



The Impact of the MIT-BIH Arrhythmia Database

History, Lessons Learned, and Its Influence on Current and Future Databases

The MIT-BIH Arrhythmia Database was the first generally available set of standard test material for evaluation of arrhythmia detectors, and it has been used for that purpose as well as for basic research into cardiac dynamics at about 500 sites worldwide since 1980. It has lived a far longer life than any of its creators ever expected. Together with the American Heart Association (AHA) Database, it played an interesting role in stimulating manufacturers of arrhythmia analyzers to compete on the basis of objectively measurable performance, and much of the current appreciation of the value of common databases, both for basic research and for medical device development and evaluation, can be attributed to this experience. In this article, we briefly review the history of the database, describe its contents, discuss what we have learned about database design and construction, and take a look at some of the later projects that have been stimulated by both the successes and the limitations of the MIT-BIH Arrhythmia Database.

Nature of the Data

Electrocardiograms (ECGs) are very widely used as an inexpensive and noninvasive means of observing the physiology of the heart. In 1961, Holter [1] introduced techniques for continuous recording of the ECG in ambulatory subjects over periods of many hours; the long-term ECG (Holter recording), typically with a duration of 24 hours, has since become the standard technique for observing transient aspects of cardiac electrical activity.

Since the mid-1970s, our research group has studied abnormalities of cardiac rhythm (arrhythmias) as reflected in long-term ECGs as well as automated methods for identifying arrhythmias. Many other research groups in academia and industry

have had similar interests. Until 1980, it was necessary for those wishing to pursue such work to collect their own data. Although the recordings themselves are plentiful, access to these data is not universal, and thorough characterization of the recorded waveforms is a tedious and expensive process. Furthermore, there is very wide variability in ECG rhythms and in details of waveform morphology, both between subjects and within individuals over time, so that a useful representative collection of long-term ECGs for research must include many recordings.

During the 1960s and 1970s, development of automated arrhythmia analysis algorithms was hampered by a lack of universally accessible data. Each group that performed such work acquired its own set of recordings and often self-evaluated their algorithms using the same data that had been used to develop those algorithms. From the earliest days, it was clear that performance of these algorithms was invariably data-dependent, and the use of different data for the evaluation of each algorithm did not permit objective comparisons of algorithms from different groups.

Selection of Data

In 1975, recognizing that we would need a suitable set of well-characterized long-term ECGs for our own research, we began collecting, digitizing, and annotating long-term ECG recordings obtained by the Arrhythmia Laboratory of Boston's Beth Israel Hospital (BIH; now the Beth Israel Deaconess Medical Center). From the outset, however, we planned to make these recordings available to the research community at large, in order to stimulate work in this field and to encourage strictly reproducible and objectively comparable evaluations of different algorithms [2]. We expected that the availability of a common database

George B. Moody and Roger G. Mark
Harvard-MIT Division of
Health Sciences and Technology

tape rewind, then writing track 3 on the first tape, etc.

During digitization, the analog signals from the playback unit were filtered to limit saturation in analog-to-digital conversion and for anti-aliasing, using a passband of 0.1 to 100 Hz relative to real time, well beyond the range of frequencies reproduced by the recordings. Since the recorders were battery-powered, most of the 60 Hz noise in the recordings was introduced during playback. This noise appears at 30 Hz (and multiples of 30 Hz) relative to real time in the recordings that were digitized at twice real time.

Four of the 48 recordings include paced beats. Pacemaker artifacts are not accurately reproduced in the original analog recordings, since most of the energy in these artifacts is at frequencies in the kilohertz range, far above the passband of the recorders. The digitized recordings in the database faithfully reproduce the analog recordings, so that software intended to analyze analog tapes containing paced beats can be evaluated using these recordings. One of the major remaining gaps in the publicly available collections of ECGs is a representative set of high-fidelity recordings of paced rhythms, which would be helpful for those designing software for real-time analysis of such signals.

In order to record 30 min of data in the available space, it was necessary to convert the digitized 11-bit samples into 8-bit first differences on the fly, which had the effect of limiting the slew rate to 225 mV/s, a limit that was exceeded by the input signals very infrequently, only on a few records during periods of severe noise. The effect of this procedure on the signal quality was negligible.

Annotation

Once the digital tapes had been prepared, we annotated them using a simple slope-sensitive QRS detector. Next, each tape was played back through a digital-to-analog converter to a thermal chart recorder that had been equipped with a pair of seven-element print heads. The playback software, written in assembly language as was all of the other software, generated the appropriate signals to form characters, printing periodic elapsed time markers on one edge of the paper, and the annotations on the other edge. Each half-hour tape was used to produce two identical 150-foot (46-m) paper chart recordings.

The charts for each recording were given to two cardiologists, who worked in-

dependently, adding additional beat labels and deleting false detections as necessary, and changing the labels for abnormal beats. The cardiologists also added rhythm and signal quality labels. The paper charts with the cardiologists' annotations were then transcribed into computer-readable form using an interactive annotation editor that displayed the waveforms on an oscilloscope using the same digital-to-analog converter board that was used to make the chart recordings. The result of this process was a tape containing two sets of cardiologist annotations.

At this point the two sets of annotations were compared automatically and another chart recording was printed, showing the cardiologists' annotations in the margin, with all discrepancies highlighted. Each discrepancy was examined and resolved by consensus. Corrections were then entered using the annotation editor, and all of the annotations were then audited using a program that checked them for consistency. (The auditing program also identified the ten shortest and longest inter-beat intervals, to identify possible false detections or missed beats.) Approximately 110,000 annotations were created and verified in this way.

Notably, six of the 48 records contain a total of 33 beats that remain unclassified, because the cardiologist-annotators were unable to reach agreement on the beat types. In these cases, as in clinical practice, it occasionally happens that some beats cannot be classified with certainty, either because of technical defects in the recording, or because there is insufficient information in the record to permit a confident choice between two or more reasonable hypotheses. It is important that a database intended to represent "real world" signals should contain the broadest possible range of waveforms, including these ambiguous cases, which may represent the most interesting challenges for automated analysis.

The annotators were instructed to use all evidence available from both signals to identify every detectable QRS complex. The database contains seven episodes of loss of signal or noise so severe in both channels simultaneously that QRS complexes cannot be detected; these episodes are all quite short and have a total duration of about 10 s. In all of the remaining data, every QRS complex was annotated, about 109,000 in all.

A database intended to represent "real world" signals should contain the broadest possible range of waveforms, including ambiguous cases, which may be the most interesting challenges for automated analysis.

Completing the Database

All of this work was done by necessity on our custom-built microcomputers. As the processing neared completion, we began transferring the data over a 9600 baud serial line to our laboratory's mini-computer, which was equipped with a nine-track tape drive; this process required several weeks. The first copies of the completed database were distributed on sets of 800 bpi nine-track tape in the summer of 1980. Our initial expectation was that perhaps as many as ten academic and industry groups might obtain copies, probably within the first six months of the database's release, and then we could exit gracefully from the mail-order business. Indeed, after six months, this prediction still seemed plausible — but orders continued to arrive steadily, at a rate averaging one per month, for the next nine years! During this period, we distributed about 100 copies of the database on nine-track half-inch digital tape at 800 and 1600 bpi, and a much smaller number of copies on quarter-inch IRIG-format FM analog tape.

The general standard of performance of commercial arrhythmia detectors improved rapidly, stimulated by the availability of the databases.

In early copies of the database, most beat annotations were placed at the R-wave peak, but manually inserted labels and those that occurred during periods of noise were not always placed consistently at the peak. In 1983, we adjusted the positions of the beat annotations using software that digitally bandpass-filtered the primary signal (usually MLII) to emphasize the QRS complexes, and then positioned each annotation at the major local extremum within 100 ms of the original location, after correcting for phase shift in the filter. We reviewed the placement of the annotations that had been repositioned by the largest amounts; for a very few of these beats, which were severely corrupted by noise, we manually repositioned the annotations. This postprocessing step allowed the beat annotations to be used as reliable fiducials for studies requiring waveform averaging, as well as for high-precision measurement of inter-beat intervals in studies of heart-rate variability (once the mechanical sources of tape speed variability were understood).

Our own use of the database, and feedback from early users, allowed us to identify errors in the roughly 109,000 beat annotations. Sixteen such errors were corrected between 1980 and 1987, and in addition, the dominant beats in record 214, which had originally been identified as normal, were relabeled as left bundle branch block beats. No other beat annota-

tion errors have been found since 1987, despite intense scrutiny by many more users, and it is likely that none remain. The roughly 1000 rhythm annotations received more revisions, and now include annotations of ventricular bigeminy and trigeminy, and of paced rhythm, that were not present in early copies.

Throughout the 1980s, we collected many more recordings to support studies of important arrhythmias that were not well represented in the original MIT-BIH Arrhythmia Database. Although we distributed a few of these, the impact of a major tape replication activity on the function of our research laboratory was a limiting factor. In 1989, we were able to produce a CD-ROM containing not only the MIT-BIH Arrhythmia Database but also seven additional ECG databases (the current edition includes two more). Approximately 400 copies of these CD-ROMs have been distributed to date [4].

In 1999, we established PhysioNet (<http://www.physionet.org/>), a web-based resource for research on complex physiologic signals [5]. More than half of the MIT-BIH Arrhythmia Database is now available via PhysioNet, making it possible for students and others to use a significant portion of these data for exploratory studies without cost.

Using the Database

We initially avoided prescribing methods for using the database to evaluate arrhythmia detectors, to allow ourselves and other users of the database an opportunity to develop performance measures that might be predictive of “real world” performance [6]. In 1984, we proposed methods for beat-by-beat and episode-by-episode comparison of reference and algorithm-generated annotation files [7]. These methods became the basis for a recommended practice for evaluating ventricular arrhythmia detectors [8] developed under the aegis of the Association for the Advancement of Medical Instrumentation (AAMI) between 1984 and 1987. More recently, we have participated in the development of the current American National Standards for ambulatory electrocardiographs [9] and for evaluating arrhythmia and ST segment measurement algorithms [10], both of which specify evaluation protocols based on those in the earlier recommended practice.

Although the details of the evaluation protocols are beyond the scope of this article, an important principle was proposed

in [7] and adopted in the recommended practice and both of the national standards. In simplest terms, this principle requires that the algorithm or device under test must produce for each test recording either an annotation file in the format of the reference annotation files supplied with the database, or an equivalent information stream that can be transformed into a such a file using an algorithm or accessory device, the details of operation of which must be fully disclosed. All performance measurements are then determined by automated comparisons of the algorithm’s annotation files with the reference annotation files, using standard comparison software specified by the standards. Regulatory agencies and end users of arrhythmia analyzers are able to verify test results [11], since all materials needed (the test data, the comparison software, and any required accessory device needed to produce annotation files) are available to anyone.

Other Long-Term ECG Databases

No discussion of the MIT-BIH Arrhythmia Database would be complete without mention of the two other important collections of long-term ECGs that are also available to researchers: the AHA Database for Evaluation of Ventricular Arrhythmia Detectors [12] and the European ST-T Database [13]. The AHA Database was created between 1977 and 1985 by a group led by G. Charles Oliver at Washington University in St. Louis. This database has many features in common with the MIT-BIH Arrhythmia Database. Notably, both databases contain two-channel Holter recordings, with each recording containing 30 minutes of signals that have been meticulously hand-annotated beat-by-beat. Close and sustained cooperation between the groups at Washington University and at MIT ensured that these databases would appear in compatible formats and that their contents would be complementary.

Recordings included in the AHA Database were chosen to satisfy one of eight sets of stringently defined selection criteria based on the severity of ventricular ectopy. As a result, the AHA Database has excellent representation of the most severe types of ventricular ectopy. Twenty recordings were chosen for each of the eight sets. Each of these was divided into equal subsets, one for algorithm development and one for evaluation of performance. The first of the “development” records were distributed in

1982, and all 80 have been distributed since 1985 by ECRI [12]. In 2000, ECRI made the 80 "evaluation" records available for the first time.

The AHA Database contains relatively few examples of supraventricular ectopy, conduction defects, and noise-contaminated waveforms, all of which are commonly encountered in clinical practice. By contrast, a number of records in the MIT-BIH Arrhythmia Database were selected specifically because they contain complex combinations of rhythm, morphologic variation, and noise that can be expected to provide multiple challenges for automated arrhythmia analyzers.

In 1985, the group headed by Carlo Marchesi at the CNR Institute for Clinical Physiology in Pisa assessed how they might contribute most usefully to the collection of reference ECG recordings available to researchers. They chose to take on the challenge of creating a database for development and evaluation of changes in the ST segment and the T wave indicative of myocardial ischemia. (The portion of the ECG waveform that follows the QRS complex in each cardiac cycle, consisting of the ST segment and the T wave, reflects ventricular repolarization. If the amount of oxygen delivered by the coronary arteries to the ventricular myocardium is insufficient to meet the demand for oxygen, ischemia results, and usually produces distinctive changes in the ST segment and the T wave.)

In the years following the creation of the MIT-BIH and AHA Databases, improvements in ambulatory ECG recorders permitted accurate reproduction of components of the ECG in the 0.01-0.10 Hz frequency range, needed in order to observe these changes. With the support of the European Society for Cardiology, the Pisa group coordinated data collection from clinical laboratories in 11 European Union nations, eventually selecting 90 two-hour excerpts of two-channel long-term ECG recordings. These were annotated in their entirety, following the protocol used for the MIT-BIH Arrhythmia Database with the addition of new annotation types to indicate episodes of ST and T-wave change. The first 50 records of the European ST-T Database were completed and made available to researchers in 1990 [13], and the remainder of the database was completed in 1991. The database is available on a CD-ROM in the same format as the MIT-BIH Arrhythmia Data-

base, a result of close cooperation between the Pisa and MIT groups.

Results

The experience of the past 20 years since the publication of the MIT-BIH Arrhythmia Database, and the AHA Database shortly thereafter, can be regarded as a grand experiment in shaping the direction of development of arrhythmia detectors. Until the databases became available, performance statistics were of little or no value since it was widely understood that each manufacturer designed its products using its own data, and designed its statistics to present the products in a favorable light. Conscientious developers who sought to make their algorithms more accurate faced pressure to match their competitors' products feature for feature, rather than spending effort and money making improvements that could not be quantified, and therefore added no perceived value to the product.

What, then, were the results of the experiment? In the early 1980s, the appearance of the databases marked a sea change in development efforts. End users and regulatory agencies began to ask manufacturers how well their devices worked on standard tests. Manufacturers had little choice but to perform the tests and report the results, and those whose algorithms did not measure up to their competitors spent their development budgets in concentrated efforts to improve performance. The general standard of performance of commercial arrhythmia detectors improved rapidly, stimulated by the availability of the databases.

It would be incorrect, however, to suggest that manufacturers were capable of producing much better products in the late 1970s and chose instead to add bells and whistles in response to their customers' apparent lack of interest in performance. Rather, a lack of well-characterized data hindered the manufacturers' progress as well as that of academic researchers. Many industry and academic groups performed tedious evaluations in which unknown data were analyzed by their algorithms, and then the outputs of the algorithms were examined for errors. This evaluation approach is superficially attractive, because the process can be started at any time, using any available data. The major flaw in this approach is not that it introduced bias (although it did); it is that at the end of the process, it was just as expensive to evalu-

The future in development of these databases may be in web-based collaborations between geographically scattered researchers.

ate the next version of the algorithm, because no investment was made in studying the data, only in characterizing the algorithm errors. Furthermore, without a reproducible test, it was impossible to know if any differences in measured performance of two algorithms were due to differences in the algorithms or differences in the data.

The successes of the MIT-BIH Arrhythmia Database, the AHA Database, and the European ST Database demonstrated the value of creating generally available, representative, and well-characterized collections of ECGs. Their limitations stimulated the development of other databases. In particular, the needs of academic researchers are frequently for much longer recordings, so that temporal patterns of change within a single subject (for example, diurnal variations over a period of 24 or 48 hours) can be observed in detail.

Advances such as inexpensive high-capacity mass storage, laser printers, color graphics displays, high-speed/high-resolution analog-to-digital converters, and digital multichannel ECG recorders can allow us to avoid many of the problems we faced 25 years ago, as well as to gather higher quality data. The effort required to annotate the data in detail remains tedious and demanding, however. The future in development of these databases may be in web-based collabo-

rations between geographically scattered researchers, as in an ongoing project to develop a long-term ST database of 24-hour recordings [14]. This project is led by Franc Jager in Ljubljana (Slovenia), with participation from the developers of the European ST Database in Pisa and from our group at MIT. Using web servers, we are able to share data with our distant colleagues and to annotate these data collaboratively. Since our goal is to characterize ST changes in detail, and not primarily to identify arrhythmias (which are less common in these recordings than in the arrhythmia databases), we can use much more automation in the annotation process than would be reasonable when developing an arrhythmia database, where the introduction of bias in beat labels would be a greater concern. We meet as a group two or three times per year to confer on annotations and to plan work schedules for the next several months. In this way, we are creating a database roughly two orders of magnitude larger than the MIT-BIH Arrhythmia Database, in about the same amount of time.



George B. Moody is a research staff scientist in the Harvard-MIT Division of Health Sciences and Technology, and the designer and webmaster of PhysioNet. His research interests include automated analysis of physiologic signals, statistical pattern recognition, heart-rate variability, multivariate trend analysis and forecasting, and applications of artificial intelligence and advanced digital signal processing techniques to multiparameter physiologic monitoring.



Roger G. Mark, M.D., Ph.D., is distinguished professor of Health Sciences & Technology and professor of electrical engineering at MIT. He is a Co-PI of the Research Resource for Complex Physiologic Signals. Dr. Mark served as the co-director of the Harvard-MIT Division of Health Sciences & Technology from 1985 to 1996. Dr. Mark's research activities include physiological signal processing and database development, cardiovascular modeling, and intelligent patient monitoring. He led the group that developed the MIT-BIH Arrhythmia Database.

Address for Correspondence: George B. Moody, MIT, Room E25-505A, Cambridge, MA 02139 USA. E-mail: george@mit.edu.

References

- [1] N.J. Holter, "New methods for heart studies," *Science*, vol. 134, p. 1214, 1961.
- [2] P.S. Schluter, R.G. Mark, G.B. Moody et al., "Performance measures for arrhythmia detectors," in *Computers in Cardiology 1980*. Long Beach, CA: IEEE Comput. Soc. Press, 1981.
- [3] R.G. Mark, P.S. Schluter, G.B. Moody et al., "An annotated ECG database for evaluating arrhythmia detectors," in *Frontiers of Engineering in Health Care-1982, Proc. 4th Annu. Conf. IEEE EMBS*. Long Beach, CA: IEEE Comput. Soc. Press, pp. 205-210.
- [4] G.B. Moody and R.G. Mark, "The MIT-BIH Arrhythmia Database on CD-ROM and software for use with it," in *Computers in Cardiology 1990*. Los Alamitos, CA: IEEE Comput. Soc. Press, 1991, pp. 185-188.
- [5] G.B. Moody, R.G. Mark, and A.L. Goldberger, "PhysioNet: A web-based resource for study of physiologic signals," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 70-75, 2001.
- [6] G.B. Moody and R.G. Mark, "How can we predict real-world performance of an arrhythmia detector?" in *Computers in Cardiology 1983*. Long Beach, CA: IEEE Comput. Soc. Press, 1984, pp. 71-76.
- [7] R.G. Mark and G.B. Moody, "Evaluation of automated arrhythmia monitors using an annotated ECG database," in *Ambulatory Monitoring: Cardiovascular System and Allied Applications*. The Hague: Martinus Nijhoff, 1984, pp. 339-357.
- [8] Testing and Reporting Performance Results of Ventricular Arrhythmia Detection Algorithms [AAMI ECar]. Assoc. for the Advancement of Medical Instrumentation, Arlington, VA, 1987.
- [9] *American National Standard for Ambulatory Electrocardiographs*. AAMI/ANSI Standard EC38:1998, 1998.
- [10] *American National Standard for Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*. AAMI/ANSI Standard EC57:1998, 1998.
- [11] G.B. Moody, C.L. Feldman, and J.J. Bailey, "Standards and applicable databases for long-term ECG monitoring," *J. Electrocardiology*, vol. 26 (suppl.), pp. 151-155, 1993.
- [12] R.E. Hermes, D.B. Geselowitz, and G.C. Oliver, "Development, distribution, and use of the American Heart Association database for ventricular arrhythmia detector evaluation," in *Computers in Cardiology 1980*. Long Beach, CA: IEEE Comput. Soc. Press, 1981, pp. 263-266. [The AHA Database is available from: ECRI, 5200 Butler Pike, Plymouth Meeting, PA 19462 USA (<http://www.healthcare.ecri.org/>); Contact: Ms. Hedda Shupack, hshupack@ecri.org.]
- [13] A. Taddei, A. Biagini, G. Distanti et al., "The European ST-T database: development, distribution, and use," in *Computers in Cardiology 1990*. Los Alamitos, CA: IEEE Comput. Soc. Press, 1991, pp. 177-180. [The European ST-T Database is available from: Dept. of Bioengineering and Medical Informatics, National Research Council (CNR) Institute of Clinical Physiology, via Trieste, 41, 56126 Pisa, Italy; Contact: Alessandro Taddei, taddei@ifc.pi.cnr.it.]
- [14] F. Jager, G.B. Moody, A. Taddei et al., "A long-term ST database for development and evaluation of ischemia detectors," in *Computers in Cardiology 1998*. Piscataway, NJ: IEEE Press, 1998, pp. 301-304.