List of Titles

Technical Information Report

AAMI TIR No. 24:1999

Acquisition and use of physiologic waveform databases for testing of medical devices

AAMI Association for the Advancement of Medical Instrumentation

AAMI TIR No. 24:1999

Acquisition and use of physiologic waveform databases for testing of medical devices

Approved 18 August 1999 by
Association for the Advancement of Medical Instrumentation

Abstract: This report defines the nomenclature, ingredients, and principles needed to develop, annotate, evaluate, and use physiologic waveform databases in developing and testing medical devices.

Keywords: waveform, physiologic, algorithms, arrhythmia

List of Titles

Published by

Association for the Advancement of Medical Instrumentation
1110 N. Glebe Road, Suite 220
Arlington, VA 22201-4795

Printed in the United States of America

ISBN 1–57020–132–3

List of Titles

AAMI Technical Information Report

A technical information report (TIR) is a publication of the Association for the Advancement of Medical Instrumentation (AAMI) Standards Board that addresses a particular aspect of medical technology.

Although the material presented in a TIR may need further evaluation by experts, releasing the information is valuable because the industry and the professions have an immediate need for it.

A TIR differs markedly from a standard or recommended practice, and readers should understand the differences between these documents.

Standards and recommended practices are subject to a formal process of committee approval, public review, and resolution of all comments. This process of consensus is supervised by the AAMI Standards Board and, in the case of American National Standards, the American National Standards Institute.

A TIR is not subject to the same formal approval process as a standard. However, a technical committee and the AAMI Standards Board do approve a TIR for distribution.

Another difference is that although both standards and TIRs are periodically reviewed, a standard must be acted upon—either reaffirmed, revised, or withdrawn—and the action must be formally approved usually every 5 years but at least every 10 years. For a TIR, AAMI consults with a technical committee about every 5 years after the publication date (and periodically thereafter) for guidance on whether the document is still useful—that is, to check that the information is relevant or of historical value. If the information is not useful, the TIR is removed from circulation.

A TIR may be developed because it is more responsive to underlying safety or performance issues than a standard or recommended practice or because achieving consensus is extremely difficult or unlikely. Unlike a standard, a TIR permits the inclusion of differing viewpoints on technical issues.

CAUTION NOTICE: This AAMI technical information report may be revised or withdrawn at any time. Because it addresses a rapidly evolving field or technology, readers are cautioned to ensure that they have also considered information that may be more recent than this document.

List of Titles

List of Titles

Contents

List of Titles

Annexes

Table

Figures

Committee representation

Association for the Advancement of Medical Instrumentation

Waveform Testing Committee

The AAMI Waveform Testing Committee developed this technical information report (TIR). Committee approval of the TIR does not necessarily imply that all committee members voted for its approval.

At the time this document was balloted, the **AAMI Waveform Testing Committee** had the following members:

Cochairs:	Carl A. Pantiskas
	Sandy Weininger
Members:	James J. Bailey, MD, National Institutes of Health
	Don Brodnick, Marquette Medical
	Richard Diefes, ECRI-MSLB
	Robert Donehoo, Critikon/Johnson & Johnson
	Stacy Gehman, Quinton Instrument
	George Moody, Massachusetts Institute of Technology
	William J. Murray, Siemens Medical
	Carl A. Pantiskas, Spacelabs Medical
	Cadathur Rajagopalan, PhD, Datascope Corp.
	William Saltzstein, Medtronic Physio-Control
	Kok-Swang Tan, PhD, Bureau of Radiation & Medical Devices Canada
	Lee Taylor, MD, Protocol Systems
	Sandy Weininger, U.S. Food and Drug Administration
Alternates:	David Browne, Critikon/Johnson & Johnson
	David Daly, U.S. Food and Drug Administration
	David Hengl, Siemens Medical
	William Midgette, U.S. Food and Drug Administration

NOTE—Participation by federal agency representatives in the development of this standard does not constitute endorsement by the federal government or any of its agencies.

Acknowledgments

The Waveform Testing Committee gratefully acknowledges the work of the Apnea Monitoring Committee in the areas of patient selection and database annotation as well as the contributions of Michael K. Leverich and Mustafa H. Sagiroglu in the area of signal processing. The committee would also like to acknowledge the works of the former cochairs: Don Brodnick of Marquette Medical, Stephen Daleo of Nellcor, and Richard Bowser, formerly of Creighton University.

Foreword

The objective of the *Acquisition and use of physiologic waveform databases for testing of medical devices* technical information report (TIR) is to provide an integrated overview of the methodology, technology, and potential pitfalls of acquiring, annotating, reconstructing, and using databases for the testing of medical devices and algorithms. It reflects the conscientious efforts of a group of concerned health care professionals, biomedical engineers, device manufacturers, and government representatives to develop a reference document for those involved in developing medical instrumentation.

Waveform databases either can be a collection of digital files stored on disk or other media or can be continuous analog data stored directly on tape media. Digital databases are often used to directly test the performance of software algorithms. Both reconstructed digital databases and analog databases are used to test subsystems or complete systems. Both database forms are valuable in the testing of medical instrumentation and algorithms and have unique advantages and disadvantages. Care must be taken to ensure that the database that is acquired or chosen to test the instrument or algorithm is suitable for its intended purpose.

Continuous patient signals can be complex in size, shape, and timing. Designing waveform sequences that totally represent all possible continuous patient signals is not feasible. Therefore, the use of waveform databases that are collected from the clinical environment is a great advantage in verifying the accuracy and performance of medical instrumentation.

This TIR is written primarily for those who collect, annotate, distribute, use, or evaluate the use of databases for developing and testing medical devices. This report is intended as a reference document for anyone involved with waveform database testing. It must be reviewed and updated periodically to integrate progressive technological developments.

Introduction

Physicians and associated health care professionals need to know how diagnostic and monitoring medical devices work for each subpopulation they encounter, including both the healthy and the diseased. Therefore, similar devices from different manufacturers must perform in a similar manner. If these health care professionals understand the common characteristics of different devices, they can predict how devices will work in particular cases and can rely on the validity of the results.

Medical device standards unify such needs and codify relevant device characteristics. A standards group evaluates a device in each aspect of its overall functionality. For standards purposes, devices are classified according to their specific medical purpose (e.g., the diagnostic electrocardiograph). Generally, such standards groups devise tests for multiple, diverse, and often quite subtle characteristics. Each individual test is then designed to quantify one or more particular aspects of the characteristics of that specific class of medical device.

All standards testing is necessarily parametric because each test quantifies some aspect of a device's characteristics. Even verification of a required label is dichotomized as "acceptable" or "not acceptable" if the device is physically inspected for conformity.

Traditional testing of medical devices has focused on certain device characteristics such as input impedance, frequency response, dynamic range, and common mode rejection. Within the confines of traditional engineering, these characteristics are easy to measure. Relevant medical device standards have codified acceptable ranges or limits for such characteristics. Assessing conformance within these limits is relatively straightforward.

Once the standard engineering parameters are measured, a device can be subjected to signals and loads that simulate actual applications to real patients, and the results can be evaluated. This process is the basis of operational testing. Even though such tests are parametric, they use simulated reality to assess actual clinical utility. Common operational tests include tests for accuracy, sensitivity and specificity of event detection, noise stress testing, and defibrillation protection and recovery.

Computerized medical devices that run diagnostic and classification algorithms on sampled analog signals are becoming increasingly prevalent. Tests for such devices demand careful extensions to the already established principles of traditional test design. For example, devices that actively sense and interpret a patient's condition must be tested for their ability to recognize certain clinical patterns (e.g., a cardiac monitor must sense the onset of ventricular fibrillation (VF) and respond with its highest priority alarm). Synthesizing waveform sequences that totally represent all possible continuous patient signals is not feasible. Hence, the need for properly documented databases of representative waveforms for testing medical devices becomes evident. For example, electrocardiogram (ECG) rhythm annotations can be determined by consensus review (e.g., MIT-BIH databases), whereas left ventricular hypertrophy requires non-ECG clinical evaluation (e.g., CSE database).

Waveforms that are used for assessment of such responses must be documented clinically or validated by expert medical opinion, according to established medical practice. Specific procedures used in validating such waveforms should be based on standards of medical practice for that purpose. Thus, if consensus review of computer-interpreted patient waveforms were medically accepted, it would be inappropriate to demand independent blind validation by isolated physicians who would be deprived of the preliminary computerized scan and interpretation. The same would be true for the CSE diagnostic database that was documented by non-ECG means, namely angiography, echocardiography, or cardiac enzyme studies.

All tests have characteristic associated costs and information yields, which are determined by the nature of each device and its intended use. Good medical device standards specify tests that appropriately evaluate the characteristics under scrutiny, yield repeatable and clinically relevant results, and impose reasonable costs on those who perform the tests.

List of Titles

AAMI Technical Information Report **AAMI TIR No. 24:1999**

Acquisition and use of physiologic waveform databases for testing of medical devices

1 Scope

This document is intended to define the nomenclature, ingredients, and principles to develop, annotate, evaluate, and use physiologic waveform databases for developing and testing medical devices. The TIR identifies issues that should be addressed in the design and development of a physiologic database. It discusses many major pitfalls that must be avoided. Annexes are included that describe several databases in detail. The database profiles that are presented here are intended to serve as a guide in the design, development, acquisition, and documentation of future waveform databases that may be used in the development and evaluation of medical devices and algorithms.

This report considers continuous electrophysiologic signals such as the electrocardiogram and electroencephalogram, as well as nonelectrophysiologic signals such as invasive blood pressure and respiratory tachograms. Medical devices that deal with intermittent data such as thermodilution cardiac output or oscillometric noninvasive blood pressure are not covered in this initial report, but may be included in a later report.

2 Normative reference

2.1 ASSOCIATION FOR THE ADVANCEMENT OF MEDICAL INSTRUMENTATION. *Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms.* AAMI EC57. Arlington (Vir.): AAMI, 1998. American National Standard.

3 Definitions

For purposes of this AAMI Technical Information Report, the following definitions apply:

3.1 ADC: Analog-to-digital converter.

3.2 aliasing: The source of distortion that may result when a signal is sampled at less than the Nyquist rate.

NOTE—Energy from higher-frequency signal components—higher than one-half the Nyquist rate—are folded into frequency components of less than one-half the Nyquist rate.

3.3 CSE: Common standards for quantitative electrocardiology.

3.4 DNRi: See instantaneous dynamic range.

3.5 FFT: Fast Fourier transform.

3.6 impedance: A measurement (in ohms) of electrical opposition to the flow of current in a given circuit.

3.7 instantaneous dynamic range (DNRi): The ratio of the largest signal a system will see to the smallest signal, usually the noise floor.

3.8 LSB: Least significant bit.

3.9 loop area: The actual surface area of a planar loop encircled by the signal pickup leads and the signal source.

NOTE—The amount of noise electromagnetically induced is directly proportional to the area of this loop. Therefore, its minimization is critical. This phenomenon applies to magnetic field coupling only.

3.10 morphology: The appearance of a waveform when viewed in the time domain or, more simply put, the waveform's shape.

3.11 Nyquist rate: The sample rate required to preserve signal content without distortion; at least two times the highest frequency component of the signal.

3.12 oversampling: An analog I/O (input oversampling) technique whereby samples are either read in or reproduced at a rate equal to some constant (the interpolation factor) times the fundamental sample rate.

NOTE—This technique is used to suppress input and output aliasing and has no bearing on ratios between signal frequency components and the recording's sample rate. It can be more clearly stated as either input oversampling or output oversampling.

3.13 quantization error: One-half the analog value range associated with the least significant bit in a system.

NOTE—For example, a 12-bit converter has 4,096 discrete states possible. If it is applied to a signal with a ± 5-volt range, there is a step size equal to 10 times the total voltage range divided by 4,096, for nearly 2.44 mV per step. Thus, any analog voltage within ± 1.22 mV of the center of each value range will be associated with exactly the same digital representation. This uncertainty is associated with finite precision quantization. Its range is twice the quantization error and is sometimes referred to as the uncertainty bit size.

3.14 power spectrum: A representation of the energy content per unit frequency. The accuracy of the estimation of a population's power spectrum depends on an ensemble average.

3.15 SNR: Signal-to-noise ratio.

3.16 sample rate: The rate at which the instantaneous value of an analog signal is quantized to one of the *N* discrete voltage levels an ADC can resolve.

3.17 VF: Ventricular fibrillation.

4 Database requirements

The following requirements have implications for database development and the use of databases use in assessing medical device performance, and should be considered when planning to build a database.

4.1 Intended use of database

The intended use of a waveform database determines its requirements and system specifications. For example, a database that is intended for training purposes needs only enough fidelity so that users can interpret the waveforms. A higher-fidelity recording system may be needed for a database that will be used for algorithm testing without reconstruction to the analog domain. If the database is used for system verification testing, then the demands on the fidelity of the recording and simulation system may be much higher. In a similar manner, the requirements for database annotation may be quite different for each intended use. A mismatch between the intended use and the resultant requirements may produce a system whose testing results are invalid or whose costs far exceed the benefits derived.

4.2 Clinical requirements

It is impractical to collect data until every expected event has been captured. Many times, it is sufficient to synthesize waveforms to augment the recorded physiologic data. This approach should be discussed before database development to facilitate appropriate decisions at each stage.

4.2.1 Population

Inclusion and exclusion criteria must be established for each clinical variable and must be traceable to the database's intended use. For example, using a database of adult signals to assess the performance of infant monitors is incorrect.

4.2.2 Study design

The study design depends on how the patients are stratified with respect to risk or disease categories. An important design problem is how to collect a sufficient number of cases in each category so that classification algorithms can be expected to achieve reasonable statistical power and accuracy. Further, if the design cannot identify signals that are pathognomonic for each category, then signal characteristics must, at least, be distinctive enough to separate the categories of patients, again, with reasonable statistical power. The design should anticipate whether parameters that are extracted from the signals will be used in a stepwise, discrete fashion or as continuous variables.

4.3 Engineering requirements

The characteristics of each signal (e.g., dynamic range, resolution, mode floor, measurement precision, and precision of temperature measurement) to be collected must be specified. The duration of data collection must be specified. Special needs in the clinical setting that may affect the architecture of the equipment must be identified.

4.4 Annotation requirements

There must be a set of rules for analyzing the recorded physiological signals, a method of administering the rules, and a system for incorporating the results into a waveform database. These rules guide the referee annotators to mark and characterize events of interest in the recorded data; these *a priori* rules must be specifically tailored to the database. The data in the database should not be used to determine the rules. The annotation methodology and implementation may place constraints on the engineering of the signal acquisition system and must be established in the planning stage before any development.

4.5 Archive requirements

The methodology for archiving the data should be considered in the planning stage before any data collection. The archive should also contain documentation of the software and hardware data acquisition systems, data formats, and methods for recovering data. Additionally, the archive database storage, maintenance, and ownership should also be addressed.

5 Waveform acquisition and synthesis

This section describes the procedures for recording waveforms within the database and developing a system to play back the waveforms.

5.1 Overview

Any database use strongly depends on the quality of the equipment that is used to record (acquire), store, and reproduce the physiological signals of interest. At each step of the process, care must be taken to ensure that the signal's integrity is maintained, particularly for databases that ultimately will be used to simulate a patient. In short, the signal that is recorded and stored in a database must be reproducible with sufficient accuracy to ensure that the device or algorithm to be tested is presented with an adequate representation of the original physiologic signal.

The validity of a multiparameter monitor test method depends on the capability to record (acquire) and reproduce high-quality physiologic signals without introducing significant error. As mentioned previously, acquisition fidelity should match the intended use of the database. A high-fidelity acquisition system would be required for reproducing physiologic signals from CD-ROM media to drive the transducer(s) of a monitor under test. However, a lower-fidelity recording system may be designed if the waveforms will be used only for display during training exercises. In either case, care must be taken at each step of the design process to ensure appropriate signal integrity.

The overall architecture of the database's data acquisition system profoundly affects the accuracy of the collected data and the reproduced waveforms during playback. Therefore, two initial tasks required to develop and to use a physiologic database are to determine the general signal requirements and to establish system specifications to meet them.

5.2 Signal issues and requirements

This section discusses the significance of general signal issues, the requirements of system design, and their effect on specifications of the system's signal acquisition, storage, and reproduction subsystems. Sufficient background theory is given where needed to justify the selected implementation.

Four signal acquisition and reproduction issues have been identified as being important for testing physiologic monitors and algorithms: distortion, relative timing (skew), maximum uninterrupted duration, and frequency translation.

5.2.1 Distortion

The most important design goal for a waveform recording and playback system is preserving the signal's waveshape. Knowledge of the distortion introduced by the signal processing system is of primary interest. A distortionless system can be described in the time domain by stating that the waveshapes must be the same between the input and output. The output of a distortionless system can be expressed as shown in equation 1,

$$y(t) = Kx(t - t_d) \quad (1)$$

where t is a term for time, K is a gain term, and t_d is a delay term and $K = 1$ and $t_d = 0$. Equation 1 demonstrates that the input and output waveforms may vary in amplitude and time reference, but the output waveshape is a magnified and time-shifted version of the input waveshape, except in the ideal distortionless case. In realizable systems, some distortion in either amplitude or time always exists. Ideally, the waveform recording and playback system should minimize the effects of these distortions.

Visual checks of waveshape are usually not sufficient to validate that a waveform was recorded and reproduced (played back) with adequate signal fidelity, because the human eye is very tolerant of distortion. A more analytical

measurement technique must be used. Time domain correlation is sometimes used, but the resulting correlation coefficient does not provide sufficient information to assess electronic system performance. The majority of signal acquisition circuits process signals according to their frequency content. Spectral distortion definitions are often used instead because they provide more meaningful information. Frequency domain measurements are sensitive enough to show signal degradation that would otherwise be undetectable using time domain measurements.

In the frequency domain, distortion refers to the alteration of signal amplitude and phase spectral content. To obtain an accurate and meaningful waveform database, the physiological signal's original spectral content must be preserved. This preservation ensures that the database is sufficiently general to test any manufacturer's medical device or algorithm. Similarly, the spectral content of any digitized signal must be preserved as much as possible upon playback in order to produce an accurate test and to obtain a fair measure of monitor performance.

To analyze signal distortion, some sort of spectral analysis must be performed on each channel during the design and testing of the signal acquisition and playback systems. Commercial spectral analyzers are available and work directly on an analog input signal. In practice, the development of signal playback systems typically lags behind development of signal acquisition systems. Thus, signals recorded with the acquisition system cannot be played back in analog form until simulators are developed, so some other means must be used for analysis. Given that the signals exist in digitized form, spectral analysis must be done using digital signal processing techniques (such as a windowed FFT) on a computer. Commercial software exists that can accomplish this task, but care must be taken to apply it properly.

5.2.2 Distortion classifications

Signal distortion can occur by changing the amplitude and phase of existing spectral components. The major types of distortion that affect waveform databases tend to be either linear or nonlinear. The primary difference between these two types of distortion is that linear distortion has a constant effect on the output signal while nonlinear distortion has varying effect on that signal. Other types of distortion (e.g., harmonic) rarely affect waveform databases and are ignored in the following discussion.

5.2.2.1 Linear distortion

Linear distortion occurs when the amplitude or phase of the signal's spectral components alter the output signal's waveshape as described in equation 1. In a distortionless system, the amplitude or phase change that occurs is 0.

Maintaining proper system frequency response is important. This maintenance places restrictions on the shape, magnitude, and phase response characteristics for the system. In particular, distortionless transmission requires flat magnitude and linear phase response within the signal passband of interest. Thus, a linear system can be described by a frequency domain transfer function.

The simplest approach to eliminate linear distortion is to design a signal acquisition system that provides a flat magnitude and linear phase response within the signal's passband of interest. Linear distortion can also be reduced or eliminated through the use of filter networks that provide equalization or spectral balancing. These techniques add complexity and are usually reserved for cases where the transmission medium is distorting and cannot be controlled.

5.2.2.1.1 Amplitude distortion

Amplitude distortion generally results when the spectral magnitude response of the signal processing system is not flat within the passband of interest, effectively filtering the signal. This filtering is commonly quantified by a "flatness" specification that can be determined by applying a swept sinusoid to the system input and by measuring the corresponding output amplitude versus the frequency. Minimizing the passband variation reduces the amount of amplitude distortion that is introduced.

5.2.2.1.2 Phase distortion

Phase distortion generally refers to the nonlinear spectral phase response within the passband of interest of a signal processing system. In the time domain, the system is often characterized by applying a step function to the input and monitoring the output step (or transient) response. Although the step response is influenced by the amplitude distortion, any overshoot or ringing (which determines the settling time) that occurs is a result of a nonlinear system phase response and, thus, provides a relative indication of the amount of phase distortion.

An alternative method that is often used for design purposes defines phase distortion in terms of delay distortion. Specifically, group delay can be determined from the derivative of the system phase response with respect to frequency:

$$T_g = -\frac{d\varphi}{d\omega} \qquad (2)$$

Equation 2 shows that a distortionless system with a linear phase response (with a negative slope) also has a constant group delay. This finding means that all spectral components are delayed by the same amount of time as they travel through the system. Minimal spectral phase distortion is desirable so that the relative skew between two signals is not frequency dependent and is a function only of time delays through the system. Having minimal spectral phase distortion allows skew to be corrected after acquisition to within the time resolution of the system.

5.2.2.2 Nonlinear distortion

Nonlinear distortion occurs when spectral components appear at the output of a system but were not present in the input signal. All linear systems have some amount of nonlinearity, so extraneous spectral content is usually present. Nonlinear distortion cannot always be reduced or eliminated once it is created, so system output spectral content must be evaluated during the design process.

The amount of distortion in a nonlinear system is generally analyzed by measuring the magnitudes of all extraneous spectral components that are present at the system output but that were not present in the input signal. This distortion includes spurious (nonharmonic) and harmonic-related components. For test purposes, the input signal that is used usually has simple spectral content so that the input and output components may be easily distinguished. Such signals include single sinusoids (for harmonic distortion testing) or summed sinusoids (for intermodulation distortion testing). Sinusoidal signals are also virtually immune from linear distortion. Note that, in each case, these sinusoidal signals must be monochromatic (e.g., each must be a pure sinusoid composed of a single frequency and without modulation).

Depending on the source of distortion, the maximum allowable magnitude of any extraneous component for each recording channel may be obtained from either the instantaneous or the total dynamic range for the associated physiologic signal. The instantaneous dynamic range refers to the maximum likely AC signal variation, whereas the total dynamic range refers to the maximum possible signal variation (including the DC component). Relative to full-scale range, any extraneous amplitude components must be attenuated below the QLSB (quantization least significant bit) level for the dynamic range of the signal of interest (this assumes that the analog-to-digital (A/D) converter resolution for the channel has been matched to the signal resolution). For example, a signal with a specified dynamic range of 12 bits should ideally have all extraneous components attenuated by at least 78 dB ([6 dB/bit x 12 bits] + 6 dB = 78 dB). Another way of stating this requirement is that the filter's reject band should be at least 78 dB below that filter's passband. Of course, this specification becomes more difficult to meet as the dynamic range increases.

The three principle sources of nonlinear distortion (e.g., amplifier saturation) in digital signal acquisition systems are nonlinear impulse responses, detector or device coupling (e.g., electrodes), and frequency aliasing (inadequate sampling rate).

5.2.2.2.1 Nonlinear transfer characteristic

Nonlinear systems have time domain transfer characteristics that alter the data's amplitude and frequency domain characteristics in a nonlinear fashion. Because of this effect, the output gain and frequency cannot be predicted from the transfer function itself. Recursive systems with feedback mechanisms are examples of this type of nonlinear system.

The amplitude of any extraneous spectral components that are generated by a nonlinear transfer characteristic should not exceed the amplitude of the input test signal. Therefore, the amplitude of all extraneous components should be below the instantaneous dynamic range of the signal of interest.

5.2.2.2.2 Coupling

Stray coupling of unwanted signals into the signal of interest may be considered a type of nonlinear distortion. The coupled signals may be from adjacent system channels (crosstalk) or from extraneous sources (internal clock noise, 60-Hz pickup, etc.). This type of interference can be minimized with proper isolation, shielding, and grounding techniques.

It is possible for the coupled amplitude level from these extraneous sources to exceed the amplitude of the desired signal. Ideally, the amplitude of any unwanted signal's spectral components should be below the total dynamic range of the signal of interest. This result is difficult to achieve for wide dynamic range systems.

5.2.2.2.3 Aliasing

Improper sampling can create nonlinear distortion when any of the rules of digital sampling theory are violated. This type of distortion is impossible to correct with postprocessing once a signal has been digitized, so the use of proper sampling techniques is very important.

The most common type of sampling distortion is called aliasing. The periodic nature of a sampled signal spectrum can lead to undesirable aliasing. If the input signal is not properly band limited with respect to the sampling rate, then high-frequency spectral components cannot be separated from low-frequency baseband components, resulting in distortion (see 5.3.1.1.2). Aliasing can also occur because stray coupling of unwanted high-frequency signals can have larger amplitudes than the signal of interest. To guarantee that no aliasing occurs, a conservative design approach requires the maximum allowable amplitude for any aliased component to be below the total dynamic range of the signal of interest. This limitation is typically accomplished using anti-aliasing filters.

5.2.3 Skew

Signal skew is the relative time displacement between any two signals. Signal skew must be controlled for two reasons. First, the selection and annotation by clinicians of significant events from recorded waveform data require signals with known time relationships. Second, many medical devices or algorithms measure multiple physiologic variables simultaneously. If these physiologic waveforms are recorded separately, they must be aligned for playback to preserve timing so that valid test results can be obtained (e.g., cardiogenic artifact rejection with impedance monitors).

As previously mentioned, signal skew can be frequency dependent if the system transfer function introduces phase distortion. Linear phase system skew takes the form of a constant time delay that is caused by time differences between signal processing paths. Time-base drift (caused by asynchronous sample clock tolerance differences between signal channels) may result in either linear or nonlinear phase system skews.

Following acquisition, constant time delays are correctable to within half a sample period by shifting the time series forward or backward as needed (de-skewing using linear interpolation). Because nonlinear time-base drift cannot be corrected after signal acquisition, it imposes additional performance and architectural requirements on any signal acquisition hardware.

The maximum uncorrectable signal skew can be determined from the acquisition system's minimum time resolution. For example, if a complex multilead signal has a minimum time resolution of 20 milliseconds (ms), then the maximum allowable signal skew is ± 10 ms. Adequate time resolution for possible data annotation purposes should be considered when determining allowable sample rate and skew. Although a ± 10 ms maximum allowable skew implies a minimum sample rate of 50 samples per second (s/s), 5.3.1.2.2 shows that interpolation methods may increase the apparent sample rate by a multiple factor of any integer. Thus, a specification for maximum uncorrectable skew does not necessarily dictate the minimum channel-sampling rate.

5.2.4 Duration

Signal duration is the maximum continuous recording length without interruption. The maximum recording duration is important for two reasons. First, it contributes to the system's storage capacity requirements (in conjunction with the number of channels and channel sample rates). Second, for systems that acquire data asynchronously, maximum recording duration places requirements on the accuracy of the sample clock frequency to meet the skew requirement above. The maximum signal duration is partly determined by the requirements of the clinical data collection protocol.

5.2.5 Frequency translation

Frequency translation is defined as a constant frequency shift over the entire signal spectrum. It is not normally a problem during signal acquisition but is a point of concern for any signal playback system. Translation occurs if a different sampling interval is used for recording a signal than is used in playing back that signal into a device.

Translation is different from distortion because it is a linear transformation that is controlled by proper specification of the sample clock frequency tolerance. Additional spectral components are not generated; the original components are merely translated in frequency (toward either zero or the Nyquist frequency). Like distortion, frequency translation is an important consideration in maintaining signal spectral accuracy.

The maximum allowable frequency translation depends on the system architecture. For a synchronous system, the allowable frequency translation should not exceed the rate accuracy of a typical medical device or algorithm for the physiological signal of interest (i.e., breathing rate, heart rate, etc.). For asynchronous systems, the maximum allowable frequency translation is determined by the maximum signal duration and skew requirements discussed above. These issues will be considered in detail in 5.3.

5.3 System design issues and requirements

This section discusses practical system implementation requirements that are related to the above signal issues. Preliminary system sampling specifications may be obtained if one uses these design guidelines.

5.3.1 Sampling theory considerations

In a waveform acquisition system, the signal of interest is represented as a time series of data points. Each point is correlated to its neighbors, so that the individual data points are not independent; together, they may be considered a signal space vector. The original sequential orientation and time spacing of the individual samples must be preserved to obtain a true and accurate representation of the waveform from which the original signal can be reproduced. For digitally acquired signals, this representation mandates that the signal data be collected and processed according to the dictates of digital sampling theory in order to avoid signal distortion. This method must also be used for reproducing and playing back signals (e.g., as in a simulator).

5.3.1.1 Nyquist's sampling theory

According to Nyquist's sampling theory, three requirements must be met when a continuous analog time signal is converted to a digital representation through an A/D conversion or ADC process:

a) The sampling must be periodic; that is, the time period (T) between any two samples must be constant.

b) The continuous time signal must be band limited.

c) The sampling rate ($1/T$) must be at least two times the largest frequency component contained in the continuous time signal.

If these requirements are met, the discrete time series that is obtained by sampling and quantizing an analog signal will be an accurate digital representation of the original analog waveform. Furthermore, near-perfect reconstruction of the original analog signal from the time series is theoretically possible by using conventional D/A (digital-to-analog) conversion (DAC) techniques.

5.3.1.1.1 Periodic sampling

Although other methods exist, periodic sampling is the most common and practical method (from a hardware implementation viewpoint) for creating a discrete representation of a continuous signal. While this method has the disadvantage of requiring band-limited signals (usually obtained using anti-alias filters), it preserves convolution and allows continuous functions to be represented with linear, discrete-time, shift-invariant systems (practical digital systems).

Periodic sampling is also the only method that allows a continuous-time signal to be accurately and completely represented by the resulting time series. In signal space, it may be said that periodic sampling produces a suitable basis function $f_n(t)$ by which a sampled time series $x(nT)$ uniquely represents the original analog signal $x(t)$, as shown in equation 3:

$$x(t) = \sum_{n=-\infty}^{+\infty} x(nT) f_n(t) \tag{3}$$

where $n = 0, \pm 1, \pm 2, \ldots$ and T is the sampling period. The basis function $f_n(t)$ is shown by equation 4.

$$f_n(t) = \frac{\sin(\pi(t - nT)/T)}{\pi(t - nT)/T} \tag{4}$$

Note that the basis function is a sinc function having the form $\sin(x)/x$. Combining equations 3 and 4 results in the following interpolation formula:

$$x(t) = \sum_{n=-\infty}^{n=+\infty} x(nT) \frac{\sin(\pi(t - nT)/T)}{\pi(t - nT)/T} \tag{5}$$

Equation 5 shows that the value of $x(t)$ at any arbitrary time t can be recovered from the discrete time points that are obtained by periodic sampling and from the corresponding basis function. No signal information is lost between sample points even though time has been quantized. This quantization equates to convolving the time series $x(nT)$ with a suitable continuous-time sinc function. The practical hardware implementation of this process consists of a D/A converter with input $x(nT)$ that is clocked at a constant rate and is followed by a low-pass smoothing filter.

The time quantization of a digitized waveform, as discussed above, should not be confused with amplitude quantization (the two concepts are radically different from a signal processing point of view). The time quantization

method determines the basis function by which a continuous signal may be represented. If this basis function is suitably chosen, it can be used with a practical hardware realization that may be used to recover the original analog waveform from its time samples. Proper time quantization (sampling) does not result in a loss of time. If this were the case, true D/A conversion would not be possible. In contrast, amplitude quantization results in a loss of information and may be thought of as a process that introduces additive noise to the signal. This noise is irreversible; only finer amplitude quantization may improve the signal fidelity.

Because of the above considerations, periodic sampling of the signals to be acquired is mandatory if an accurate database is to be produced. This requirement does not mean, however, that periodic sampling must be used to digitize the full input signal. It needs only to be used in the region of the signal of interest.

5.3.1.1.2 Periodic spectra, aliasing, and acquisition

Periodic sampling does not produce a unique signal representation unless the acquired waveform is properly band limited. Specifically, a phenomenon called aliasing occurs when a signal that is periodically sampled at a rate of $f_S = 1/T$ (also called the Nyquist rate) contains resolvable spectral components above f/2 (refer to 5.2.2.2.3). When aliasing occurs, the spectral components between f/2 and f_s "fold over" the f/2 frequency (also called the Nyquist or folding frequency) and become indistinguishable from the original base-band spectral components. Furthermore, the spectrum between zero and f_s is replicated between f_s and $2f_s$, between $2f_s$ and $3f_s$, and so on. The result is signal distortion unless the signal is properly bandlimited. This concept is depicted in figure 1.

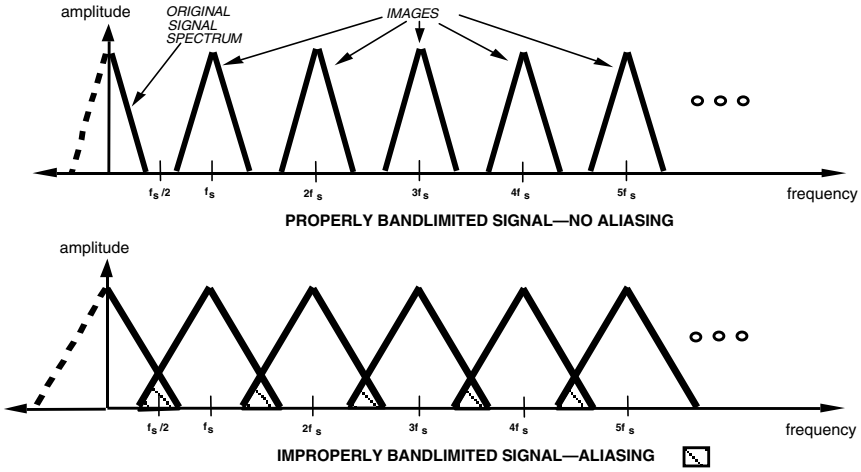


Figure 1—Sampled signal spectra and aliasing

Signal bandlimiting is generally accomplished by using an anti-alias filter before the sampling operation to prevent aliasing from occurring. Conceptually, bandlimiting is accomplished by filtering the signal to be acquired with a perfect "brick-wall" low-pass filter (no transitionband) that has a cutoff frequency equal to f/2. The process is equivalent to sampling at two times the highest frequency component in the signal. However, realizable filters have finite transitionbands, so modifications to this approach are often recommended to produce a practical system with no aliasing. This is depicted in figure 2. High-resolution acquisition systems must use oversampling techniques; otherwise, the demands placed on the anti-alias filters become insurmountable.

Real-world filters can be accommodated by incorporating oversampling techniques. If a signal is sampled at a rate much higher than twice its highest frequency component, then the folding frequency is pushed well beyond the required system bandwidth (BW). This action produces a system transitionband that is larger than the passband, which allows lower-order filters with minimal delay distortion characteristics to be used. Beyond the folding frequency, the filter must attenuate all spectral components below the resolution of the acquisition system to prevent potential distortion resulting from aliasing. This region is known as the system stopband. A diagram depicting this process is shown in figure 3. The chief disadvantage of this method of oversampling is that after A/D conversion, the input data rate to the digital system is substantially larger than the theoretical minimum, thus increasing the processing burden. This method of oversampling is also susceptible to extraneous spectral content (noise) in the system transitionband region that could corrupt the signal. This disruption can occur as a result of noise coupling to or within the acquisition system itself.

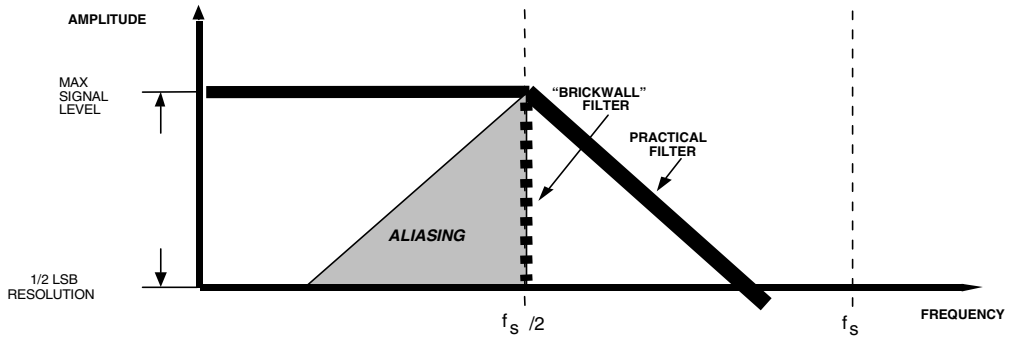


Figure 2—Ideal vs. practical anti-alias filter characteristics

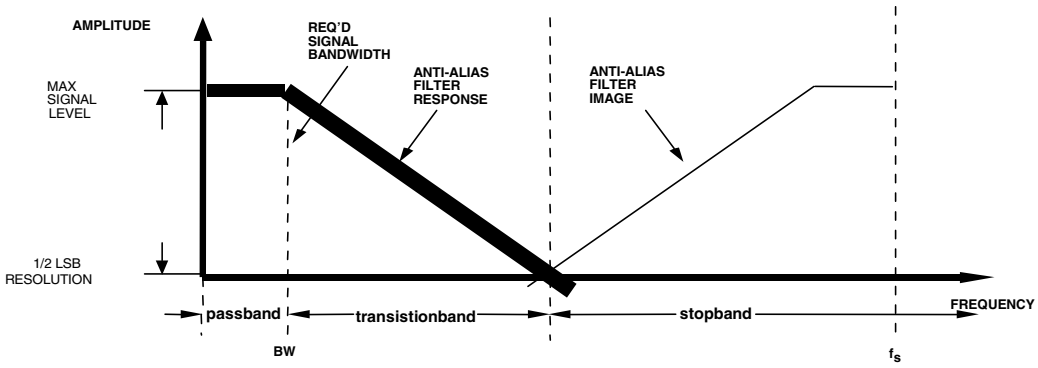


Figure 3—Anti-alias filtering for signal oversampling

An alternative method to oversampling exists and retains the advantages of oversampling but reduces the output data rate from the A/D converter. This approach filters the A/D converter output using a very high-order digital filter and then decreases the data rate at the filter output by discarding points periodically. This process is depicted in figure 4.

Output data rates that are close to the theoretical limit can be obtained using this technique because high-order digital filters can achieve very steep rolloff rates (narrow transitionbands). Linear phase is easily maintained by implementing the filter using a finite-impulse response (FIR) topology. The sample rate can be changed after filtering by periodically dropping points (decimation) because the FIR filter acts as an anti-alias filter. Although this method is complex, it is available as a complete integrated circuit known as a sigma-delta A/D converter. The one disadvantage is the long time delay of the digital filter, a delay that typically exceeds the output sample period. This delay will introduce signal skew if it is not accounted for when acquiring signal data from different channels. In summary, care must be taken when using a data acquisition system that performs signal conditioning using sigma-delta A/D converters because this technology is based on frequency translation (refer to 5.2.5).

The preceding discussion explained why spectral techniques must be applied to the evaluation and design of a digital signal acquisition system. The periodic nature of the digitized signal spectrum mandates careful design and control of the system spectral response to prevent signal distortion that results from aliasing. This design and control must be done for each channel according to the characteristics of the signal of interest. Furthermore, high-resolution acquisition methods are based on techniques that require the knowledge of system spectral content for proper system design. Ignoring those aspects will produce a system with a resolution that is lower than the specified resolution of the individual A/D converter component.

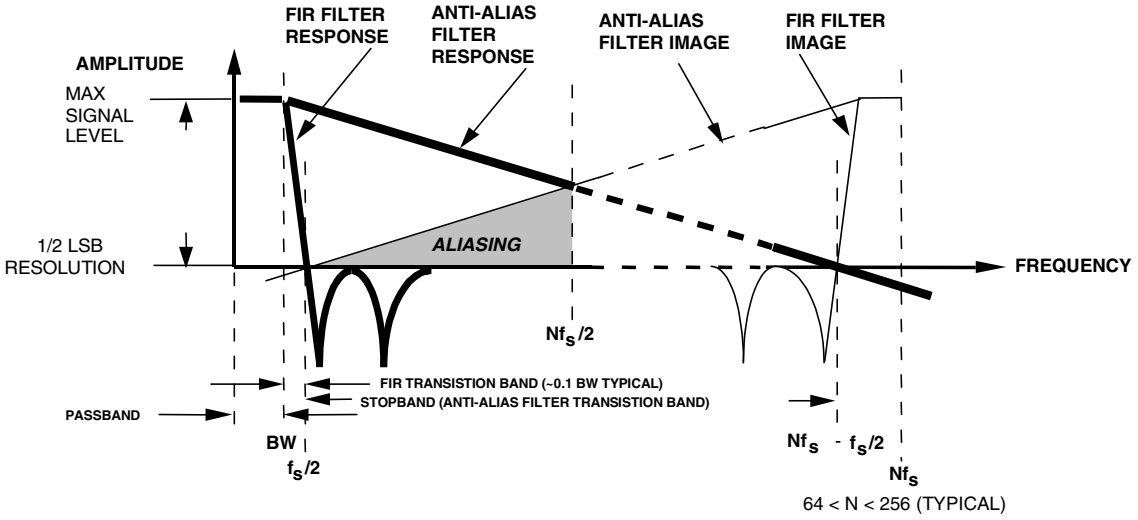


Figure 4—Anti-alias filtering for sigma-delta converters

5.3.1.2 Time series sample rate changes

When a time series is acquired by periodically sampling a signal, data points cannot be arbitrarily dropped from or added to the series. Doing so would effectively stretch or contract a waveform, thereby distorting the signal. Furthermore, the time series cannot be arbitrarily played back at a different rate than that at which it was originally acquired because the reproduced signal will be translated in frequency. Figures 5, 6, 7, and 8 illustrate this issue.

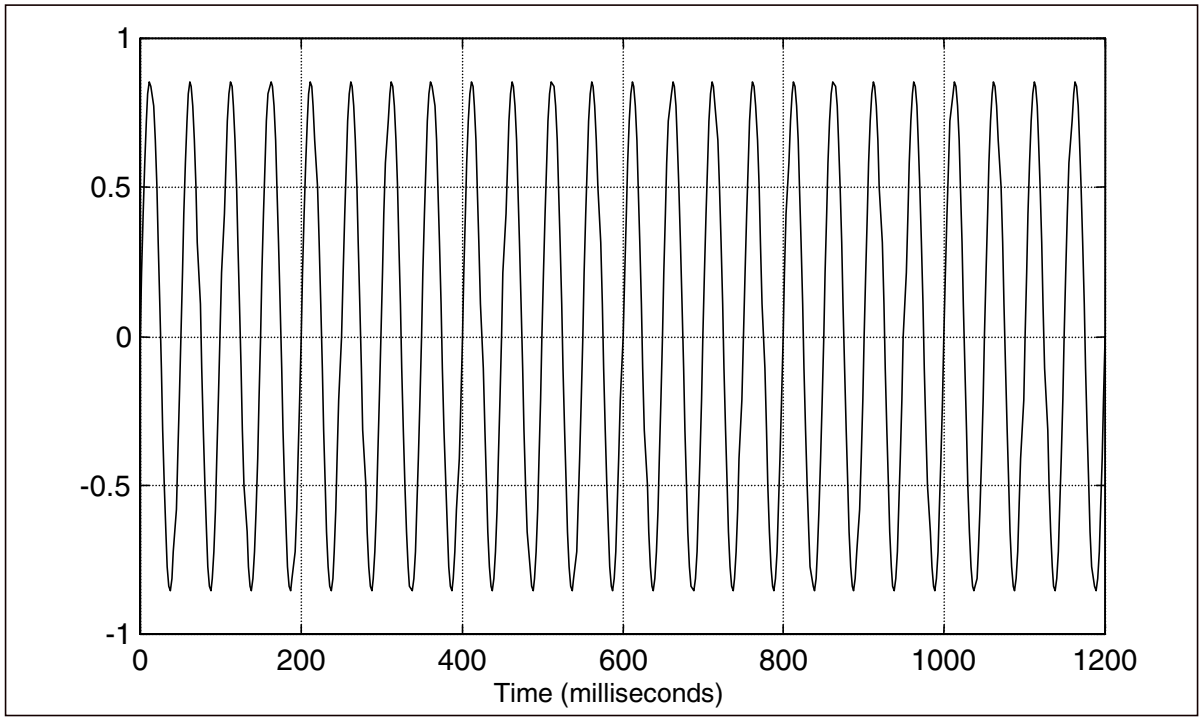


Figure 5—Monochromatic, 20-Hz sine wave sampled at 500 samples per second

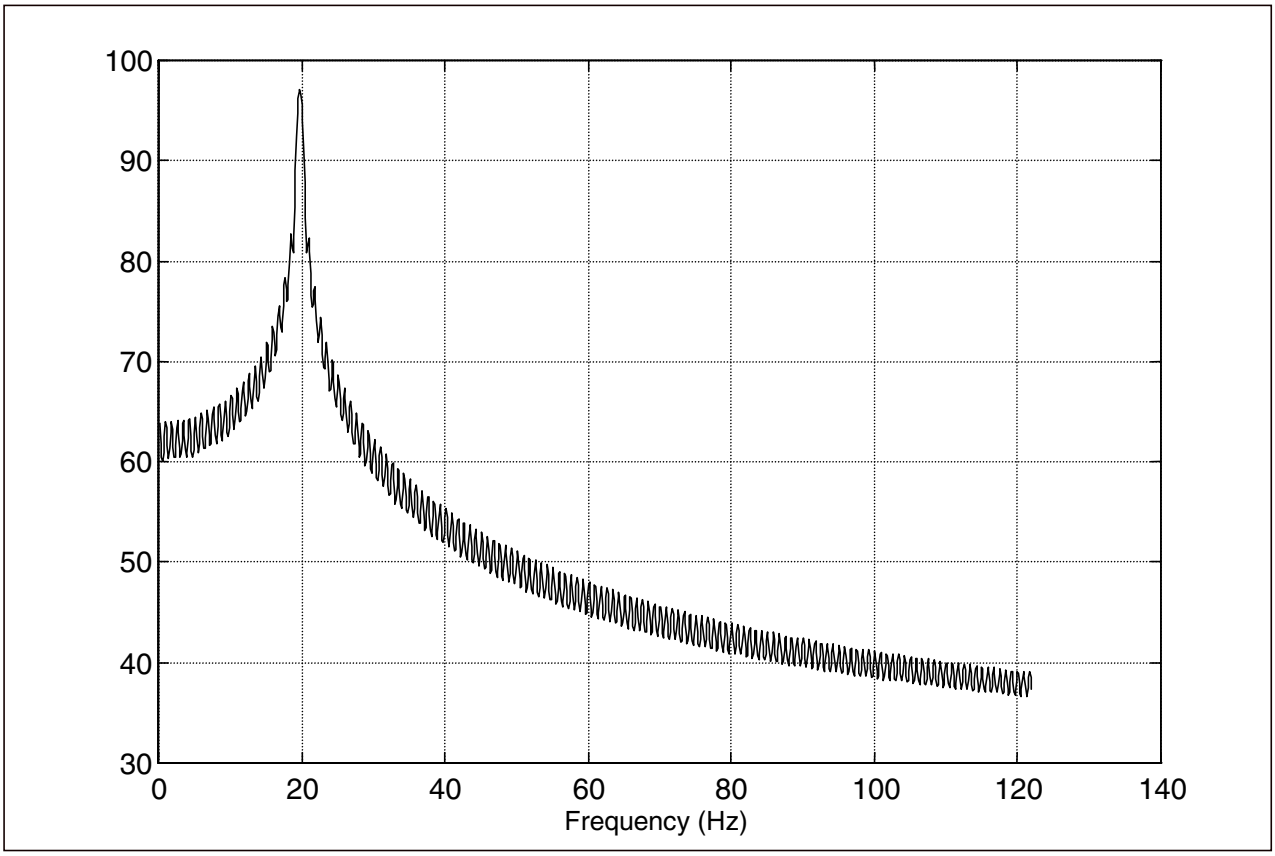


Figure 6—Amplitude spectrum for Figure 5

To understand the effect of using a nonconstant sampling rate, consider the monochromatic sinusoidal motion depicting 20-Hz oscillation. If the waveform produced by this motion is sampled and recorded at a constant rate of 500 samples per sec, the time-and-frequency domain descriptions of this wave correctly display the characteristic 20-Hz constant frequency (figures 5 and 6).

However, sampling this waveform at a nonconstant rate (either by dropping samples or introducing extra samples into the time series, and then treating the sampled data as though it were sampled at a constant rate), distorts the signal and produces erroneous results. Figure 7 shows how this process effectively modulates figure 5's waveform (a monochromatic, 20-Hz sinusoid sampled at a constant 500 samples per sec).

Figure 7 displays three distinct 400 ms windows that each contain 200 samples. Each window shows the effect of sampling figure 5's monochromatic waveform at sample rates of 500 s/s, 1,000 s/s, and 250 s/s, respectively.

The data in the 0 ms to 400 ms window depicts figure 5's original 20-Hz sinusoid. The window between 400 ms and 800 ms was obtained by resampling the previous window's data at 1,000 samples per sec (e.g., by linear interpolation that adds one extra data sample between every adjacent sample point). This 400 ms to 800 ms window gives the appearance of a 10-Hz sinusoid. The window between 800 ms and 1,200 ms appears to be a 40-Hz sinusoid, but was obtained by skipping every other data point from the original 20-Hz, 500 s/s data.

The effect that these resampling processes have on the waveform's frequency spectrum can be seen by comparing the amplitude spectra from figures 6 and 8. Figure 8 clearly shows additional frequency peaks at 10 Hz and 40 Hz.

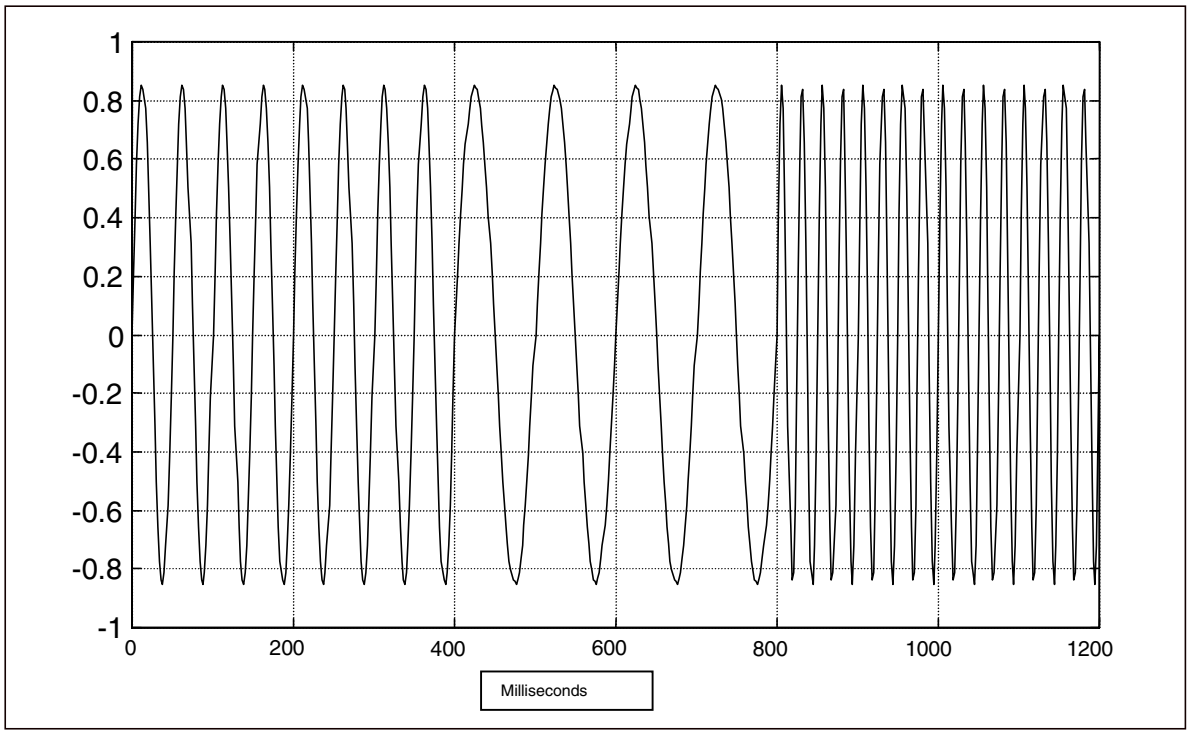


Figure 7—Effect of irregular sampling of a monochromatic waveform

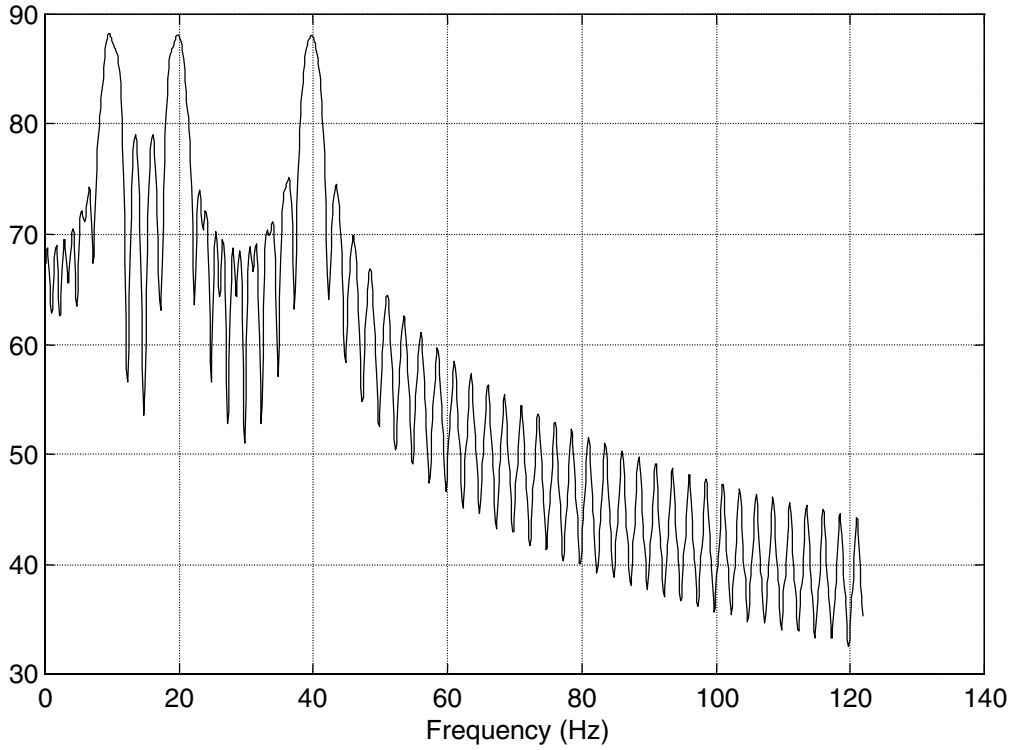
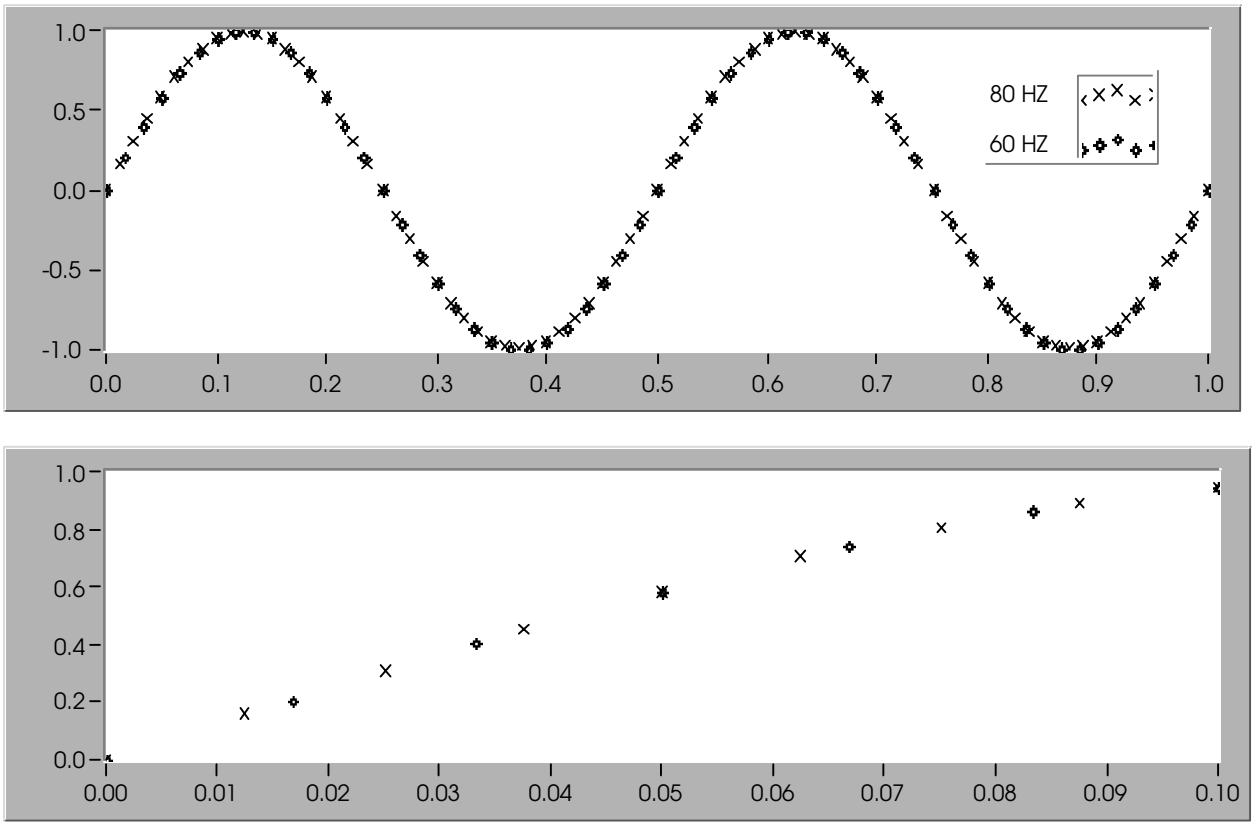


Figure 8—Amplitude spectrum for figure 7

5.3.1.2.1 Improper methods of sample rate change (resampling)

The following two examples demonstrate improper methods for signal sampling and illustrate the spectral results.



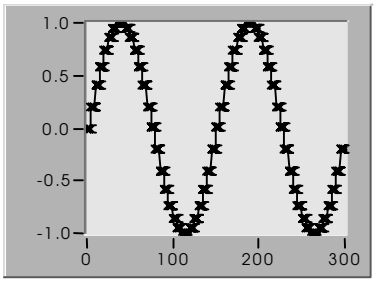
Example 1:

Two sinusoids (each previously sampled asynchronously at different rates) are to be synchronized to a common time base. Consider a 2-Hz sinusoid that was previously sampled at 60 s/s and 80 s/s (see below). Signal 1 was sampled at 60 s/s, signal 2 was sampled at 80 s/s, and the desired common sampling interval is 20 ms (which corresponds to a new sample rate of 50 Hz). Further, assume that the digital samples for signals 1 and 2 are written to a pair of latches at rates of 60 s/s and 80 s/s, respectively. This pair of latches is read every 20 ms, and the resulting values are assumed to represent the original signals at the new sample rate of 50 s/s.

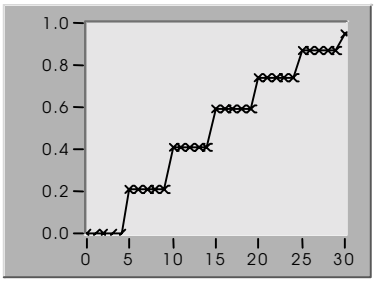
Figure 9 demonstrates this improper sample rate reduction. A 2-Hz sine wave, sampled at 60 s/s, has its output latched at a rate of 50 s/s. This sampling can be accomplished by upsampling (sample rate expansion) to 300 s/s with a zero-order hold of four additional data points, then decimating by a factor of 6 (choose every sixth data point) to achieve a 50 s/s data rate. No filtering is performed before decimation; hence, the harmonics generated during the sample rate expansion are mapped into the final signal as seen in the power spectrum in figure 9f.

NOTE—All spectra are 10-log(x).

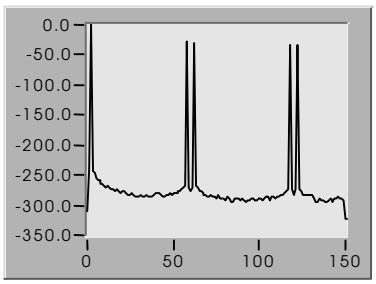
Figures 9a and 9b show the original 60 s/s sampled signal with each data point repeated four additional times, thus representing the action of the latch. The resulting data rate is 300 samples per sec. The *x* axis is shown as “points.” Figure 9c is the power spectrum of this upsampled signal as explained in the next example. Note the artifacts around the original sample rate (60 s/s) plus or minus the fundamental’s component (2 Hz). The artifacts in the decimated signal (figure 9f) are centered around the frequency where the harmonics occurred, divided by the decimation factor, 60 s/s / 6 = 10 Hz and 120 s/s / 6 = 20 Hz.



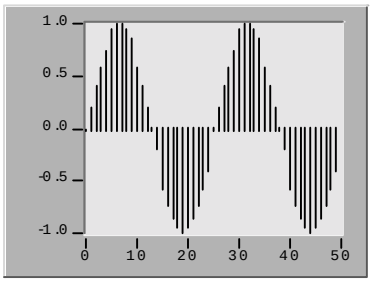
9a—Original 2-Hz sine wave (f_s = 60 s/s) resampled at 300 s/s with zero-order hold to simulate latch. Plot shows 1 second of data (300 pts).



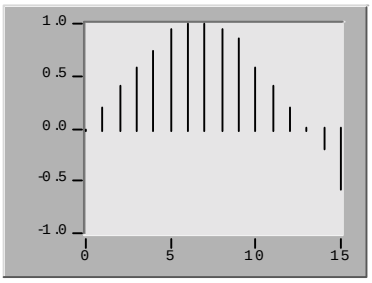
9b—Blowup of 9a to illustrate zero-order hold (i.e., latched data). Plot shows 1/10 sec of data (30 pts).



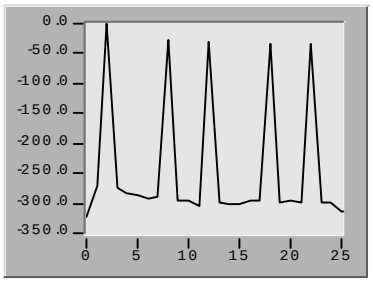
9c—Power spectrum of upsampled signal. Sample rate is expanded from 60 s/s to 300 s/s. Harmonics are (58 Hz, –29 dB), (62 Hz, –29 dB) and (118 Hz, –33 dB), and (122 Hz, –33 dB).



9d—Signal in 9b decimated by a factor of 6 to produce a data rate of 50 s/s.



9e—Blowup of 9d to show detail.



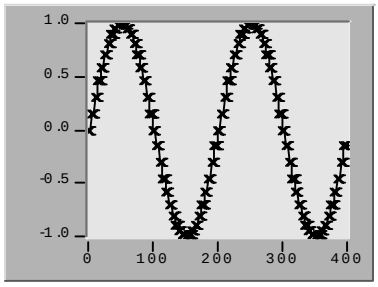
9f—Power spectrum of decimated signal (9d). Harmonics are (8 Hz, –29 dB), (12 Hz, –29 dB) and (18 Hz, –33 dB), and (22 Hz, –33 dB).

Figure 9—Improper sample rate reduction

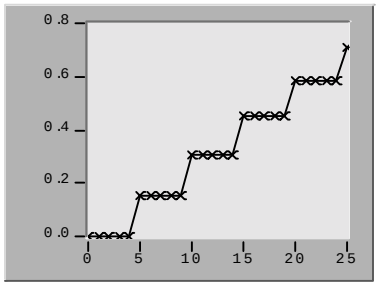
The process in figure 10 is the same sample as in figure 9, but the 2-Hz sine wave, originally sampled at f_s = 80, is expanded to 400, then decimated by a factor of 8 to 50 s/s.

NOTE—All spectra are 10-log(x).

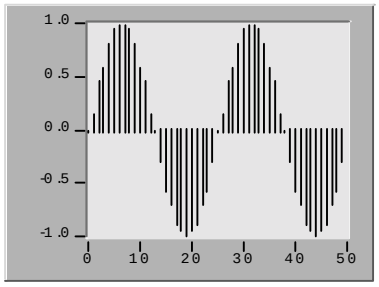
Ideally, signals that are produced by either resampling (60 s/s to 50 s/s or 80 s/s to 50 s/s) should produce a single spike in the amplitude spectrum. This spike should be at the desired signal's fundamental frequency (e.g., 2 Hz in our examples). However, because an erroneous resampling method was used here, both signals have large amplitudes at the harmonic frequencies of 8 Hz, 12 Hz, 18 Hz, and 22 Hz and have amplitudes that are only –29 dB to –33 dB down from the fundamental frequency's peak amplitude. If the instantaneous dynamic range of the original signal is 400 dB, then neither signal 1 nor 2 has acceptable fidelity because the extraneous spectral components are above the noise floor.



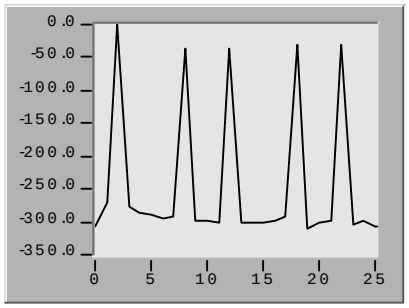
10a—Original 2-Hz sine wave (f_s = 80 s/s) resampled at 400 s/s with zero-order hold to simulate latch. Plot shows 1 sec of data (400 pts).



10b—Blowup of 10a to illustrate zero-order hold (i.e., latched data). Plot shows 1/16 sec of data (25 pts).



10c—Signal in 10b decimated by a factor of 8 to produce a data rate of 50 s/s.



10d—Power spectrum of decimated signal. Harmonics are (8 Hz, –36 dB), (12 Hz, –36 dB) and (18 Hz, –32 dB), and (22 Hz, –32 dB).

Figure 10—Improper sample rate reduction

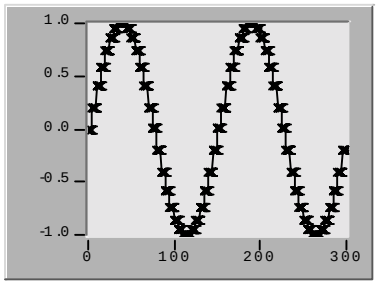
Example 2:

Two sinusoids (each previously sampled asynchronously at different rates) are to be acquired by oversampling with an asynchronous clock. Assume that signal 1, signal 2, and their respective sample rates are the same as given in example 1. Further, assume that the asynchronous sample clock is 300 Hz and that the sample values are latched and read as in example 1.

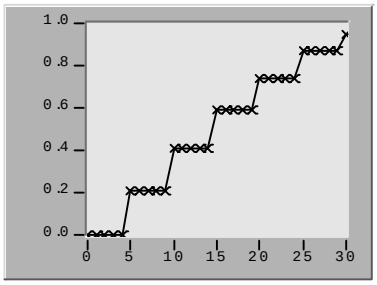
This situation is an example of improper sample rate expansion. Again, ideally, spectral signals 1 and 2 should appear as a single peak at 2 Hz (signal 1's spectra is shown in figure 11). The process in figure 11 is the same as in figures 9a through 9c.

NOTE—All spectra are 10-log(x).

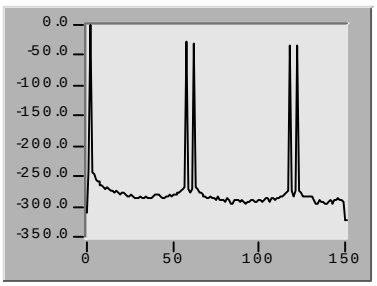
Extraneous spectral content appears at 58 Hz, 62 Hz, 118 Hz, and 122 Hz. The extraneous spectral content for signal 2 is even greater (presented in figure 12). If the instantaneous dynamic range of the original signal is again 40 dB, then neither of these acquired signals would have adequate fidelity because the extraneous components are all above the noise floor.



11a—Original 2-Hz sine wave (f_s = 60 s/s) resampled at 300 s/s with zero-order hold to simulate latch. Plot shows 1 sec of data (300 pts).

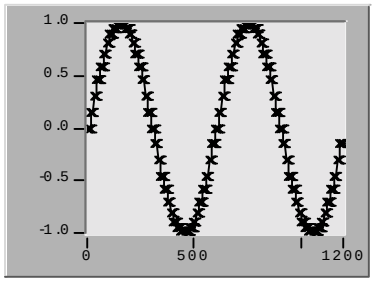


11b—Blowup of 11a to illustrate zero-order hold (i.e., latched data). Plot shows 1/10 second of data (30 pts).

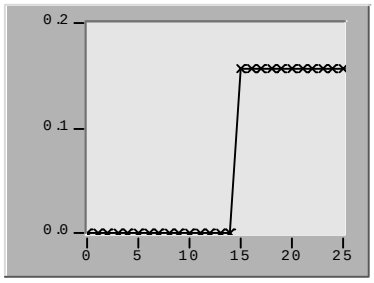


11c—Power spectrum of upsampled signal (sample rate expansion). Harmonics are (58 Hz, –29 dB), (62 Hz, –29 dB) and (118 Hz, –33 dB), and (122 Hz, –33 dB).

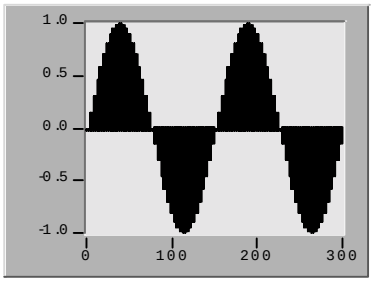
Figure 11—Improper sample rate expansion of a 2-Hz sine wave sampled at 60 s/s expanded to 300 s/s and as if it were latched



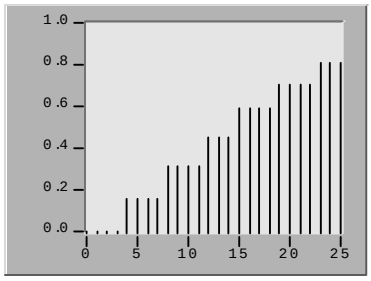
12a—Original 2-Hz sine wave (f_s = 80 s/s) resampled at 1,200 s/s with zero-order hold to simulate latch.



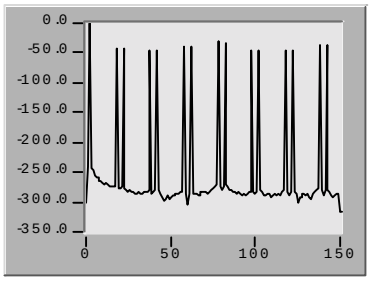
12b—Blowup of 12a showing inserted data points forming a stairstep that represents latching.



12c—Original sine wave decimated by a factor of 4 to f_s = 300 s/s.



12d—Blowup of signal in 12c to show effects of latching.



12e—Power spectrum of decimated latched data. Largest harmonics are (78 Hz, –32 dB) and (82 Hz, –32 dB).

Figure 12—Improper sample rate expansion

It is desired to read a 2-Hz sine wave sampled at 80 s/s (latched) at a rate of 300 s/s. This reading can be simulated by upsampling to 1,200 s/s with a zero-order hold of 14 additional data points. This signal is then decimated by a factor of 4 to achieve a data rate of 300 samples per sec. No filtering is performed before decimation, hence the harmonics seen in figure 12e.

NOTE—All spectra are 10-log(x).

These examples demonstrate that brute-force decimation of sampled signals can produce unacceptable levels of distortion. Spectral techniques must be used to analyze and predict the magnitude of this potential problem.

5.3.1.2.2 Decimation-interpolation methods for sample rate changes

Sample rate changes may be accomplished in a controlled manner with standard digital decimation, interpolation methods, or both. These methods maintain the relative time positions of the original sampled data points and use digital filtering to eliminate distortion that is caused by spectral imaging or aliasing effects.

The process of decimation decreases the sampling rate by an integer factor. Because the signal was originally band limited and was based on a higher sampling rate, decimation may produce aliasing unless the signal is digitally prefiltered. This digital filter is essentially a second anti-alias filter that appropriately bandlimits the signal for the new sample rate. After filtering, sample points may be dropped from the time series in a periodic manner to decrease the sample rate.

Interpolation increases the sampling rate by adding sample points between existing sample points in a periodic manner in the time series. These sample points are usually linearly interpolated from adjacent, true data values. Sampling rate changes alter the sampled data's intrinsic sampling frequency. Care must be taken during later signal processing (i.e., application of a digital filter) to ensure that this sampling rate change is properly reflected throughout the rest of the data acquisition system.

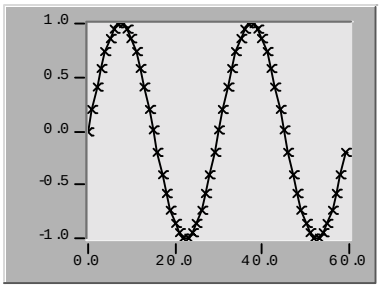
Noninteger sample rate changes may be obtained by a cascade interpolation—digital filtering—decimation operation. However, it must be possible to represent the composite fractional rate change with integer numerator and denominator values because the individual interpolation and decimation processes can produce only integer factor rate changes. That is, the numerator value is represented by the interpolation factor, and the denominator value is represented by the decimation factor. Arbitrary sample rate changes will result in arbitrary signal spectral content as illustrated in 5.3.1.2.1.

Given this background information, the proper way to change the sample rate in example 1 above using decimation-interpolation methods is as follows:

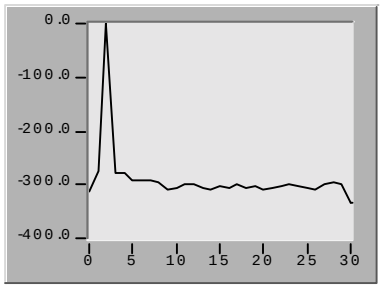
a) Interpolate (increase) the sample rate of signal 1 by a factor of 5 s/s to 300 s/s by adding four zero-valued data points between each original data point in the series. In the same manner, interpolate the sample rate of signal 2 by a factor of 5 s/s to 400 s/s (see figures 13a–d and 14a–d for signals 1 and 2, respectively).

b) Process the resulting signals with an appropriate filter to remove the harmonics generated by the interpolation process. This filter produces a sinc function in the time domain. Its characteristics are determined from knowledge of the original signal spectral content and imaging-aliasing considerations that result from the decimation-interpolation process (see figures 13e and 14e for signals 1 and 2, respectively).

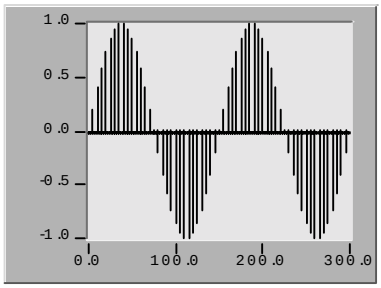
c) Decimate (decrease) the 300 s/s sample rate signal by a factor of 6 s/s to 50 s/s by retaining every sixth point in the series and throwing away all the other points. In a similar manner, decimate the 400 s/s sample rate signal by a factor of 8 s/s to 50 s/s by retaining every eighth point in the series and throwing away all of the other points (see figures 13f–g and 14f–g for signals 1 and 2, respectively).



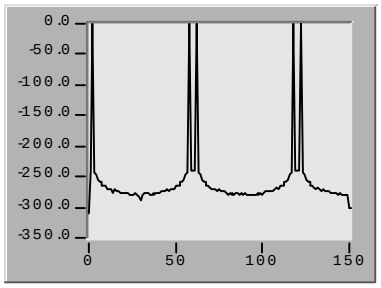
13a—Original 2-Hz sine wave sampled at 60 s/s. Plot shows 1 sec of data (300 pts).



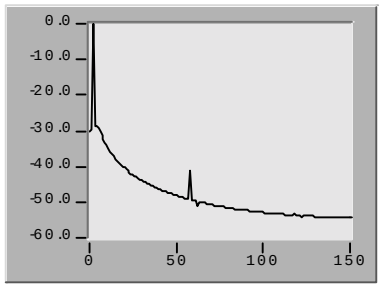
13b—Power signal in 13a.



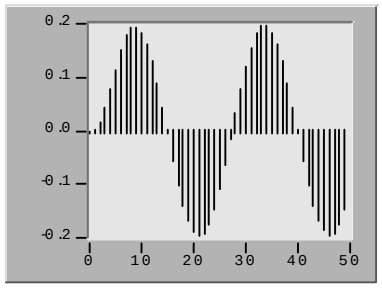
13c—Sample rate is increased by adding four zero-valued points between each original point in the series.



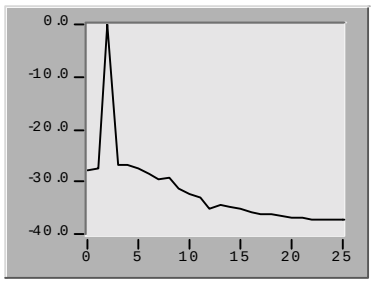
13d—Power spectrum of interpolated 300 s/s signal.



13e—Power spectrum after application of low-pass filter with cutoff at 5 Hz.



13f—Signal in 13b, decimated by a factor of 6 to produce a data rate of 50 s/s.

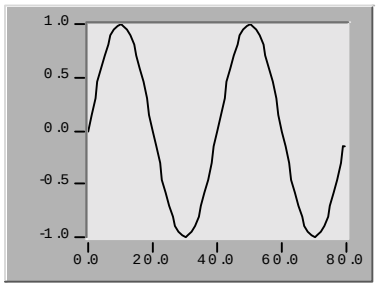


13g—Power spectrum of signal at the desired sample rate of 50 s/s.

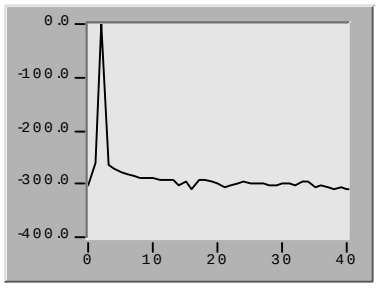
Figure 13—Proper sampled rate reduction

A 2-Hz sine wave sampled at 60 s/s is upsampled to 300 s/s by inserting four zero-valued points after each point of the original signal. This change is passed through a low-pass filter with a cutoff of 5 Hz to remove the harmonics generated by the insertions. The filtered signal is decimated by a factor of 6 to obtain the desired 50 s/s data rate.

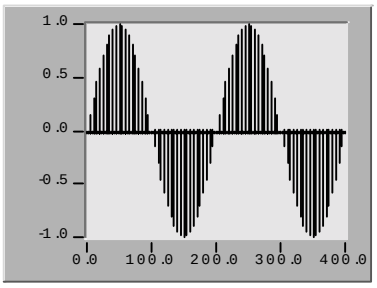
NOTE—All spectra are 10-log(x).



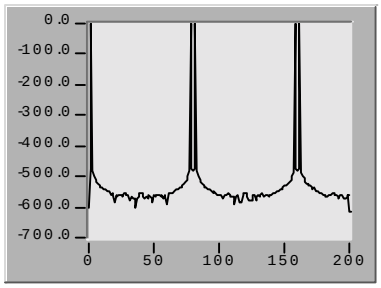
14a—Original 2-Hz sine wave sampled at 80 s/s. Plot shows 1 sec of data (80 pts).



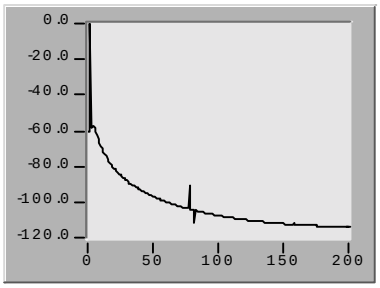
14b—Power spectrum of signal in 14a.



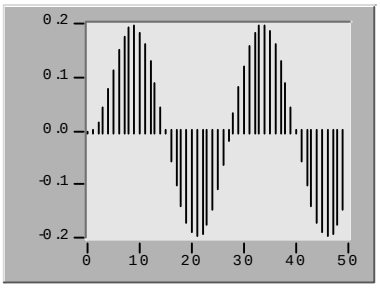
14c—Sample rate is increased to 400 s/s by adding four zero-valued points between each original point in the series.



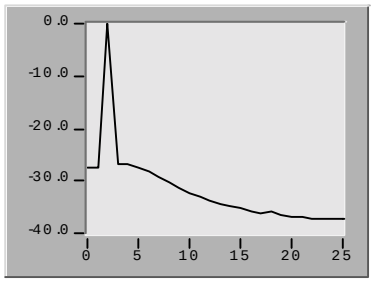
14d—Power spectrum of 14c. Harmonics are at 58 Hz and 62 Hz, and at 118 Hz and 122 Hz.



14e—Power spectrum after application of low-pass filter with cutoff at 5 Hz.



14f—Signal in 14c, decimated by a factor of 8 to produce a data rate of 50 s/s.



14g—Power spectrum of signal in 14f.

Figure 14—Proper sampled rate reduction

A 2-Hz sine wave sampled at 80 s/s is upsampled to 400 s/s by inserting four zero-valued points after each point of the original signal. This change is passed through a low-pass filter to remove the harmonics generated by the insertion. The filtered signal is decimated by a factor of 8 to obtain the desired 50 s/s data rate.

NOTE—All spectra are 10-log(x).

A practical application of a discrete decimation process was presented in 5.3.1.1.2 when the sigma-delta A/D converter was discussed. This type of converter achieves high resolution by oversampling, filtering, and decimating

the acquired data. A practical example of continuous interpolation is given by equation 5, which allows practical D/A conversion. Discrete interpolation results if the substitution $t = mT'$ is made in equation 5, where $T' < T$ and where m is the new interpolated series index.

5.3.1.3 Sample rate determination and compatibility among databases

Many existing physiologic signal databases were acquired with sample rates at multiples of 60 s/s. If users are to maintain compatibility with these databases, other databases should be acquired at similar sample rate multiples, but identical sampling rates are not required to maintain compatibility. The decimation-interpolation process described above, coupled with sampling rates that are integer multiples or submultiples, can be applied to the signal data after acquisition to obtain equivalent sample rates. Sample rate determination must be based on more pressing technical design and performance issues as described in 5.3.1.1.2.

5.3.1.4 Nonideal sampling effects

Two main sources of ideal sampling degradation must be considered in the system design: aperture jitter and aperture delay.

5.3.1.4.1 Aperture jitter

All practical periodic sample clocks have short duration timing variations from period to period. These variations are referred to as aperture (or phase) jitter. If this jitter is very small relative to the nominal sampling period, then periodic sampling may still be assumed, and the amplitude error that results from the aperture jitter can be modeled as an additive noise process.

Aperture jitter presents uncertainty in the exact sampling time. This uncertainty may be thought of as an error window (called an aperture window) that is centered on each periodic sampling occurrence. The signal to be acquired is changing in amplitude within the aperture window, and this amplitude uncertainty represents a quantization error. The maximum slew rate (dV/dt) of the sampled signal multiplied by the mean aperture window duration (or aperture jitter parameter t_a) yields the maximum expected quantization amplitude error. This amplitude error may be considered a signal noise source and can be used to calculate the signal-to-noise ratio (SNR). Assuming a sinusoidal input with frequency f, the SNR resulting from aperture jitter is

$$SNR = 20 \log\left[\frac{1}{2\pi f t_a}\right] \text{dB} \tag{6}$$

This equation shows the root mean square (RMS) signal to RMS aperture jitter noise ratio, so t_a is in units of RMS time. From equation 6, the worst-case SNR should be at the physiologic signal passband edge (cutoff frequency).

The maximum allowable sample clock aperture jitter may be determined from the instantaneous dynamic range (DNRi) of the signal of interest. The SNR from equation 6 is related to the instantaneous dynamic range according to:

$$SNR = \text{DNRi} + 1.76 \text{ dB} \tag{7}$$

If the signal is acquired with an A/D converter that is matched to the desired signal resolution, substituting equation 7 into equation 6 and solving for t_a yields:

$$t_a = \frac{1}{2\pi f 10^{\left(\frac{\text{DNRi} + 1.76 \text{ dB}}{20}\right)}} \tag{8}$$

Equation 8 can be used to specify the maximum tolerable sample clock phase jitter for a signal with a bandwidth of f (in Hz) and an instantaneous dynamic range of DNR (in –dB). Again, note that this equation is valid only for $t << T$ (sample period).

For a synchronous system, the signal with the largest bandwidth and instantaneous dynamic range will determine the overall system jitter specification. For example, substituting an ECG signal bandwidth of f = 150 Hz and a dynamic range of 72 dB into equation 8 yields a maximum allowable jitter of 218 ns RMS, which is easily attainable with present technology.

5.3.1.4.2 Aperture delay

Aperture delay is the length of time between the issuance of a sample command and the actual occurrence of the sampling operation. Signal skew occurs if the aperture delay time varies significantly between channels, because

this variation complicates the signal processing requirements. As a rule of thumb, it is acceptable to have a total signal skew value of less than or equal to the sampling interval across a multichannel set of signals. Signal skews larger than this range typically introduce a lag that affects the crosscorrelation between the acquired data channels.

Aperture delay should be characterized and included with other sources of skew delay error.

5.3.2 Architectural issues

Two types of sampling architectures may be considered for system implementation: synchronous and asynchronous. Each one has distinct advantages and disadvantages and will be considered separately.

5.3.2.1 Asynchronous sampling

An asynchronous system is relatively easy to realize because each signal is sampled independently with its own local clock. The main advantage of this architecture is hardware simplification. With this type of system, clock distribution circuitry is not required, and the signals are considered independent.

For example, the signals that are acquired for physiologic waveform databases are rarely independent because the annotation process typically requires several signals that have known time relationships. In an asynchronous system, if the phase relationships between the asynchronous sample clocks were known and remained constant over time, then the signal skew would remain constant, and the maximum skew could be controlled by carefully selecting the sample clock frequencies. This condition would allow the signals to be de-skewed to within half of a sample period for the slowest clock by simply shifting the time series after acquisition. However, the sample clocks are independent, so that no two have exactly the same frequency. The clock phases will drift relative to one another over time, resulting in a time-varying signal skew. The magnitude of this accumulated skew could eventually exceed the value that can be corrected by de-skewing. This possibility does depend on the maximum recording duration. Time-shifting the series at this point would de-skew the end of the series but would also skew the beginning of the series.

The maximum allowable clock frequency tolerance may be determined from the maximum unadjustable signal skew and the maximum recording duration specifications. Assume that any constant time delay skew that is present is negligible and that the maximum specified signal skew is equally split between two signals of interest. As an example, a signal skew of less than ± 10 ms at the end of a 5-hour (h) recording session requires that the signal sample clock frequency has individual tolerances of less than ± 0.28 parts per million or ± 5 ms/5 h. This tolerance must also be split between the recording and playback clocks, producing an accuracy of better than ± 0.14 parts per million.

Typical commercial, crystal-based oscillators have frequency tolerances of ± 200 parts per million. Ovenized oscillators can achieve tolerances of better than ± 1 parts per million, but the high cost and large size of these devices makes their use impractical for a multiple-clock, asynchronous system. Thus, time correlation between signals cannot be reasonably maintained with an asynchronous system. For this reason, signal acquisition and reproduction systems must use synchronous sampling.

5.3.2.2 Synchronous sampling

A synchronous system acquires signals with sample clocks that are derived from a single source. This system is a considerably more complex use of hardware than that used for asynchronous systems because the clock routing and the distribution circuitry require dividers and buffers. This approach's main advantage is the elimination of time-base drift between signals. This drift occurs because any frequency drift by the master clock results in the same magnitude and direction of drift for all of the sample clocks. This type of system can experience only constant skew errors because skew cannot accumulate over time.

As previously discussed, constant skew errors can result from filter mismatch, aperture delay, or postconversion processing circuits (such as FIR filters or communications hardware). Of these three sources of error, only the postprocessing errors should be significant because they may be large compared to the sampling rate. If the delay for each channel is characterized, the signals may be de-skewed after acquisition in a synchronous system. The magnitude of relative skew between channels does not matter as long as it is constant and can be corrected to within the maximum signal skew specification after acquisition.

Two methods exist for sampling in synchronous systems: simultaneous and multiplexed. With simultaneous sampling, all of the channels are sampled at the same instant. With multiplexed sampling, channel sampling is interleaved within each sample period. Multiplexed systems are usually less complex and less expensive than simultaneous sampled systems, but they suffer from signal skew (each channel sample point is offset from its neighbors by T/n, where T is the sample period and n is the number of channels).

Multiplexing skew is an important issue for annotation reproduction purposes because it produces a controlled, periodic skewing of the acquired signal points. The acquired points need not be de-skewed for graphics presentation

during annotation if they are appropriately staggered on a time grid. This grid will provide a more accurate visual presentation of the signals than de-skewing, although it will create additional graphics processing overhead. Similarly, multiplexed sampling is not a problem during playback if the signal sample points are clocked into the D/A converters using the same clock-phase relationships that were used for acquisition. In summary, signals that are acquired in a multiplexed fashion should be displayed and output in the same multiplexed fashion rather than displayed after they have been de-skewed.

The maximum allowable clock tolerance may be determined from the typical rate-accuracy of physiologic monitors. As an example, ECG signals typically have the highest repetition rate of the signals of interest. Assume a maximum rate in a polysomnography signal acquisition system of 300 beats per minute (bpm) and a rate accuracy of 1 bpm (0.0167 beats per second or Hz). Splitting the difference between recording and playback yields a maximum repetition rate error of 0.0083 Hz. A ± 200 parts per million, crystal-based clock then produces a worst-case repetition rate shift of 0.001 Hz (200*200 bpm/1E6 = 0.06 bpm). This shift is just 12% of the allowable repetition rate shift, so commercial, crystal-based clocks should be more than adequate for use.

5.3.3 Channel acquisition guidelines

General channel acquisition guidelines can be established according to the previous material presented. Because the sampling parameters for each physiologic channel are unique and depend on the signal characteristics and the A/D conversion technique, the suggested specifications that follow may have to be modified to suit the particular application.

5.3.3.1 Analog channel frequency response requirements

To prevent linear distortion, each channel must have a flat amplitude and linear phase frequency response. For example, the use of a Bessel anti-alias filtering characteristic provides a maximally flat group delay and a relatively flat magnitude response. Low-order (2 pole–3 pole) Bessel filters do not provide flat group delay across their entire passband. Higher-order filters do provide this delay and also have flatter magnitude responses. Use of low-order filters should not be a problem if the filter cutoff frequency is considerably larger (five times or more) than the physiologic bandwidth. For example, if the capture of motion artifact and noise are desired, then low-order filters can be used to acquire the wideband data. If the filter bandwidth must be reduced because wideband acquisition is not feasible, then fourth-order or greater Bessel function filters are recommended for anti-aliasing purposes.

Note that interchannel delay variation of the anti-alias filter group will contribute to signal skew errors. This variation arises because different channel filter orders are used and because component tolerances are mismatched. Generally, the magnitude of this variation should not be significant because it is normally much lower than the slowest sampling rate. However, this variation should be characterized and included with other sources of delay error.

5.3.3.2 Analog channel oversampling guidelines

The oversampling ratio (f_s /BW) and dynamic range of each channel determine the order of the anti-alias filter required for acquisition. High-order filters are difficult to design and implement because of their large sensitivities. A useful technique to limit these problems is to limit the maximum filter order and then to select an oversampling ratio that will provide the desired stop-band attenuation.

For example, assuming a Bessel characteristic, a maximum eighth-order filter is recommended. The minimum oversampling ratios needed to prevent aliasing for various channel dynamic ranges are presented in table 1. Oversampling ratios for fourth- and sixth-order filters are presented for comparison purposes only. These oversampling ratios apply to classic oversampling methods as depicted in figure 3. When one uses the sigma-delta conversion technique that is depicted in figure 4, analysis should be conducted to determine the filter order because the FIR filter characteristics and oversampling ratios vary among component vendors. Oversampling ratios that are greater than those from table 1 reduce the filter order requirements.

This table represents data for suggestive prosecution. Lower filter orders and f_s/BW ratios can be used with sigma-delta ADCs.

Table 1—Minimum f_s/BW ratios vs. A.A. filter order (for Bessel response characteristics)

DYNAMIC RANGE	f_s/BW RATIO FOR A.A. FILTER ORDER		
	4 POLES	6 POLES	8 POLES
48.2 dB (8 BITS)	14.0	10.0	8.4
54.2 dB (9 BITS)	17.4	11.0	9.0
60.2 dB (10 BITS)	20.0	13.0	9.6
66.2 dB (11 BITS)	24.0	14.0	10.6
72.3 dB (12 BITS)	28.0	15.6	12.0
78.3 dB (13 BITS)	35.1	18.0	13.0
84.3 dB (14 BITS)	43.1	20.0	14.0
90.3 dB (15 BITS)	51.1	22.0	15.0
96.3 dB (16 BITS)	59.1	25.0	16.4
102.4 dB (17 BITS)	67.2	28.0	18.0
108.4 dB (18 BITS)	75.2	32.4	20.4
MAX. RECOMMENDED FILTER ORDER			

5.3.3.3 Digital data acquisition

Some databases must acquire digital signals and event data that are generated by systems with separate clocks. This data must then be synchronized with the acquisition system sample clock. This process is not a major concern for event data provided the event can be captured with adequate time resolution. Synchronization is a problem for signal data, however, because data resampling is required, which may introduce distortion.

Two methods exist for obtaining time synchronization of such signal data: brute-force resampling and reacquisition. Brute-force resampling (discussed in 5.3.1.2.1) simply accepts the digital data at a new rate and may sample some of the original data points multiple times or miss some entirely. This rate change for the brute-force sample, therefore, introduces distortion that may be partially avoided by using reacquisition, which uses a D/A converter and smoothing filter on the digital data to obtain an analog signal. Then it reacquires the signal at the new rate using an anti-alias filter and A/D converter. This procedure is more complex than resampling and still introduces some distortion, but the distortion can be controlled better than with resampling. The best method to use ultimately depends on the characteristics of the signal of interest.

5.4 Storage

Even when the most exacting standards are adhered to, compromises must be made in both obtaining and reproducing the stored data. Compromises are required because of the limitations in the analog circuitry that interfaces with the A/D converter, the uncertainty inherent in digital storage of analog signals, and the noise floor in the analog circuitry. These problems can be reduced if the acquisition and storage systems provide resolutions well beyond the needs of the stored data.

Easy acquisition uses analog storage systems as a source for some databases. For example, the MIT-Beth Israel (MIT-BIH) Arrhythmia Database includes ECG waveforms that were obtained using Holter recording systems. Although it provides good ECG waveform types, this database cannot fully reproduce the range of noise seen in clinical settings. The disadvantage is that it was acquired using heavily AC-coupled acquisition circuitry that eliminated most slow baseline shifts and biased the data toward stable, unshifting baselines and toward few apparent body movements.

It is important that the specifications of the data acquisition, storage, and data reproduction systems be adequate for the intended use. The first consideration in collecting high-resolution physiologic data that may be used in medical instrument design and algorithm development is to reliably reproduce the data to guarantee the exact behavior of the test subjects. This accurate reproduction depends on:

a) The acquisition system's quality. (The system should collect signals that are electronically noise free over a frequency range that is much broader than the linear response range of the acquisition circuitry used in any commercial system that acquires similar signals.)

b) Data storage that provides for exact (lossless) reproduction of copies of the data without adding electronic noise to the copy.

c) The response range, accuracy, and speed of response of the system that reproduces and outputs the data.

d) Annotation and interpretation of the data by accepted clinical criteria for interpreting the patterns stored in the database. (This interpretation may require that the definitions of the behavior being reproduced be developed far more precisely and quantitatively than is usually sufficient for clinical decision making.)

An additional concern that should be addressed is the ease of distributing and reproducing copies of the data in a format that is practical, inexpensive, and yet reliable.

Once the signal data is acquired and digitized, it must be stored on some medium that allows ready access. In recent years, digital data storage on various magnetic or optical media has become commonplace. However, some existing databases are stored in analog form on types of magnetic media (usually tape).

5.4.1 Raw signals—analog databases

An analog database is composed of continuous signal waveforms that are stored on an analog medium (usually tape). Recording analog signals directly onto the analog medium avoids sampling and its resulting quantization error. Event annotations, if available, may be in the database documentation or on a parallel marker channel. Analog databases are less desirable than digital databases because they are more complex, lack a common analog medium standard, and depend on specific recording devices. Several digital databases, however, derived their recordings by digitizing an intermediary analog database.

Originally, analog tape was the only bulk storage medium available. The original efforts at recording with analog tape have been restricting and frustrating because the medium is not stable for long-term storage and may change slightly each time it is replayed. Data reproduction is also not exact because the analog circuitry adds noise on each data transfer. The potential for noise at the tape head makes exact reproduction of each playback from the same tape uncertain with respect to other individual playbacks. This last concern can be reduced by requiring multiple playbacks and by averaging the response of the equipment being tested. However, practical concerns exist: the cost in time and the need for automated methods to score instrument response that match the presented data. Analog tape is also limited by the need to hand-adjust and calibrate the input and output amplifiers, by the inconvenience when accessing data at the “middle” or “far end” of the tape, and by the inability to easily alter the sequence of data to be reproduced. Dirt on the tape or tape heads, wearing of the tape heads, and realigning and adjusting all make exact data reproduction less reliable with tape systems.

5.4.2 Waveforms—digital databases

A digital database is composed of quantized and sampled representations of continuous signal waveforms that are stored on a digital medium. Digitized records may come either from the source of the physiologic signal or from an analog database. The digitized sample values may represent a physiologic signal or some derived function of one or more physiologic signals. Event annotations, if available, are in a digital format. An event’s time of occurrence can be related to a particular digital sample. Most digital databases are contained on media (floppy disk, hard disk, CD-ROM, etc.) that can store many megabytes of data and can allow easy, rapid access to the data in any sequence desired. Digital storage media are generally considered to be a more desirable and stable means of archiving a large database because this method provides excellent retrieval with minimum loss or contamination.

Digital acquisition, storage, and reproduction systems are well suited to meeting most signal acquisition and storage requirements in normal physiologic ranges. Virtually identical digital copies can be produced at very high speeds using bulk transfer technology that also validates the exact reproduction of data files. In addition, the response of the medical device or algorithm being tested can be stored and validated by the computer that reproduces the database signals, thus simplifying automated testing.

5.5 Archive

Archive media type, drives, and access methods must be documented. Programs used exclusively to access archive media must be identified. Archive media access can occur directly through operating system commands, archive software commands, and application-specific storage, recall, and replay software. If replay or reproduction programs exist, descriptions and features of those programs must be made available. If used, the version or release number, date, manufacturer, and any special options or accessories that are used to access archive software must be documented.

5.5.1 Environmental considerations

Operating environmental conditions must be documented so that data can be reproduced, duplicated in its original form, or both. The hardware configuration must be explicitly identified with respect to the type of machine used, the

manufacturer, and the configuration. Configuration details must be documented that identify generic and name-brand specific parts. Operating system details that are necessary for archive access must be described. The version or release number, date, manufacturer, and any special options or accessories that are necessary to access archive media must be disclosed.

5.5.2 Format

Because the archive is part of a hierarchical hardware-software system, several layers of formatting are typically recommended to establish and access archive media. The operating system uses file names to access data that is stored on media; those file names generally use some established convention to identify what data is contained in the archive files. These file-naming conventions should be disclosed.

File record formats can consist of different parts, each of which should be described as part of archive documentation. File header records can contain items such as identification, physical location, channel configuration information, time, and date. Data records can contain header, identification, time information, channel, sequence information, and check data. File record formats must be documented as part of the archive.

Data formats must be described and include a detailed byte description of each type of record that is used. Code representations, encryption, and encoding must be disclosed. Formulas used for data checks also must be documented with algorithm detail.

5.5.3 Data integrity

Considerations for detection and correction of damaged data need to be documented. Damage can result from physically damaged media, damaged file directories, damaged files, and damaged header and data records. Written guidelines should detail how archive data is affected by each type of damage.

Application methods should be included to establish confidence that archive data is valid and correct. Recommended methods should be given and should explain how to identify damage, as well as how application software can identify and recover from damage. If available, methods or programs for damage recovery should be detailed. If damage recovery is not possible, recommendations should explain how to ignore or skip over bad sections of data without losing the whole archive.

5.5.4 Applicable standards

Applicable formal or de facto industry standards can be used to describe storage and maintenance requirements—environmental, media, format, and error recovery—for archived data.

5.6 Database annotation

5.6.1 Annotation process

The assessment of device or algorithm performance is made possible with the use of database annotations. Annotations are indicators of significant events occurring at a specific point in the database. For example, (a) a normal heart beat was detected 97,324 samples into the database record, (b) signal processing was interrupted because of excessive noise between samples 125,249 and 126,125, or (c) a ventricular fibrillation episode was indicated between samples 592,748 and 648,000. These database annotations are accepted as the “gold standard” for the database record and are termed “truth,” or reference, annotations.[1]

Database annotations make device or algorithm performance assessments possible because comparing the database annotations with the device or algorithm’s annotations can produce performance results.[2] It is also important to use the same set of rules when comparing performance results from different iterations of an algorithm or device because the performance comparisons are meaningful only if the results are based on a consistent set of annotations and rules.[3] A detailed description of the waveform annotation methodology is beyond the scope of this document. However, for the database to be of any practical use, the methodology must be clearly described and documented before the actual annotation. The use of multiple annotators is desirable. This will require that a

[1] The “gold standard” may change over time because errors may be noticed in the reference annotations during the database’s use. An annotation error may cause a medical device or algorithm to be evaluated incorrectly. As these errors are noticed, the database provider may update the reference annotations, so it is important for reference annotations to be identified by revision identification.

[2] The performance results that are obtained using a specific revision of database annotations should also indicate which revision of database annotations were used to obtain the results.

[3] Annotation types vary greatly (e.g., a CSE diagnostic database where a non-ECG-based diagnostic label is attached to an entire record in contrast to monitoring-type databases where individual events in each record are labeled).

methodology for resolving differences between the annotators be defined. The complexity of the data will dictate the rigor of the annotation method and the qualifications that are required of the annotators.

5.6.2 Annotation rules

Referee annotators should be provided with recordings of the studies to be annotated in either printed or computer-readable form (preferred).

Annotators should also receive specific rules and details relating to the process of performing this annotation. These rules and details must be developed before any data is acquired. The annotation rules and details identify the events that will be annotated, as well as the signals of interest, the relevant artifactual signals, and the instrumentation fault conditions. The latter items add significant value to the resulting database because they allow periods of artifact or fault to be properly included or excluded depending on the characteristics of the signal of interest.

Artifactual signals include any signal that a medical device might misinterpret. The two most commonly recognized types are motion and electromagnetic artifact. Other types exist for specific artifactual signals (e.g., cardiogenic artifact on respiration signals). Motion artifact is the term for signal changes that are produced by movement of the subject's body or body parts. Extreme care must be taken during data acquisition to identify when such movement is causing artifact. Electromagnetic artifact is typically identified and eliminated before data is initially acquired, so it is not of concern during data acquisition, and referee annotators need not consider it. Other types of artifacts include instrumentation fault conditions such as sensor malfunctions, sensor- or leads-off conditions, equipment malfunctions, and so forth. Identification of such faults also adds significant value to the resulting database.

For each physiological signal of interest, a definition of the signal's identifying characteristics should be provided, specifying the signal's minimum duration and the signal characteristics that should be used to identify the beginning and end of the signal. If several distinct signals must be considered to identify a signal of clinical interest, then the order that should be used to analyze data from different data signals should also be specified. Similar information for each type of artifact that is anticipated should be provided. There is no minimum duration for artifact or instrumentation fault conditions that occur during physiologic events. In essence, the specification of minimum duration for such events is a specification of the maximum duration for such an event that does not prevent the medical device from performing its intended function.

Any additional guidance that the annotators should consider when annotating the data should be provided. This information should include specific guidance about how to annotate periods, including signal artifacts and instrumentation malfunctions. As a general rule, data acquired from malfunctioning sensors or equipment should not be annotated.

5.6.3 Data presentation specifications

5.6.3.1 Time resolution

The time resolution, chart speed, and apparent display resolution for each type of data that is presented to an annotator must be specified. Different values may be appropriate for different signals of interest, and specific values of a particular data set may vary, depending on specific characteristics of that data. For example, respiration data may be displayed at 5 mm/s, ECG data may be displayed at 25 mm/s, EEG data may displayed at 30 mm/s or 60 mm/s, and so forth. The annotator's needs and preferences (both cultural and experiential) must be considered when choosing the time resolution that is used for individual and multiple signals of interest.

5.6.3.2 Signal synchronization

When several physiologic signals are acquired simultaneously, the resulting data presentation must faithfully reproduce the specific temporal relationships of each signal with respect to the others. For example, if body surface ECG and pulse oximetry oxygen saturation signals are acquired simultaneously, the annotator's presentation must reflect the distinct physiologic delays of transmission of the pulse oximetry saturation signal through the vascular system and the electrical transmission of the ECG signal to the body surface.

5.6.3.3 Polarity

The polarity of all signals must be specified. If possible, associate positive changes in the *y* direction with increases in amplitude (some signals such as respiration may use alternate conventions such as a positive change in the *y* direction indicating inspiration).

5.6.3.4 Signal gain of each channel

All gain factors that were applied to the data must be specified. These factors should include initial amplification and size or scaling adjustments before display.

5.7 Maintenance and distribution

Issues such as ownership and maintenance of the data, as well as of the signal acquisition and simulation system during and after testing, should be addressed early in the development stages. An appropriate person should be assigned the responsibility to correct errors and introduce new information into the database. Procedures for these tasks should be spelled out and taken into account during the database's design phase.

6 Application of waveform databases to testing

In general, databases can be very valuable within specific areas of application. The information content affects the range of uses for that database. In many cases, existing databases have neglected everything other than the immediate signal of interest (e.g., ECG) because other signals that are present are not measured or annotated. For example, the baseline ST measurement in the European-ST Database is not annotated. Shortcomings such as this example limit the usefulness of existing databases for purposes beyond their original design. However, re-annotation could correct such shortcomings.

Actual medical devices are end products. They are entirely hardware specific. After parametric measurements are finished, the device should be tested for performance in the manner in which it is used (e.g., ECG monitors should be tested with both normal and abnormal ECG input signals). For this reason, analog playback of digitally recorded waveforms from actual patients is highly desired. The analog signal usually must be reconstructed from the database's digitized data before it can be presented to the monitor. Such a playback system would constitute a "simulated patient" and could optimally test equipment under scrutiny. To accomplish this "simulation," the analog signal must be reconstructed from the digital data.

6.1 Evaluation of performance

Performance results for algorithms or devices are based on the test conditions. These test conditions include the revision that is used during the test of the algorithm or device, the reference annotations, the data files, and the set of rules used to generate the results. For performance results to be compared between the same or different devices, it is, therefore, important to generate the performance results under identical circumstances.

During testing, the algorithm or device generates an annotation record called a "test annotation" in response to the database records. Performance results are generated by comparing the database's reference (or truth) annotations with the test annotations from the algorithm or device.

This annotation comparison uses a set of rules to provide performance measures that determine how well the tested device's test annotations match the reference annotations. When algorithms or devices are to be compared using database performance results, the same set of rules must be used to generate those results. One example of such a rule is the match window found in the recommended practice for testing arrhythmia performance (see AAMI EC57:1998, *Testing and reporting performance results of cardiac rhythm and ST-segment measurement algorithms*). The requirement in EC57 for the heartbeat detection match window is that valid heartbeat test annotations (algorithm beat labels) must occur within 150 ms of the reference annotations (truth beat label). Without such a requirement, it is not possible to meaningfully compare performance results of different algorithms. For example, realistically comparing the test results of one device using a 150-ms window with those of another device using a 250-ms window is not possible. Even if the numeric values of the results are the same, the accuracy of each set of results is very different.

A database's reference annotations can change over the lifetime of a database as errors are noticed during use of the reference annotations. These errors could adversely affect the performance results. If an annotation is in error, a device could perform poorly and yet be graded well, or a device that performs correctly could be penalized for that good performance. As these errors are noticed, the database provider may generate updated reference annotations. As these updates are generated, reference annotations should be identified according to revision number or date and according to the expected change in database testing results.

Meaningful performance comparisons can be made only if the results are based on the same set of reference annotations. Similarly, errors may be noted in the data files themselves. Such errors may be corrected by the database provider (resulting in different revisions of the data files that contain the waveform data samples). As with the reference annotations, it is equally important to note that different revisions of the data files may also exist.

6.2 Test objectives

A database methodology must take into account the target of testing, whether the intent is to test either the medical device system or the diagnostic algorithm. It is desirable that systems under test allow for fully automated testing. Minimum characteristics of this system test include the following:

a) readable data from a long-term storage medium;

b) preparation of the data for output;

c) presentation of signal to the algorithm or medical device, allowing it to “operate” on the signal;

d) capture of the test record and comparison of this captured record with the database’s annotated record;

e) presentation of summary reports.

If an existing database does not include all of the information needed to meet all of these characteristics, then additional information should be “merged” into the database signal before providing it to the medical device. The database user must be extremely careful when designing this “merger” to preserve all of the information from the existing database, as well as the additional data that must be provided to the medical device. For example, the American Heart Association and Massachusetts Institute of Technology databases cannot currently be used to test cardiac monitor response to pacing because these databases do not include the information that is needed to reproduce the original pacing signal. Because of the wide variety of pacemaker detection systems, great care must be taken when evaluating the possible addition of pacing data to these databases so that all devices can be tested using the updated database. Casual misuse of databases can lead to subtle inaccuracies in the conclusions drawn from such use and can lead potentially to scientifically unsound results. All results from such casual database use beyond the original intent are suspect until proven valid. The burden of proof is on the database user.

Several databases exist that are used for the development and performance verification of arrhythmia monitoring devices and algorithms. The complete AHA (annex F) and MIT-BIH ECG (annex C) arrhythmia databases encompass roughly 1 gigabyte of data. These databases and the CSE databases (annexes A and B) may serve as the prototype for developing other databases of physiologically acquired waveforms.

6.3 Algorithm versus device testing system

Diagnostic and classification algorithms are conventionally tested as off-line, stand-alone modules using exactly repeatable data. In particular, medical detection and classification algorithms are often evaluated with recorded waveform data (see AAMI EC57:1998).

It is often convenient to use a digital database to test algorithms directly. However, several issues must be considered to ensure that the results of the test are valid. The methodology that is used to collect the data must be understood. If the bandwidth of the circuitry that is used to collect the data is lower than the frequency response of the front-end electronics that will be used to feed the algorithm in the target system, then the algorithm may not behave as expected. If the sampling rate of the system that is used to collect the data is not the same as the sampling rate of the target system, algorithm performance may not be predictable. However, resampling filters can be used to effectively change the sampling rate of the database to match the target. If the resolution of the database (number of bits in the data) is different than the target, steps must be taken to ensure that false conclusions are not drawn from the tests. The database characteristics must be well-documented so that the user can make the appropriate judgment concerning the usefulness of the database for the contemplated testing.

Algorithm testing serves well to evaluate the familiar statistical performance measures such as sensitivity, specificity, positive predictivity, and so forth. Nevertheless, this testing cannot easily assess local algorithm stability (i.e., algorithmic behavior in the presence of small changes in input measures), which can be a big problem with an inadequate algorithm.

The practical implication is that the recorded signal is simpler (i.e., cleaner, easier, smoother) than the original. If this simpler analog signal is applied to the front end of a data-gathering system, it will be retransduced, recompensated, and reconverted to a digital signal. This second transduction, compensation, and conversion process simplifies (smoothes) the signal even more. Thus, the final signal approximates the original signal but does not duplicate it.

In devices that sample analog signals, the sample clock essentially has a random-phase relationship with the physiologic processes of an actual patient. The measurements will inherently be somewhat variable. Even if exact ideal analog replicas of actual patient waveforms are applied to such devices, repetitive measurements will show small variations each time. Consequently, device repeatability may tend to reflect the local stability of its diagnostic and classification algorithms.

6.4 Sufficiency and validity

6.4.1 Sufficiency

A suite of tests within a standard is sufficient if, and only if, it establishes the following:

a) All conditions that are required for a device to operate safely and effectively for its intended purpose are satisfied.

b) No hazards that could adversely affect the safe and effective operation of the device are ignored.

This latter requirement is rarely satisfied completely. Usually, a standard will attempt to treat the most likely hazards, not remote possibilities. The estimation of hazards is usually done by groups of individual experts using their combined, accumulated experience. In practice, those groups work to reach a consensus regarding their best estimation of a sufficient suite of valid tests to properly characterize that specific class of medical device for safety and efficacy.

No individual expert is omniscient, and, therefore, neither is any standards group. As standards are applied in practice over time, it may become evident that a particular standard is not as good an approximation as it needs to be. For this reason and because medical device technology continues to evolve, standards groups should convene periodically to reevaluate their existing standards.

6.4.2 Validity—correct conclusions from results

A valid test is one that is both accurate and appropriate for its intended purpose. Both of these conditions are necessary because a failure in either regard renders a test invalid.

A test must control for all potential sources of error, at least well enough that the results are usable for the intended purpose. For example, it would not be a good idea to decide if a letter needs one or two stamps by weighing it on a highway truck scale. The test has insufficient resolution.

Tests for use in medical equipment standards will measure device suitability for an indicated use. Consequently, all common experimental precautions should be observed. For example, tests for a device's ability to recognize medically documented patterns must be substantially free from errors when the device reproduces those patterns. Each test that is applied by a standard should be carefully analyzed to ensure that it is free of any experimental error in the particular circumstances in which it will be applied.

A test must measure the characteristic under scrutiny. For example, it would be inappropriate to test devices for risk currents by weighing them.

Medical device standards should set minimums for product safety and efficacy. Any tests used to determine safety and efficacy should be valid. Therefore, contributors to medical device standards should assess both the accuracy and appropriateness of each proposed test in terms of the device purpose, operating environment, and error sensitivity. Also, it may be difficult to obtain large numbers of clinically useful waveforms (e.g., different ECG arrhythmias). This difficulty may occur, in part, because of the transient and unpredictable nature of the arrhythmias, as well as the cost, size, and repeatability of the recording equipment.

6.4.3 Precautions, limitations, interpolations, and interpretations (limited scope of results)

The application of the principles of this TIR to guide the design of tests for a particular device or algorithm must be performed by a competent, informed technologist who analyzes the limitations of the particular database that is intended for use and who determines its suitability to the task at hand. Further, the results of the testing require proper analysis to avoid drawing improper and inappropriate interpretations and conclusions about the accuracy and predictive utility of the device or algorithm.

The results of database performance represent a grading of the combined decisions that were made concerning the device and that affect the device's behavior. Although it is tempting to use these results to imply a certain level of clinical performance, it should be recognized that databases represent inherently small and biased sample sets. Because databases do not model exactly the clinical environment in all situations, the clinical performance of a device is likely to differ from database performance results. Tests have shown, for example, that an algorithm that is described as 55% sensitive and 95% specific in a hospital population for left ventricular hypertrophy will be only 28% to 32% sensitive in the Framingham population. The relevance of database test results is, therefore, inherently limited. It is common practice to use statistical descriptions of the results to address this limitation and justify clinical relevance.

However, the reliance on statistical descriptions can be overemphasized and lead to interpretation errors. For example, a device may perform well during some portions of the database and poorly during others. Similarly, it may perform better with one database than with another. If this anomaly occurs, performance results for that particular data set will likely identify the situation, though gross summary statistics are not likely to identify these difficulties unless they happen on a large scale. The challenge represented by that section of the database may be important in a given clinical situation. If this situation is true, then clinical performance for the intended use of the device may not be fairly represented in the gross summary statistics. Careful examination of individual test results will show, with greater accuracy, the actual performance in a given situation. This detailed examination will likely show that the device works better in some situations than in others. This example shows only one illustration of why summary statistics are generalizations of the device performance and cannot be treated as accurate in each and every clinical circumstance. Other situations that challenge the interpretation of performance results may arise depending on the device, the database used, and the procedure used to generate results.

Some devices allow users to optimize performance through various settings. These settings allow the device or algorithm to establish a trade-off between sensitivity and specificity. A lower threshold may achieve higher sensitivity, but at the expense of a lower specificity. These trade-offs can be evaluated only by considering the device's characteristics and its actual intended use.

The specific application of the device or algorithm must be compared against the database relevancy. Consider that in a coronary care unit, changes over a period of hours or days are important. Yet, in an outpatient monitoring setting, changes should be assessed that occur over weeks to months. An epidemiological study may span many years. The same database may be used in the initial testing, but the relevance of the results could be vastly different for each application.

Statistical tools classically have been able to produce useful insight when used correctly and total confusion when improperly applied. Results of database testing are no different. Careful, thoughtful experimental design, as well as careful, thoughtful analysis and interpretations of results, is necessary before drawing any useful conclusions.

Annex A
(informative)

CSE ECG Reference Library (Measurement Database)

Purpose: To provide a uniform database for testing ECG computer program detection and measurement of ECG waveforms.

Completion date: Spring 1985

Source: Paul Rubel, Directeur
Unite de Recherches sur
L'Activité Electrique du Coeur
Hopital Cardiologique
I.N.S.E.R.M. Unit 121
BP Lyon-Montchat
69394 Lyon Cedex 3
France

Telephone: 33-4-723-57372
Telefax: 33-4-723-41876
E-mail: rubel@insa.insa-lyon.fr

A.1 Acquisition methodology

Five or ten seconds of simultaneous 3-channel, 8-channel (12-lead), or 11-channel (15-lead) ECG data were supplied by five European cardiology centers. The 3-channel recorders grouped the leads as follows: (I, II, II); (aVR, aVL, aVF); (V1, V2, V3); (V4, V5, V6); and (X, Y, Z). The sampling rate was 500 per sec with a minimum resolution of 5 µV (at least 11 bits per sample).

Artificial ECGs were created by picking one beat of each of the lead groups of the original ECG recordings and by constructing strings of identical beats with a constant RR interval. The selected beats were chosen by eye in such a way as to be close to the dominant beat with a signal-to-noise ratio as high as possible; in other words, with as little baseline shift, noise, or artifact as possible.

The original ECG recordings and their corresponding artificial, "uni-beat" ECGs were divided into two data sets. The first is available in detail to be used as a learning set. Only overall statistics will be released for the second, allowing it to be used as a testing set.

For researchers to study the effect of beat-to-beat variability, two additional beats were selected. The first was selected as above and its adjacent beat was selected. Another set of artificial "duo-beat" ECGs was constructed. Again, the set was divided into learning and testing data sets.

To study the effects of noise, a limited data set of 10 selected original ECGs and their corresponding artificial ECGs with seven levels of noise for each set was constructed. Level 1 added no noise to the original data. Levels 2–4 added 15 µV, 25 µV, and 35 µV RMS high-frequency noise. Level 5 added 50 µV point to point (approximately 18 µV RMS) of 50-Hz sinusoidal signal. Level 6 added sinusoidal 0.3 Hz, 0.5 µV peak. Level 7 added trianguloid 0.3 Hz, 0.5 µV peak. Level 8 added sawtooth 0.3 Hz, 0.5 µV peak.

A.2 Annotation methodology

Each of five referees (board-certified cardiologists with extensive experience in reading ECGs) received a hard copy on Mingograph paper of 5 beats per ECG (one for each lead group). Beats were enlarged five and ten times in scale and time, and a time-reference channel was written on a fourth channel. Referees were asked to indicate point estimates, as well as upper and lower confidence limits, for the onsets and offsets of P, QRS, and T waves. In addition, the referees had to provide a wave morphology description; for example, (P + QRSR' T+ i.e., positive P and T waves and an R' after the QRS complex). Together with the beats written out at 500 mm/s paper speed, recordings were delivered of low-pass filtered ECGs (3 dB point at 15 Hz and zero output at 35 Hz) at a paper speed of 250 mm/s. A "standard" ECG recording of the whole tracing was also given, with an indication of the selected beats. Referees made all beat markings on the 500 mm/s paper speed recordings, whereby sample indications in the fourth channel were at 1 mm apart. The time locations of these markings were manually read in the coordinating center and transferred into a computer for statistical analysis.

A.3 Storage and distribution methodology

Raw, digital ECG data was stored on digital tapes. The transfer of raw, digital ECG data to media that were appropriate for each user was accomplished at the CSE Data Center in Leuven, Belgium. Results of referees'

measurements were also stored in digital media. For the learning data set, these measurements were distributed to the participants (users).

A.4 Signal reconstruction methodology

If the user's program used a digital-sampling rate other than 500 samples per sec, the user had to reconstruct the digital ECG using appropriate interpolation schemes. (No program used sampling rates above 500 s/s.)

A.5 Strengths and weaknesses of the database

Many European medical centers contributed data. Hence, the CSE database contains a rich collection of cases with both pathologic electrophysiology and normal cases. The uni-beat artificial ECGs allow users to test the reproducibility of their algorithms with identical digital data. The duo-beat artificial ECGs allow testing for the effects of beat-to-beat variability. The artificial ECGs (uni-beat and duo-beat) are also adaptable for testing programs that attempt to select a single representative beat, programs that take statistics on the measurements of several beats, and programs that form a template beat from many beats in order to extract measurements. After tuning their programs on the learning data set, users could determine how well their programs would ultimately perform by receiving summary statistics from the performance on the testing data set.

A.6 References

THE CSE WORKING PARTY, Willems JL, *et al.* Common Standards for Quantitative Electrocardiology: CSE project phase I. *Computers In Cardiology*, IEEE 82CH1814–3, Los Angeles (Cal.): 1982, pp. 69–74.

WILLEMS JL, ARNAUD P, VAN BEMMEL JH, *et al.* Assessment of the performance of ECG computer programs with the use of a reference database. *Circulation*, 1985a, vol. 71, pp. 523–534.

WILLEMS JL, ARNAUD P, VAN BEMMEL JH, DEGANI R, *et al.* Establishment of a reference library for evaluating computer ECG measurement programs. *Comut. Biomed. Res.*, 1985b, vol. 8, pp. 439–457.

ZYWIETZ CHR, ALRAUN W, and WILLEMS JL. (on behalf of the CSE working party). Results of ECG program noise tests within the CSE project. *Computers In Cardiology*, IEEE 84CH2078–4, Los Angeles (Cal.): 1984, pp. 377–380.

Annex B
(informative)

CSE ECG Reference Library (Diagnostic Database)

Purpose: To provide a uniform database for the assessment of the diagnostic performance of ECG computer interpretation programs.

Completion date: Fall 1990

Source: CSE Data Center
I.N.S.E.R.M. Unit #121
Hopital Cardiologique
BP Lyon-Montchat
69394 Lyon Cedex
France

B.1 Acquisition methodology

The CSE Diagnostic Database consists of 1,220 well-validated, multilead ECG recordings that were collected from adults—831 men and 389 women, all white with a mean (± SD) age of 52 ± 13 years—that are categorized into seven diagnostic groups: one group including normal (NL) and no structural abnormalities (NSA); three groups including left, right, and biventricular hypertrophy (LVH, RVH, and BVH); and three groups including anterior, interior, and combined myocardial infarction (AMI, IMI, and MIX). Also included were a limited number of cases with both myocardial infarction and ventricular hypertrophy.

Cases were selected entirely on the basis of clinical information that was unrelated to electrocardiographic findings. Those in the normal group were free of significant cardiopulmonary disease on the basis of either a health screening examination (75% of cases) or invasive cardiac studies (25% of cases). The diagnosis of LVH, RVH, or BVH was based on cardiac catheterization, echocardiographic findings, or both. Akinesia or dyskinesia in seven different segments of the ventriculogram, coded according to American Heart Association rules, was used as the main selection criterion for infarct classes. Also included were some cases of acute myocardial infarction with a typical history and enzyme changes.

The ECGs were recorded on digital tape, 15 leads simultaneously (i.e., the standard 12 leads plus leads X, Y, and Z). The lead length was 10 sec. Sampling rate was 500 per sec with a minimum resolution of 5 µV (at least 11 bits per sample). The data were collected in five European cardiology centers. A review board, consisting of three cardiologists, has checked the clinical information; a consensus was required for case selection. Excluded were ECGs showing major ventricular conduction defects (i.e., complete left or right bundle branch block and WPW) and those of poor technical quality.

B.2 Annotation methodology

The ECGs were analyzed by fifteen different computer programs and nine cardiologists from seven European Union (EU) member states. Eight of these cardiologists have interpreted the ECGs, and five have independently interpreted the VCGs (i.e., the vector loops and the scalar XYZ leads). Nine of the programs used the standard 12-lead ECG; six used the VCG. Except for age and sex, no prior clinical information was provided to the processing center or to the cardiologists.

The processing centers and the cardiologists were asked to apply a scheme for translating statements into a common set of diagnostic codes. Each statement could be qualified as definite, probable, or possible. When a program or cardiologist listed minor abnormalities without referring to any of the seven primary categories listed above, the case was classified in the NL or so-called “normal” group. A combined result that was based on the interpretations of all of the cardiologists was derived by means of a weighted average. For every diagnostic statement, each occurrence of “definite” was assigned a value of 3 points; “probable,” 2 points; and “possible,” 1 point. For each ECG, the scores were averaged for all of the cardiologists. A combined result between 1 and 3 was thus obtained for each diagnostic category and then reconverted into one of the levels of certainty. This result is equivalent to that obtained by panel review or by a majority vote of several readers. A combined result was similarly obtained for the programs. The average cardiologist was defined as the one whose total accuracy level was closest to the group average. The average program was similarly defined.

The annotation data available for the CSE Diagnostic Database, thus, consists of (a) the clinical truth (the diagnoses that are based on the independent clinical data); (b) the “interpretation” that was inferred by the average and the

combined cardiologists (the group interpretation) for the ECGs, the VCGs, and the combined ECGs and VCGs; and (c) the interpretations inferred by the average and the combined programs for the ECGs, the VCGs, and the combined ECGs and VCGs.

B.3 Storage and distribution methodology

Raw, digital ECGs were stored on digital tapes. The transfer of raw ECG digital data to media appropriate for each category of user was accomplished at the CSE data center in Leuven, Belgium. Since the end of 1991, the raw data has been available on CD-ROM that is compatible with the High Sierra standard (PC, not Macintosh). Only the digital ECG data and the demographic data (sex and age) will be released, not the independent clinical information or the interpretations of the cardiologists, the average, or the combined programs. The latter are kept secret at the CSE data center for independent evaluation purposes.

Program interpretation results may be submitted either on floppy disk or by e-mail to the CSE data center for assessment. Users who submit their results for evaluation must provide the full printouts of their interpretation results for control of the mapping scheme.

B.4 Signal reconstruction methodology

If the user's program employed a digital sampling rate other than 500 s/s, it was up to the user to reconstruct the digital ECG by using appropriate interpolation schemes. (No program used sampling rates greater than 500 s/s.)

B.5 Program evaluation methodology

B.5.1 Comparison with the "clinical truth"

Different misclassification matrices are computed for each program to be tested. Two-by-two matrices are computed first for each of the diagnostic categories versus all of the rest. Then the CSE Data Center proceeds to a 3-by-3 classification in which LVH, RVH, and BVH are pooled into one common "hypertrophy" class. Similarly, all infarctions are put together into one category, "infarction," independent of the infarction location. In the 5-by-5 classification matrices, the classes NSA, LVH, RVH, DVH, and "infarction" are counted. In the final 7-by-7 misclassification matrix, all diagnostic entities are considered separately.

With respect to the classification tables, absolute figures are given for each cell first, followed by percentages. The results refer to the analysis obtained after additional BFH or MIX tests are otherwise specified. Multiple diagnostic statements (N) on the highest probability level are each counted as 1/N in the respective cells of the different matrices.

Summaries of sensitivity, specificity, total and partial accuracy, and predictive values for positive and negative test results are given for each program to be tested. For the calculation of the partial accuracy figures, BVH is counted as partially correct where the case is classified as LVH or RVH. Similarly, for combined myocardial infarction, the output is taken into account if the diagnosis of either AMI or IMI is provided and vice versa.

B.5.2 Comparison with the combined referee results

Although the primary objective of the CSE diagnostic study was to compare the study with the clinical truth as reference, program results are compared with the interpretation results that were obtained by the cardiologists. To this end, the combined results that were obtained after weighted averaging of all of the 8 ECG and 5 VCG interpretations are used. Program results are separately compared with the combined ECG and combined VCG results of the cardiologists.

The evaluation strategy is based on the McNemar approach to compare the diagnostic results. In this analysis, results are basically still compared with the "clinical truth." Comparison is restricted to a 2-by-2 table that reports the presence or absence of the "true" category. For example, for cases with true LFH, LVH can be diagnosed by the reference and the program to be tested. It can be missed by both, or it can be diagnosed by one and not by the other. In case the true diagnosis is missed by both, the test does not give any indication as to whether both agree or provide the same diagnosis.

B.6 Strengths and weaknesses of the database

Many European medical centers contributed data. Hence, the CSE database contains both a collection of cases with pathologic electrophysiology and normal cases. All 15 leads of the ECG and VCG (i.e., the standard 12 leads plus the X, Y, and Z leads) were recorded simultaneously. Two-thirds of the XYZ data was recorded with the Frank system; the remaining one-third was recorded by using a hybrid lead system. The latter recordings are of the order of 8 seconds only and have been padded at the end with the value of the last valid ECG sample up to a total length of 10 sec. Case selection was restricted to ECGs from patients with one of seven main diagnoses that could be validated on the basis of nonelectrocardiographic evidence, such as the results of cardiac catheterization,

echocardiography, and cardiac enzyme measurements. Neither the evaluation of rhythm and conductance disturbances (for which the ECG itself is the standard) nor the assessment of ST-segment and T-wave changes or other descriptive statements was performed.

Comparative evaluation of computer programs presents several problems. Almost all programs use different diagnostic terminology. Some generate only one or two diagnostic statements, whereas others provide a list of the most likely interpretations in decreasing order of likelihood. The requirement to convert the statements from the program to be tested into a common code (mapping)—even when applied consistently—may not allow for all of the differences in reporting. The fact that 87% of the database consists of patients with single diseases may favor programs that, by definition, select from one of seven or eight diagnostic categories. For a proper evaluation and comparison of a program's diagnostic classification ability of patients with combined diseases, several hundred validated ECGs are required for each disease combination. The cases should present a whole spectrum of disease severity, which is not easy to quantify.

B.7 References

WILLEMS JL. *Common standards for quantitative electrocardiography: 10th and final CSE progress report*. Acco: Leuven, Belgium 1990.

WILLEMS JL. Assessment of diagnostic ECG results using information and decision theory. Results from the CSE diagnostic study. *J. Electrocardiol.*, 1992, vol. 25, suppl., pp. 120–125.

WILLEMS JL, ABREU-LIMA C, ARNAUD P, *et al.* Evaluation of ECG interpretation results obtained by computer and cardiologists. *Meth. Inform. Med.*, 1990, vol. 29, pp. 308–316.

WILLEMS JL, ABREU-LIMA C, ARNAUD P, *et al.* The diagnostic performance of computer programs for the interpretation of electrocardiograms. *N. Engl. J. Med.*,1991, vol. 325, pp. 1767–1773.

WILLEMS JL, ARNAUD P, VAN BEMMEL JH, *et al.* Development of a reference library for multi-lead ECG measurement programs. *J. Electrocardiol.*, 1987, vol. 20, suppl., Oct., pp. 56–61.

WILLEMS JL, ARNAUD P, VAN BEMMEL JH, *et al.* Common standards for quantitative electrocardiography: goals and main results. *Meth. Inform. Med.*, 1990, vol. 29, pp. 63–271.

Annex C
(informative)

MIT-BIH Arrhythmia Database

Purpose: To aid in the development and evaluation of real-time ECG rhythm analysis algorithms.

Completion date: 1980

Source: George B. Moody
Harvard-MIT Division of Health Sciences and Technology Biomedical Engineering Center
MIT Room 20A–113
77 Massachusetts Avenue
Cambridge, MA 02139 USA
Telephone: 1-617-253-7424
Telefax: 1-617-253-2514

Beth Israel Hospital
Biomedical Engineering Division, KB–26
330 Brookline Avenue
Boston, MA 02215 USA
Telephone: 1-617-735-4553

C.1 Acquisition methodology

The MIT-BIH Arrhythmia Database contains 48 records, each slightly more than 30 min long. The source of the ECGs is a set of over 4,000 long-term Holter recordings that were obtained by the Beth Israel Hospital arrhythmia laboratory between 1975 and 1979. Twenty-three records were chosen at random from this set. Twenty-five records were selected to include a variety of rare, but clinically important, phenomena that would not be well represented by a small random sample of Holter recordings.

The original analog recordings were made using Del Mar Avionics Model 445 two-channel Holter recorders. During the digitization process, the analog recordings were played back on a Del Mar Avionics Model 660 playback unit. The analog outputs of the playback unit were filtered to limit analog-to-digital converter saturation and anti-aliasing.

Samples were acquired at 360 samples per sec from each channel almost simultaneously, using a unipolar ADC with 11-bit resolution over a ± 5 mV range. Sample values range from 2 to 2,047, inclusive, with a value of 1,024 corresponding to zero volts. The 11-bit samples were recorded in an 8-bit, first-difference format.

C.2 Annotation methodology

An initial set of beat labels was produced by a simple slope-sensitive ORS detector, which marked each detected event as a normal beat. Two identical chart recordings were given to two cardiologists, who worked on them independently. The cardiologists added beat labels where the detector missed them, deleted false detections as necessary, and changed the labels for all abnormal beats. They also added rhythm labels, signal quality labels, and comments. All discrepancies between cardiologists were reviewed and resolved by consensus.

Annotations generally appear at the R-wave peak. This pattern was accomplished both by moving each beat label to the major local extrema after band pass filtering to emphasize the ORS complex and by correcting for the phase shift in the filter.

C.3 Storage and distribution methodology

CD-ROMs containing the MIT-BIH Arrhythmia Database, supplementary databases, documentation, and development tools are maintained, stored, and distributed by MIT and Beth Israel Hospital.

C.4 Strengths and weaknesses of the database

This database is well established, well used, and well understood. Most arrhythmia algorithms have used the database for development. Most algorithm comparisons have been based on this database. The database consists of real waveforms from real clinical environments.

The weaknesses of the database result from the method of signal acquisition. The analog acquisition using standard Holter recorders places limitations on the use of the database because of the limited recording bandwidth, tape

flutter, and so forth. The quality of the second channel is sometimes low. The database documentation includes discussion of these issues.

Because this database has been used extensively for both algorithm development and testing, there is much controversy regarding performance comparisons using this database.

C.5 References

MIT-BIH Arrhythmia Database Directory. 2nd ed. (BMEC TRO10), August 1988.

MARK RG, SCHLUTER PS, MOODY GB, *et al.* An annotated ECG database for evaluating arrhythmia detectors. *Frontiers of Engineering in Health Care: Proceedings of the 4th Annual Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 205–210. New York: IEEE Press, 1982.

Annex D
(informative)

Noise Stress Database

Purpose: To aid in the development and evaluation of real-time algorithms of ECG rhythm analysis.

Completion date: 1984

Source: George B. Moody
Harvard-MIT Division of Health Sciences and Technology Biomedical Engineering Center
MIT Room 20A–113
77 Massachusetts Avenue
Cambridge, MA 02139 USA
Telephone: 1-617-253-7424
Telefax: 1-617-253-2514

Beth Israel Hospital
Biomedical Engineering Division, KB–26
330 Brookline Avenue
Boston, MA 02215 USA
Telephone: +1-617-735-4553

D.1 Acquisition methodology

This database is based on the MIT-BIH Arrhythmia Database. Three records, “bm,” “em,” and “ma,” contain noise that is typically observed in ECG recordings. They were obtained using a Holter recorder on an active subject with leads placed so that the subject’s ECG was not visible. (Record “bm” contains baseline wander, “em” contains electrode motion artifact, and “ma” contains muscle noise.) MIT-BIH Arrhythmia Database records 118 and 119 are mixed with various levels of the “em” noise.

D.2 Storage and distribution methodology

The Noise Stress Database is distributed with the MIT-BIH Arrhythmia Database.

D.3 Strengths and weaknesses of the database

The strength of the Noise Stress Database is its derivation from the MIT-BIH Arrhythmia Database, making it easy to obtain and use. It provides a standard method to quantify an arrhythmia algorithm’s robustness in the presence of noise.

The noise records are not pure noise; they contain cardiac signals. Some algorithms can reliably detect the heartbeats within the noise records. Mixing such “noise” signals with MIT-BIH Arrhythmia Database records mixes two detectable cardiac signals.

As the signal-to-noise ratio increases, the database records become less clinically relevant.

D.4 References

MOODY GB. *MIT-BIH Arrhythmia Database*. File: readme.doc, July 1989.

MOODY GB, MULDROW WK, and MARK RG. A noise stress test for arrhythmia detectors. *Computers in Cardiology*, 1984, vol. 11, pp. 381–384.

Annex E
(informative)

European ST-T Database

Purpose: Evaluation of algorithms that will be used for the analysis of ST and T-wave changes.

Completion date: April 1991

Source: Alessandro Taddei
National Research Council (CNR) Institute of Clinical Physiology
Computer Laboratory
via Trieste, 41
56100 P Italy
Telephone: 39-50-502771
Telefax: 39-50-589038

George B. Moody
Harvard-MIT Division of Health Sciences and Technology Biomedical Engineering Center
MIT Room 20A–113
77 Massachusetts Avenue
Cambridge, MA 02139 USA
Telephone: 1-617-253-7424
Telefax: 1-617-253-2514

E.1 Acquisition methodology

This database consists of 90 annotated excerpts of ambulatory ECG recordings from 79 subjects. Myocardial ischemia was diagnosed or suspected for each subject; additional selection criteria were established to obtain a representative selection of ECG abnormalities in the database. The database includes 368 episodes of ST segment change and 401 episodes of T-wave change, with durations ranging from 30 sec to several min and with peak displacements ranging from 100 µV to more than one mV.

Each record is 2 hours in duration and contains two leads, each sampled at 250 samples per second with 12-bit resolution over a 20-mV input range. The header files include information about the leads used; the patient's age, sex, and medications; the clinical findings; and the recording equipment.

The files on this disk are in a format similar to those on the MIT-BIH Arrhythmia Database CD-ROM.

E.2 Annotation methodology

Two cardiologists worked independently to annotate each record beat by beat and look for changes in ST segment and T-wave morphology, rhythm, and signal quality. ST segment and T-wave changes were identified in both leads, using predefined criteria that were applied uniformly in all cases. Their onsets, extrema, and ends were annotated. The separate annotations made by the two cardiologists were compared, the coordinating group in Pisa, Italy resolved disagreements, and the reference annotation files were prepared. Altogether, these files contain 802,866 annotations.

Several annotation codes were newly defined for the European ST-T Database and have been added to those previously defined for the MIT-BIH Arrhythmia Database and the AHA Database for Evaluation of Ventricular Arrhythmia Detectors. Software that was provided by the developers of the MIT-BIH Arrhythmia Database (some of which is included on the distribution media) has been revised accordingly so that it can be used with any of these databases.

The files on this disk are in the same format as those on the MIT-BIH Arrhythmia Database CD-ROM.

E.3 Storage and distribution methodology

CD-ROMs containing the European ST-T Database and the VALE Database documentation and development tools are maintained, stored, and distributed by National Research Council (CNR) Institute of Clinical Physiology Computer Laboratory and MIT.

E.4 Strengths and weaknesses of the database

The strength of the European ST-T Database is that it is in the same format as the MIT-BIH Arrhythmia Database. The database is easy to obtain and use and provides a standard method to quantify an ST segment algorithm's ST and T-wave performance.

The notable shortcomings of the database relate to the method of ST segment annotation. By design, the ST segment values are annotated once per episode only at the point of maximum change. There are no beat-to-beat ST level annotations. Additionally, there is no known baseline value for the ST annotations. The ST level annotations are not related to the isoelectric line, but to a reference beat early in the database record. As a result, it is difficult to compare algorithm results of truth annotations from the database.

E.5 References

MARCHESI C. The European Community concerted action on ambulatory monitoring. *J. Med. Eng. and Techn.*, 10:131–134, 1986.

MOODY G, and TADDEI A. *European ST-T Database*. File: readme.doc, April 1991.

TADDEI A, BENASSI A, BIAGINI A, *et al.* ST-T change analysis in ECG ambulatory monitoring: a European standard for performance evaluation. *Computers in Cardiology*, 1987, vol. 14, pp. 63–68.

TADDEI A, BIAGINI A, DISTANTE G, *et al.* An annotated database aimed at performance evaluation of algorithms for ST-T change analysis. *Computers in Cardiology*, 1989, vol. 16.

Annex F
(informative)

American Heart Association Database for Evaluation of Automated Ventricular Arrhythmia Detectors

Purpose: To provide system users and designers with a uniformly acceptable method for evaluating automated arrhythmia detection system performance using digital data.

Completion date: 1980

Source: Emergency Care Research Institute (ECRI)
5200 Butler Pike
Plymouth Meeting, PA 19462 USA

F.1 Acquisition methodology

The American Heart Association ECG database consists of 155 3-hour segments of two-channel Holter recordings.

The segments are equally divided among eight arrhythmia classes: no premature ventricular contractions (PVCs), isolated uniform PVCs, isolated multiform PVCs, bigeminy, couplets, R-on-T beats, ventricular rhythms, and ventricular fibrillation or ventricular flutter beat.

Each channel is sampled at 250 samples per sec with 12-bit resolution. The span is approximately ± 5 mV.

F.2 Annotation methodology

The last 30 min of each record were annotated by three independent, expert electrocardiographers. Each ORS complex was labeled and time aligned. All disagreements among annotators were reconciled. Note that the annotations generally appear at the QRS onset.

F.3 Storage and distribution methodology, strengths, and weaknesses of the database

The database has been split into two series. Copies of both series are available on CD-ROM. Abridged segments, consisting of the last 35 min of each 3-hour segment, are available on CD-ROM and floppy disks. Other media are available upon request.

Content includes the dual-channel waveforms, annotation (beat type and time), annotation summaries for each segment, and source code and executable software for plotting and analyzing the waveforms and annotation.

The database is recognized and accepted for its organization and broad range of ECGs and has been used by many researchers and developers. It provides a standard method to quantify an algorithm's performance and has served as a model for creation of other, sometimes more extensive, databases.

Subtle errors have been reported in a few of the annotations.

The first series in the database has been used extensively by industry for algorithm development and testing. Consequently, there is a great controversy regarding its use to perform algorithmic comparisons. The second series, however, has not been available to the public until recently.

F.4 References

HERMES RE, GESELOWITZ DB, and OLIVER GC. Development, distribution, and use of the American Heart Association database for ventricular arrhythmia detector evaluation. *Computers in Cardiology*, 1980, pp. 263–266.

RIPLEY KL, and OLIVER GC. Development of an ECG database for arrhythmia detector evaluation. *Computers in Cardiology*, 1977, pp. 203–209.

Annex G
(informative)

Massachusetts General Hospital/Marquette Foundation Waveform Database

Purpose: To provide a collection of electronic recordings of hemodynamic and electrocardiographic waveforms of patients in critical care units, surgery, cardiac catheterization, and other electrophysiology studies.

Completion date: Not yet complete

Source:
Massachusetts General Hospital
Anesthesia Bioengineering Unit
Fruit Street
Boston, MA 02114 USA
Telephone: 1-617-724-3150
E-mail: info@abu.mgh.harvard.edu

G.1 Acquisition methodology

The database consists of 250 90-min recordings typically containing three ECG and three pressure, respiratory impedance, and CO_2 channels. The database contains 225 adult and 25 pediatric recordings that represent both physiologic and pathophysiologic states.

G.2 Annotation methodology

Annotations are provided for heartbeats and relevant events. Header file annotations include patient data, ECG interpretation, technical comments, hemodynamic data, and clinical synopsis.

G.3 Storage and distribution methodology, strengths, and weaknesses of the database

The data is sampled at 360 samples per sec and is stored in the MIT-BIH format. The database is distributed on 10 CD-ROMs that are categorized according to events of interest.

G.4 Reference

MGH/MF Waveform Database Documentation.

Annex H
(informative)

Creighton University Ventricular Tachyarrhythmia Database

Purpose: To assess ECG monitoring algorithm behavior in response to the onset of potentially malignant ventricular tachyarrhythmias.

Completion date: August 1992

Source: MIT-BIH Database Distribution
MIT Room 20A–113
Cambridge, MA 02139 USA

H.1 Acquisition methodology

The late Floyd M. Nolle, D.Sc., collected this database at the cardiac center of Creighton University as part of his work on ventricular fibrillation in the surface electrocardiogram. It contains 35 single-channel records, each of which shows the onset of ventricular fibrillation. Record CU01 was obtained from a Holter recording played back at real time for digitization. All other records were digitally recorded directly to disk in real time. Records came from high-level analog outputs from patient monitors at 1 V/mV nominal gain. All signals were preprocessed by an active second-order Bessel low-pass filter with a 70-Hz cutoff and then digitized with 12-bit resolution (2s complement coding) over a ± 5 V range (± 5 mV nominal, relative to the unamplified signals) at 250 Hz. Each record contains 127,232 samples for slightly less than 8.5 min of ECG.

Please note that in episodes of cardiac failure, fibrillation is almost always preceded by a run of ventricular tachycardia, which eventually gives way to the fibrillation itself. The onset of fibrillation is extremely difficult to pinpoint in many cases. Any clinically useful detector should respond to the runs of tachycardia that precede fibrillation, because medical intervention is needed at the earliest opportunity. However, responding to these runs means that any detector responding to the premonitory tachycardia can exhibit a negative “time to alarm” compared to the onset of fibrillation itself. For this reason, the database is defined as a tachyarrhythmia database rather than a fibrillation database.

In these records, the minimum number of non-VF beats before the onset of a VF episode is 61. The mean time interval from the beginning of the record to the onset of VF is 5:47 ± 2:01 min. Five records (CU12, CU15, CU24, CU25, and CU32) were from paced patients (in some cases, pacing artifacts are not visible, and pacing is apparent only from the regularity of the rhythm). Repeated defibrillation attempts are visible in many records.

H.2 Annotation methodology

Three board-certified cardiologists with extensive experience reading ECGs were chosen as referees. Each referee received a high resolution (Versatec graph) hard copy showing each complete record.

Each referee first classified all beats and then isolated and marked all episodes of ventricular tachyarrhythmias. Returned casebooks were examined for areas of consensus and areas of disagreement. Revised casebooks were then circulated with discrepancies identified. Final consensus was reached by means of a conference call.

H.3 Storage and distribution methodology

Gain-adjusted digital ECG data was recorded directly to disk for each patient included in the database. The binary data was transferred to MIT by diskette, and the finished CD-ROM was scrutinized in detail for adherence to the original ECG records. The CD-ROMs are maintained, stored, and distributed by MIT.

H.4 Signal reconstruction methodology

Because the database was originally designed for algorithm testing, no method of analog signal reconstruction was specified, and no method for sample frequency conversion was provided. Users who need a sampling rate other than 250 Hz have to reconstruct the digital ECG on their own, using appropriate interpolation schemes. (The MIT-BIH Database includes a sample rate-conversion utility that makes solving the problem easy.)

H.5 Strengths and weaknesses of the database

The database's greatest strength is its freedom from time-base distortion, which occurs when direct-to-disk recording is used in the waveform acquisition phase. Its greatest weaknesses are the lack of supplementary lead signals and the low number of patient ECG records that are included.

H.6 Reference

NOLLE FM, BADURA FK, CATLETT JM, *et al.* CREI-GARD, a new concept in computerized arrhythmia monitoring systems. *Computers in Cardiology*, 1986, vol. 13, pp. 515–518.

