MIT-BIH Arrhythmia Database Directory

Next: Records **Up:** Contents **Previous:** Foreword

Introduction

This introduction describes how the database records were obtained, and discusses the characteristics of the recorded signals. Following these notes are annotated ``full disclosure'' plots of the entire database. These can be useful for obtaining an overall impression of the contents of individual records. Following the ``full disclosure'' plots are sample ECG strips. These strips were chosen to illustrate the salient features of each record. Next are notes on the important features of each record. These notes also include background information on the subjects, including their medications. At the end of the book are tables of rhythms and annotations, which summarize the contents of the database. These tables can be helpful in finding a record with a specific set of characteristics.

Selection criteria

The source of the ECGs included in the MIT-BIH Arrhythmia Database is a set of over 4000 long-term Holter recordings that were obtained by the Beth Israel Hospital Arrhythmia Laboratory between 1975 and 1979. Approximately 60% of these recordings were obtained from inpatients. The database contains 23 records (numbered from 100 to 124 inclusive with some numbers missing) chosen at random from this set, and 25 records (numbered from 200 to 234 inclusive, again with some numbers missing) selected from the same set to include a variety of rare but clinically important phenomena that would not be well-represented by a small random sample of Holter recordings. Each of the 48 records is slightly over 30 minutes long.

The first group is intended to serve as a representative sample of the variety of waveforms and artifact that an arrhythmia detector might encounter in routine clinical use. A table of random numbers was used to select tapes, and then to select half-hour segments of them. Segments selected in this way were excluded only if neither of the two ECG signals was of adequate quality for analysis by human experts.

Records in the second group were chosen to include complex ventricular, junctional, and supraventricular arrhythmias and conduction abnormalities. Several of these records were selected because features of the rhythm, QRS morphology variation, or signal quality may be expected to present significant difficulty to arrhythmia detectors; these records have gained considerable notoriety among database users.

The subjects were 25 men aged 32 to 89 years, and 22 women aged 23 to 89 years. (Records 201 and 202 came from the same male subject.)

ECG lead configuration

In most records, the upper signal is a modified limb lead II (MLII), obtained by placing the electrodes on the chest. The lower signal is usually a modified lead V1 (occasionally V2 or V5, and in one instance V4); as for the upper signal, the electrodes are also placed on the chest. This configuration is routinely used by the BIH Arrhythmia Laboratory. Normal QRS complexes are usually prominent in the upper signal. The lead axis for the lower signal may be nearly orthogonal to the mean cardiac electrical axis, however (i.e., normal beats are usually biphasic and may be nearly isoelectric). Thus normal beats are frequently difficult to discern in the lower signal, although ectopic beats will often be more prominent (see, for example, record 106). A notable exception is record 114, for which the signals were reversed. Since this happens occasionally in clinical practice, arrhythmia detectors should be equipped to deal with this situation. In records 102 and 104, it was not possible to use modified lead II because of surgical dressings on the patients; modified lead V5 was used for the upper signal in these records.

Analog recording and playback

The original analog recordings were made using nine Del Mar Avionics model 445 two-channel recorders, designated *A* through *I*:

Recorder	Records
A	102, 107, 111, 115, 121
B	212
C	203
D	118, 124, 217
E	101, 103, 106, 108, 112, 117, 119, 122, 209, 219, 220, 223, 233
F	104, 109, 123, 205, 207, 210, 215, 221
G	100, 105, 114, 116, 213, 214, 222, 228
H	113, 201, 202, 231
I	200, 230, 232, 234

(It is not known which recorder was used for record 208.)

During the digitization process, the analog recordings were played back on a Del Mar Avionics model 660 unit. The analog tapes used for records 112, 115 through 124, 205, 220, 223, and 230 through 234 were played back and digitized at twice real time; the rest were played back at real time using a specially constructed capstan for the model 660 unit. Skew between the two signals was found to be as great as 40 milliseconds for some of the analog recorders. In addition to the fixed skew that results from extremely small differences in the orientations of the tape heads on the recorder and the playback unit, microscopic vertical wobbling of the tape, either during recording or playback, introduces a variable skew, which may be comparable

in magnitude to the fixed skew. This problem (which also occurs on the AHA database) may present difficulties for certain two-channel analysis methods designed for real-time applications.

Minor tape speed variations should not pose problems for typical arrhythmia detectors. It is difficult to avoid tape sticking or slippage during low-speed playback, and several episodes of tape slippage were noted and marked with comment annotations. Wow and flutter should be studied carefully in the context of heart-rate variability studies, since flutter compensation was not possible in these recordings. A number of frequency-domain artifacts have been identified and related to specific mechanical components of the recorders and the playback unit:

Frequency (Hz)	Source
0.042	Recorder pressure wheel
0.083	Playback unit capstan (for twice real-time playback)
0.090	Recorder capstan
0.167	Playback unit capstan (for real-time playback)
0.18-0.10	Takeup reel (frequency decreases over time)
0.20-0.36	Supply reel (frequency increases over time)

The most significant of these artifacts by far is the 0.167 Hz artifact on recordings that were played back at real time. The next largest is the 0.090 Hz artifact; the 0.083 Hz artifact on recordings that were played back at twice real-time is of roughly the same magnitude as the 0.090 Hz artifact. The 0.042 Hz artifact is of much lower magnitude. Other frequencies related to the drive train (at 0.42 Hz, 1.96 Hz, 9.1 Hz, and 42 Hz) do not appear as noticeable artifacts. The frequencies of the last two artifacts listed in the table depend on how much tape is on the supply and takeup reels; the supply reel causes a much more noticeable artifact than does the takeup reel. Other frequency-domain artifacts generated by the supply reel appear in the 0.10-0.18 Hz and 0.30-0.54 Hz bands.

Four of the 48 records (102, 104, 107, and 217) include paced beats. The original analog recordings do not represent the pacemaker artifacts with sufficient fidelity to permit them to be recognized by pulse amplitude (or slew rate) and duration alone, the method commonly used for real-time processing. The database records reproduce the analog recordings with sufficient fidelity to permit use of pacemaker artifact detectors designed for tape analysis, however.

Digitization

The analog outputs of the playback unit were filtered to limit analog-to-digital converter (ADC) saturation and for anti-aliasing, using a passband from 0.1 to 100 Hz relative to real time, well beyond the lowest and highest frequencies recoverable from the recordings. The bandpass-filtered signals were digitized at 360 Hz per signal relative to real time using hardware constructed at the MIT Biomedical Engineering Center and at the BIH Biomedical Engineering Laboratory. The sampling frequency

was chosen to facilitate implementations of 60 Hz (mains frequency) digital notch filters in arrhythmia detectors. Since the recorders were battery-powered, most of the 60 Hz noise present in the database arose during playback. In those records that were digitized at twice real time, this noise appears at 30 Hz (and multiples of 30 Hz) relative to real time.

Samples were acquired from each signal almost simultaneously (the intersignal sampling skew was on the order of a few microseconds). As noted above, analog tape skew was several orders of magnitude larger. The ADCs were unipolar, with 11-bit resolution over a ±5 mV range. Sample values thus range from 0 to 2047 inclusive, with a value of 1024 corresponding to zero volts.

The 11-bit samples were originally recorded in 8-bit first difference format (this was necessary because of limited mass storage capacity). Given the sampling frequency and the resolution of the ADC, the difference encoding implies a maximum recordable slew rate of ±225 mV/s. In practice, this limit was exceeded by the input signals very infrequently, only during severe noise on a small number of records. The effect on the quality of the recorded signals is totally negligible. On this CD-ROM, the samples have been reconstructed from the first differences and stored as pairs of 12-bit amplitudes packed in triplets of consecutive bytes (for details on the storage format, see signal(5)).

Annotations

An initial set of beat labels was produced by a simple slope-sensitive QRS detector, which marked each detected event as a normal beat. Two identical 150-foot chart recordings were printed for each 30-minute record, with these initial beat labels in the margin. For each record, the two charts were given to two cardiologists, who worked on them independently. The cardiologists added additional beat labels where the detector missed beats, deleted false detections as necessary, and changed the labels for all abnormal beats. They also added rhythm labels, signal quality labels, and comments.

The annotations were transcribed from the paper chart recordings. Once both sets of cardiologists' annotations for a given record had been transcribed and verified, they were automatically compared beat-by-beat, and another chart recording was printed. This chart showed the cardiologists' annotations in the margin, with all discrepancies highlighted. Each discrepancy was reviewed and resolved by consensus. The corrections were transcribed, and the annotations were then analyzed by an auditing program, which checked them for consistency and which located the ten longest and shortest R-R intervals in each record (to identify possible missing or falsely detected beats).

In early copies of the database, most beat labels were placed at the R-wave peak, but manually inserted labels were not always located precisely at the peak. In copies of the database made since 1983, the beat labels have been shifted from their original locations. The ECG (usually the upper signal) was digitally bandpass-filtered to emphasize the QRS complexes, and each beat label was moved to the major local

extremum, after correction for phase shift in the filter. A few noisy beats were manually realigned. This process was applied to all records except record 117 in 1983; the beat labels for record 117 were not realigned until March 1998, however. The result is that annotations generally appear at the R-wave peak, and are located with sufficient accuracy to make the reference annotation files usable for studies requiring waveform averaging and for heart rate variability studies (but note the comments with respect to analog tape wow and flutter above). In the annotated ECG plots produced by `psfd` and `pschart`, and in printed copies of this directory, each label is placed so that the fiducial mark for the annotation corresponds to the left edge of the label.

The database contains approximately 109,000 beat labels. Sixteen were corrected in the first seven years after the database was released in 1980 (in records 104, 108, 114, 203, 207, 217, and 222); in addition, all of the left bundle branch block beats in record 214 were originally labelled as normal beats. The rhythm labels have been more substantially revised and now include notations for paced rhythm, bigeminy, and trigeminy, which were missing in early copies.

In October 1998, a rhythm label in record 203 was corrected. In October 2001, a seventeenth error in the beat labels was discovered and corrected (in record 209). In April 2003, 26 PVC annotations in record 119 were manually realigned by small amounts (up to 74 ms). In May 2003, an eighteenth error in the beat labels was discovered and corrected (in record 214). In April 2005, many of the episodes previously labelled as atrial fibrillation in record 222 were partially or completely relabelled as atrial flutter. In April 2008, three beat labels were corrected (two in record 108, and one in record 215). In June 2010, the 22nd and 23rd errors in the beat labels were found and corrected (both in record 203). Thanks to Bob Bruce, Pat Hamilton, Yin Dengfeng, Roger Mark, Sebastian Vasquez, and Mariano Llamedo Soria for finding and reporting these errors.

Symbols used in plots

[An expanded and updated version of the table below can be found at http://www.physionet.org/physiobank/annotations.shtml.]

Symbol	Meaning
· or N	Normal beat
L	Left bundle branch block beat
R	Right bundle branch block beat
A	Atrial premature beat
a	Aberrated atrial premature beat
J	Nodal (junctional) premature beat
S	Supraventricular premature beat

V	Premature ventricular contraction
F	Fusion of ventricular and normal beat
[	Start of ventricular flutter/fibrillation
!	Ventricular flutter wave
]	End of ventricular flutter/fibrillation
e	Atrial escape beat
j	Nodal (junctional) escape beat
E	Ventricular escape beat
/	Paced beat
f	Fusion of paced and normal beat
x	Non-conducted P-wave (blocked APB)
Q	Unclassifiable beat
|	Isolated QRS-like artifact
Rhythm annotations appear *below* the level used for beat annotations:	
(AB	Atrial bigeminy
(AFIB	Atrial fibrillation
(AFL	Atrial flutter
(B	Ventricular bigeminy
(BII	2° heart block
(IVR	Idioventricular rhythm
(N	Normal sinus rhythm
(NOD	Nodal (A-V junctional) rhythm
(P	Paced rhythm
(PREX	Pre-excitation (WPW)
(SBR	Sinus bradycardia
(SVTA	Supraventricular tachyarrhythmia
(T	Ventricular trigeminy
(VFL	Ventricular flutter
(VT	Ventricular tachycardia
Signal quality and comment annotations appear *above* the level used for beat annotations:	
qq	Signal quality change: the first character (`c' or `n') indicates the quality of the upper signal (clean or noisy), and the second character indicates the quality of the lower signal
U	Extreme noise or signal loss in both signals: ECG is unreadable

M (*or* MISSB)	Missed beat
P (*or* PSE)	Pause
T (*or* TS)	Tape slippage

[George B. Moody](george@mit.edu)

24 May 1997
Revised 24 June 2010