
| 104 | 94.12% | 99.47% | 50.00% | 100.00% | 1 |
| 105 | 99.44% | 97.82% | 31.58% | 42.86% | 12 |
| 106 | 99.85% | 99.95% | 86.54% | 98.68% | 450 |
| 107 | 99.76% | 100.00% | 28.81% | 94.44% | 17 |
| 108 | 99.42% | 99.01% | 42.86% | 75.00% | 6 |
| 109 | 100.00% | 100.00% | 60.53% | 100.00% | 23 |
| 111 | 99.95% | 99.95% | 0.00% | -- | 0 |
| 112 | 100.00% | 100.00% | -- | -- | 0 |
| 113 | 100.00% | 100.00% | -- | 0.00% | 0 |
| 114 | 99.78% | 99.84% | 100.00% | 97.73% | 43 |
| 115 | 100.00% | 100.00% | -- | -- | 0 |
| 116 | 98.77% | 100.00% | 96.33% | 100.00% | 105 |
| 117 | 100.00% | 100.00% | -- | -- | 0 |
| 118 | 100.00% | 100.00% | 20.00% | 75.00% | 3 |
| 119 | 100.00% | 100.00% | 100.00% | 100.00% | 431 |
| 121 | 99.89% | 100.00% | 0.00% | -- | 0 |
| 122 | 100.00% | 100.00% | -- | -- | 0 |
| 123 | 99.80% | 100.00% | 0.00% | -- | 0 |
| 124 | 100.00% | 100.00% | 44.68% | 100.00% | 21 |
| 200 | 99.92% | 99.65% | 90.86% | 100.00% | 736 |
| 201 | 99.27% | 100.00% | 78.79% | 71.23% | 156 |
| 202 | 99.86% | 99.95% | 38.89% | 58.33% | 7 |
| 203 | 98.50% | 98.73% | 64.32% | 92.18% | 283 |
| 205 | 99.88% | 100.00% | 92.96% | 100.00% | 66 |
| 207 | 98.73% | 99.78% | 23.94% | 88.24% | 45 |
| 208 | 99.59% | 99.83% | 97.53% | 100.00% | 949 |
| 209 | 100.00% | 99.90% | 100.00% | 100.00% | 1 |
| 210 | 99.54% | 99.73% | 67.02% | 100.00% | 128 |
| 212 | 100.00% | 100.00% | -- | -- | 0 |
| 213 | 99.97% | 100.00% | 96.82% | 100.00% | 213 |
| 214 | 99.86% | 99.95% | 62.85% | 98.76% | 159 |
| 215 | 100.00% | 99.97% | 95.65% | 100.00% | 154 |
| 217 | 99.17% | 99.91% | 76.40% | 96.85% | 123 |
| 219 | 100.00% | 100.00% | 85.71% | 100.00% | 54 |
| 220 | 100.00% | 100.00% | -- | -- | 0 |
| 221 | 99.87% | 100.00% | 97.93% | 100.00% | 379 |
| 222 | 98.89% | 99.75% | -- | 0.00% | 0 |
| 223 | 100.00% | 100.00% | 58.47% | 99.64% | 276 |
| 228 | 99.85% | 99.30% | 98.00% | 100.00% | 343 |
| 230 | 100.00% | 100.00% | 100.00% | 100.00% | 1 |
| 231 | 100.00% | 100.00% | 100.00% | 100.00% | 2 |
| 232 | 100.00% | 99.83% | -- | -- | 0 |
| 233 | 99.93% | 99.93% | 95.59% | 100.00% | 781 |
| 234 | 100.00% | 100.00% | 100.00% | 100.00% | 3 |

HEWLETT PACKARD

Figure 10-3: PVC Detection Accuracy for Individual Trials



## 10.5 Other Performance Measures

An important question which should be answered by any evaluation is "how well will the algorithm work for any given patient?". The average sensitivity and positive predictivity for an entire data base do not reveal the variability in performance for individual tapes, as shown in Table 10-6 and Figure 10-3. How can this variability be communicated to the potential user?

In a paper by Cox and Hermes (Cox,1980) discussing performance measures and models for PVC detectors, an alternate format for presenting individual trial variability was presented. The contribution of individual trial results to the "gross detection ratio" (sensitivity) is shown by the cumulative distribution in Figure 10-4a, where each tape trial is weighted by the number of truth PVCs. This distribution is the integral of a series of impulses (one for each trial) where the position of each impulse along the X-axis corresponds to the sensitivity for a given tape, and the area of the impulse is proportional to the number of truth PVCs on the tape. Since each trial is weighted by the number of truth PVCs, the area to the upper left of the curve (84.1%) is identical to the overall sensitivity shown in Table 10-5b. The gross sensitivity of 84.1% is the expected detection ratio for PVCs on a beat-to-beat weighting, and is a estimate of the probability of correctly identifying a PVC from the universe of all beats (in the data base).

Another aggregate sensitivity measure is the "average detection ratio" in which the results for each tape are weighted equally (Figure 10-4b). The average detection ratio is the expected value of PVC sensitivity for a single patient selected at random (from a patient population comparable to that used in compiling the data base). The average detection ratio for the bedside monitor was 67.8%, which is somewhat lower than the gross detection ratio.

Similar plots can be made for PVC positive predictivity. Figure 10-5a shows the cumulative positive predictivity, where each trial is weighted according to the number of PVCs detected by the algorithm. The "gross positive predictivity" is 95.4%, which is identical to the value derived in the previous section. If all trials are weighted equally, the "average positive predictivity" is 77.7%.

The reason for the difference between the two averages is very simple. The "gross" detection ratio weighs equally all occurrences of PVCs whether or not they are frequent in a record whereas the "average" detection ratio weighs all records equally regardless of PVC count. The bedside monitor apparently has a higher sensitivity for tapes that have a larger number of PVCs. One reason stems from the fact that the morphology classifier often requires a particular morphology to be seen several times before it is labeled a PVC. For a record with many PVCs, there will be many active PVC classes, and therefore the sensitivity of the algorithm will be high. On the other hand, records that

contain only a few PVCs have relatively fewer active PVC classes (with lower class populations), reducing sensitivity. In the extreme case where a record has a single PVC, the bedside monitor would most likely call it artifact or a SVPB (since it wasn't seen before) since very strict prematurity and width thresholds are used to classify a brand-new morphology.

The question of which average to use is difficult to answer. The "gross" detection ratio tends to be unduly weighted by records that have many PVCs that are correctly classified, whereas a trial's contribution to the "average" detection ratio is the same for a tape that has only two PVCs and one is missed, or a tape with 1000 PVCs, and 500 are missed. Perhaps a trial weighting based on the number of true PVCs in the record can be devised that reflects the clinical cost of making a mistake.

HEWLETT PACKARD

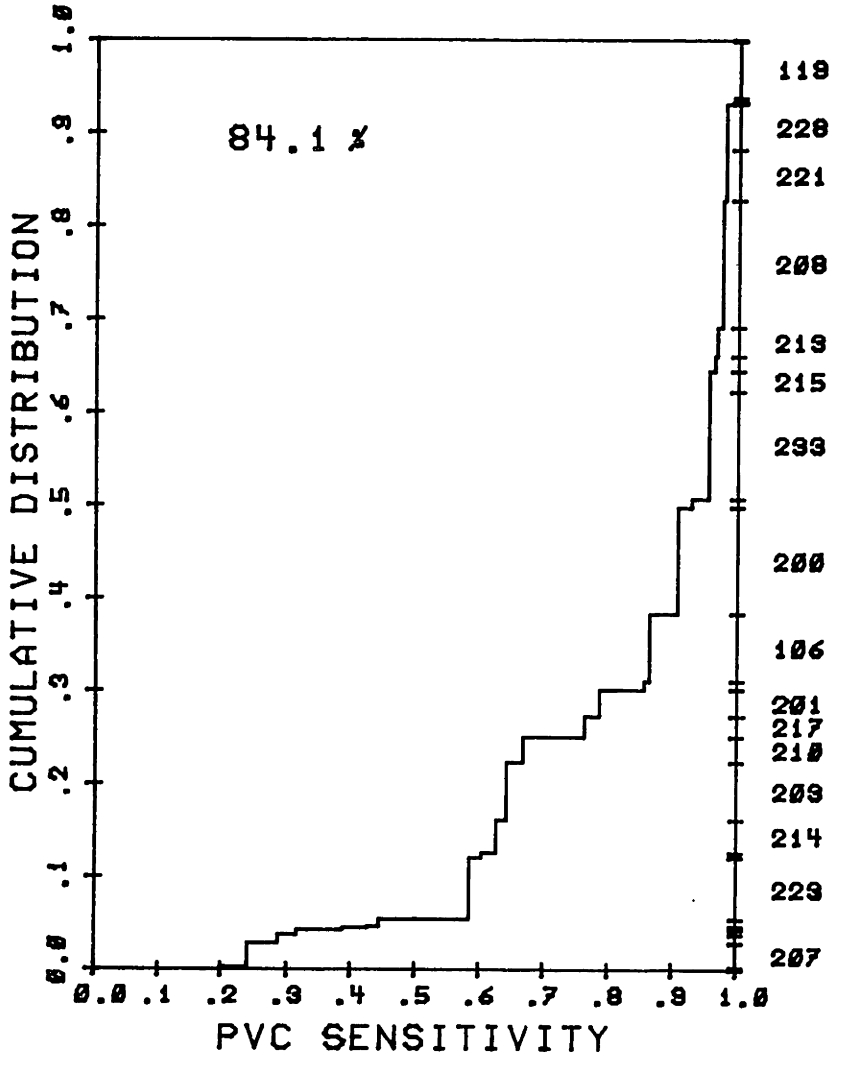



Figure 10-4a
Cumulative Gross Detection Ratio

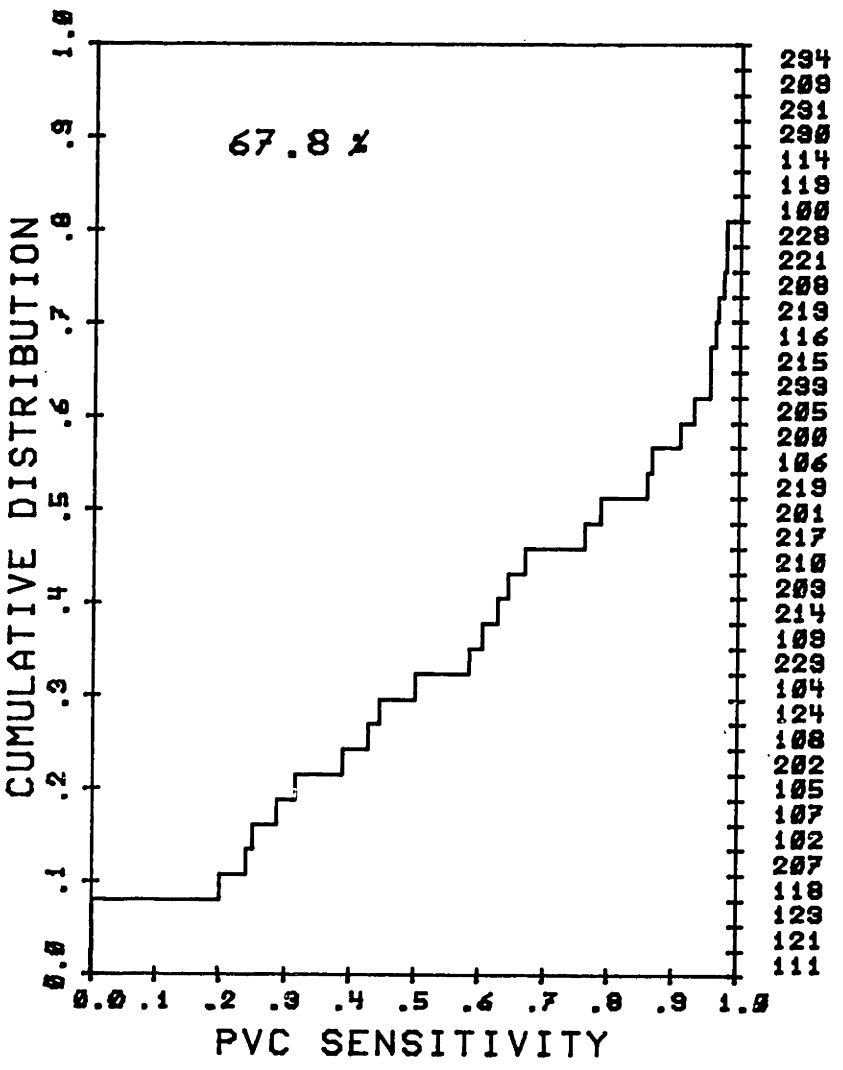



Figure 10-4b
Cumulative Average Detection Ratio

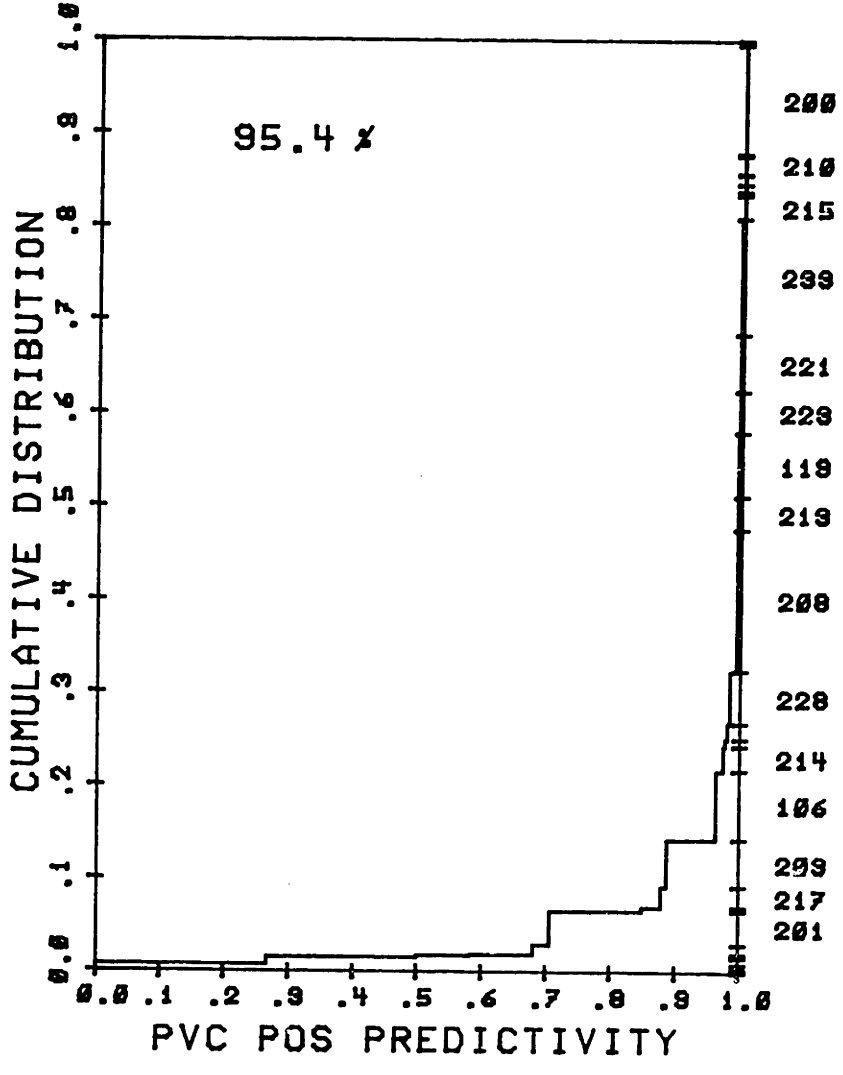



Figure 10-5a
Cumulative Gross Positive Predictivity

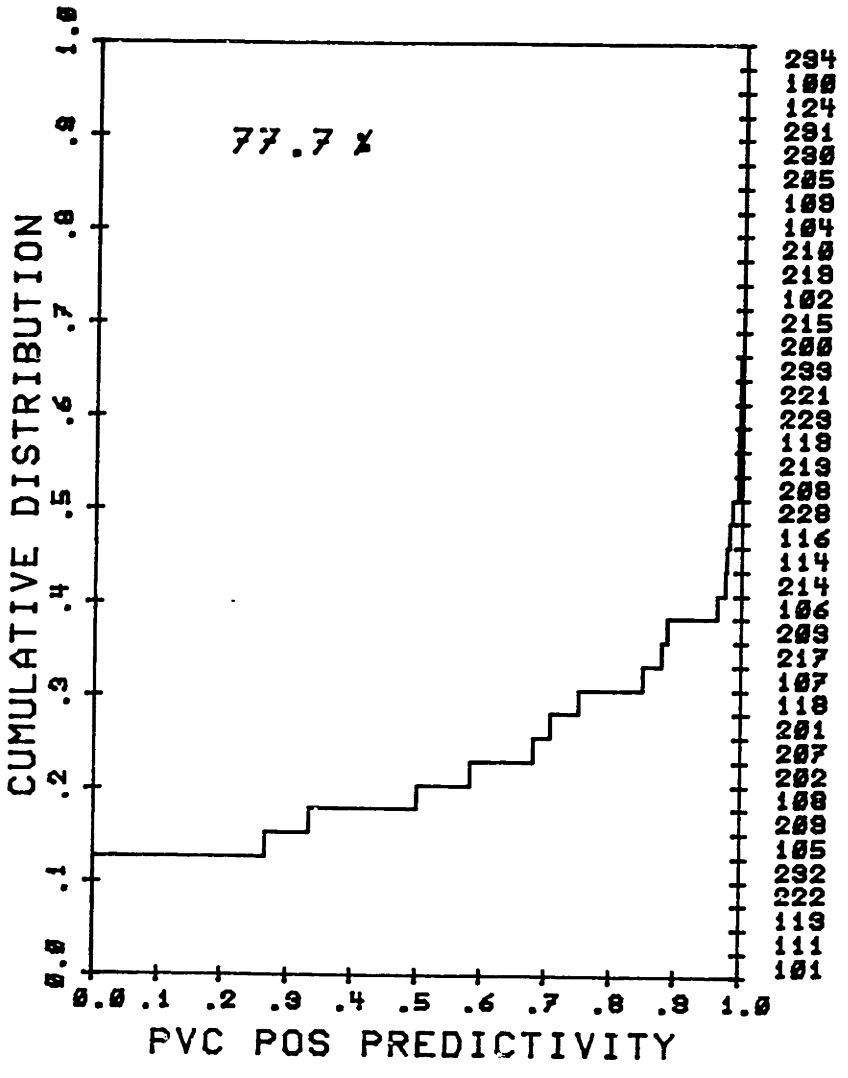



Figure 10-5b
Cumulative Average Positive Predictivity

## 10.6 Sensitivity for Sequential Events

The beat-by-beat performance measures do not reveal another important aspect of algorithm performance: the detection of runs of ventricular and supraventricular premature beats. The run length comparison is performed by noting the maximum run lengths for overlapping ventricular runs. A few examples are shown below:

Beat sequence

```
Truth:      xxVxxVVxxVVVVVVxxVVVVVVVxxxVxxxxxx xx
Algorithm:  xxVxxVVxxVVVVxxxxVVVxVVVVxxxxxVVxxVxx
```

Run-Length

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Truth: | 1 | 2 | 6 | 7 | 1 | 0 | 0 |
| Algorithm: | 1 | 2 | 4 | 4 | 0 | 2 | 1 |

The run length comparison algorithm maintains two state variables for both the truth and algorithm annotation channels:

$CRL_{t,a}$: the length of a current active run

$PRL_{t,a}$: the length of a previous run that terminated, but hasn't been processed yet.

For each truth (t) and algorithm (a) annotation, the two run-state variables are updated according to the following rules:

IF (annotation$_{t,a}$ is PVC or VESC) AND
(annotation$_t$ is not FUSION)

THEN

$CRL_{t,a} = CRL_{t,a} + 1$

ELSE

$PRL_{t,a} = MAX(PRL_{t,a}, CRL_{t,a})$

$CRL_{t,a} = 0$

After this has been done, the confusion matrix entry is updated for the $PRL_t$ and $PRL_a$ run length pair if both $CRL_a$ and $CRL_t$ are

zero (there are no active runs), and that either $PRL_t$ or $PRL_a$ is is non-zero (there was a run in at least one previously). If there is a run length pair to report, the appropriate confusion matrix entry is updated, and $PRL_t$ and $PRL_a$ are set to zero.

The overall performance of the bedside monitor for ventricular run-length detection is shown in Table 10-8. Detector sensitivity for couplets and salvos is readily computed, and is an important performance measure.

Table 10-8
PVC Run Length Comparison for Entire Data Base
Bedside Monitor, August 1981

| | | ALGORITHM PVC RUN LENGTH | | | | | Sensitivity |
|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3-5 | >5 | |
| DATA BASE | 0 | - | 170 | 11 | 2 | 0 | |
| | 1 | 642 | 4872 | 23 | 2 | 0 | 88% |
| | 2 | 23 | 92 | 394 | 4 | 1 | 77% |
| | 3-5 | 9 | 14 | 4 | 22 | 0 | 45% |
| | >5 | 1 | 1 | 1 | 2 | 11 | 69% |

notes:

(1) Regions of ventricular flutter were excluded
(2) Fusion beats were ignored

The most striking result is the poor sensitivity for short (3 to 5 beat) ventricular runs (this result suggests that the detection sensitivity for an arbitrary PVC is independent of the length of the run in which it is embedded). Fortunately the detection sensitivity for longer runs is almost 70%. One possible reason is that long runs are often recognized by the overlapping triangular morph test for ventricular tachycardia, bypassing the usual classifier. The test for ventricular tachycardia and flutter requires an overlapping morph duration of at least 2 seconds, and is usually not activated by short runs.

These results indicate the need for specialized pattern recognition algorithms (such as the overlapping morph test) to detect ventricular runs, instead of a classical event detector and classifier approach. Power spectrum analysis (Nygards,1977, Nolle,1980) and time-domain analysis (Kuo,1978) are other possible approaches.

## 10.7 SVPB Detection Accuracy

The algorithm exhibited rather poor performance for SVPB detection. Based on the results in Table 10-5a:

$$\text{SVPB sensitivity} = \text{SVPB Se} = \frac{\text{true positive SVPBs}}{\text{number of true SVPBs}} = 19.1\%.$$

$$\text{SVPB positive predictivity} = \text{SVPB P+} = \frac{\text{true positive SVPBs}}{\text{SVPBs detected by algorithm}} = 45.7\%.$$

One reason for the low SVPB sensitivity is that several tapes have long runs of SVTA which the algorithm labeled NORMAL after the first five to ten beats. In fact, one of the data base tapes (tape 232) has a predominant rhythm of SVTA, and was responsible for 1,338 of the missed SVPBs. Other errors (both false positive and negative classifications) were due to transitions between atrial fibrillation and normal sinus rhythm. Many of the normal beats that were declared SVPB occurred during periods of atrial fibrillation that were not recognized.

The detection accuracy for isolated SVPBs and couplets was also evaluated during normal sinus rhythm only, which removed all runs of three or more SVPBs, and episodes of atrial fibrillation and flutter (but not transitions). In this case, the SVPB sensitivity was 56%, and SVPB positive predictivity was 43%.

## 10.8 Rhythm and Noise Classification Accuracy

Other than PVC run length evaluation, the performance of rhythm (atrial fibrillation detection) and noise detection stages of the algorithm was not measured. The atrial fibrillation detector and its performance (based on the condensed data base developmental evaluation) were discussed in Chapter 7, section 4, and the results are repeated below:

BIN AF detector (8%-15%)

| truth | | $\overline{AF}$ | AF |
|---|---|---|---|
| | $\overline{AF}$ | 15383 | 1912 |
| | AF | 3551 | 8605 |

The transitions between atrial fibrillation and normal sinus rhythm on several of the tapes were responsible for many of the errors. The running average (over the previous 16 beats) of the atrial fibrillation indicator could not respond rapidly to the frequent transitions.

## 10.9 Analysis of Algorithm Errors

There were several records that were responsible for a major fraction of algorithm errors. Specific cases are discussed below, and examples are shown in Figure 10-6.

### Missed beats MN and MP

Two tapes with paced rhythm (102 and 104) were responsible for most beats missed by the event detector. The triangular morph delineator included only the pacemaker spike, and not the ventricular response to the pacemaker impulse. These (very narrow) morphs were rejected as artifact, and were thus missed. This problem (as far as performance on the data base is concerned) could be easily remedied by reducing the width threshold in the family classifier so that it would accept very narrow complexes. This was not done, however, to handle the potential clinical situation where there is no ventricular capture of the pacemaker impulse. Here we DO want the bedside monitor to declare asystole, rather than recognize the pacemaker impulse as a QRS complex. Of course, the correct approach to the pacemaker artifact problem is to add a pacemaker spike detection circuit to the front end of the amplifier. Other missed beats usually had low amplitude or slope, and thus were rejected by various slope threshold tests in the algorithm. Others were "missed" because of algorithm shutdown due to artifact and noise.

Misclassification of SVPBs

The algorithm exhibited rather poor performance for SVPB detection. This is relatively surprising in light of the experiments performed with the condensed data base in chapter 7, where error rates for false positive and negative SVPB detections (expressed as a percentage of the number of SVPBs) were much lower: 3% and 25% for the 100 series tapes, and 12% and 40% for the 200 series tapes. There are two differences, however, between the condensed data base experiments and the evaluation described in this chapter. First, a "perfect" event detector (data base annotations) were used in the experiments in Chapter 7, whereas the SVPB classification in the actual evaluation had to cope with event detector error. Second, the SVPB detection thresholds in the bedside monitor are based on the observed variance of the prematurity-ratio histogram, rather than the fixed prematurity threshold (84%) used in the experiments. Also, tape 232 ("normal" rhythm is SVTA) was not included in the data set used for the RR-interval predictor experiments. To explore this problem further, another data base evaluation using a fixed 84% prematurity-ratio threshold (rather than an adaptive one based on the prematurity-ratio histogram) will have to be performed.

PVC classification error

There were several factors that caused PVC classification error:

(1) Aberrated atrial premature beats are often labeled PVC by algorithms that use only morphology and prematurity to classify beats. Of the 170 false-positive PVC classifications, 66 (39%) were aberrated atrial premature beats. Sixty-three of the improperly classified aberrated beats were on tape 201, all of which have a morphology distinctly different from NORMALs, and are consistently premature. An example is shown in Figure 10-6a.

(2) Fusion beats often exhibit morphologies that are "between" morphologies for NORMALs and PVCs, and can cause clusters to drift, or be improperly labeled. Tape 213 has many examples of fusion beats that are almost indistinguishable from normals (for a single channel analysis algorithm), but are somewhat more distinguishable on the second ECG channel. An example is shown in Figure 10-6b.

(3) Wide normal beats (paced or bundle branch block) tend to reduce PVC sensitivity, since the family classifier decision regions "close in" more slowly for wide NORMAL beats. Thus, it may take several instances of the same morphology (assuming it's different from the NORMAL, and premature) before it is labeled PVC. An example is shown in Figure 10-6c for tape 107, which has wide paced "normal" beats. The first few times the PVC is seen (at 13:05), it is declared NORMAL or WIDE until it is finally labeled PVC (at 13:20).

(4) ECG records with substantial morphologic variability can rapidly fill the class catalogue with many templates, and prevent them from aquiring sufficient class population so that they are correctly labeled. In other cases, the morphology for NORMAL beats can overlap the morphology of PVCs, and dramatically increase classifier error rates. The dreaded tape 203 has many examples of this, one of which is shown in Figure 10-6d.

(5) Another ECG record for which the classifier missed many PVCs was tape 223. The poor PVC sensitivity was due to several long runs of bidirectional ventricular tachycardia. In this case, the earlier ectopic beat was properly labeled PVC, but the later one was incorrectly labeled NORMAL. An example is shown in Figure 10-6e.

(6) A clinically important aspect of algorithm performance is its ability to detect runs of ventricular flutter and fibrillation. The onset of the run of ventricular flutter is shown in Figure 10-6f, and indicates that the run was successfully detected. However, it also illustrates how the event detector "flips" between labeling positive or negative deflections, and can cause errors in computing the rate of the run. In Figure 10-6g, the ventricular flutter briefly degenerates into ventricular fibrillation, which was of such low amplitude that it was not detected, and would normally trigger an "asystole" alarm. Towards the end of Figure 10-6g, the amplitude of the flutter again decreases to a point where it is briefly missed by the event detector. The run terminates with a salvo of three PVCs in

Figure 10-6h. Some of the beats in the run were labeled ABERR (WIDE) by the algorithm because they did not pass the overlapping triangular morph test, which bypassed the labeling based on prematurity and width.

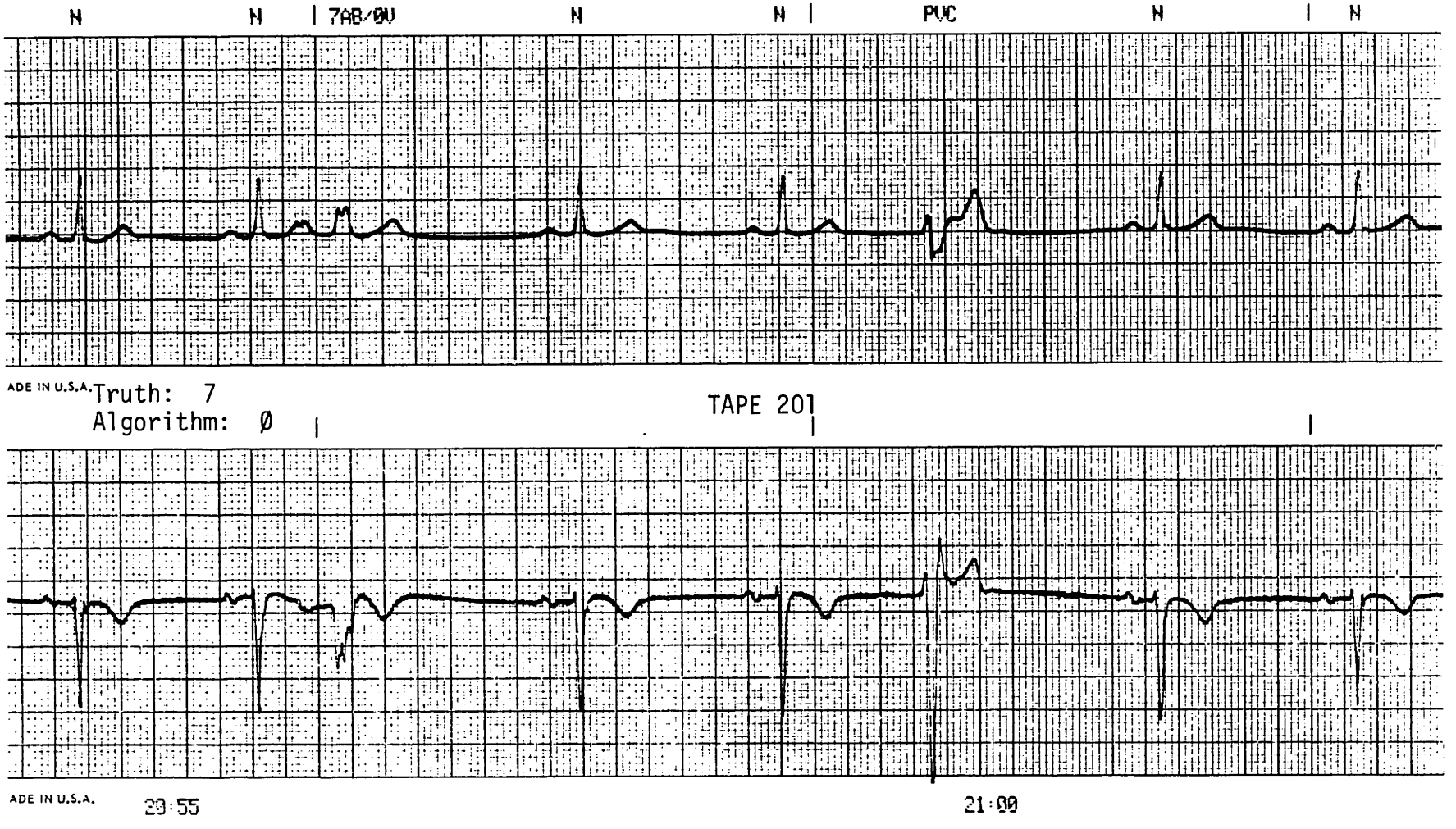



Figure 10-6a: Tape 201 - Aberrated Atrial Premature Beats

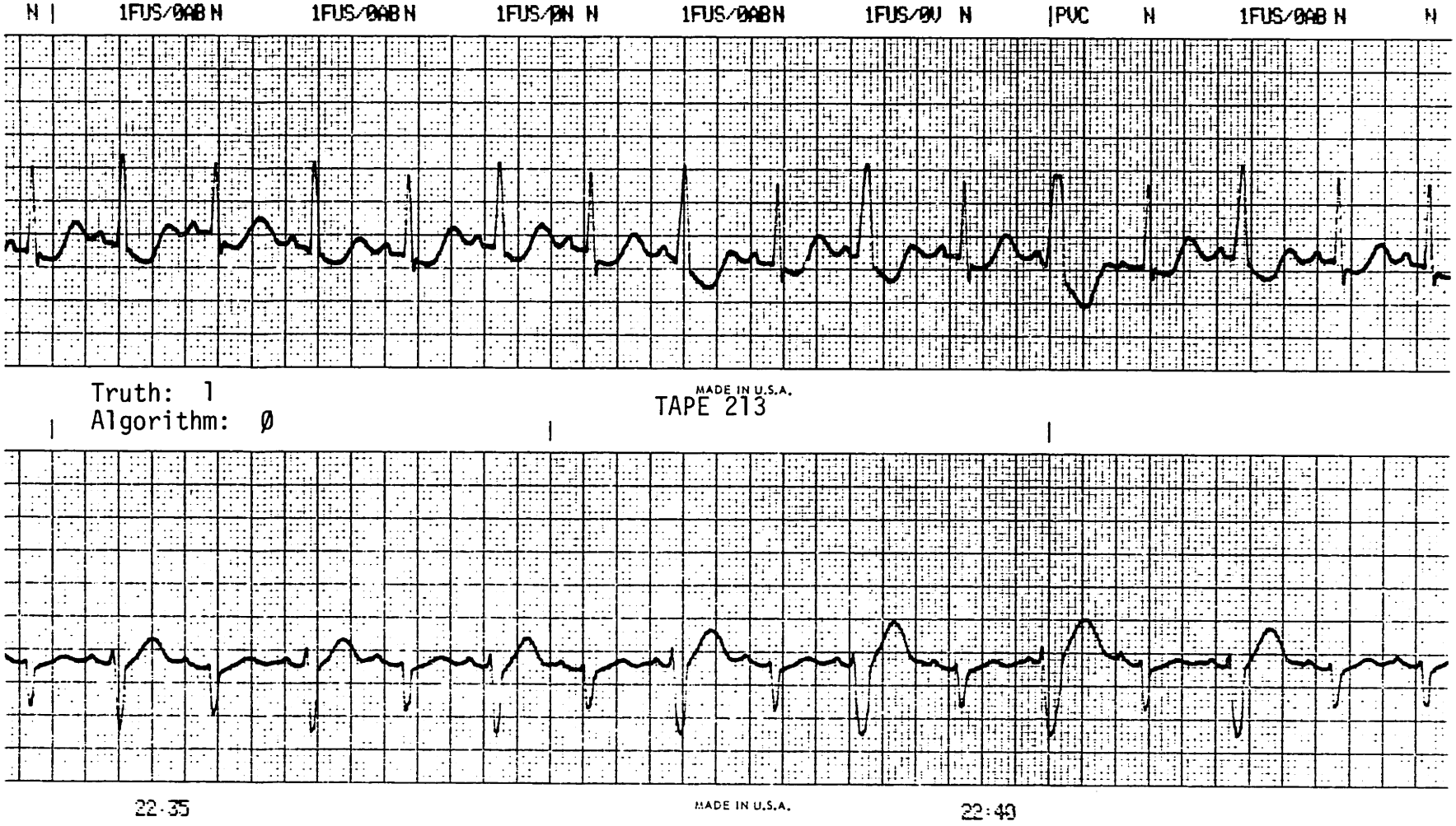



Figure 10-6b: Tape 213 - Fusion Beats

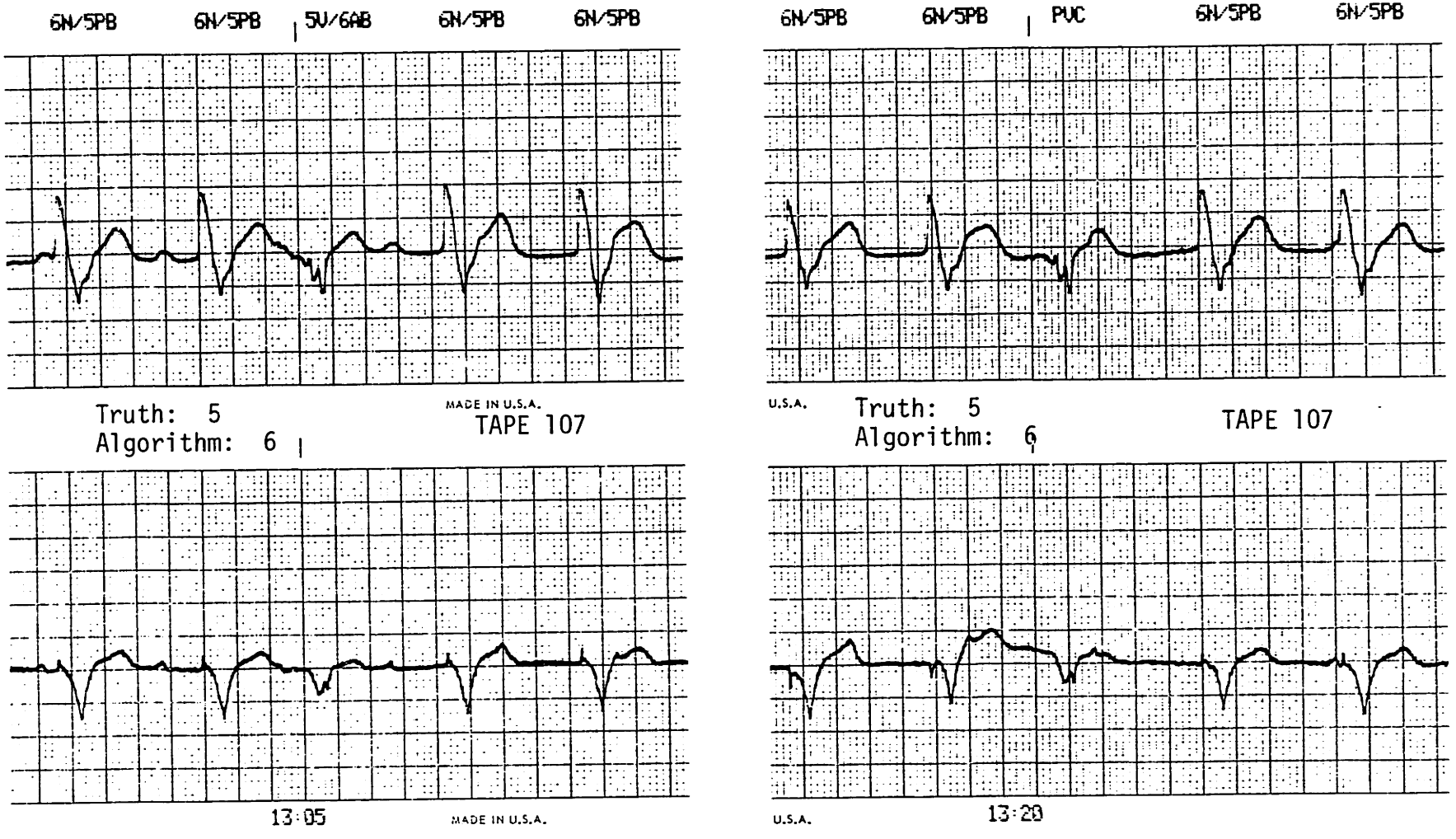



Figure 10-6c: Tape 107 - Wide Normal Beats

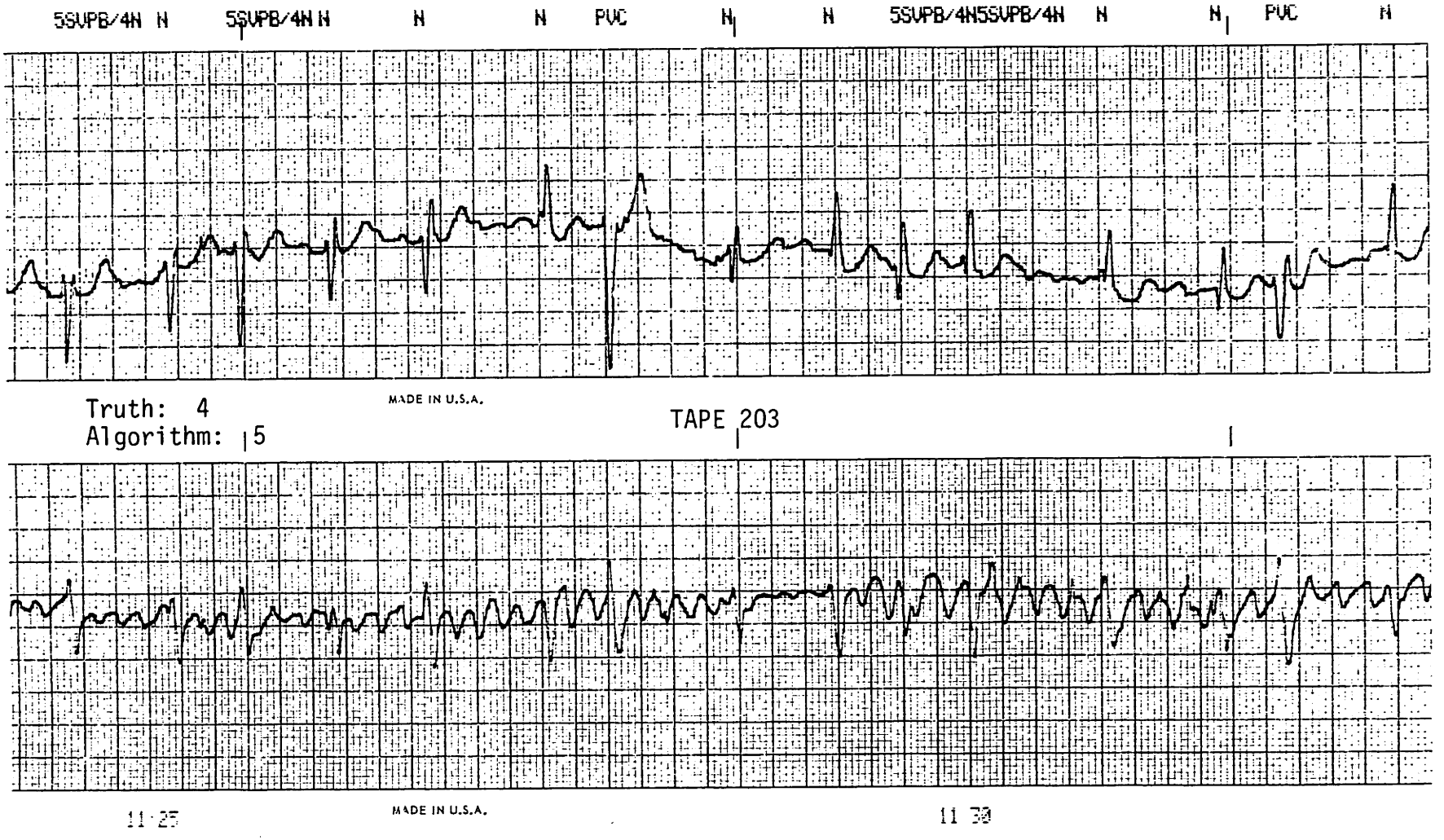



Figure 10-6d: Tape 203 - Variable NORMAL Morphology

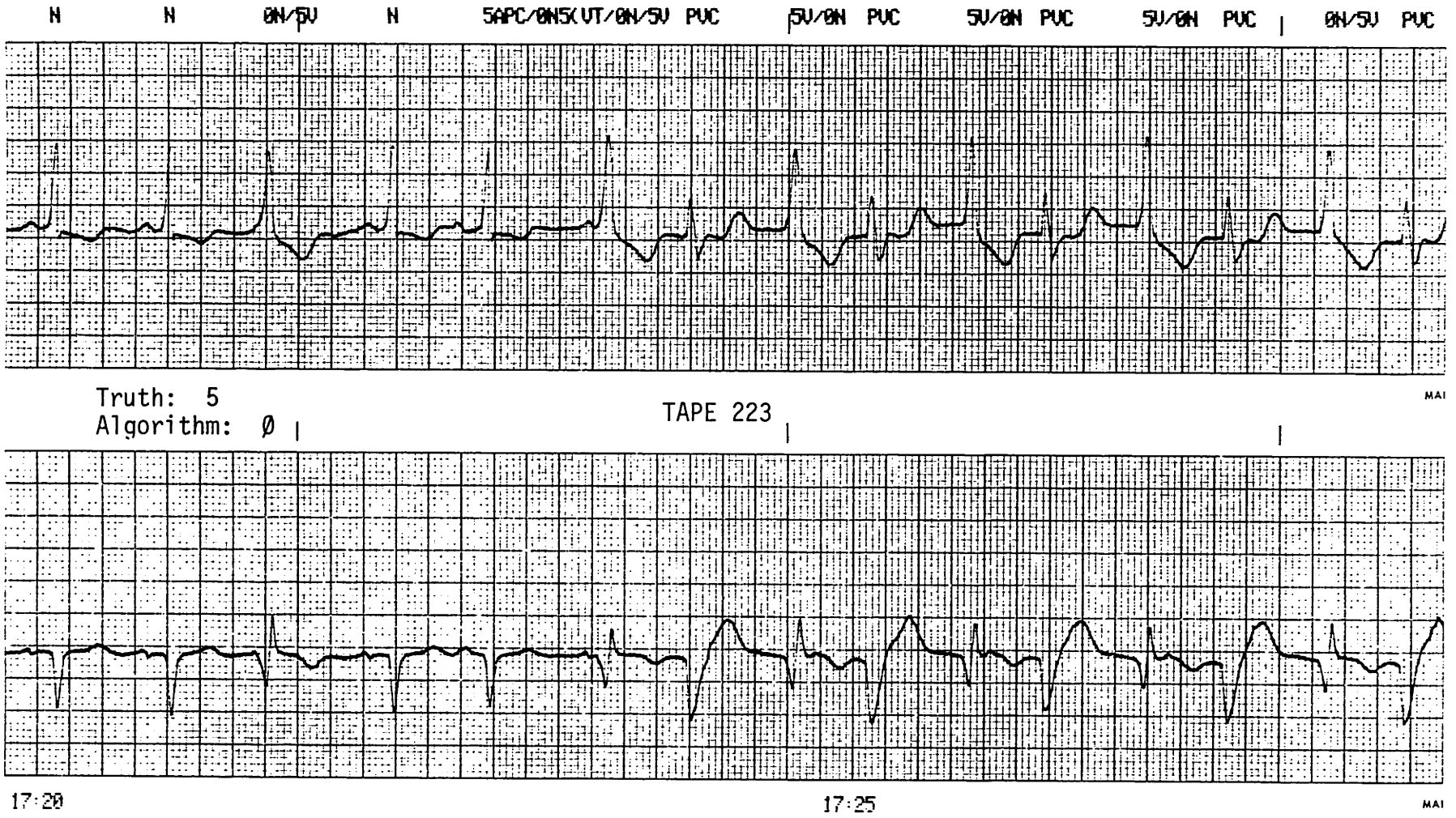



Figure 10-6e: Tape 223 - Bidirectional Ventricular Tachycardia

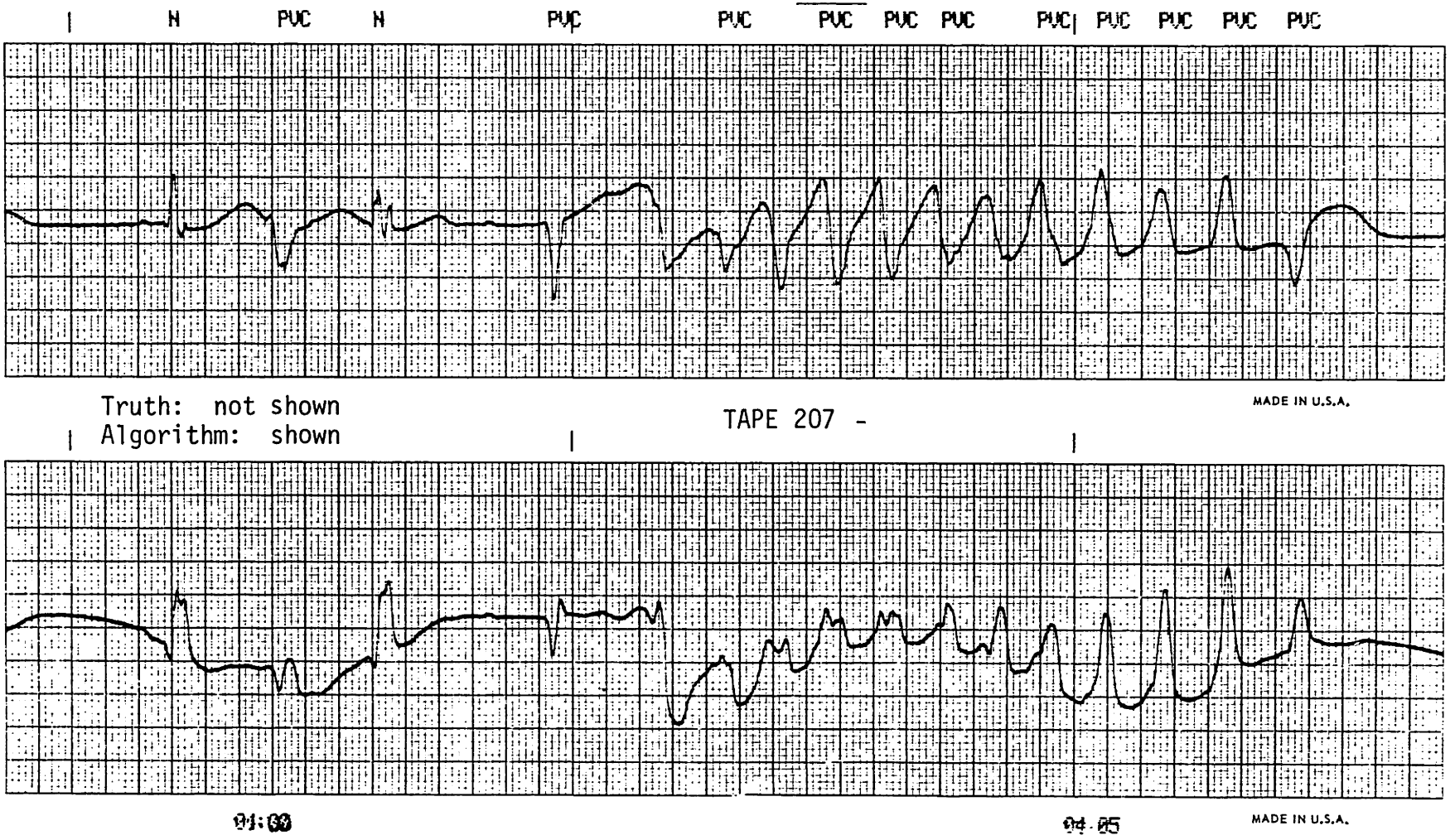



Figure 10-6f: Tape 207 - Onset of Ventricular Flutter

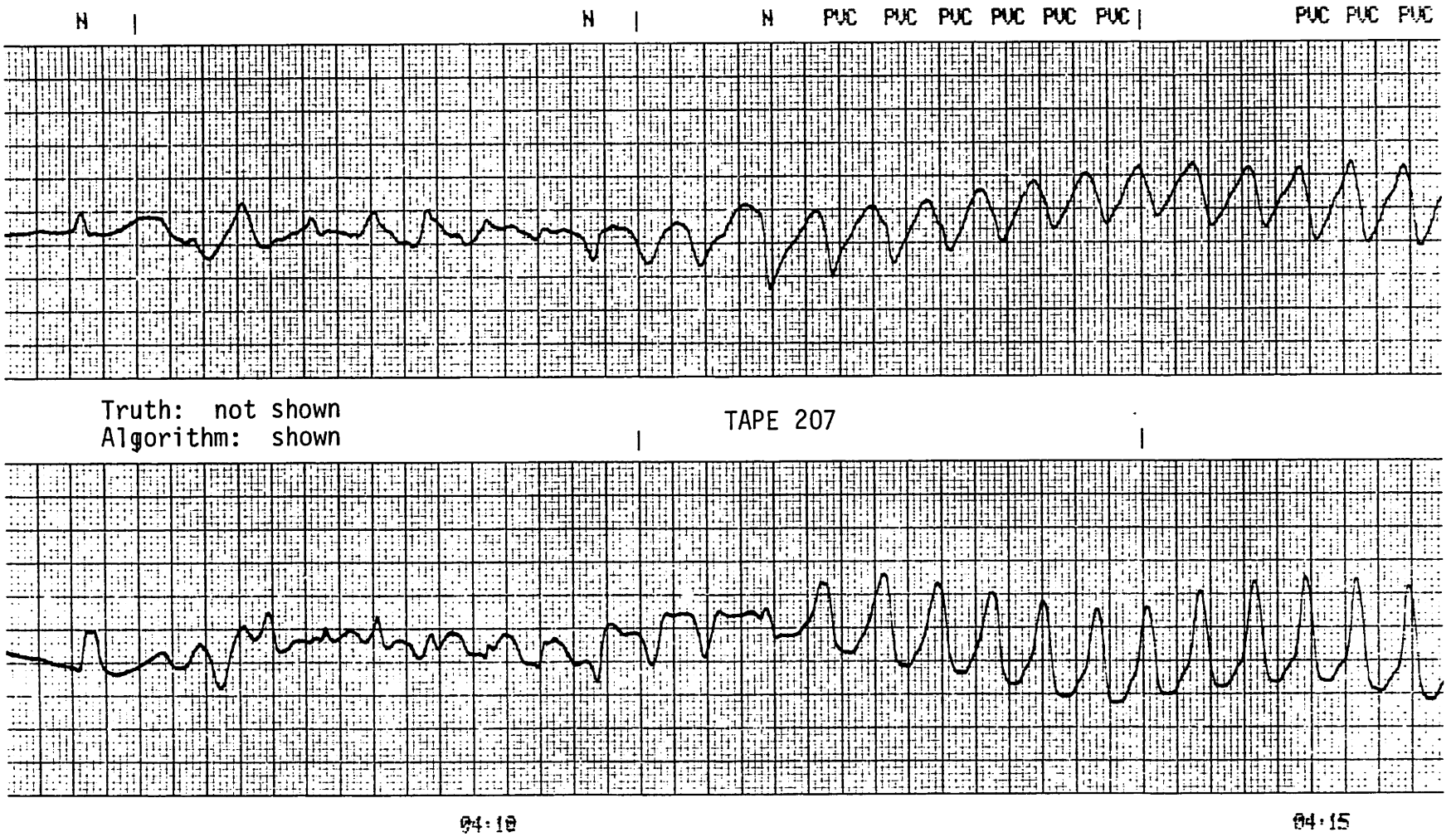



Figure 10-6g: Tape 207 - Ventricular Fibrillation

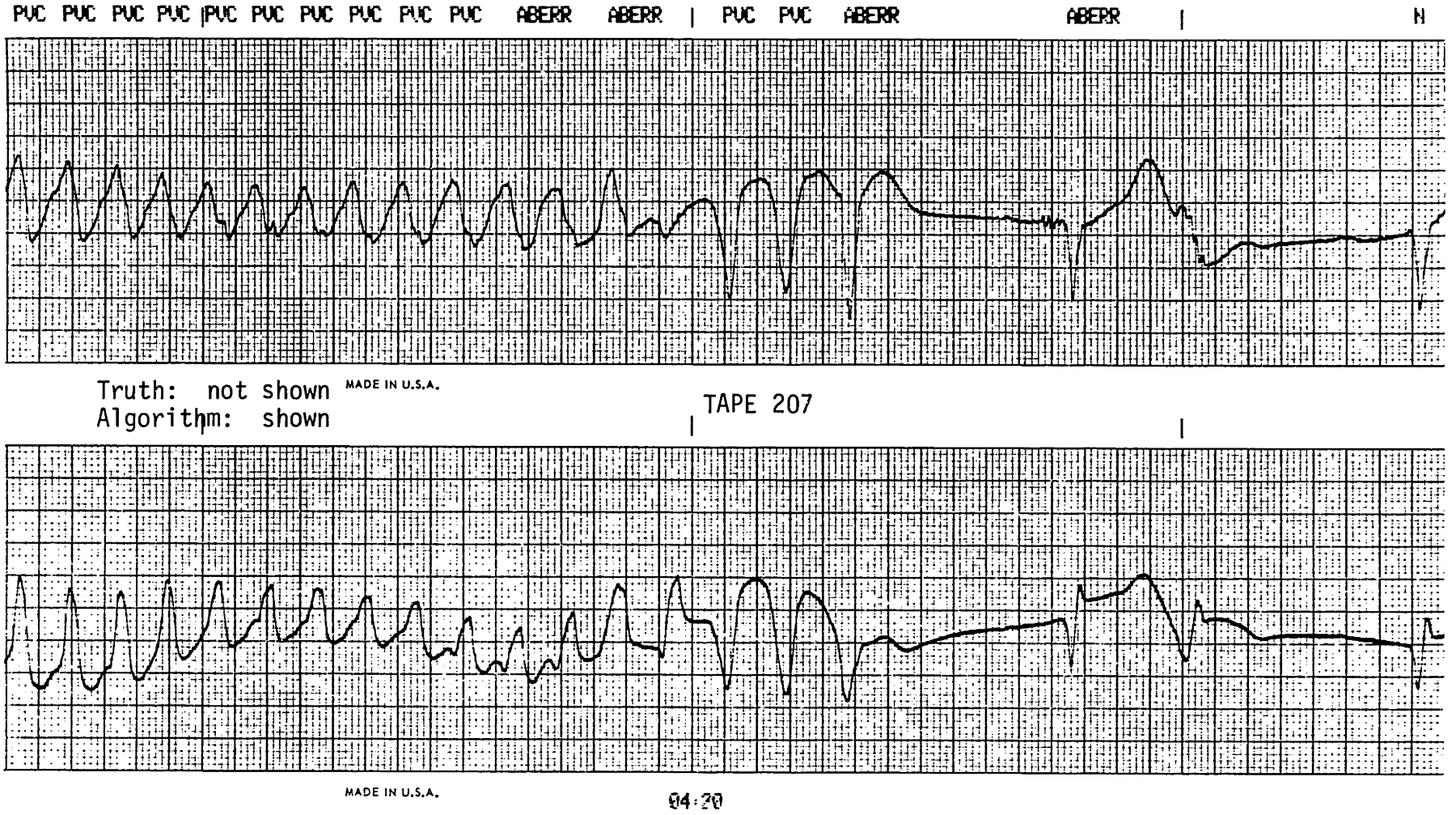



Figure 10-6h: Tape 207 - Termination of Ventricular Flutter

## 10.10 Comparison With Other Algorithms

One logical question to ask is "how does the bedside monitor compare with other systems?". The question is difficult to answer, since performance is highly dependent on the input data base, making it hard to confidently compare algorithms. At the time of this writing, there was no other publically available data base that could be used for an evaluation, and for which there were published results. Hopefully this situation will be remedied with the availability of the American Heart Association data base, and to a lesser extent, with the availability of the MIT-BIH data base.

### Why intersystem comparison is difficult

The primary reason for the difficulty in comparing algorithm performance is the lack of a widely accepted "golden standard" ECG data base with machine readable annotations that would permit automatic beat-by-beat evaluations. Most evaluations were performed with locally aquired data that is difficult for others to obtain. In other instances, the data is available only in analog form, and thus it is difficult to perform reproducable experiments, or make high-quality copies of the original data.

Evaluations performed with local data bases suffer from other flaws as well. The test data set is often too small, and does not expose the system to a sufficiently wide variety of background rhythms and morphology. This is especially acute where the same data base is used for both development and

evaluation, since it would allow the system developer to "tune" his algorithm to the data set, making it possible to bias the detector and classifier for optimal performance. A large data base would make this much more difficult to do.

Another difficulty arises from the different evaluation protocols and performance measures that have been used. The evaluation raw data is often not published, making it difficult to derive alternate performance measures. The system evaluation may depend on the skills of a trained technician or investigator to adjust algorithm parameters, thus tailoring the algorithm to each patient's waveform. In clinical trials, extraordinary care in electrode placement (to minimize artifact, or enhance the morphologic difference betweeen PVCs and normals) can be used to tailor the waveform to the algorithm. In either case, it leads to an evaluation of the combined performance of the technician and system, and makes it impossible to tell whether another technician could achieve comparable results.

Performance figures for other systems

Evaluation results for several ECG analysis systems are shown in Table 10-8. The ARGUS system (feature extraction) developed at Washington University was evaluated using a data base of approximately 50,000 QRS complexes taken from 39 patients. Of the 3880 VPBs identified by the cardiologist, 3045 (78.4%) were correctly detected by the computer. There were 835 PVCs incorrectly labeled for a false negative rate of 21.6%, and 174 non-PVCs were incorrectly labeled PVC giving a false positive

rate of 0.4% of total beats. Very good performance for the Electrodyne (correlation based) multipatient monitoring system is indicated in Table 10-8b. Evaluation results for other systems are summarized by Ripley (Ripley,1975).

The evaluation results for the bedside monitor are shown in Table 10-8c, and indicate performance similar to the ARGUS monitoring system. However, it is difficult to say which approach really works the best, since all three systems were evaluated on different data bases. Both the ARGUS and Electrodyne systems were evaluated in a real-time clinical setting, wherease the bedside monitor was evaluated using an annotated digital ECG data base.

Table 10-9: Performance Results for Other Systems

| | | Algorithm | |
|---|---|---|---|
| | | $\overline{PVC}$ | PVC |
| Truth | $\overline{PVC}$ | 43621 (99.6%) | 174 (0.4%) |
| | PVC | 835 (21.6%) | 3045 (78.4%) |

Table 10-9a: Original ARGUS (Ripley,1975)

| | | Algorithm | |
|---|---|---|---|
| | | $\overline{PVC}$ | PVC |
| Truth | $\overline{PVC}$ | 140780 (99.9%) | 146 (0.1%) |
| | PVC | 262 (4.6%) | 5467 (95.4%) |

Table 10-9b: Electrodyne System (Shah and Arnold,1977)

| | | Algorithm | |
|---|---|---|---|
| | | $\overline{PVC}$ | PVC |
| Truth | $\overline{PVC}$ | 98458 (99.7%) | 170+116 (0.3%) |
| | PVC | 1046+83 (15.9%) | 5973 (84.1%) |

Table 10-9c: Bedside Monitor

## 10.11 Discussion

The algorithm's overall performance on the data base was satisfactory. Event detector sensitivity was 99.43%, and positive predictivity was 99.72%. PVC detection sensitivity was 84.1%, and positive predictivity was 95.4%. SVPB detection performance was considerably less accurate, and suggests that a fixed 15% prematurity threshold would be superior to the adaptive prematurity threshold currently used by the algorithm. Detection sensitivity for isolated PVCs was 88%, couplets 77%, short runs (3 to 5 beats) 45%, and longer runs 69% suggests that specialized pattern recognition algorithms can improve detection accuracy for ventricular tachycardia and flutter.

The annotated ECG data base proved to be an extremely useful tool both for algorithm development and evaluation. As a development aid, the data base provided an easy-to-use source of ECG data that could be played back at any speed (or even stopped) to allow detailed observation of the algorithm. Algorithm annotations could be automatically compared with truth annotations in real-time, and if there were a disagreement, special diagnostic routines could be executed while internal algorithm parameters were still available (something difficult to due with a retrospective evaluation). Repeatable experiments were possible, permitting comparison of different draft versions of the algorithm.

The annotated ECG data base was, of course, unquestionably useful for performing beat-by-beat evaluations of the algorithm. It permitted automatic measurement of detector sensitivity and specificity for QRS complexes, PVC's, SVPB's, couplets, ventricular runs, etc. In addition to providing quantitative results, hardcopy rhythm strips including both the truth and algorithm annotations could be made for detailed review. The digital data base also provides a basis for intersystem comparison, and to this date over 14 other organizations have purchased the MIT-BIH data base.

One criticism of our use of the data base was that is was used both for algorithm development and evaluation. Thus, there exists the possibility that the algorithm is "tuned" to the data base, and would perform less well on other data sets. This is certainly a valid point, BUT, the development of the data base was a very time and resource consuming project, and we simply felt that we couldn't afford to produce separate development and evaluation data sets. The imminent availability of the AHA data base also made this a mute point. Furthermore, it is the author's opinion that the MIT-BIH database provides a sufficiently wide variety of background rhythms and morphologies that it is very difficult to tune an algorithm to perform well on all tapes. Increasing the sensitivity for a few tapes will inevitably increase the false positive rate for others, and is a judgment that should be guided by clinical experience as well as by laboratory evaluation results.

## 11. LONG-TERM ANALOG TAPE EVALUATION

The principal quantitative evaluation of algorithm performance was performed using the annotated digital ECG data base. However, it was not at all clear that 48 half-hour recordings were an adequate representation of the "universe" of ECG waveforms that the system might encounter in the clinical environment. Also, the beat-by-beat evaluation did not test other functions such as the alarm and event documentation algorithms that also govern what the clinician sees, and play a critical role in overall instrument performance.

Thus, the second phase of the evaluation of the bedside monitor used eight hour segments of long term Holter (analog) recordings as the test data. Reports produced by the bedside monitor were compared with those previously prepared by the Holter tape scanning technicians on an hour by hour basis by noting the severest level of ectopic activity documented by each technique. Although this procedure is not as precise as a beat-by-beat evaluation, it expands the scope of the evaluation in several areas:

(1) it exposes the algorithm to far more data than is available on the annotated ECG data base,

(2) it compares the bedside monitor with a widely used, clinically accepted technique (Holter tape scanning),

(3) it exercises the entire arrhythmia monitor system - the event detector, classifier, and episode documentation strategy, and

(4) it may reveal long-term aspects of algorithm behavior (if

any) that may not be apparent when using short test records.

Selection of Analog Tapes

Forty analog tapes were selected from a library of Holter tape recordings obtained by the Arrhythmia Laboratory of the Beth Israel Hospital in Boston, Massachusetts. The data was selected from three categories that were considered important for testing. The categories and number of selected tapes were:

1) 13 tapes with "mild" ventricular ectopic activity (PVCs occurring singly at any rate)

2) 21 tapes with "severe" ventricular ectopic activity (couplets and runs), and

3) 6 tapes with a background rhythm of atrial fibrillation (with both levels of ventricular ectopic activity)

The tapes were randomly selected within each category, but were discarded if excessively noisy or otherwise unreadable.

Comparison Protocol

A six to eight hour excerpt from each tape was played into the bedside monitor at real-time. In addition, a hard copy recording of one of the hours was also produced on a strip chart recorder. The hour strip was later reviewed by hand, and served as a "gold standard" to evaluate the absolute performance of both the Holter technicians and the bedside monitor. At the conclusion of each tape trial, the bedside monitor log report and all strip chart recordings were saved.

The bedside monitor and Holter reports were compared by noting the severest level of ventricular ectopic activity detected in each hour by each technique. The following "grades" of ectopic activity were used:

(NONE) no ventricular ectopic activity,

(PVC) occasional single PVCs (less than 30 per hour)

(FPVCS) frequent single PVCs (more than 30 per hour)

(CPLTS) couplets, and

(RUNS) runs of 3 or more PVCs.

Although a comparison based on only these five grades is relatively crude compared to the more quantitative results possible with a beat-by-beat evaluation, the level of severest ectopic activity is a clinically important piece of information. Furthermore, it is easily derived from both the Holter and bedside monitor reports, making it feasible to compare many hours of data. The comparison by "grade" is suitable for an alarm documentation strategy that records the severest rhythm(s) that occur within each hour. This method of comparison has also been used to compare Holter tape scanning techniques (Stein,1980).

The bedside monitor reports were also reviewed for false positive errors, which permitted the Holter reports to be compared with:

(1) the "uncorrected" raw reports from the bedside monitor, which would reflect the error rates if the findings of the bedside monitor were used without review (e.g. relying only on trend graphs without verification of significant episodes), and

(2) the "corrected" bedside monitor reports, where human review permitted the exclusion of all false positives (the usual mode of clinical operation).

11.1 Hour-by-hour comparison

Table 11-1a compares the uncorrected data from the bedside monitor with the findings of the Holter laboratory in confusion matrix form. It is hard to draw any conclusions from this table alone, since both false negatives and positives can cancel. Table 11-1b shows the comparison results for the corrected bedside monitor reports, for which all documented episodes were reviewed and corrected, all all false positives eliminated. This table indicates that for 270 hours of test data, there were 16 hours in which the bedside monitor found a more severe episode than the Holter technique, and 69 hours where the Holter technique found an episode more severe than the bedside monitor. [A note: since not all bedside monitor hours were correctable (because the monitor's strip chart recorder ran out of paper), there are slight differences between the total number of hours in otherwise similar comparisons. For example, the uncorrected comparison in Table 11-1a has a total of 272 hours, and the corrected comparison in Table 11-1b has a total of 270 hours.]

The data in Table 11-1b suggests that the Holter technique is more sensitive than the bedside monitor. Are the differences (69 versus 16 hours) statistically significant, or could they be due to random chance? The Wilcoxon signed rank test for paired

replicates (Lehman,1975) was used to test the statistical significance of the data. If ties (hours where there was agreement) are randomly assigned, we obtain:

BAMC > HOLTER: 16 + 185/2 = 109.5

BAMC < HOLTER: 69 + 185/2 = 161.5

Determining whether the difference is statistically significant is similar to asking whether, out of 270 tosses of a coin, would 161 heads come up? The expected value for the number of heads (H) or tails is:

E(H) = N/2 = 270/2 = 135

and the variance is

Var(H) = N/4 = 270/4 = 67.5

For large N, the distribution of

$$\frac{H - N/2}{1/2*\mathrm{sqrt}(N)}$$

tends to the standard normal distribution.

Thus, the probability

$$P(H>161.5) = P\left( \frac{H - N/2}{1/2*\mathrm{sqrt}(N)} > \frac{161.5 - 135}{1/2*\mathrm{sqrt}(270)} \right)$$

$$= Q\ (\ 3.225\ ) = 0.007$$

Thus, the Wilcoxon test shows that the differences between the Holter analysis and the bedside monitor analysis are statistically significant ($p < 0.007$), and that the Holter analysis was more sensitive in detecting ventricular ectopic events.

A derived "truth grade" for each hour

Since all false positives have been removed from both the corrected bedside monitor and Holter reports, the severest episode noted by either technique can be used as the "truth" (assuming both techniques did not miss the same episode). The results for this comparison are shown in Table 11-2a for the uncorrected bedside monitor report, and for the corrected bedside report in Table 11-2b. False positives for the uncorrected bedside monitor report were noted in 100 hours. After correction, all false positives were removed, and in 29 hours (69-40) reports were corrected downward (in severity).

Table 11-1a: Uncorrected Bedside Monitor vs Holter

| | | BAMU | | | | |
|---|---|---|---|---|---|---|
| | | NONE | PVCS | FPVCS | CPLTS | RUNS |
| HOLTER | NONE | 10 | 26 | 2 | 6 | 4 |
| | PVCS | 7 | 35 | 4 | 13 | 12 |
| | FPVCS | - | 6 | 20 | 18 | 8 |
| | CPLTS | 1 | 8 | 6 | 45 | 18 |
| | RUNS | 1 | 1 | - | 9 | 12 |

BAMU > HOLTER: 111 (40%)

BAMU = HOLTER: 122 (44%)

BAMU < HOLTER: 39 (14%)

Table 11-1b: Corrected Bedside Monitor vs Holter

| | | BAMC | | | | |
|---|---|---|---|---|---|---|
| | | NONE | PVCS | FPVCS | CPLTS | RUNS |
| HOLTER | NONE | 41 | 5 | 1 | - | 1 |
| | PVCS | 17 | 52 | 2 | - | - |
| | FPVCS | - | 14 | 32 | 4 | - |
| | CPLTS | 2 | 10 | 16 | 47 | 3 |
| | RUNS | 2 | - | 2 | 6 | 13 |

BAMC > HOLTER: 16 (5%)

BAMC = HOLTER: 185 (68%)

BAMC < HOLTER: 69 (25%)

Table 11-2a: Uncorrected BAM versus BAMC+HOLTER

| | | BAMU | | | | |
|---|---|---|---|---|---|---|
| | | NONE | PVCS | FPVCS | CPLTS | RUNS |
| BAMC+HOLT | NONE | 10 | 23 | 1 | 4 | 3 |
| | PVCS | 7 | 37 | 3 | 15 | 12 |
| | FPVCS | - | 6 | 21 | 16 | 6 |
| | CPLTS | 1 | 8 | 6 | 47 | 17 |
| | RUNS | 1 | 2 | - | 9 | 15 |

BAMU > BC+H: 100 (37%)

BAMU = BC+H: 130 (48%)

BAMU < BC+H: 40 (14%)

Table 11-2a: Corrected BAM versus BAMC+HOLTER

| | | BAMC | | | | |
|---|---|---|---|---|---|---|
| | | NONE | PVCS | FPVCS | CPLTS | RUNS |
| BAMC+HOLTER | NONE | 41 | - | - | - | - |
| | PVCS | 17 | 57 | - | - | - |
| | FPVCS | - | 14 | 35 | - | - |
| | CPLTS | 2 | 10 | 16 | 51 | - |
| | RUNS | 2 | - | 2 | 6 | 17 |

BAMC > BC+H: 0 (0%)

BAMC = BC+H: 201 (74%)

BAMC < BC+H: 69 (25%)

Verification using the Gold Hour

As an additional check on both the bedside monitor and Holter technique, a randomly selected hour from each trial was recorded on a strip chart, and was carefully reviewed by hand. This hour served as a "gold standard" to evaluate the absolute performance of both the Holter technicians and the bedside monitor. The results of this comparision are shown in Table 11-3a for the corrected bedside monitor report, and in Table 11-3b for the Holter scanning technique. These results also indicate that the Holter technique was more sensitive, and in fact was in almost perfect agreement with the carefully reviewed hours. This result suggests that the Holter technique (as practiced at the BIH Arrhythmia Laboratory) can be used as a reliable reference for comparison. The bedside monitor, on the other hand, reported less severe episodes for 15 (38%) of the hours, which can be contrasted with the 25% missed episodes for the entire 270 hours in Table 11-1b.

Table 11-3a: Corrected Bedside Monitor versus Gold Hour

| | | BAMC | | | | |
|---|---|---|---|---|---|---|
| | | NONE | PVCS | FPVCS | CPLTS | RUNS |
| GOLD HOUR | NONE | 4 | - | - | - | - |
| | PVCS | 5 | 4 | - | - | - |
| | FPVCS | - | - | 3 | - | - |
| | CPLTS | - | 1 | 2 | 6 | - |
| | RUNS | 1 | - | 3 | 3 | 7 |

BAMC > GOLD: 0 (0%)

BAMC = GOLD: 24 (61%)

BAMC < GOLD: 15 (38%)

Table 11-3b: Holter versus Gold Hour

| | | HOLTER | | | | |
|---|---|---|---|---|---|---|
| | | NONE | PVCS | FPVCS | CPLTS | RUNS |
| GOLD HOUR | NONE | 4 | - | - | - | - |
| | PVCS | - | 9 | - | - | - |
| | FPVCS | - | - | 3 | - | - |
| | CPLTS | - | - | - | 10 | - |
| | RUNS | - | - | - | 1 | 13 |

HOLTER > GOLD: 0 (0%)

HOLTER = GOLD: 39 (97%)

HOLTER < GOLD: 1 (2%)

Hour-by-hour comparison with the annotated data base

A similar comparison can be performed using the annotated ECG data base as the "gold standard", and noting the severest rhythms documented by both the bedside monitor and the "truth" annotation channel of the data base. The results for this comparison are shown in Table 11-4a for the uncorrected bedside monitor, and in Table 11-4b for the corrected bedside monitor reports. Comparing the corrected bedside monitor results for the entire Holter tape series (Table 11-2b), the "gold hours" (Table 11-3a), and the data base (Table 11-4b) show alarm episode false negative rates of 25%, 38%, and 16% respectively. This suggests that the annotated data base was actually somewhat easier than the set of Holter tapes played into the bedside monitor. Another interpretation is that the algorithm was better "tuned" to the data base than the Holter tapes used for the analog tape evaluation.

Table 11-4a: Uncorrected Bedside Monitor versus Data Base

| | | ALGU | | | | |
|---|---|---|---|---|---|---|
| | | NONE | PVCS | FPVCS | CPLTS | RUNS |
| DATA BASE | NONE | 6 | 2 | 2 | 1 | - |
| | PVCS | 2 | 7 | - | - | - |
| | FPVCS | - | 3 | 3 | 2 | 1 |
| | CPLTS | - | - | 1 | 3 | 1 |
| | RUNS | - | - | - | 2 | 12 |

ALGU > DBASE: 9 (18%)

ALGU = DBASE: 31 (64%)

ALGU < DBASE: 8 (16%)

Table 11-4b: Corrected Bedside Monitor versus Data Base

| | | ALGC | | | | |
|---|---|---|---|---|---|---|
| | | NONE | PVCS | FPVCS | CPLTS | RUNS |
| DATA BASE | NONE | 11 | - | - | - | - |
| | PVCS | 3 | 6 | - | - | - |
| | FPVCS | - | 3 | 6 | - | - |
| | CPLTS | - | - | - | 5 | - |
| | RUNS | - | - | - | 2 | 12 |

ALGC > DBASE: 0 (0%)

ALGC = DBASE: 40 (83%)

ALGC < DBASE: 8 (16%)

## 11.2 Episode Documentation

Another important aspect of the bedside monitor is the number of arrhythmia episodes it documents per hour, and the relative fraction of correct and incorrect strips. If too few strips are made, and many are incorrect, the clinician's confidence in the monitor is reduced; and if too many strips are made, the clinican may be less likely to review them. The ideal, based on our experience, seems to be five to ten strips per hour. In most monitoring situations, this number provides enough data to verify the bedside summary reports, but does not burden the clinician with redundant strips.

The bedside monitor log and strip chart recordings were carefully reviewed after each tape trial. Each incorrect strip (e.g. a falsely detected PVC due to artifact) was noted on the log report. The corrected log was then used as input to a program that tabulated the number of strips produced each hour, and the real (correct) reason for the strip. Table 11-5 lists the strip titles produced by the bedside monitor. Listed also are other "true" reasons for strips, such as BASE (baseline artifact), LOS (loss of signal), and other conditions that can initiate an "incorrect" strip chart recording.

The total results for the entire series of Holter tapes is shown in Table 11-6. A total of 1732 strips were produced during the entire 248 hour period. Of these, 867 (50%) were correct, and 865 (50%) were incorrect.

Not surprisingly, artifact (which includes BASE, NOISE, 60HZ, ART, I-ART, T+ART, V+ART, N+ART, and P-ART) was the most common single cause of false positive strips, accounting for 238 (28%) of all incorrect strips. It is interesting to note that high frequency noise (NOISE and 60HZ) did not cause any problems, and that almost all errors were due to rapid baseline shifts and other low-frequency artifact. This observation suggests that an artifact rejection strategy based on morphology, rather than aggressive signal filtering, will have the best chance to reduce false positive episodes due to artifact.

Of the 348 "missed" beat episodes falsely declared by the bedside monitor, (40% of all incorrect strips) 139 were due to pauses during atrial fibrillation. Although one cannot fault the monitor for declaring long pauses "missed beats", they are superfluous alarms during atrial fibrillation. This finding suggests a future modification that imposes stricter requirements for declaring a missed beat during atrial fibrillation. Event detector error (missed low amplitude PVCs and NORMALs) accounted for only 66 of the missed beat episodes. Other errors were due to loss of signal or amplifier saturation that was not long enough to be recognized as a poor signal (NSIG) condition.

Classifier error accounted for most of the remaining mistakes. There were 61 instances where isolated PVCs were labeled as couplets, which often happened when the trailing edge of the T-wave of interpolated PVCs was included in the morphologic description of the normal beat that followed. The

corrupted normals would become a new class that was premature, and sufficiently different from the true normal to be labeled as PVCs. T-waves themselves were falsely labeled as isolated PVCs in 18 instances.

Table 11-5: Bedside Monitor Documentation Episodes

Bedside Monitor Episode Categories

| | |
|---|---|
| MISS | missed beat. |
| PVC | isolated PVC. |
| TRI | trigeminy. |
| BI | bigeminy. |
| SVTA | run of SVTA. |
| CPLT | ventricular couplet. |
| AIVR | slow VT. |
| VT | ventricular tachy (rate > 100). |
| TCHY | tachycardia. |
| BRDY | bradycardia. |
| NSIG | prolonged poor signal. |
| ASYS | asystole. |

Other (actual) reasons for generating a strip recording.

| | |
|---|---|
| N | Normal beat (called something else). |
| SVPB | Supraventricular premature beat. |
| ABERR | Aberrated atrial premature beat. |
| T | T-wave (falsely detected). |
| V+T | PVC plus T-wave (falsely called couplet) |
| BASE | Baseline shift. |
| NOISE | High frequency noise. |
| 60HZ | Powerline interference. |
| ART | Artifact (falsely detected). |
| I-ART | Iterpolated artifact. |
| T+ART | T-wave corrupted by artifact. |
| V+ART | PVC and artifact (detected as CPLT or VT) |
| N+ART | Normal corrupted by artifact. |
| P-ART | Post-artifact error. |
| SLOW | Non-abrupt heart rate slowdown. |
| AFPSE | Pause due to AF. |
| V-PSE | (Compensatory) pause after a PVC. |
| S-PSE | Pause after SVPB or run of SVTA. |
| LOAMP | Low amplitude beat. |
| LOS | Loss of signal. |
| OTHER | Error due to unknown causes. |

Table 11-6: Episode Documentation for Holter Tape Evaluation

| | | BEDSIDE MONITOR STRIPS | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | MISS | PVC | TRI | BI | SVTA | CPLT | AIVR | VT | TCHY | BRDY | NSIG | ASYS |
| T | MISS | 134 | - | - | - | - | - | - | - | - | - | - | - |
| R | PVC | 34 | 452 | 2 | - | 1 | 61 | 1 | 3 | - | - | - | - |
| U | TRI | - | - | 42 | - | - | - | - | - | - | - | - | - |
| T | BI | - | 1 | - | 25 | - | 2 | - | 1 | - | - | - | - |
| H | SVTA | 1 | - | - | - | 16 | 1 | - | 1 | - | - | - | - |
| | CPLT | 7 | 5 | - | - | 1 | 137 | - | 5 | - | - | 1 | - |
| | AIVR | - | 1 | - | - | - | - | 3 | - | - | - | - | - |
| | VT | 1 | 1 | - | - | - | 3 | 1 | 11 | - | - | - | - |
| | TCHY | - | - | - | - | - | - | - | - | - | - | - | - |
| | BRDY | - | - | - | - | - | - | - | - | - | 13 | - | - |
| | NSIG | - | - | - | - | - | - | - | - | - | - | 21 | - |
| | ASYS | - | - | - | - | - | - | - | - | - | - | - | 13 |
| | N | 6 | 61 | 1 | - | 3 | 34 | 16 | 5 | - | - | - | - |
| | SVPB | - | 11 | 2 | - | - | - | - | - | - | - | - | - |
| | ABERR | - | 17 | 1 | - | - | 24 | - | 10 | - | - | - | - |
| | T | 6 | 18 | 2 | 4 | - | 4 | - | - | - | - | - | - |
| | V+T | - | - | - | - | - | 11 | - | - | - | - | - | - |
| | BASE | 15 | 21 | - | - | - | 11 | - | 1 | - | - | - | - |
| | NOISE | - | - | - | - | - | - | - | - | - | - | - | - |
| | 60HZ | - | - | - | - | - | - | - | - | - | - | - | - |
| | ART | 9 | 48 | - | 3 | 3 | 28 | 2 | 11 | - | - | - | - |
| | I-ART | - | 7 | - | - | - | 1 | - | - | - | - | - | - |
| | T+ART | - | 2 | - | - | - | 1 | 1 | - | - | - | - | - |
| | V+ART | - | - | - | - | - | 5 | - | 4 | - | - | - | - |
| | N+ART | - | 6 | - | - | - | 11 | 8 | - | - | - | - | - |
| | P-ART | 40 | - | - | - | - | - | - | - | - | - | - | - |
| | SLOW | 2 | - | - | - | - | - | - | - | - | - | - | - |
| | AFPSE | 139 | - | - | - | - | - | - | - | - | - | - | - |
| | V-PSE | 6 | - | - | - | - | - | - | - | - | - | - | - |
| | S-PSE | 13 | - | - | - | - | - | - | - | - | - | - | - |
| | LOAMP | 17 | - | - | - | - | - | - | - | - | - | - | - |
| | LOS | 44 | - | - | - | - | - | - | - | - | 5 | - | 19 |
| | OTHER | 8 | 1 | - | - | 1 | - | - | - | - | 1 | - | - |

Total hours: 248
Correct strips: 867 (50%)
Incorrect strips: 865 (50%)
Total strips: 1732 7 strips/hour

Other observations

Another important question to ask is how many strips were produced per hour. This is summarized in the histogram shown in Table 11-7a, which indicates the number of hours for which a number (0 to 19) of strips were produced. The median strip production rate is six per hour, and from the data in Table 11-6, the average production rate is seven strips per hour. The histogram also indicates the relative fraction of correct and incorrect strips as a function of the total strips produced in each hour, and suggests the the fraction is independent of the total number of strips. It was initially suspected that false positive strips would account for the majority of strips in hours that had frequent strips, but this does not seem to be the case.

Another parameter that was recorded by the bedside monitor was the duration of analyzable signal in each hour. Table 11-7b shows the relative distribution of hours for which a certain amount of time was lost due to the monitor declaring a "poor signal" condition. Out of the 248 hours, 153 hours (62%) had no period of declared poor signal, and 233 hours (94%) had less than 1 minute of signal discarded due to artifact.

Table 11-7a: Distribution of Strip Production Rate

| #/hr | #hrs | +++ correct --- incorrect |
|---|---|---|
| 0 | 0 | ******* |
| 1 | 15 | +++++++------- |
| 2 | 16 | +++++++++++++--- |
| 3 | 21 | +++++++++++++++------ |
| 4 | 17 | ++++++++++------- |
| 5 | 24 | +++++++++++++++--------- |
| 6 | 27 | +++++++++++++++---------- |
| 7 | 22 | ++++++++++++--------- |
| 8 | 21 | ++++++++++++--------- |
| 9 | 19 | ++++++++++--------- |
| 10 | 12 | +++++------- |
| 11 | 17 | +++++++++-------- |
| 12 | 13 | +++++++------ |
| 13 | 8 | +++++++-- |
| 14 | 7 | ++----- |
| 15 | 2 | +- |
| 16 | 1 | + |
| 17 | 3 | +++ |
| 18 | 0 | |
| >18 | 1 | + |

Table 11-7b: Poor Signal Duration

| Duration | #hrs | histogram |
|---|---|---|
| = 0 sec | 153 | **********>> |
| < 5 sec | 30 | ****************************** |
| < 10 sec | 18 | ****************** |
| < 15 sec | 11 | ********** |
| < 30 sec | 11 | ********** |
| < 1 min | 10 | ********** |
| < 2 min | 8 | ******** |
| < 5 min | 3 | *** |
| < 10 min | 4 | **** |
| < 20 min | 0 | |
| < 30 min | 0 | |
| < 40 min | 0 | |
| < 60 min | 0 | |

## 11.3 Discussion

One conclusion that can be drawn from the Holter tape trials is that the bedside monitor is less sensitive than the Holter technique as practiced at the Arrhythmia Laboratory at Beth Israel Hospital. As shown in Table 11-1b, there were 16 hours in which the bedside monitor found a more severe episode than the Holter technique, and 69 hours where the Holter technique found an episode more severe than the bedside monitor. There was also outstanding agreement (97%) between the Holter technique and the carefully reviewed "gold hours", shown in Table 11-3b, which suggests that the BIH Holter readings are a good reference to use for this type of comparison.

One likely reason for the superior Holter performance is that human operators can operate at a much higher "sensitivity" than can the bedside monitor. There is a much lower cost associated with a "false positive" during Holter scanning, since the event can be quickly and accurately reviewed, and discarded if it isn't worth documenting. The cost of this compulsiveness may vary in different laboratories. For example, productivity pressures might increase the premium on scanner technician time and result in lower sensitivities in some commercial scanning services. On the other hand, the cost of false positives is very high for the bedside monitor, since they create an additional burden for the clinician who must review them, and they can cause the alarm processor to inhibit true episodes.

## 12. CLINICAL EVALUATION

The bedside monitor was evaluated clinically for over two years at Beth Israel Hospital in Boston, Massachusetts. Because of the instrument's portability, it was primarily used on general medical and surgical floors. During the early phases of the evaluation, patients with ventricular arrhythmias were identified through the Arrhythmia Laboratory and the hospital's cardiologists. As the bedside monitor gained exposure in the hospital, however, unsolicited requests for its use were received from the clinical staff.

Informed consent was obtained from the patients and their physicians. Initially, written consent was used, but later verbal consent was considered preferable by the hospital Committee on Clinical Investigations (on New Procedures and New Forms of Therapy).

### Bedside Monitor Setup

The bedside monitor was usually placed just outside the patient's room, so that the storage display and hardcopy plotter would be less annoying to the patient. Usually the bedside monitor was connected in parallel with existing hospital monitors. Thus, the patient remained connected to the central monitoring display, and did not require the application of an extra set of electrodes.

In general, Holter monitoring electrodes (disposable NDM Silvon silver/silver-chloride electrodes) were used, in a modified lead V1 configuration. Sometimes it was necessary to vary the electrode location in order to aquire a high-amplitude QRS complex, or to enhance the morphologic difference between normal and ectopic beats.

Daily Procedure

After the bedside monitor was attached to the patient, its operation was left to the clinical staff. The investigators' role was to prepare a daily report that summarized the patient's ectopic activity. The format and content of the report was very similar to that report prepared by the Holter technicians, and included:

(1) A physician or Holter technician summary of the patient's ectopic activity for the previous day,

(2) An hour-by-hour tabulation of heart rate, PVC rate, and numbers of couplets and runs.

(3) Sample strip chart recordings (usually 6 to 24 per day) of significant arrhythmias. All strips included in the report were "corrected" (false positives eliminated).

These reports eventually became a permanent part of the patient's medical record. The machine readable log files were also saved for subsequent analysis.

The primary purpose of the written report was to provide a daily summary of the patient's ectopic activity, and to check on the accuracy of the bedside monitor. Of course, the clinical staff had immediate access to the summary trends, statistical displays, and strip chart recordings.

## 12.1 Clinical Application of the Bedside Monitor

This section discusses some of the clinical trial results for patients who were on the bedside monitor. One interesting observation is the distribution of lengths of monitoring periods for individual patients. Data from the most recent 16 patients is shown in Table 12-1. For seven patients, the monitor was used for only one day (primarily as a screening test). Longer monitoring periods were usually to support drug titrations.

Table 12-1: Patient Connect Time to Monitor (in days)

| Days | Number of Patients | | |
|---|---|---|---|
| 1 | 7 | \|******* | 16 patients, 39 days |
| 2 | 3 | \|*** | |
| 3 | 2 | \|** | |
| 4 | 1 | \|* | |
| 5 | 2 | \|** | |
| 6 | 1 | \|* | |
| 7 | 0 | \| | |
| 8 | 0 | \| | |

Aggregate Observations

Tables 12-2 (a-e) present the overall results for 12 patients who were on the bedside monitor, and for whom a machine readable patient log file was maintained. In the 565 hours of data that were reviewed (Table 12-2a), 3576 strips were generated by the bedside monitor, implying an average strip production rate of 6.3 strips per hour. Of the total, 2595 strips (73%) were correct, and 981 (27%) strips were incorrect. Most of the incorrect strips were due to the relatively few patients that had ECG waveforms that were difficult to analyse. According to Table 12-2b, there does not appear to be any relationship between the

number of strips produced per hour, and relative fraction of correct and incorrect strips.

The relative fraction of correct strips (73%) is significantly higher than the 50% obtained during the Holter tape trials of the bedside monitor. There are two possible reasons for this: (1) the Holter tapes may have contained more complex data than that in the clinical trials, and (2) in the clinical trials, it was possible to adjust the ECG electrodes to optimize signal quality and morphologic differences between normal QRS complexes and PVCs - an advantage not possible during the Holter tape trials. One experiment that would bypass these issues would be to simultaneously monitor the patient with a Holter recorder and the bedside system.

Table 12-2c indicates the amount of time lost to a "poor signal" condition. There were 305 hours (54%) with no lost time, and 525 hours (93%) with less than one minute lost due to poor signal.

Table 12-2d shows the distribution of keyboard requests per hour that were made during patient monitoring. No keyboard interactions were made during 298 (53%) of the hours, although there were many hours where frequent use was made of the monitor's displays.

Table 12-2e shows the distribution of keyboard requests by type (the log report example shown in Figure 9-10 can be used as a key for this table). The most frequently requested display was the incoming data display (RAWDSP), which is the default display for the bedside monitor. The one-hour and three-hour PVC run-length (VS-1H and VS-3H) and PVC rate trends (VR-1H and VR-3H) were those most frequently requested by the clinical staff. The RR-interval (RR-HST) and PVC coupling interval (VC-HST) histograms were used relatively infrequently.

Table 12-2a: Episode Documentation Summary for All Patients)

| | | BEDSIDE MONITOR STRIPS | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | MISS | PVC | TRI | BI | SVTA | CPLT | AIVR | VT | TCHY | BRDY | NSIG | ASYS |
| T | MISS | 326 | - | - | - | - | - | - | - | - | - | - | - |
| R | PVC | 43 | 1098 | - | 1 | - | 146 | 2 | 15 | - | - | 4 | - |
| U | TRI | - | 1 | 48 | - | - | - | 2 | 4 | - | - | - | - |
| T | BI | - | 7 | - | 66 | - | - | 14 | 2 | - | - | - | - |
| H | SVTA | 3 | - | - | - | 3 | - | - | - | - | - | - | - |
| | CPLT | 10 | 15 | 2 | - | - | 788 | - | 47 | - | - | - | - |
| | AIVR | - | - | - | - | - | - | 3 | - | - | - | - | - |
| | VT | 4 | 2 | - | - | - | 29 | 1 | 145 | - | - | - | - |
| | TCHY | - | - | - | - | - | - | - | - | - | - | - | - |
| | BRDY | - | - | - | - | - | - | - | - | - | 1 | - | - |
| | NSIG | - | - | - | - | - | - | - | - | - | 1 | 72 | 1 |
| | ASYS | - | - | - | - | - | - | - | - | - | - | - | 45 |
| | N | 32 | 22 | - | - | 2 | 60 | 34 | 28 | - | - | 2 | 1 |
| | SVPB | - | - | - | - | - | - | - | - | - | - | - | - |
| | ABERR | - | - | - | - | - | 1 | 3 | - | - | - | - | - |
| | T | - | 8 | 1 | 5 | - | 1 | - | - | - | - | - | - |
| | V+T | - | - | - | - | - | 34 | - | - | - | - | - | - |
| | BASE | 1 | 17 | 1 | - | - | 4 | - | 1 | - | - | - | 2 |
| | NOISE | - | 1 | - | - | - | 1 | - | - | - | - | - | - |
| | 60HZ | - | - | - | - | - | - | - | - | - | - | - | 2 |
| | ART | 25 | 95 | 4 | 2 | 18 | 75 | 1 | 24 | 1 | - | - | 1 |
| | I-ART | - | 3 | - | - | - | - | - | - | - | - | - | - |
| | T+ART | - | 3 | - | - | - | 8 | - | - | - | - | - | - |
| | V+ART | - | - | - | - | 1 | 12 | - | 2 | - | - | - | - |
| | N+ART | 1 | 2 | - | - | - | 8 | - | 2 | - | - | - | - |
| | P-ART | 9 | - | - | - | - | - | - | - | - | - | - | - |
| | SLOW | 1 | - | - | - | - | - | - | - | - | - | - | - |
| | AFPSE | 3 | - | - | - | - | - | - | - | - | - | - | - |
| | V-PSE | 26 | - | - | - | - | - | - | - | - | - | - | - |
| | S-PSE | 10 | - | - | - | - | - | - | - | - | - | - | - |
| | LOAMP | 18 | - | - | - | - | - | - | - | - | 1 | - | - |
| | LOS | 1 | - | - | - | - | - | - | - | - | - | - | - |
| | OTHER | 2 | - | 1 | - | - | - | - | - | - | 1 | - | - |

Total hours: 565
Correct strips: 2595 (73%)
Incorrect strips: 981 (27%)
Total strips: 3576 6.3 strips/hour

Table 12-2b: Distribution of Strips per Hour for All Patients

```
#/hr  #hrs   |+++ correct  --- incorrect    || = 2

  0    66    |**********************************
  1    45    |+++++++++++++++-------
  2    28    |++++++++++---
  3    35    |+++++++++++++---
  4    40    |++++++++++++++----
  5    65    |++++++++++++++++++++++++++-------
  6    48    |++++++++++++++++++++----
  7    39    |++++++++++++++++---
  8    32    |++++++++++++----
  9    29    |++++++++++----
 10    23    |+++++++++--
 11    24    |+++++++++---
 12    22    |+++++++----
 13    16    |++++++--
 14    17    |++++++--
 15    10    |+++--
 16    14    |+++++--
 17     3    |+
>18     4    |+
```

Table 12-2c: Poor Signal Duration for All Patients

```
Dur        #hrs   || = 5

=  0 sec    305   |**************************************************>>
<  5 sec     98   |********************
< 10 sec     46   |*********
< 15 sec     29   |*****
< 30 sec     24   |****
<  1 min     23   |****
<  2 min     14   |**
<  5 min     11   |**
< 10 min      9   |*
< 20 min      6   |*
< 30 min      0   |
< 40 min      0   |
< 60 min      0   |
```

Table 12-2d: Keyboard Requests per Hour for All Patients

```
#/hr   #hrs    || = 2

  0    298     |************************************************>>
  1     18     |*********
  2     32     |****************
  3     20     |**********
  4     27     |*************
  5     20     |**********
  6     23     |***********
  7     11     |*****
  8     21     |**********
  9      9     |****
 10      8     |****
 11      6     |***
 12     11     |*****
 13      4     |**
 14      4     |**
 15      6     |***
 16      7     |***
>16     13     |******
```

Table 12-2e: Keyboard Requests by Type for All Patients

```
Type     #reqs    || = 5   duration = 565 hours

 HR-15M     82    |*****************
  HR-1H     87    |******************
  HR-3H     89    |******************
 HR-12H     72    |**************
 VR-15M     82    |*****************
  VR-1H    137    |****************************
  VR-3H    147    |******************************
 VR-12H     90    |******************
 VS-15M     98    |********************
  VS-1H    136    |****************************
  VS-3H    181    |*************************************
 VS-12H    113    |***********************
   PLOT    234    |************************************************
 RAWDSP    662    |************************************************>>
 FREEZE     28    |*****
 ALMENU     52    |**********
 RR-HST     17    |***
 VC-HST     26    |*****
 PR-HST      3    |*
 RW-HST      5    |*
CLS-DSP      7    |*
DST-DSP      0    |
   USER     52    |**********
          ----
          2400 requests, average = 4.2 requests/hour
```

Results for Individual Patients

The aggregate results do not reflect the variability of system utilization and performance for individual patients. Table 12-3 shows the results for the 12 patients who were included in the summary. Episode strip accuracy ranged from 20% to 90%, and are a reflection more the the signal quality (morphology and artifact) than the inherent variability in the algorithm itself. In instances where algorithm performance was poor (patients K and L), there were considerably fewer requests for trend displays and hardcopy plots.

The large number of requests for patient F were, in fact, requests by the patient, who took an extraordinary interest in what was happening to him during his stay at the hospital. He soon learned how to operate the bedside monitor and make his own hardcopy trends. In other instances, he exercised (doing deep knee bends) in front of the bedside monitor, using the 15-minute heart-rate and PVC-rate trends to monitor his ectopic activity with exercise. For example, he discovered that abrupt cessation of physical effort brought on PVCs, whereas a more gradual slowing did not. Because the patient took such an active interest in the monitor, the number of display and plot requests was much higher than the typical patient.

Table 12-3: Individual Patient Summaries

| Patient | #hrs | #cor | #inc | #req | #s/hr | %cor | #req/hr |
|---|---|---|---|---|---|---|---|
| A: | 31 | 141 | 26 | 155 | 5.4 | 84% | 5.0 |
| B: | 23 | 103 | 25 | 53 | 5.6 | 80% | 2.3 |
| C: | 62 | 178 | 156 | 187 | 5.4 | 53% | 3.0 |
| D: | 23 | 145 | 73 | 173 | 9.5 | 67% | 7.5 |
| E: | 65 | 412 | 104 | 147 | 7.9 | 80% | 2.3 |
| F: | 82 | 209 | 28 | 867 | 2.9 | 88% | 10.6 |
| G: | 92 | 150 | 69 | 122 | 2.4 | 68% | 1.3 |
| H: | 68 | 470 | 44 | 357 | 7.6 | 91% | 5.2 |
| I: | 71 | 710 | 245 | 197 | 13.4 | 74% | 2.8 |
| J: | 15 | 19 | 1 | 89 | 1.3 | 95% | 5.9 |
| K: | 20 | 29 | 113 | 46 | 7.1 | 20% | 2.3 |
| L: | 13 | 29 | 97 | 7 | 9.7 | 23% | 0.5 |

Column Headings:

#hrs: total number of hours of reviewed strips and log reports
#cor: total number of correct strips
#inc: total number of incorrect strips
#req: total number of requests for displays and plots

#s/hr: average number of strips made per hour
%cor: percentage of correct strips
#req/hr: average number of requests per hour

## 12.2 Interviews with the Clinical Staff

One indicator of the usefulness of a new instrument is its acceptance by the clinical staff. A structured interview was conducted with several of the personnel who used the bedside monitor. The questions, and responses by nurses (N), interns (I), residents (R), Holter and arrhythmia lab technicians (H), and cardiologists (C) are summarized below. Note that most of the comparisons were made vis-a'-vis the Holter technique, since there were no other computer monitoring systems available on the general medical and surgical floors when the bedside monitor clinical trials were performed.

1. How useful did you find the Bedside Arrhythmia Monitor?

| no use (1) - extremely useful (5) | N | I | R | H | H | H | C | C | C | |
|---|---|---|---|---|---|---|---|---|---|---|
| Strips | 5 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | (4.9) |
| Trends | 4 | 2 | 5 | 3 | 4 | 3 | 1 | 5 | 4 | (3.4) |
| Technician Report | 3 | 4 | 5 | 5 | 5 | 5 | 5 | x | x | (4.6) |

2. How much did you trust the results?

| never (1) - always (5) | N | I | R | H | H | H | C | C | C | |
|---|---|---|---|---|---|---|---|---|---|---|
| Strips | 5 | 5 | 3 | 5 | 4 | 5 | 5 | 4 | 5 | (4.6) |
| Trends | x | 2 | 3 | 3 | 4 | 3 | 1 | 3 | 4 | (2.9) |
| Real Time Display | 5 | 4 | 3 | 5 | 5 | 5 | 5 | 5 | 4 | (4.6) |
| Technician Report | x | 4 | 5 | 5 | x | x | 5 | x | x | (4.8) |

3. Which features did you rely on the most?

| never (1) - always (5) | N | I | R | H | H | H | C | C | C | |
|---|---|---|---|---|---|---|---|---|---|---|
| Strips | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | (5.0) |
| Trends | 3 | 2 | 5 | 3 | 3 | 4 | 2 | 3 | 5 | (3.3) |
| Real Time Display | 5 | 5 | 3 | 5 | 3 | 3 | 5 | 5 | 3 | (4.1) |
| Technician Report | 3 | 5 | 5 | 5 | 5 | 5 | 3 | x | x | (4.4) |

4. When using the system, did you personally check the displays and review the strips, or did you rely on the nursing staff and the bedside monitor technician to obtain results?

   N: ... I checked the strips every hour in order to determine the validity and usefulness of the trends.
   I: ... check them myself.
   R: yes.
   H: [I prepared the reports]
   H: [I prepared the reports]
   H: [I prepared the reports]
   C: I reviewed the strips.
   C: I did [review the strips].
   C: I, personally, reviewed the strips.

5. Did the use of the system alter the patient's medical treatment? How?

   N: It helped facilitate the titration of antiarrhythmic medications ... and real-time documentation.
   I: No.
   R: Yes. [Helped with] choice of antiarrhytmic [drugs].
   H: Yes. Usually the detection of arrhythmias led the physician to adjust or change a patient's medications.
   H: x
   H: x
   C: Yes. [Indicated need for pacemaker].
   C: Yes. Identified VT, couplets, etc, without delay.
   C: Certainly - medication adjustments and evaluation of relation of arrhythmia to symptoms made more rational.

6. Did the system have any physical or psychological effects on the patient? What, if any?

   N: A few people became very concerned when the machine broke down and was not operating well.
   I: Improved. Felt "watched".
   R: No. System was out of sight [outside patient's room].
   H: It depended on the patient: to those who were anxious, they were anxious about everything. To others it was a reassuring new technological development.
   H: Physical - limits activity of patient [unlike Holter]. Psychological - varies from dependence and security, to increasing anxiety [in the patient].
   H: No significant physical side-effects.
   I have seen patients feel safe with the system but also can become dependent and anxious when it is discontinued. Also, some patients [became] fearful - but not because of this system - any machine.
   C: Beneficial.
   C: No.
   C: Probably not.

7. What effect, if any, did the use of this system have on the patient's length of stay?

N: Decreased it substantially when used in conjunction with agressive [nursing] participation. Eliminated the need to book Holters, wait for Holter, wait for Holter report, etc. Also - we can see what is happening when it happens.
I: None.
R: None.
H: It depended upon both the physician and the patient's total condition, but it appeared to shorten most of the patients' stay.
H: x
H: Decreased the length of stay.
C: Decreased [the patient's stay].
C: Shortened [the patient's stay].
C: Didn't lengthen and at times shortened by a day or two [because of] Holter turn-around time.

8. What do you consider unnecessary on the system, and what would you suggest be added?

N: [no comment]
I: [no comment]
R: [Don't need] RR-interval histograms.
H: The most important addition should be alarms. Without alarms, the patients must still be hooked up to a floor monitor.
H: Needs alarms.
H: Alarms.
C: Make the trends [the algorithm] more accurate.
C: [no comment]
C: Nothing is really unnecessary. An arrhythmia alarm system is essential. Also, detection of pacemaker spike and telemetry capacity and event [drug change] might be helpful.

9. Do you feel confident enough in the system to use in in place of a Holter. If not, why?

N: I would use it. The only time I would not want to use it is if there is a question of SVTA vs VT where a second lead system [two channel analysis] is necessary.
I: [I don't feel confident] with the computer interpretations. Seems okay, but called to many [false positive] VEA.
R: Yes.
H: Yes.
H: Yes.
H: x
C: Yes.
C: Yes.
C: Yes. [Its] only significant limitation is its inablility to allow for patient mobility.

10. In what clinical situations would you find the system most applicable?

N: Drug testing,
acute management of post surgical patients.
I: [Monitoring] patients who have various arrhythmias.
R: Incidence of VEA.
H: Drug trials, arrhythmia verification as well as ruling out arrhythmias as a cause for a patient's symptoms.
H: [Monitoring] high-grade VEA.
H: Drug trials,
Patients with a known history of highgrade VEA to document [runs of] VT.
C: x
C: Drug testing.
C: Drug trials,
Post MI for patients not in CCU,
Recovery room,
Emergency vehicles.

11. Would you recommend the system's use for other patients?

N: x
I: Yes.
R: Yes.
H: Yes.
H: Yes.
H: Yes.
C: Yes.
C: Yes.
C: Yes!

Observations

In general, the overall impressions about the bedside monitor were quite favorable. Most respondents felt confident enough in the system to use it in place of a Holter recorder, provided that its interpretations were accurate, and that it did not limit the patient's mobility. About half of those interviewed felt that the system reduced the patient's stay at the hospital, primarily by eliminating delay for scheduling and processing a Holter tape recording.

Almost all users said that the strip chart recordings were the most useful output of the bedside monitor, since the strips provided direct evidence of arrhythmias. The trend displays were considered less useful, however, and most felt that it was necessary to verify their accuracy with the strip chart recordings. The daily technician report (which included an hour-by-hour summary and selected rhythm strips) was considered valuable, especially when there was substantial artifact or other problems that caused the monitor to be inaccurate.

The primary use of the bedside monitor was for monitoring patients who had various arrhythmias, and for antiarrhythmic drug trials. Several users cited the need for an alarm system, which was not present on the version used in this evaluation. One cardiologist suggested that the bedside monitor could be used in other areas of a large hospital (emergency and recovery rooms); in smaller community hospitals that cannot justify or afford the purchase of a multipatient monitoring system; small clinics; military bases; emergency vehicles; and other situations that require a self-contained, portable arrhythmia monitor.

## 12.3 Cost Comparison of Arrhythmia Monitoring Systems

The cost of new medical technology is another question that needs to be addressed. In this section, the cost of an arrhythmia monitoring service based on the bedside monitor is compared to the comparable costs of the Holter scanning technique and multipatient monitoring systems.

The bedside arrhythmia monitoring service would include scheduling, attaching patients to the monitor, and preparing daily reports similar to Holter reports. Of course, the principal advantage of this service over a Holter recording is the immediate availability of trend summaries and hardcopy rhythm strips which document important clinical arrhythmias. Daily reports are prepared by a technician, and include an hour-by-hour summary of ectopic activity, and ECG strips that document important arrhythmic episodes. Thus, the clinician gets the benefit of real-time reports, with a follow-up report the next day which includes reviewed, selected strips that characterize the patient's ectopic activity.

### Estimated budget for a bedside arrhythmia monitoring service

The estimated budget for a bedside arrhythmia monitoring service is shown in Table 12-4a. The budget includes the salary for a full-time arrhythmia technician who would be responsible for putting the monitors on patients, periodically checking the performance of the system, replacing or adjusting electrodes, and preparing daily summary reports. Support services (laboratory

administration, engineering, and clerical) are also budgeted.

The capital expenses are based on what it currently costs to fabricate the bedside monitor in small quantities. If mass produced, the manufacturing costs would be considerably less, but the final selling price would probably be about the same ($10,000 to $12,000). The maintenance budget includes the cost for spare circuit boards, and replacement tubes for the storage displays. Finally, the cost of daily supplies is listed, assuming a Polaroid camera is used for making hardcopy records of the displays (other hardcopy options include an X-Y plotter or a strip chart recorder capable of recording both graphical and analog data). The overall estimate for the bedside monitoring service is about 81 dollars per patient-day.

Holter monitoring costs

The estimated annual budget for scanning services in the Arrhythmia Laboratory at Beth Israel Hospital is shown in Table 12-4b. The lab processes from 1200 to 1800 tapes a year using two Avionics scanners. The estimated cost for a 24 hour tape is 94 dollars. Note that hospital overhead costs have been exluded from both budgets. The overhead rate would be expected to be the same for both services - probably in the vicinity of 43% of total direct costs less equipment. If these costs were included, the bedside monitor cost would be $109.50 per patient-day, and the Holter would be $129.60 per 24 hour tape. Physician interpretation fees have not been included in calculating the costs of either service, but might be quite comparable.

Table 12-4a: Estimated Budget for the Bedside Arrhythmia Monitor

Assumptions: Two units and 480 patient-days of monitoring.
Service includes preparation of daily reports.

| | | |
|---|---|---|
| **SALARIES** | | |
| Arrhythmia Technician (100%) | | $ 12,375 |
| Support Services (Lab director, engineering, clerical) | | $ 7,000 |
| Fringe Benefits at 20% | | 3,875 |
| Total Salaries and Fringe | | $ 23,250 |
| **CAPITAL EQUIPMENT** | | |
| 2 Bedside Arrhythmia Monitors @ $12,000 amortize over 3-years per year cost | | $ 8,000 |
| **MAINTENANCE** | | |
| Spare circuit boards, Storage display tubes, etc. per year cost | | $ 3,000 |
| **SUPPLIES** | | |
| ECG Paper (1 roll per pt day) | $ 1.00 | |
| Electrodes | $ 2.10 | |
| Polaroid Film (for trends) | $ 5.00 | |
| Leads, tape, Xerox | $ 2.00 | |
| Total Supplies | | $ 4,848 |
| TOTAL SERVICE COST PER YEAR | | $ 39,098 |
| per patient-day: | | $ 81.45 |
| **OVERHEAD** | | |
| Total direct costs less equipment 43% x $31,098 | | $ 13,372 |
| TOTAL COST WITH OVERHEAD | | $ 52,470 |
| per patient-day: | | $109.50 |

Table 12-4b: Estimated Budget for Holter Arrhythmia Laboratory

Assumptions: 1800 24-hour tapes per year.

| | | |
|---|---|---|
| **SALARIES** | | |
| 3 full-time arrhythmia technicians | | $ 48,400 |
| Support services (Lab director, engineering, clerical) | | $ 38,100 |
| Fringe Benefits at 20% | | $ 17,300 |
| Total Salaries and Fringe | | $103,800 |
| **CAPITAL EQUIPMENT** | | |
| New Scanner - $39,000 | | |
| Amortize over 3 years (per year) | | $ 13,000 |
| Recorder replacements (3 per year) | | $ 9,000 |
| Total Equipment (per year) | | $ 22,000 |
| **MAINTENANCE** | | |
| Parts and supplies | | $ 5,000 |
| **SUPPLIES** | | |
| Electrodes | $ 3.50 | |
| ECG Paper | $ 8.00 | |
| Recording Tape | $ 4.67 | |
| Miscellaneous | $ 5.47 | |
| | $21.64 per tape | |
| Total Supplies | | $ 38,952 |
| TOTAL HOLTER COST PER YEAR | | $169,752 |
| per 24 hour tape: | | $94.30 |
| **OVERHEAD** | | |
| Total direct costs less equipment 43% x $147,752 | | $ 63,533 |
| TOTAL COST WITH OVERHEAD | | $233,285 |
| per patient-day: | | $129.60 |

Multipatient monitoring costs

Multipatient monitoring systems are also used to assist in the management of patients with arrhythmias. Although the initial capital expense for these systems is quite large, they can be amortized over multiple patients, and make more effective use of the nursing staff than would a decentralized monitoring system.

It was difficult to precisely identify the operating costs for multipatient monitors, since they are often lumped together with the other costs of an intensive care unit, and are not separately charged to the patient. However, one may compare the hardware costs of a centralized monitoring system with a distributed network of bedside monitors that are located in the patients' rooms. Such a distributed system could still have a central console, but the ECG aquisition and analysis would be performed at the bedside. Thus, the central facility would not require a powerful minicomputer, and would only need to call up trend and other information from the bedside units.

The expense of capital equipment for both systems can be easily compared. The total initial cost of a (Hewlett-Packard) centralized multipatient monitoring system is approximately $280,000 for eight beds. Eight bedside monitors would cost 8 x $12,000, or $96,000, which leaves enough money for a central station (with printers, disk storage for trends, etc.). Thus, a distributed monitoring system would be financially attractive, and can be quite competitive with a centralized monitor.

A distributed system would have other financial advantages as well. Smaller hospitals could avoid the major initial expense of a centralized multipatient facility by starting with one or two bedside systems. If their needs grow, more bedside units could be purchased along with a central station.

Since multipatient monitoring is usually only available in intensive care units, the daily room charge must also be compared. For example, the daily charge in the Medical Intensive Care Unit (MICU) at the Beth Israel Hospital (which recently aquired a multipatient monitor) is $565 a day. This can be contrasted with $260 a day for a private room, or $225 for a semi-private room on a general medical or surgical floor (the difference is due primarily to staffing costs).

# 13. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

## 13.1 Conclusions

The development and evaluation of the bedside cardiac arrhythmia monitor proved to be a much more challenging task than initially anticipated. Pattern recognition tasks that the human brain performs almost subconsciously were very difficult to describe in algorithmic form. The waveforms that were encountered during the clinical trials were often less than "textbook perfect", and frequently required completely new approaches to process them.

The overall results, however, indicate that the bedside monitor can indeed play a useful role in the treatment of patients with ventricular arrhythmias. Its acceptance by the clinical staff has been quite positive, and it appears that the bedside monitor will soon be offered as a regular clinical service by the Arrhythmia Laboratory at Beth Israel Hospital. The bedside monitor offers sophisticated ECG waveform processing comparable to that used in multipatient monitors, but has a much lower minimum system cost, and can be used anywhere within the hospital. Its principal advantage over in-patient Holter recording is the immediate availability of trends and documented ECG episodes. The reduced turn-around time should result in shorter hospital stays, and increase safety by noting dangerous trends that would otherwise not be apparent until long after they have occurred.

ECG analysis algorithm

The development of a real-time ECG analysis algorithm was central to this work. Several novel approaches were developed:
(1) A data compression technique that robustly partitions the incoming ECG into a series of straight line segments that describe the ECG in a compact form. This step renders the data into a more easily parsable representation, and performs an initial feature extraction stage for later waveform analysis.
(2) An event detector that searches for the largest (signal to noise ratio) events first, and uses the contextual relationships between previously detected events and the candidate to identify it.
(3) A QRS morphology feature set that embodies a temporal description of the QRS complex in terms of triangular submorphs.
(4) A morphology labeling procedure that can adapt to patients with different underlying rhythms and morphology, reducing the need for operator adjustment of the monitor.

Evaluation using the annotated data base

Evaluation of sophisticated medical instrumentation such as the bedside monitor was a difficult task, and required almost as much time as the development of the monitor itself. The annotated digital data base served as a basis for evaluating draft versions of the algorithm in a repeatable and quantitative manner. The digital data base may be easily duplicated, and thus encourage meaningful intersystem comparisons.

The data base evaluation of the bedside monitor indicated an acceptable level of performance on a reasonably difficult set of ECG waveforms. Event detector sensitivity was 99.43%, and positive predictivity was 99.72%. PVC detection sensitivity was 84.1%, and positive predictivity was 95.4%. Detection sensitivity for isolated PVCs was 88%, couplets 77%, short runs (3 to 5 beats) 45%, and longer runs 69% suggests that specialized pattern recognition algorithms can improve detection accuracy for ventricular tachycardia and flutter.

The existence of standard data base(s), however, still does not totally answer the question of "What defines good performance?" for an algorithm. As evidenced in the data base evaluation, there are many performance measures that can be used. Overall sensitivity for isolated PVCs, couplets, and runs are indeed important measures, but additional approaches are required to convey the substantial tape-to-tape (patient-to-patient) variability in performance.

Evaluation using long term analog tape recordings

The multifaceted evaluation used in this work was designed to test all aspects of the system's performance. The data base evaluation was the principal quantitative evaluation, to be sure, but there were other components of the system such as the alarm episode documentation strategy that had to be tested as well. This aspect of system performance (both the algorithm and report generation strategies) was evaluated using long term Holter analog tape recordings. Although less quantitative than the

beat-by-beat evaluation using the annotated data base, it exposed the algorithm to far more data, and compared the bedside monitor to a clinically accepted arrhythmia documentation technique (Holter recording).

Clinical Evaluation

The clinical trials constituted the final phase of the system's evaluation, and tested the many factors that govern real-world performance: the performance of algorithm and documentation procedures, human factors in instrument design, and whether or not the instrument fulfills a real need in the clinical setting. Here, the bedside monitor has been quite successful, and appears to be playing an ever increasing role in the management of patients with ventricular arrhythmias.

## 13.2 Directions for Future Work

During the course of this work, it became evident that the analysis of the ECG waveform was an extremely difficult task. Not every aspect of the problem could be thoroughly explored, and the possibilities for future work are seemingly endless. Listed below are a few areas that warrant further exploration.

Artificial pacemaker rhythms

The electrocardiogram of patients with an artificial pacemaker contains an additional component, the pacemaker discharge artifact. The analysis algorithm currently does not recognize the pacemaker spike as a separate waveform entity, and

in most cases it is included in the morphologic description of the QRS complex. This normally does not cause problems for the event detector and classifier: the detector generally will detect the ventricular response rather than the spike based on the triangular morph width. Since all timing and width measurements are based on the major QRS deflection, the occassional inclusion of the pacemaker spike will not produce false positive PVCs. For ventricular pacemakers, the QRS complexes are usually wide and aberrant, and the family classifier automatically adapts to the wide "normal" QRS complex just as it would with bundle branch block.

To reliably detect the pacemaker artifact in order to evaluate artificial pacemaker function, additional analog circuitry must be added to the ECG amplifier. This could include a high-pass filter (prior to the low-passed ECG) and voltage comparator with a software controlled detection threshold to detect the pacemaker artifact. Upon detection of the pacemaker spike, the event detector and delineator could take special procedures to avoid inclusion of the pacemaker artifact in the description of QRS morphology.

P-wave detection

Complete analysis of cardiac rhythm requires detection of P-waves, which are often difficult to detect on the surface ECG. P-waves are typically an order of magnitude smaller than the QRS complex, and thus are often buried in noise. Several approaches can be attempted to recognize P-waves. The first would be to use

an esophageal electrode (Jenkins,1976, Jenkins,1978), and to incorporate an additional analysis channel in the algorithm. Another approach is to attempt P-wave detection on the surface ECG lead. One approach (Gustafson,1979) attempts P-wave recognition by first detecting and then removing the QRST complexes from the signal, followed by searches for possible P-waves on the QRST-free waveform. This technique, however, requires a relatively noise free ECG (which is often not the case in clinical practice), or a reliable noise-level estimator to determine if the ECG is indeed analyzable for P-waves. Given the artifact and noise present in typical monitoring applications, surface P-wave analysis techniques would be helpful in only a fraction (30-50%) of cases.

Detection of ventricular flutter and fibrillation

Rhythms such as ventricular tachycardia, flutter, and fibrillation are clinically important, and should be reliably detected by a monitoring system. Although these rhythms may be recognized using the classical event detector and classifier approach, it may be more advantageous to use specialized pattern recognition algorithms to recognize these specific waveforms, given the measured sensitivities in this study. The bedside monitor, for example, examines the compressed ECG for a sequence of overlapping triangular morphs characteristic of ventricular tachycardia and flutter. Power spectrum analysis (Nygards,1977, Nolle,1980) and time-domain analysis (Kuo,1978) are other possible approaches.

Multiple channel analysis

Additional information about the ECG can be obtained from multiple lead recordings. By analyzing a second channel, it may be possible to improve algorithm accuracy and noise immunity. For example, an ectopic beat may have a morphology similar to normals in one lead, but a distinctly different morphology in another. In other cases, one lead may have artifact, whereas the other does not. Several approaches are possible for multiple lead analysis:

(1) use a single channel algorithm which switches between different leads depending on noise and background morphology,

(2) combine the two leads into one, and analyse it,

(3) do a complete multiple lead analysis by maintaining separate event detector and classifiers for each channel (which would require a voting scheme or "confidence" measure for each channel), and (4) use a single channel algorithm for event detection, but the morphology features from both, depending on noise.

Other morphology measures

The selection of a feature set, distance measure, and morphology clustering algorithm figure prominently in the design of an ECG analysis algorithm. To this date, there has not been a definitive comparison of the different techniques, and those that have been performed often omitted crucial steps (such as optimum voltage and time alignment between the compared waveforms). Although there are many ways to compare classifiers, there is an

especially simple one which makes use of an annotated ECG data base, and would allow meaningful comparison. The beat annotations would serve as the (perfect) event detector for the algorithm under test (AUT). The AUT would delineate the complex (if necessary), compute its feature vector, and assign it to a morphology cluster. If the complex is considered to have a new morphology, then the true beat label on the data base can be used to label the new class. The classifier performance can be based on false negative and positive errors, just as in a conventional algorithm evaluation. In this evaluation, however, the functions of the event detector and morphology class labeling rely on the data base annotations.

ST-segment and other physiologic parameters

Other physiologic measurements, such as ST-segment, blood pressure, respiration, and temperature can also be incorporated into the bedside monitor. Although there appeared to be no demand for monitoring these variables during the clinical trial of the bedside system, their inclusion would extend the scope of applications for the bedside monitor.

APPENDIX A-1: ECG Data Base Format on 1/4" Tape Cartridges

The tape format described in this appendix was the original format developed for ECG sampling, editing, and evaluations on 1/4" tape cartridges. It permits the digital recording of physiologic data, interspersed with machine readable annotations. The format supports multiple data channels and up to eight independent annotation channels.

CONVENTIONS:
---

Storage conventions:

| | |
|---|---|
| Byte: | Smallest unit of addressable storage at location ADDR. |
| Word: | 2 bytes, low-order byte at ADDR, high-order byte at ADDR+1. |
| DblWord: | 4 bytes, low-order byte at ADDR, ascending bytes. |
| Arrays: | ADDR points to first byte, ascending bytes. |

Text Strings: ASCII text, parity (msb) is zero.

| | |
|---|---|
| string<0> | ASCII text terminated with null, ADDR points to first character. |

| | |
|---|---|
| Time-of-day: | Double, # of seconds past midnight. |
| Date: | 4 bytes, ADDR: day (byte), ADDR+1: month (byte), ADDR+2: year (word). |

| | |
|---|---|
| xxxH | hexadecimal number |
| xxx | decimal |
| xxx. | decimal |

[xxx] following a symbolic value, gives numerical value.

| | |
|---|---|
| BYTE<7:0> | specific bits in a byte |
| WORD<15:00> | specific bits in a word |

BLOCK TYPES:
---

Data, annotations, and other data are sequentially recorded on the tape in fixed length physical blocks [1024 bytes long]. The block types currently implemented include:

1) HDRBLK Header Block.
Specifies data sampling rate, number of data channels, data, time of original data recording, and annotator identification.

2) DDBLK Difference Data Block.
Contains the original physiologic data. These are recorded sequentially in time on the tape. Multiple channels are permitted.

4) ANNBLK Annotation Block.
Contain machine readable waveform annotations and other user comments. An ANNOTATION BLOCK only describes the DATA BLOCKS that follow it.

5) ETRKBLK End of Track Block.
Specifies the end of the current track on multiple track media, such as 3M tape cartridges.

6) ETPBLK End of Tape Block.
Specifies the end of all data on the media.

7) ENDBLK End of Record Block.
Specifies the end of entire data record.

BLOCK SEQUENCE:
---

The annotation block(s) and corresponding data block (or blocks, for multiple channels) form a complete epoch in time.
The waveform editor and other programs using serially formatted magnetic tape must buffer the entire epoch in memory.
Since several time epochs are processed concurrently, a maximum of EPMBLKS [5] blocks ( annotation blocks + data blocks ) for each time epoch must be observed to avoid buffer overflow.

Examples:

SINGLE CHANNEL DATA
---

```
HDRBLK                  An example of single channel data,
ANNBLK 0                with at most EPMBLKS data blocks
DATABLK chan0, time0    per annotation block.
DATABLK chan0, time1
DATABLK chan0, time2
DATABLK chan0, time3
ANNBLK 1                ***
DATABLK chan0, time4      *
DATABLK chan0, time5      *  an epoch
DATABLK chan0, time6      *
DATABLK chan0, time7    ***
ANNBLK 2
DATABLK chan0, time8
  :
ETRKBLK                 end-of-track block
```

TWO CHANNEL DATA
---

```
HDRBLK
ANNBLK0
DATABLK chan0, time0
DATABLK chan1, time0
DATABLK chan0, time1
DATABLK chan1, time1
ANNBLK1                 ***
DATABLK chan0, time2      *
DATABLK chan1, time2      *  an epoch
DATABLK chan0, time3      *
DATABLK chan1, time3    ***
ANNBLK2
DATABLK chan0, time4
DATABLK chan1, time4
DATABLK chan0, time5
DATABLK chan1, time5
:
ETRKBLK
```

TAPE BLOCK FORMAT:
---

Each block on the tape is BLKLEN [1024] bytes long,
and has the following format:

```
BLK:              BLOCK TYPE       ;word:  identifies block type.
BLK+2:            -BLOCK TYPE      ;for error check (2's complement).

BLK+4:            CHECKSUM         ;word sum (no end-around-carry)
                                   ; of all words in block.

BLK+6:            block dependent information follows
   .
   .
   .
   .
   .
   .
   .
   .
   .
   .
   .

BLK+BLKLEN-4:     BLOCK LINK       Linkage to next block.
BLK+BLKLEN-2:     -BLOCK LINK      used by mass storage device.
```

The BLOCK LINK is used by the mass storage device driver routine,
and is different for different mass storage devices.

3M TAPE CARTRIDGES:
  BLOCK LINK<15:14> tape track number
  BLOCK LINK<13:00> block number on tape, starts with 1.
  BLOCK LINK<15:00> = 0 :  don't care.

DISK
  BLOCK LINK<15:00> ptr to next block on disk.

SERIAL LINK
  BLOCK LINK<15:00> = 0.

HEADER BLOCK
---

The header block contains machine readable data specifying the sampling rate, number of channels, time and data of recording, annotator identification, and other information.

| | | |
|---|---|---|
| BLK: | HDRBLK*100H [HDRBLK=1] | |
| BLK+2: | -(HDRBLK*100H) | |
| BLK+4: | CHECKSUM | |
| BLK+@IDN [6]: | RECORD ID NUMBER (word) | Record identification. Assigned by user during digitization. |
| BLK+@GEN [8]: | RECORD GEN NUMBER (word) | Nth copy number. When a copy of a record is made, this number is incremented on the new record. |
| BLK+@REVST [10]: | RECORD REV STATUS (word) | 'History' of record. Bits in this word indicate processing status of this record. |
| BLK+@ATTIME [16]: | ANALOG TAPE TIME ( 4 bytes: SECONDS ) | Time, and |
| BLK+@ATDATE [20]: | ANALOG TAPE DATE ( 4 bytes: YY,M,D ) | Date analog tape was made. |
| BLK+@DTIME [24]: | DIGITIZATION TIME | Time, and |
| BLK+@DDATE [28]: | DIGITIZATION DATE | Date of digitizaiton. |
| BLK+@LTIME [32]: | LAST TIME REVISED | Time, and |
| BLK+@LDATE [36]: | LAST DATE REVISED | Date of this revision. |
| BLK+@SFREQ [48]: | SAMPLING FREQUENCY (word) | Samples per second. |
| BLK+@NCHAN [50]: | NUMBER OF CHANNELS (word) | < 255 |
| BLK+@NDBLK [52]: | NUMBER OF DATA BLOCKS RECORDED FOR A SINGLE CHANNEL (word) | |

HEADER BLOCK specification, continued

Each annotation has a byte that identifies the annotator(s) who made (agree with) the annotation.
Every annotator is assigned a bit position in the byte.

```
---------------
7 6 5 4 3 2 1 0        ANNOTATION I.D. BYTE
---------------        in annotation block.
```

By convention:

ANNOTATOR 0 is a preliminary MACHINE ANNOTATION of events. This is used to automatically advance cursor for 'blinded' human interpretation of beats.

The remaining annotation channels 1-7 may be used for human or machine annotations.

---

ANNOTATOR INFORMATION in HEADER BLOCK

If the first byte of the entry is 0, the annotator channel is inactive.
For annotator N (N = 0-7):

BLK+@HATBL+(N*HALEN):

| | |
|---|---|
| Bytes(0:HADATE-1) | INITIALS (recognized by program) initials<0> |
| Bytes(HADATE:HATEXT-1) | DATE (YY/M/D) |
| Bytes(HATEXT:HALEN-1) | FULL NAME OR TITLE name<0> |

[@HATBL=64] Unused bytes are zeroed.
[HALEN=32]
[HADATE=6]
[HATEXT=10]

BLK+@SNAME: SUBJECT NAME SNMLEN bytes long
[@SNAME=320] name<0> Unused bytes are zeroed.

.
.
.
.
.

BLK+BLKLEN-4: BLOCK LINK
BLK+BLKLEN-2: -BLOCK LINK

DATA BLOCK FORMAT (DIFFERENCE REPRESENTATION)
---

Data blocks are recorded time-sequentially on the tape.
If multiple channels are recorded, then the different channels are recorded sequentially (channel 0, channel 1, channel 2, etc.), and this is repeated for each time frame.

DDNSAMP [1000] number of samples per block, every block.

| | | |
|---|---|---|
| BLK: | (DDBLK*100H + CHANNEL NUMBER) ( 0, 1, 2, ... ) | |
| BLK+2: | - (DDBLK*100H + CHANNEL NUMBER) ( 0, 1, 2, ... ) [DDBLK=16] | |
| BLK+4: | CHECKSUM | |
| BLK+6: | SAMPLE NUMBER (DblWord) | (for this channel, from beginning of tape) |
| BLK+10: | GAIN (DblWord) | Gain factor for this block. Unknown if 0. |
| BLK+14: | OFFSET (Word) | Offset (add to array values, then scale by GAIN). |
| BLK+16: | ABSOLUTE VALUE (Word) | Absolute value of sample immediately before first difference. |
| BLK+18: | . | |
| | . | |
| | . | |
| | . | |
| | 8-BIT SIGNED DIFFERENCES ( -128 to +127 ) | |
| | . | |
| | . | |
| | . | |
| | . | |
| BLK+18+DDNSAMP: | 0 | Remainder of block filled with zeroes. |
| | . | |
| | . | |
| | 0 | |
| | . | |
| | . | |
| BLK+BLKLEN-4: | BLOCK LINK | |
| BLK+BLKLEN-2: | -BLOCK LINK | |

## ANNOTATION BLOCK FORMAT

Annotation blocks allow the user to insert machine readable waveform identifications and comments. The annotation block(s) always preceed the data block(s) that they describe.

The annotation block(s) are always followed by data channel 0, and other data channels in the same time frame.

The annotation block(s) and corresponding data block (or blocks, for multiple channels) form a complete epoch in time. The waveform editor and other programs using serially formatted magnetic tape must buffer the entire epoch in memory. Since several time epochs are processed concurrently, a limit of EPMBLKS [5] (annotation blocks + data blocks) for each time epoch must be observed to avoid buffer overflow.

| | | |
|---|---|---|
| BLK: | ANNBLK*100H [ANNBLK=3] | |
| BLK+2: | -(ANNBLK*100H) | |
| BLK+4: | CHECKSUM | |
| BLK+6: | SAMPLE NUMBER | from beginning of tape. Equals the first sample number of the data block that follows. |
| BLK+10.: | NUMBER OF ANNOTATIONS | in this block. |
| BLK+12.: | . | |
| | . | |
| | . | |
| | ANNOTATIONS | format specified on next page ... |
| | . | |
| | . | |
| | 0,0 | zero word terminates annotation list in a given block. |
| | . | |
| | . | |
| | . | |
| BLK+BLKLEN-4: | BLOCK LINK | |
| BLK+BLKLEN-2: | -BLOCK LINK | |

ANNOTATION FORMAT

| | | |
|---|---|---|
| ANN: | ANNLEN (word) | ANNOTATION LENGTH in bytes. (at least 8 bytes) |
| ANN+2: | ANNOTATION TYPE | PVC, NORMAL, TEXT, etc. |
| ANN+3: | ANNOTATION SUBTYPE | for subtypes (multiform PVC's, etc.) |
| ANN+4: | READER IDENTIFICATION | a bit in this byte is set for annotator(s) |
| ANN+5: | DATA CHANNEL NUMBER | ECG data channel |
| ANN+6: | TIME (word) | relative to SAMPLE NUMBER of annotation block. |
| ANN+8.: | . . | |
| | Optional annotation arguments. | |
| | . | |
| ANN+ANNLEN-1: | . | |

... subsequent annotation(s) ...

The word following the last byte of the last annotation must be zero. This identifies the end of the annotation list in a given block (or, NUMBER OF ANNOTATIONS can be used to determine the end of the list).

notes:

TYPE codes are listed in Table 10-1a.

SUBTYPE convention for PVCs:
  0: morphology classes are not noted
  1,2,3,... for each PVC morphology class

SUBTYPE convention for NOISE and QUIET:
  0: moderate noise, beats are labeled
  1: severe noise, beats are not labeled

COMMENT ANNOTATION:
  The text string format is:
  ANN+8: STRING<0>
  The annotation length includes the terminating null.

ONSET/OFFSET ANNOTATION CONVENTIONS:
  Rhythm Onset annotation strings are listed in Table 10-1b.

## END OF TRACK / TAPE / RECORD BLOCK

The END-OF-TRACK block indicates that the last block of the current track has been read. This is used only on multiple-track media, such as 3M tape cartridges.

The END-OF-TAPE block indicates the end of the entire recording media.

The END-OF-RECORD block indicates the end of the entire record, which may span several physical volumes of the recording media.

| | |
|---|---|
| BLK: | ETRKBLK*100H / ETPBLK*100H / ENDBLK*100H |
| BLK+2: | -(ETRKBLK*100H) / -(ETPBLK*100H) / -(ENDBLK*100H)<br>[ETRKBLK=4, ETPBLK=5, ENDBLK=6] |
| BLK+4: | CHECKSUM |
| BLK+6: | 0<br>.<br>.<br>.<br>Remainder of block filled with zeroes.<br>.<br>.<br>.<br>0 |
| BLK+BLKLEN-4: | BLOCK LINK |
| BLK+BLKLEN-2: | -BLOCK LINK |

## TAPE BLOCK RECORDING FORMAT FOR 1/4" TAPE CARTRIDGE

The Biomedical Instrumentation Center for Clinical Instrumentation uses 1/4 inch tape cartridges (3M's DC 300A) for tape mass storage. The recording format is essentially identical to the proposed American National Standard for 4 track, 1/4", 1600 BPI, phase encoded tape cartridges. The only difference is that we record a checksum, rather than the proposed cyclic redundancy check (CRC). Since our tape controller doesn't generate the CRC automatically, we elected to compute a byte checksum rather than the CRC because it is easier to calculate.

For example, assume a tape controller that writes bytes onto the tape with the least significant bit transmitted first. The format of the tape block is shown below:

```
interblock gap

00H     preamble          first byte written/read on/from block
80H     preamble
xx      first data byte
xx
.
.
.
xx
xx      last data byte
chk     2's complement of ([sum of all data bytes],mod 100H)
0       upper byte of chksum is zero.
01H     postamble
00H     postamble

interblock gap
```

## APPENDIX A-2: Specification for FM Analog Data Base Tapes Based on the MIT-BIH Digital ECG Data Base

This appendix describes the format of the analog version of the MIT-BIH ECG data base. The analog data base was developed in response to requests by potential users who needed a relatively inexpensive and easy-to-use data base for beat by beat evaluations.

### SPECIFICATIONS, IN BRIEF

MEDIA: 1/4" instrumentation tape, 2300 foot reels.
This holds 4 half-hour records at 3-3/4 ips.

FORMAT: IRIG (Inter-Range Instrumentation Group) Document 106-71
4 channels FM
3-3/4 ips for real-time playback: FM bandwidth is 1250 Hz.

CHAN1: Channel 1 analog ECG data
CHAN2: Channel 2 analog ECG data
Channels 1 and 2 is scaled approximately 1 volt per millivolt.
Baseline (average DC value) is 0 volts.

CHAN3: Tape Identification and Annotations.
Recorded at EIA RS232 compatible levels, 300 baud.

CHAN4: Time Marks.
Recorded at EIA RS232 compatible levels, 300 baud.

### INTER-RECORD GAP

There is a 10 second inter-record gap preceeding the first record, between adjacent records, and after the last record.
The FM carrier is always on:
CHAN1 and CHAN2 are at their nominal baseline value (0 volts).
CHAN3 and CHAN4 are in the MARK (negative) state.

TAPE IDENTIFICATION -- Channels 3 and 4

Each half-hour record is preceeded by a null-terminated ASCII string containing the tape number and copyright notice, recorded on CHAN3 and CHAN4.

MIT205

<0>

ANNOTATIONS -- Channel 3

General Notes --

Annotations are recorded as a sequence of 8-bit bytes, and are terminated with a null <0>.

Annotations preceed event fiducial, so that second stop bit of terminating null is aligned with time of event.

Beat labels will always be exactly on time. Rhythm, comment, and noise labels are adjusted (if necessary) to avoid annotation overlap.

Annotation format, each entry --

| | |
|---|---|
| 1st byte: | Annotation type and ECG channel number. Lower six bits <5:0> are the MIT-BIH annotation code. Upper two bits <7:6> specify analog channel number that the annotation refers to (usually channel 1). |
| 2nd byte: [optional] | Annotation subtype. Noise level subtype or QRS morphology class number, and has a numeric value less than ASCII (space). The annotation subtype is assumed to be 0 if the terminating null immediately follows the annotation type. |
| aux bytes: [optional] | ASCII rhythm or comment text string. |
| last byte: | Terminating null. |

Annotation type codes are listed in Table 10-1a, and Rhythm Onset annotations strings are listed in Table 10-1b.

TIME TICKS -- Channel 4
---

The tape identification and time is recorded every 5 seconds on CHAN4 as ASCII characters. Time ticks are aligned with the second stop bit of the terminating null.

e.g.: MIT205 00:00<CR><LF><0> at the beginning of the record,
MIT205 00:05<CR><LF><0> five seconds into record, etc.

The time-ticks can be directly listed on a terminal. This format also allows automatic searches for a particular tape and section for repeat playings, etc.

DISCUSSION
---

The analog FM tape format will make relatively inexpensive and easy to perform beat-by-beat evaluations. The data format is universal: most, if not all, ECG monitoring systems can accept analog data. The annotation data, coded in RS232 bit-serial format, is compatible with most serial input ports available on micro and mini computers. The availability of the data base in an analog format would certainly promote its use, and increase the likelihood of intersystem comparisons.

The principal virtue of the analog format is economic. The cost of a 9-track tape drive to play back the digital tapes is prohibitive: it costs approximately 8 to 10 thousand dollars to add a drive and controller to a minicomputer. The situation is even worse for microcomputer based systems, since 9-track tape controllers are almost non-existent for microcomputers.

For these reasons, many investigators would find an analog tape format reasonable. The data bandwidth is relatively low, and would permit recording/playback speeds down to 15/16 or even 15/32 ips, allowing up to 8 or 16 hours of data on a single tape.

The analog data can be used for program development. It has, however, a number of drawbacks compared to digital tapes:

1) data will usually have to be played back in real-time, and gracefully stopping playback is difficult.

2) data playback is not exactly reproducable.

Real-time evaluation should be not be too difficult, and can be performed with a "comparator" running in the same system as the algorithm, or in a separate system that receives events from the algorithm under test. In either case, the algorithm need only transmit event labels and time relative to the current time (that is, "NORMAL beat 1.23 seconds ago"). The time-ticks are not necessary for event time alignment, and can be ignored by the algorithm or comparator program.

APPENDIX A-3: "Compressed" Annotation Format Description

This appendix describes the "compressed" annotation format which saves only the truth annotations in a very compact format, yet allows their exact reproduction. This format is useful for experiments that require only beat labels and their time of occurrence.

Each file consists of a sequence of bytes.
All annotations are word (double byte) aligned;
the least significant byte is the first.

The basic format for annotation words is:

```
   high byte       low byte
7            0 7            0
A A A A A A I I I I I I I I I I
```

where A is the annotation type code
I is the number of samples after the previous annotation

cases A = { 0 < A < AXLMX }, I = interval
are the annotation codes defined in Table 10-1a.

A = 0, I = interval <> 0
is a "null" R-R interval so that annotation intervals greater than 3FFH sample counts can be represented.

A = STP [61.]
I = Subtype for current annotation only
otherwise, assume subtype = 0.

A = DCH [62.]
I = Data channel for current and subsequent annotations
otherwise, assume previous data channel.
Initially zero.

A = AUX [63.]
I = Number of bytes of auxilliary information
an extra null, not included in the byte count, is appended for word alignment.
Rhythm Onset annotation strings are listed in Table 10-1b.

A = I = 0
End of file.

The ordering of annotation words in a file is:

```
[ "null" R-R intervals, if necessary ]
Annotation word
[ STP ]
[ DCH ]
[ AUX,
  and subsequent auxilliary information ]
.
.
.
0 word:  indicates End-of-File
```

# BIBLIOGRAPHY

(AHA,1975)
H. V. Pipberger, et. al., "Recommendations for Standardization of Leads and Specifications for Instruments in Electrocardiography and Vectorcardiography," Report of the Committee on Electrocardiography, American Heart Association, Circulation, vol. 52, News from the American Heart Association, pp. 11-31, August 1975.

(AHA,1978)
American Heart Association, Circulation Research, vol. 42, no. 4, pg. 572, April 1978.

(AHA,1980)
American Heart Association, Heart Facts, 1980.

(Akselrod,1981)
S. Akselrod, D. Gordon, F. A. Ubel, D. C. Shannon, A. C. Barger, and R. J. Cohen, "Power Spectrum Analysis of Heart Rate Fluctuation: A Quantitative Probe of Beat-to-Beat Cardiovascular Control", Science v. 213, pp. 220-222, July 10, 1981.

(Alexander,1968)
D. C. Alexander and D. Wortzman, "Computer Diagnosis of Electrocardiograms. I. Equipment", Computers in Biomedical Research, vol. 1, pp. 348-365, 1968.

(Arnold,1972)
J. M. Arnold, Time Domain Filtering of Electrocardiograms, Thesis, Cambridge, MA, Massachusetts Institute of Technology, 1972.

(Arnold,1975)
J. M. Arnold, P. M. Shah, and W. B. Clarke, "Artifact Rejection in a Computer System for the Monitoring of Arrhythmias," Computers in Cardiology, 1975, pp. 163-167, 1975.

(Balm,1967)
G. J. Balm, "Crosscorrelation Techniques Applied to the Electrocardiogram Interpretation Problem", IEEE Transactions on Biomedical Electronics, vol. BME-14, no. 4, pp.258-262, October 1967.

(Beck,1947)
C. S. Beck, W. H. Pritchard, and H. S. Feil, "Ventricular Fibrillation of Long Duration Abolished by Electric Shock," Journal of the American Medical Association, vol. 135, pp. 985-986, 1947.

(Bernard,1974)
R. Bernard, W. Rey, H. Vainsel, M. Boothroyd, and M. Demeester, "Computerized Dysrhythmia Monitoring with an Intra-Auricular Lead, Computers in Cardiology, IEEE Computer Society, pp. 17-20, 1974.

(Berni,1975)
A. J. Berni, D. E. Dick and M. W. Luttges, "Detection of Digitalis Toxicity by Computerized Electrocardiogram Monitoring", IEEE Transactions in Biomedical Engineering, vol. BME-22, no. 1, pp. 29-34, January 1975.

(Berson,1976)
A. S. Berson, "Bandwidth, Sampling, and Quantizing for Automated ECG Processing," Computers in Cardiology, IEEE Computer Society, pp. 295-301, 1976.

(Birman,1978)
K. P. Birman, L. M. Rolnitzky, and J. T. Bigger, "A Shape Oriented System for Automated Holter ECG Analysis", Computers in Cardiology, pp. 217-220, September 1978.

(Bleifer,1973)
S. B. Bleifer, H. L. Karpman, J. J. Sheppard, et al, "Relation Between Premature Ventricular Complexes and Development of Ventricular Tachycardia," American Journal of Cardiology, vol. 31, pp. 400-403, 1973.

(Bonner,1968a)
R. E. Bonner and H. D. Schwetman, "Computer Diagnosis of Electrocardiograms. II. A Computer Program for ECG Measurements," Computers in Biomedical Research, vol. 1, pp. 366-386, 1968.

(Bonner,1968b)
R. E. Bonner and H. D. Schwetman, "Computer Diagnosis of Electrocardiograms. III. A Computer Program for Arrhythmia Diagnosis," Computers in Biomedical Research, vol. 1, pp. 387-407, 1968.

(Brown,1963)
K. W. Brown, R. L. MacMillan, N. Forbath, F. Mel'Grano and J. W. Scott, "Coronary Unit, an Intensive Care Centre for Acute Myocardial Infarction," Lancet, vol. II, pp.349-352, 1963.

(Burns&Schluter,1978)
S. K. Burns and P. S. Schluter, "The Microprocessor Matrix: A Base for Instrument Development," Technical Report TR008, Harvard-MIT Biomedical Engineering Center for Clinical Instrumentation, Cambridge, Mass., 1978.

(Burton,1975)
C. E. Burton, W. M. Portnoy, and H. Diritten, "An Algorithm for On-Line Real-Time Computer Detection of ECG Changes", International Journal of Biomedical Computing, v. 6, pp. 23-32, January 1975.

(Cohen,1978)
R. J. Cohen, "Random Process Analysis of Beat-to-Beat Fluctuations in Electrocardiographic Parameters", National Science Foundation Research Proposal, MIT, 1978.

(Cox,1968)
J. R. Cox and F. M. Nolle, "AZTEC, a Preprocessing Program for Real-Time ECG Rhythm Analysis," IEEE Trans. Bio-Med Electron., vol. BME-15, no. 2, pp. 128-129, April 1968.

(Cox,1969)
J. R. Cox, H. A. Fozzard, F. M. Nolle, and G. C. Oliver, "Some Data Transformations Useful in Electrocardiography," in Computers in Biomedical Research, vol. 3, R. W. Stacy and B. D. Waxman, Eds., New York: Academic Press, pp. 181-206, 1969.

(Cox,1972)
J. R. Cox, F. M. Nolle, and R. M. Arthur, "Digital Analysis of the Electroencephalogram, the Blood Pressure Wave, and the Electrocardiogram," Proceedings of the IEEE, vol. 60, no. 10, pp. 1137-1164, October 1974.

(Cox,1981)
J. R. Cox, R. E. Hermes, and K. L. Ripley, "Performance Evaluation of Ventricular Arrhythmia Detector"s, in Ambulatory Electrocardiographic Recording, eds. N. K. Wenger, M. B. Mock, and R. Ringquvist, pp. 183-198, Year Book Medical Publishers, 1981.

(Day,1963)
H. W. Day, "An Intensive Coronary Care Area," Diseases of the Chest, vol. 44, pp. 423-427, 1963.

(DeSoyza,1974)
N. DeSoyza, J. K. Bissett, J. J. Kane, et al, "Ectopic Ventricular Prematurity and its Relationship to Ventricular Tachycardia in Acute Myocardial Infarction in Man," Circulation, vol. 50, pp. 529-533, 1974.

(Dell'osso,1973)
L. F. Dell'osso, "An Arrhythmia-Anomalous Beat Monitoring System," IEEE Trans. on Biomedical Eng., vol. BME-20, no. 1, pp. 43-50, January 1973.

(Dillman,1978)
R. Dillman, N. Judell, and S. Kuo, "Replacement of AZTEC by correlation for more accurate VPB detection," Computers in Cardiology, pp. 29-32, 1978.

(Feldman,1971)
C. L. Feldman, P. G. Amazeen, M. D. Klein and B. Lown, "Computer Detection of Ventricular Ectopic Beats," Computers and Biomedical Research, vol. 3, pp. 666-674, 1971.

(Feldman,1974)
C. L. Feldman, "Evaluation of Arrhythmia Detectors," Computers in Cardiology, 1974, pp. 21-27, 1974.

(Feldman,1977)
C. L. Feldman and M. Hubelbank, "Cardiovascular monitoring in the Coronary Care Unit," Medical Instrumentation, vol. 11, no. 5, pp. 288-292, Sept.-Oct. 1977.

(Feldman,1979)
C. L. Feldman, M. Hubelbank, c. I. Haffajee, and P. Kotilainen, "A New Electrode System for Automated ECG Monitoring," Computers in Cardiology, IEEE Computer Society, pp. 285-288, 1979.

(Frankel,1975)
P. Frankel, J. Rothmeier, D. James and N. Quaynor, "A computerized System for ECG Monitoring", Computers and Biomedical Research, v. 8, pp. 560-567, 1975.

(Gersch,1970)
W. Gersch, P. M. Eddy, and E. Dong, Jr., "Cardiac Arrhythmia Classification: A Heart-beat Interval--Markov Chain Approach," Computers and Biomedical Research, vol. 4, pp. 385-392, 1970.

(Gersh,1975)
W. Gersch, P. Lilly and E. Dong, Jr., "PVC Dectection by the Heart-Beat Interval Data-Markov Chain Approach," Computers and Biomedical Research, vol. 8, pp. 370-378, 1975.

(Goldman,1973)
M. J. Goldman, Principles of Clinical Electrocardiography, 8th edition, Los Altos, California: Lange Medical Pubilications, 1973.

(Goldstein,1974)
S. Goldstein, Sudden Death and Coronary Heart Disease, Mt. Kisco, New York: Futura Publishing Company, 1974.

(Gradman,1978)
A. H. Gradman and J. W. Lewis, "YALECG: A New System for Computer Analysis of Ambulatory Electrocardiograms", Computers in Cardiology, pp. 211-214, 1978.

(Gustafson,1974)
D. E. Gustafson, T. L. Johnson, and A. Akant, "Cardiogram Analysis and Classification Using Signal Analysis Techniques," C. S. Draper Lab Report R-853, September, 1974.

(Gustafson,1975)
D. E. Gustafson, A. Akant, and S. K. Mitter, "Cardiogram Analysis - Feature Extraction and Clustering Studies", Charles Stark Draper Lab Report R-936, December 1975.

(Gustafson,1975a)
D. E. Gustafson, A. S. Willsky, and J-Y. Wang, "Cardiac Arrhythmia Detection and Classification Through Signal Analysis," C. S. Draper Lab Report R-920, July, 1975.

(Gustafson,1975b)
D. E. Gustafson, A. S. Willsky, and J-Y. Wang, "Detection and Identification of Transient Cardiac Arrhythmias Using Signal Analysis," C. S. Draper Lab Report R-935, December, 1975.

(Gustafson,1978a)
D. E. Gustafson, A. S. Willsky, J-Y. Wang, M. C. Lancaster, and J. H. Triebwasser, "ECG/VCG Rhythm Diagnosis Using Statistical Signal Analysis-- I. Identification of Persistent Rhythms," IEEE Trans. Bio. Med. Electronics, vol. BME-25, No. 4, pp. 344-353, July 1978.

(Gustafson,1978b)
D. E. Gustafson, A. S. Willsky, J-Y. Wang, M. C. Lancaster, and J. H. Triebwasser, "ECG/VCG Rhythm Diagnosis Using Statistical Signal Analysis-- II. Identification of Transient Rhythms," IEEE Trans. Bio. Med. Electronics, vol. BME-25, No. 4, pp. 353-361, July 1978.

(Gustafson,1979)
D. E. Gustafson, J. Y. Wang, and S. Gelfand, "P-wave Detection and Identification using Statistical Signal Analysis", JACC, 1979, pp. 420-425.

(Han,1969)
J. Han, "Mechanisms of Ventricular Arrhythmias Associated with Myocardial Infarction," American Journal of Cardiology, vol. 24, p. 128, 1969.

(Haywood,1970)
L. J. Haywood, V. K. Murthy, G. A. Harvey and S. Saltzberg, "On-Line Real Time Computer Algorithm for Monitoring the ECG Waveform," Computers and Biomedical Research, v. 3, pp. 15-25, February 1970.

(Hermes,1980a)
R. E. Hermes, "American Heart Association Database -- BCL Working Note No. 5", Biomedical Computer Laboratory, 700 S. Euclid Avenue, St. Louis, MO 63110.

(Hermes,1980b)
R. E. Hermes and J. R. Cox, Jr., "A Methodology for Performance Evaluation of Ventricular Arrhythmia Detectors", Computers in Cardiology, 1980, pp. 3-8.

(Hermes,1980c)
R. E. Hermes, D. B. Geselowitz, and G. C. Oliver, "Development, Distribution, and Use of the American Heart Association Database for Ventricular Arrhythmia Detector Evaluation", Computers in Cardiology, 1980.

(Higgins,1978)
S. B. Higgins, R. L. Woosley, C. B. Herrin, J. L. Compton, and T. R. Harris, "A Minicomputer Based System for the Quantification of Ventricular Arrhythmias", Computers in Cardiology, pp. 355-358, 1978.

(Horowitz,1975)
S. L. Horowitz, "A Syntactic Algorithm for Peak Detection in Waveforms with Applications to Cardiography," Communications of the ACM, vol. 18, no. 5, pp. 281-284, May 1975.

(Horth,1969)
T. C. Horth, "Premonitory Heartbeat Patterns Recognized by Electronic Monitor," Hewlett-Packard Journal, v. 21, no. 2, pp. 12-20, October 1969.

(Hubelbank,1978)
M. Hubelbank, C. L. Feldman and J. B. Morrison, "A Multimorphology Computerized Arrhythmia Monitoring System", Computers in Cardiology, pp. 351-354, 1978.

(Jelinek,1974)
M. V. Jelinek, L. Lohrbauer, B. Lown,, "Antiarrhythmic Drug Therapy for Sporadic Ventricular Ectopic Arrhythmias," Circulation, vol. 49, p. 659, 1974.

(Jenkins,1976)
J. M. Jenkins, and R. C. Arzbaecher, "Computer-Based Arrhythmia Classification Utilizing the PR Interval," Computers in Cardiology, IEEE Computer Society, pp. 149-156, 1976.

(Jenkins,1978)
J. M. Jenkins, D. Wu, and R. C. Arzbaecher, "Computer diagnosis of abnormal cardiac rhythms employing a new P-wave detector for interval measurement," Computers in Biomedical Research, vol. 11, pp. 17-33, 1978.

(Julian,1964)
D. G. Julian, P. A. Valentine, and G. G. Miller, "Routine Electrocardiographic Monitoring in Acute Myocardial Infarction," Medical Journal of Australia, vol. 12, pp. 433-436, 1964.

(Kao,1980a)
R. Kao, Morphological Feature Extraction for Arrhythmia Monitoring," S.B. and S.M. Thesis in E.E., MIT, June 1980.

(Kao,1980b)
R. Kao and W. H. Olson, "Comparitive Study of QRS Morphology Classification for Arrhythmia Analysis," Computers in Cardiology, 1980, pp. 201-204.

(Kuo,1978)
S. Kuo and R. Dillman, "Computer detection of ventricular fibrillation," Computers in Cardiology, pp. 347-350, 1978.

(LeBlanc,1973)
A. R. LeBlanc and F. A. Roberge, "Present State of Arrhythmia Analysis by Computer," Canadian Medical Association Journal, vol. 108, pp. 1239-1251, May 1973.

(Lee,1978)
Y. H. Lee, unpublished ECG Data Base Annotator Channel Comparator, 1978.

(Lehman,1975)
E. L. Lehman, Nonparametric Statistical Methods Based on Ranks, San Francisco: Holden Day, Inc., 1975.

(Lindsay,1969)
A. E. Lindsay and A. Budkin, The Cardiac Arrhythmias, Chicago: Year Book Medical Publishers, 1969.

(Lindsay,1975)
J. Lindsay, Jr., and N. V. Bruckner, "Conventional Coronary Care Unit Monitoring," Journal of the American Medical Association, vol. 232, pp. 51-53, 1975.

(Lipman,1972)
B. S. Lipman, E. Massie, and R. E. Kleiger, Clinical Scalar Electrocardiography, 6th edition, Chicago: Yearbook Medical Publishers, 1972

(Lovelace,1976)
D. E. Lovelace, S. B. Knoebel, and D. P. Zipes, "Recognition of Ventricular Extrasystoles in Sedentary Versus Ambulatory Populations", Computers in Cardiology, pp. 9-11, 1976.

(Lown,1968)
B. Lown, "Intensive Heart Care," Scientific American, vol. 219, no. 1, pp. 19-27, 1968.

(Lown,1969a)
B. Lown, "The Philosophy of Coronary Care," Archives of Clinical Medicine, vol. 216, pp. 201-241, 1969.

(Lown,1969b)
B. Lown, M. D. Klein, and P. I. Hershberg, "Coronary and Precoronary Care," American Journal of Medicine, vol. 46, pp. 705-724, 1969.

(MIT-BIH,1980)
"MIT-BIH Arrhythmia Database: Tape Directory and Format Specification," Technical Report No. 10, Harvard-MIT Biomedical Engineering Center for Clinical Instrumentation, Cambridge, Mass., 1980.

(Mangiola,1974)
S. Mangiola and M. C. Ritota, Cardiac Arrhythmias, Philadelphia and Toronto: J. B. Lippincott Company, 1974.

(Mark,1979)
R. G. Mark, G. B. Moody, W. H. Olson, S. K. Peterson, P. S. Schluter, and J. B. Walters, Jr., "Real-time Ambulatory Arrhythmia Analysis with a Microcomputer", Computers in Cardiology, 1979.

(Mark,1980a)
R. G. Mark, P. S. Schluter, G. B. Moody, and S. K. Peterson, "Real-time arrhythmia monitoring using microprocessors", in Microcomputers in Patient Care, H. S. Eden, M. Eden, eds. Park Ridge NJ: Noyes Medical Publishers, 1980.

(Mark,1980b)
R. Mark, G. Moody, P. Schluter, W. Olson, and S. Peterson, "Real-time cardiac arrhythmia monitoring using microprocessors"., in Changes in Health Care Instrumentation due to Microprocessor Technology, F. Pincirolli and J. Anderson, eds., Amsterdam: North Holland Publishing Co., 1981. (presented at conference of same name in Rome, Italy, May 1980).

(Mark,1980c)
R. Mark, G. Moody, W. Olson, and S. Peterson, "Event Recorders and Future Systems in Ambulatory Monitoring," NHLBI Ambulatory ECG Recording Workshop, 1980 (Washington, D.C.).

(McClelland,1976)
K. M. McClelland and J. M. Arnold, "A QRS Detection Algorithm for Computerized ECG Monitoring," Computers in Cardiology, 1976, pp. 447-450, 1976.

(Mead,1975)
C. N. Mead, T. Ferriero, K. W. Clark, L. J. Thomas, Jr., J. R. Cox, Jr., and G. C. Oliver, "An improved Argus/H system for high-speed ECG analysis," Computers in Cardiology, pp. 7-13, 1975.

(Mead,1978)
C. N. Mead, S. M. Moore, B. F. Spenner, R. E. Hitchens, K. W. Clark, and L. J. Thomas, Jr., "Detection of Multiform PVCs Using a Combination of Time-Domain and Frequency-Domain Information", Computers in Cardiology, pp. 343-346, 1978.

(Mead,1979)
C. N. Mead, K. W. Clark, S. J. Potter, S. M. Moore, L. J. Thomas, Jr., "Development and Evaluation of a New QRS Detector/Delimeter," Computers in Cardiology, pp. 251-254, 1979.

(Meltzer,1964)
L. E. Meltzer, "The Concept and System for Intensive Coronary Care," Academy of Medicine of New Jersey Bulletin, vol. 10, pp. 304-311, 1964.

(Meltzer,1969)
L. E. Meltzer, "The Present Status and Future Direction of Intensive Coronary Care," Coronary Heart Disease, vol. 1, pp. 177-189, 1969.

(Meyer,1977)
C. R. Meyer and H. N. Keiser, "Electrocardiogram Baseline Noise Estimation and Removal Using Cubic Splines and State-Space Computation Techniques," Computers and Biomedical Research, vol. 10, pp. 459-470, 1977.

(Moody,1980)
G. B. Moody, R. G. Mark, W. H. Olson, S. K. Peterson, P. S. Schluter, and J. B. Walters, Jr., "A Microprocessor-based ECG Monitor for Ambulatory Patients" Proceedings 8th Annual Northeast Bioengineering Conference, v. 8, (1980), pp. 371-375.

(Mueller,1978)
W. C. Mueller, "Arrhythmia Detection Software for an Ambulator ECG Monitor," Biomedical Sciences Instrumentation, vol. 14, 1978.

(Nolle,1971)
F. M. Nolle and K. W. Clark, "Detection of Premature Ventricular Contractions Using an Algorithm for Cataloging QRS Complexes," in Proc. San Diego Biomedical Symp., San Diego, CA., pp. 85-97, 1971.

(Nolle,1972)
F. M. Nolle, ARGUS, A Clinical System for Monitoring Electrocardiogram Rhythms, D.Sc. Dissertation, Washington University, St. Louis, 1972.

(Nolle,1980)
F. M. Nolle, K. L Ryschon and A. E. Zencka, "Power Spectrum Analysis of Ventricular Fibrillation and Imitative Artifact", Computers in Cardiology, pp. 209-212, 1980.

(Nygards,1977)
M. E. Nygards and J. Hulting, "Recognition of ventricular fibrillation utilizing the power spectrum of the ECG," Computers in Cardiology, pp. 393-397, 1977.

(Okajima,1963)
M. Okajima, L. Stark, G. Whipple and S. Yasui, "Computer Pattern Recognition Techniques: Some Results with Real Electrocardiographic Data", IEEE Transactions on Biomedical Electronics, pp. 106-114, July 1963.

(Oliver,1977)
G. C. Oliver, K. L. Ripley, J. P. Miller, and T. F. Martin, "A Critical Review of Computer Arrhythmia Detection", in Computer Electrocardiography: Present Status and Criteria", L. Pordy, ed., Mount Kisco, NY: Futura Press, 1977, pp. 319-360.

(Otterstrom,1977)
L. Otterstrom, "Automatic Monitoring of Cardiac Arrhythmias Using Template Matching", Proceedings of the IV Nordic Meeting on Medical and Biological Engineering", pp. 66-1,66-2, June 1977.

(Pavlidis,1973)
T. Pavlidis and S. L. Horowitz, "Piecewise Approximation of Plane Curves," Proc. 1st Int. Joint Conf. on Pattern Recognition, Washington, D.C., October 1973, pp. 396-405.

(Pavlidis,1974)
T. Pavlidis and S. L. Horowitz, "Straight Line Segmentation of Plane Curves," IEEE Trans. on Computers, vol. C-23, no. 8, pp. 860-870, August 1974.

(Peterson,1980)
S. K. Peterson, unpublished STOIC data base comparator routines, 1980.

(Pordy,1968)
L. Pordy, H. Jaffee, K. Chesky, C. K. Friedberg, L. Fallowes and R. E. Bonner, "Computer Diagnosis of Electrocardiograms. IV. A Computer Program for Contour Analysis with Clinical Results of Rhythm and Contour Interpretation," Computers and Biomedical Research, vol. 1, pp. 408-433, 1968.

(Quinn,1975)
M. L. Quinn, O. M. Haring and F. J. Lewis, "Evaluation of Computer Diagnosis of Ectopic Beats Encountered in Routine Patient Monitoring", Computers in Biology and Medicine, v. 5, pp. 235-244, September 1975.

(Riordan,1976)
W. J. Riordan, "Techniques for the Detection and Classification of Cardiac Arrhythmias, with Application to Holter Tape Monitoring," S. M. Thesis, Cambridge, MA., Massachusetts Institute of Technology, 1976.

(Ripley,1975)
K. L. Ripley and R. M. Arthur, "Evaluation and Comparison of Arrhythmia Detectors," Computers in Cardiology, 1975, pp. 27-32, 1975.

(Ripley,1976)
K. Ripley and J. R. Cox, Jr., "A Computer System for Capturing Transient Electrocardiographic Data," Computers in Cardiology, IEEE Computer Society, pp. 439-445, 1976.

(Ripley,1977)
K. L. Ripley and G. C. Oliver, "Development of an ECG data base for arrhythmia detector evaluation," Computers in Cardiology, 1977, pp. 203-209, 1977.

(Ripley,1978)
K. L. Ripley, D. B. Geselowitz, and G. C. Oliver, "The American Heart Association Database: A Progress Report," Computers in Cardiology, 1978.

(Romhilt,1973)
D. W. Romhilt, S. S. Bloomfield, T. Chou, and N. O. Fowler, "Unreliablity of Conventional Electrocardiographic Monitoring for Arrhythmia Detection in Coronary Care Units," American Journal of Cardiology, vol. 31, pp. 457-461, 1973.

(Ruttiman,1976)
U. E. Ruttiman, A. S. Berson, and H. V. Pipberger, "ECG Data Compression by Linear Prediction," Computers in Cardiology, IEEE Computer Society, pp. 313-315, 1976.

(Sanders,1974)
W. Sanders, E. Alderman, P. Tedklenberg, and D. C. Harrison, "The Stanford Computer-based Arrhythmia Monitoring System," Computers in Cardiology, IEEE Computer Society, pp. 199-200, 1974.

(Sanders,1975)
W. J. Sanders, E. L. Alderman, and D. C. Harrison, "Alarm Processing in a Computerized Patient Monitoring System," Computers in Cardiology, 1975, pp. 21-26, 1975.

(Sanders,1976)
W. J. Sanders, E. L. Alderman, D. C. Harrison, and R. F. Dillman, "Computerized Processing of Artificially Paced ECG Rhythms," Computers in Cardiology, IEEE Computer Society, pp. 435-438, 1976.

(Schluter,1976)
P. S. Schluter, "An On-Line Debugging Facility for Microcomputer Systems," Technical Report TR002, Harvard-MIT Biomedical Engineering Center for Clinical Instrumentation, Cambridge, Mass., 1976.

(Schluter,1978)
P. S. Schluter, S. K. Peterson, L. C. Siegel, J. B. Walters, and R. G. Mark, "A Data Base for Arrhythmia Detector Evaluation," Proceedings of the 31st Annual Conference on Engineering in Medicine and Biology, v. 20, (1978), pg. 58.

(Schluter,1980a)
P. Schluter and R. G. Mark, "A Bedside Cardiac Arrhythmia Monitor", Proceedings 8th Annual Northeast Bioengineering Conference, v. 8, (1980), pp. 381-385.

(Schluter,1980b)
P. Schluter, R. Mark, G. Moody, W. Olson, and S. Peterson, "Performance Measures for Arrhythmia Detectors," Computers in Cardiology, 1980, pp. 267-270.

(Shah,1977)
P. M. Shah, J. M. Arnold, N. A. Haberern, D. T. Bliss, K. M. McClelland, and W. B. Clarke, "Automatic Real Time Arrhythmia Monitoring in the Intensive Care Unit," The American Journal of Cardiology, vol. 39, pp. 701-708, May 1977.

(Siegel,1977)
L. Siegel, unpublished ECG Waveform Editor, 1977.

(Specht,1974)
D. F. Specht, "Gould 9500 Arrhythmia Monitor, Computers in Cardiology, IEEE Computer Society, pp. 223-224, 1974.

(Spitz,1977)
A. L. Spitz, J. W. Fitzgerald and D. C. Harrison, "Ambulatory Arrhythmia Quantification by a Correlation Technique", Computers in Cardiology, pp. 225-231, 1977.

(Stein,1980)
I. M. Stein, J. Plunkett, and M. Troy, "Comparison of techniques for examining long-term ECG recordings," Medical Instrumentation, vol. 14, no. 1, pp. 69-72, Jan.-Feb. 1980.

(Steiner,1971)
C. Steiner, "Arrhythmia Monitoring in the Community Hospital," Presented at the Conference on Computers in Cardiology, Bethesda, Maryland, 1971.

(Stockman,1976)
G. Stockman, L. Kanal, and M. C. Kyle, "Structural Pattern Recognition of Carotid Pulse Waves Using A General Waveform Parsing System," Communications of the ACM, vol. 19, no. 12, pp. 688-695, December 1976.

(Swenne,1977)
C. A. Swenne, J. H. Bemmel, T. F. M. Relik and B. Versteeg, "A Computerized Interactive Coronary Care Unit Monitoring System", IEEE Transactions on Biomedical Engineering, pp. 63-67, January 1977.

(Thomas,1979)
L. J. Thomas Jr., K. W. Clark, C. N. Mead, K. L. Ripley, B. F. Spenner, and G. C. Oliver, Jr., "Automated Cardiac Dysrhythmia Analysis", Proceedings of the IEEE, vol. 67, no. 9, pp. 1322-1337, September 1979.

(Tomek,1974)
I. Tomek, "Piecewise-Linear Approximation with a Bound on Absolute Error," Computers and Biomedical Research, vol. 7, pp. 64-70, 1974.

(Tomek,1975)
I. Tomek, "More on Piecewise Linear Approximation," Computers and Biomedical Research, vol. 8, pp. 568-572, 1975.

(Tsui,1975)
E. T. Tsui and E. Wong, "A Sequential Approach to Heart-Beat Interval Classification," IEEE Trans. on Information Theory, pp. 596-599, September 1975.

(Vetter,1975)
N. J. Vetter and D. G. Julian, "Comparison of Arrhythmia Computer and Conventional Monitoring in the Coronary Care Unit," The Lancet, vol. 9, no. 24, pp. 1151-1154, 1975.

(Walters,1976)
J. B. Walters, Jr., "A Microprocessor Based Arrhythmia Monitor for Ambulatory Subjects," Technical Report TR004, Harvard-MIT Biomedical Engineering Center for Clinical Instrumentation, July 1976.

(Walters,1977)
J. B. Walters, Jr., "ASM80-NOVA Cross Assembler for the 8080," Technical Report TR009, Harvard-MIT Biomedical Engineering Center for Clinical Instrumentation, Cambridge, Mass., 1976.

(Watanabe,1970)
Y. Watanabe, "Automated Diagnosis of Arrhythmias by Small Scale Digital Computer," Japanese Heart Journal, vol. 11, pp. 223-,1970.

(Wheeler,1971)
L. A. Wheeler and J. V. Candy, "An On-Line ECG Waveform Monitor", Twenty First Annual Southwestern IEEE Conference and Exhibition, pp. 238-242, April 1971.

(White, 1976)
C. C. White III, "Note on a Markov chain Approach to Cardiac Arrhythmia Classification," Computers and Biomedical Research, vol. 9, pp. 503-506, 1976.

(Whiteman,1974)
J. R. Whiteman, F. A. Siegel and J. B. Brennig, "A Computer System for Continuous Real Time Monitoring of Electrocardiographic Arrhythmias," in Computers in Biomedical Engineering, v. 4, R. Stacy and B. Waxman, eds., New York: Academic Press, 1974, pp. 89-114.

(Wigertz,1964)
O. Wigertz, P. Blomqvist, J. Hulting, G. Matell, M. E. Nygards, and G. Tornkvist, "A Computer-Based System for Continuous ECG Monitoring", World Conference on Medical Informatics, pp. 761-766, July 1964.

(Wiggers,1957)
C. J. Wiggers, The Heart, Scientific American, May 1957, pp. 75-87.

(Willems,1972)
J. L. Willems and H. V. Pipberger, "Arrhythmia Detection by Digital Computer," Computers and Biomedical Research, v. 5, pp. 263-278, 1972.

(Winkle,1976)
R. A. Winkle, E. L. Alderman, J. W. Fitzgerald, and D. C. Harrison, "Treatment of Recurrent Symptomatic Ventricular Tachycardia," Ann. Int. Med., vol. 85, p. 1, 1976.

(Yanowitz,1974)
F. Yanowitz, P. Kinias, D. Rawling, and H. A. Fozzard, "Accuracy of a Continuous Real-Time ECG Dysrhythmia Monitoring System," Circulation, vol. 50, pp. 65-72, July 1975.

(Yanowitz,1976)
"A Computerized ECG Alarm System for the Coronary Care Unit," Computers in Cardiology, 1976, pp. 431-433, 1976.

(Young,1964)
Y. Young and W. Huggins, "Computer Analysis of Electrocardiograms Using a Linear Regression Technique," IEEE Trans. Bio. Med. Electronics, vol. BME-11, pp. 60-67, July 1964.

(Zoll,1956)
P. M. Zoll, A. J. Linenthal, W. Gibson, M. H. Paul, and L. R. Norman, "Termination of Ventricular Fibrillation in Man by Externally Applied Electric Countershock," New England Journal of Medicine, vol. 254, pp. 727-732, 1956.

# BIOGRAPHICAL SKETCH

Paul Schluter was born in New York (May 22,1951), and raised in California. After graduating from Pasadena High School in 1969, he attended the California Institute of Technology and received a B.S. in Engineering and Applied Science in 1973. After a summer in Europe, he then attended the Massachusetts Institute of Technology, and received the M.S. and E.E. degrees in Electrical Engineering and Computer Science in 1976. He subsequently worked for a year as the Principal Engineer of the Biomedical Engineering Division of Beth Israel Hospital in Boston, the following year as a Project Engineer at the Biomedical Engineering Center for Clinical Instrumentation at MIT, and then resumed his career as a graduate student. He is a member of Tau Beta Pi, Sigma Xi, and the IEEE.