

A Database for Evaluation of Algorithms for Measurement of QT and Other Waveform Intervals in the ECG

P Laguna¹, RG Mark^{2,3}, A Goldberg³, GB Moody^{2,3}

¹ Dept. Ingeniería Electrónica y Comunicaciones. Centro Politécnico Superior, Univ. de Zaragoza, Spain

² Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, USA

³ Cardiovascular Division, Beth Israel Deaconess Medical Center, Boston, MA USA.

Abstract

We present here a QT database designed for evaluation of algorithms that detect waveform boundaries in the ECG. The database consists of 105 fifteen-minute excerpts of two-channel ECG Holter recordings, chosen to include a broad variety of QRS and ST-T morphologies. Waveform boundaries for a subset of beats in these recordings have been manually determined by expert annotators using an interactive graphic display to view both signals simultaneously and to insert the annotations. Examples of each morphology were included in this subset of annotated beats; at least 30 beats in each record, 3622 beats in all, were manually annotated in the database. In 11 records, two independent sets of annotations have been included, to allow inter-observer variability studies. The QT Database is available on a CD-ROM in the format previously used for the MIT-BIH Arrhythmia Database and the European ST-T Database, from which some of the recordings in the QT Database have been obtained.

1. Introduction

Measurements of the width or duration of waves in the ECG are widely used to define abnormal electrical conduction in the heart, to detect myocardial damage, and to stratify patients at risk of cardiac arrhythmias[1, 2]. A number of automated methods for making such measurements, especially of the QT interval, have been designed. Several studies have been undertaken to evaluate the performance of such algorithms, but they have been hampered by a lack of standardized databases containing a sufficiently large number of carefully annotated heartbeats with manually-made measurements of waveform boundaries. This situation probably reflects the tremendous effort required for a clinician to manually annotate a statistically significant set of QRST complexes. We have attempted to address this problem by constructing an annotated reference database (the QT Database) which includes a wide variety of ECG morphologies, and a significant number of patient records. This paper will describe the design and construction of the QT Database and associated utility software.

2. The QT Database

*This work was supported by grant TIC94-0608-C02-02, from CI-CYT, and PIT06/93 from CONAI[†]. Spain

The QT Database includes ECGs which were chosen to represent a wide variety of QRS and ST-T morphologies, in order to challenge QT detection algorithms with real-world variability. The records were chosen primarily from among existing ECG databases, including the MIT-BIH Arrhythmia Database[3], the European Society of Cardiology ST-T Database[4] (courtesy of Prof. Carlo Marchesi), and several other ECG databases collected at Boston's Beth Israel Deaconess Medical Center. The existing databases are an excellent source of varied and well characterized data, to which we have added reference annotations marking the location of waveform boundaries. The additional recordings were chosen to represent extremes of cardiac (patho)physiology. We gathered new data from Holter recordings of patients who experienced sudden cardiac death during the recordings, and age-and-gender matched patients without diagnosed cardiac disease.

The QT Database contains a total of 105 fifteen-minute excerpts of two channel ECGs, selected to avoid significant baseline wander or other artifacts. (We note that a consequence of this selection procedure is that heart rates during these excerpts tend to be relatively low, probably since higher rates are frequently associated with noisy signals that would have failed to satisfy our selection criteria.) Table 1 shows the sources of the data.

Table 1: Distribution of the 105 records according to the original Database.

MIT-BIH Arrhyt.	MIT-BIH ST DB	MIT-BIH Sup. Vent.	MIT-BIH Long Term	ESC STT	MIT-BIH NSR DB	Sudden Death
15	6	13	4	33	10	24

Within each record, between 30 and 100 representative beats were manually annotated by cardiologists, who identified the beginning, peak and end of the P-wave, the beginning and end of the QRS-complex (the QRS fiducial mark, typically at the R-wave peak, was given by an automated QRS detector), the peak and end of the T-wave, and (if present) the peak and end of the U-wave. In order to permit the study of beat-to-beat variations such as alternans, 30 consecutive beats of the dominant morphology were annotated in each case if possible. In records with significant QRS morphology variation, up to 20 beats of each non-dominant morphology

were also annotated. Only “normal” beats were annotated. The criteria for beat selection were: 1) all are classified as “normal” by ARISTOTLE[5], and 2) the preceding and following beats are also normal. Beats were only annotated during the final 5 minutes of the excerpts in order to allow analysis algorithms a minimum of 10 minutes for learning. In all, 3622 beats have been annotated by cardiologists. These annotations have been carefully audited to eliminate gross errors, although the precise placement of each annotation was left to the judgment of the expert annotators. The current edition of the QT Database includes two independently derived sets of annotations for 11 records (to permit study of inter-observer variability). The remaining 94 records contain only a single set of expert annotations.

All records were sampled at 250 Hz. Those which were not originally sampled at that rate were converted using `xform`, an application from the MIT Waveform Database Software Package [3], which is provided with the QT Database.

3. The Records

The records are provided in the MIT-BIH Database format. QT Database records have record names of the form `selnnnn`, where `nnnn` is the name of the original record in the source database. Each record includes a signal file (`record.dat`, where `record` is the record name), a header file (`record.heg`, describing the format of the signal file), and several annotation files: `record.atr` contains the original annotations from the source database; `record.ari` contains QRS annotations obtained automatically by ARISTOTLE[5]; `record.pu0` and `record.pu1` contain the automatic waveform onsets and ends in signals 0 and 1 respectively, as detected using the differentiated threshold method by `ecgpuwave`[6]; and `record.man`, `record.qtn`, and `record.qnc` contain the manual annotations, as described in the next section, for annotators $n=1,2$. The `record.atr` files do not exist for the 24 sudden death records, since no reference annotations were available in this case. ARISTOTLE classifies the detected beats as normal (N) or belonging to other beat types[5]. Those beats considered normal are clustered into different groups specified by the `num` fields of the `ari` annotations. In the `num` fields of the `pu*` annotations, `ecgpuwave` classifies the T waves as normal (0), inverted (1), only upwards (2), only downwards (3), biphasic negative-positive (4), or biphasic positive-negative (5). Waveform onset (“ and end “) annotations specify the waveform type in their `num` fields (0 for a P-wave, 1 for a QRS complex, 2 for a T wave, or 3 for a U-wave). All of these files are in the `/qtdb` directory of the QT Database CD-ROM.

Table 2 documents the make-up of the QT Database, showing the record name, the source database, the time interval excerpted (shown relative to the beginning of the original record in the source database), the number of manually annotated beats, and the pattern of the dominant waveform in each record.

Table 2. A: Records from MIT-BIH Arrhythmia Database

Record	Original interval time	annot beats	Waveform Pattern (p) (QRS) (t) (u)
sel100	7:00 → 22:00	30	(p) (N) t)
sel102	6:00 → 21:00	85	(N) t) u)
sel103	1:00 → 15:00	30	(p) (N) t) u)
sel104	15:00 → 30:00	77	(N) t)
sel114	0:00 → 15:00	50	(p) (N) t)
sel116	0:00 → 15:00	50	(p) (N) t) u)
sel117	15:00 → 30:00	30	(p) (N) t) u)
sel123	3:00 → 18:00	30	(p) (N) t) u)
sel213	0:00 → 15:00	71	(p) (N) t)
sel221	0:00 → 15:00	30	(N) t)
sel223	13:30 → 28:30	30	(p) (N) (t)
sel230	7:00 → 22:00	50	(p) (N) t)
sel231	3:00 → 18:00	50	(p) (N) t)
sel232	15:00 → 30:00	30	(A) t)
sel233	0:00 → 15:00	30	(p) (N) t)
15	3:45:00	673	

Table 2. B: Records from MIT-BIH ST Change Database

Record	Original interval time	annot beats	Waveform Pattern (p) (QRS) (t) (u)
sel301	14:00 → 29:00	30	(p) (N) (t)
sel302	8:00 → 23:00	30	(p) (N) t)
sel306	0:00 → 15:00	36	(p) (N) t)
sel307	0:00 → 15:00	30	(p) (N) t)
sel308	13:00 → 28:00	50	(p) (N) t)
sel310	4:00 → 19:00	30	(N) t) u)
6	1:30:00	206	

Table 2. C:

Records from MIT-BIH Supraventricular Arrhythmia Database

Record	Original interval time	annot beats	Waveform Pattern (p) (QRS) (t) (u)
sel803	0:00 → 15:00	30	(p) (N) t)
sel808	0:00 → 15:00	30	(p) (N) t) u)
sel811	7:30 → 22:30	30	(p) (N) t)
sel820	0:00 → 15:00	30	(p) (N) t)
sel821	0:00 → 15:00	30	(p) (N) (t)
sel840	14:50 → 29:50	70	(p) (N) (t)
sel847	14:22 → 29:22	33	(p) (N) t)
sel853	5:00 → 20:00	30	(p) (N) t)
sel871	4:00 → 19:00	70	(p) (N) t)
sel872	0:00 → 15:00	30	(p) (N) t)
sel873	0:00 → 15:00	33	(p) (N) t)
sel883	13:30 → 28:30	30	(p) (N) t)
sel891	14:59 → 29:59	71	(p) (N) t)
13	3:15:00	517	

Table 2. D: Records from MIT-BIH Normal Sinus Rhythm Database

Record	Original interval time	annot beats	Waveform Pattern (p) (QRS) (t) (u)
sel16265	10:37:30 → 10:52:30	30	(p) (N) t)
sel16272	11:13:10 → 11:28:10	30	(p) (N) t)
sel16273	11:22:36.5 → 11:37:36.5	30	(p) (N) t)
sel16420	14:24:00 → 14:39:00	30	(p) (N) t)
sel16483	13:32:00 → 13:47:00	30	(p) (N) t)
sel16539	20:56:00 → 21:11:00	30	(p) (N) (t)
sel16773	10:23:40 → 10:38:40	30	(p) (N) t) u)
sel16786	18:22:00 → 18:37:00	30	(p) (N) (t)
sel16795	15:00:00 → 15:15:00	30	(p) (N) (t)
sel17453	14:15:00 → 14:30:00	30	(p) (N) (t) u)
10	2:30:00	300	

Table 2. E: Records from European ST-T Database

Record	Original interval time	annot beats	Waveform Pattern (p) (QRS) (t) (u)
sele0104	1:35:00 → 1:50:00	30	(p) (N) t
sele0106	1:31:00 → 1:46:00	30	(p) (N) (t) u
sele0107	1:9:30 → 1:24:30	34	(p) (N) t u
sele0110	1:32:00 → 1:47:00	30	(p) (N) t
sele0111	1:19:00 → 1:34:00	30	(p) (N) t u
sele0112	1:32:10 → 1:47:10	50	(p) (N) t u
sele0114	1:38:00 → 1:53:00	30	(p) (N) (t) (u)
sele0116	00:37:30 → 00:52:30	30	(p) (N) t
sele0121	1:07:30 → 1:22:30	30	(p) (N) t
sele0122	40:00 → 55:00	30	(p) (N) t u
sele0124	1:41:00 → :56:00	50	(p) (N) (t)
sele0126	1:36:40 → 1:51:40	30	(p) (N) t u
sele0129	00:26:30 → 00:41:30	30	(p) (N) t
sele0133	00:37:30 → 00:52:30	30	(p) (N) t
sele0136	00:56:00 → 1:11:00	30	(p) (N) (t)
sele0166	00:52:30 → 1:07:30	36	(p) (N) (t)
sele0170	00:22:30 → 00:37:30	30	(p) (N) t
sele0203	00:52:30 → 1:07:30	30	(p) (N) t
sele0210	1:45:00 → 2:00:00	30	(p) (N) (t)
sele0211	1:45:00 → 2:00:00	30	(p) (N) (t)
sele0303	00:34:30 → 00:49:30	30	(p) (N) (t)
sele0405	00:22:30 → 00:37:30	30	(p) (N) t
sele0406	1:00:00 → 1:15:00	31	(p) (N) t
sele0409	1:20:00 → 1:35:00	30	(p) (N) (t)
sele0411	1:30:00 → 1:45:00	30	(p) (N) t
sele0509	00:15:00 → 00:30:00	30	(p) (N) t
sele0603	00:17:00 → 00:32:00	30	(p) (N) t u
sele0604	00:00:30 → 00:15:30	30	(p) (N) t u
sele0606	1:31:00 → 1:46:00	30	(p) (N) t
sele0607	1:00:10 → 1:15:10	30	(p) (N) t
sele0609	00:35:30 → 00:50:30	30	(p) (N) (t)
sele0612	00:19:30 → 00:34:30	30	(p) (N) (t)
sele0704	1:16:30 → 1:31:30	30	(p) (N) t u
33	8:15:00	1041	

Table 2. F: Records from "sudden death" patients from BIH

Record	Original interval time	annot beats	Waveform Pattern (p) (QRS) (t) (u)
sel30	7:39:30 → 7:54:30	30	(p) (N) (t)
sel31	12:57:00 → 13:12:00	30	(p) (N) (t)
sel32	20:52:20 → 21:07:20	30	(p) (N) (t)
sel33	14:57:00 → 15:12:40	30	(p) (N) (t)
sel34	5:53:00 → 6:08:00	30	(p) (N) (t)
sel35	4:52:00 → 5:07:00	31 (AF)	(N)
sel36	18:43:00 → 18:58:00	31	(B) (t) u
sel37	1:14:10 → 1:29:10	50	(Q)
sel38	5:00:00 → 5:15:00	30	(p) (N) (t)
sel39	00:04:30 → 00:19:30	30	(p) (N) (t)
sel40	7:07:00 → 7:22:00	30	(p) (N) (t)
sel41	00:05:00 → 00:20:00	30	(p) (N) (t)
sel42	10:03:00 → 10:18:00	30	(p) (N) (t)
sel43	08:00:00 → 08:15:00	30	(p) (N) (t)
sel44	21:30:00 → 21:45:00	30	(p) (N) (t)
sel45	17:30:30 → 17:45:30	30	(p) (N) (t) u
sel46	03:10:00 → 03:25:00	30	(p) (N) (t)
sel47	8:43:40 → 8:58:40	30	(p) (N) (t) (u)
sel48	07:11:00 → 07:26:00	30	(p) (N) (t)
sel49	23:10:40 → 23:25:40	30	(p) (N) (t)
sel50	10:53:00 → 11:8:00	32	(N) (t) u
sel51	00:55:00 → 01:10:00	30	(p) (N) (t)
sel52	00:08:00 → 00:23:00	30	(p) (N) (t)
sel17152	1:04:00 → 1:19:00	30	(p) (N) (t)
24	8:00:00	744	

Table 2. G: Records from MIT-BIH Long-Term ECG Database

Record	Original interval time	annot beats	Waveform Pattern (p) (QRS) (t) (u)
sel14046	9:14:13 → 9:29:13	31	(p) (N) t
sel14157	7:58:12 → 8:13:12	30	(p) (N) t u
sel14172	17:38:43 → 17:53:43	50	(p) (N) (t) (u)
sel15814	21:57:00 → 22:12:00	30	(p) (N) t
4	1:00:00	141	

Table 2: Data excerpts used in the QT Database

Holter records from the sudden death group are not calibrated with respect to amplitude; thus the signal gains recorded in the header files for these records are only estimates. Records *sel102* and *sel104* come from paced ECG recordings.

4. Manual Annotation

Manual annotations were made by two experts using a SUN workstation using WAVE[3]. The extraction of the annotation file which contains the QRS annotations was made automatically, generating the *record.man* annotation file. Each expert added estimated waveform boundaries to a copy of the *record.man* file. In this way we have for each *record* two manual annotation files denoted by *record.qtn*, where $n=1,2$ denotes expert who performed the annotations. A screen similar to that shown in Fig. 1 was presented to the expert annotators to support the editing process. Both leads were displayed simultaneously, and then a decision was made about the time location of the fiducial point. Fiducial points were marked by a ']' symbol, movable by a cursor, and the times were added to the annotation file. These annotations were audited to correct the inconsistencies detected (e.g., misplaced or missing annotations) and changed to the regular annotation symbols '(', ')', 't', 'p', and 'u', to produce the *record.qnc* files, containing the final manual annotations. For 11 records the procedure was repeated for a second annotator ($n=2$).

5. Software for the QT Database

The QT Database CD-ROM includes the MIT-BIH Waveform Database Software Package, containing many of the tools needed to work with this and other compatible databases (including the MIT-BIH Arrhythmia Database, the MIMIC Database, the AHA Database, the European ST-T Database, and the MGH Database). The major components of the DB Software Package are the DB library (subroutines that may be included in user-written C, C++, or Fortran software for reading and writing QT Database files), format-conversion software, and a variety of signal-processing applications. The DB library is provided in C source form; the other software in the package is provided in precompiled form for MS-DOS and several popular versions of UNIX. The CD-ROM also includes a demonstration version of WAVE, a graphical environment for manipulating sets of digitized signals with optional annotations. Samples and further details are available from <http://ecg.mit.edu/>.



Figure 1: WAVE screen display as used by the expert annotators for marking the locations of wave boundaries. The software automatically stores the chosen locations, and moves ahead to the next beat to be marked.

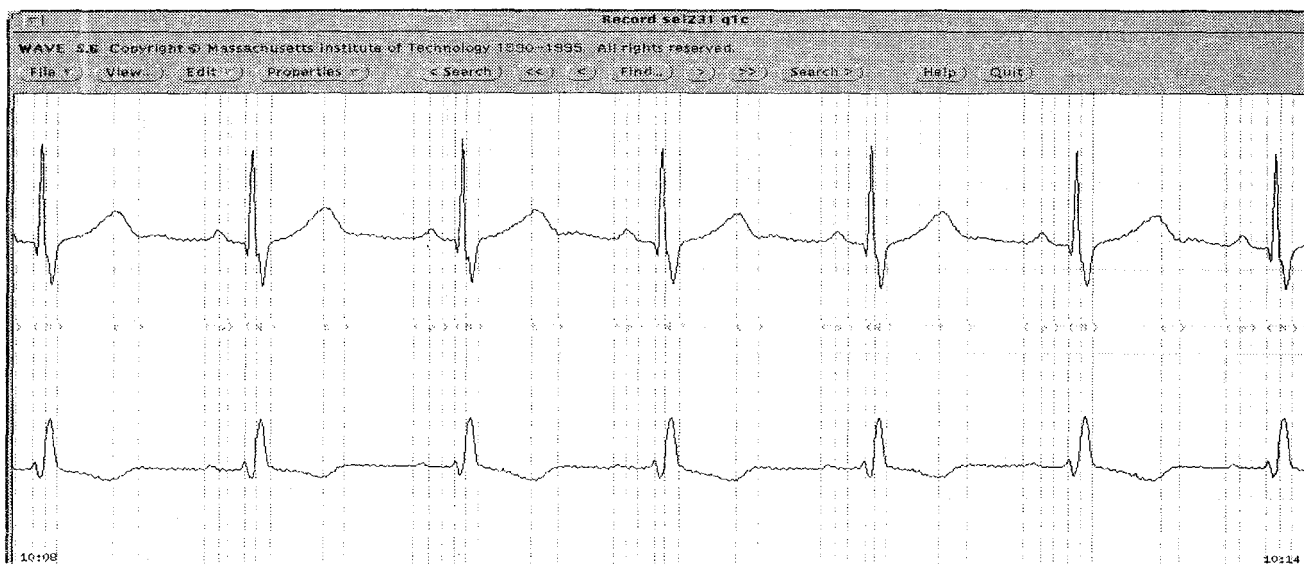


Figure 2: Screen display of the record in Fig. 1 after the manual annotations have been completed.

References

- [1] M.E. Nygard and L. Sörnmo, "Delineation of the QRS complex using the envelope of the ECG," *Med. Biol. Eng. Comput.*, vol. 21, pp. 538-547, March 1983.
- [2] The CSE Working Party, "Recommendations for measurement standards in quantitative electrocardiography," *European Heart Journal*, vol. 6, pp. 815-825, 1985.
- [3] 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*. IEEE Computer Society Press, 1990, pp. 185-188.
- [4] A. Taddei, A. Biagini, et al., "The European ST-T database: Development, distribution and use," in *Computers in Cardiology*. IEEE Computer Society Press, 1991, pp. 177-180.
- [5] G. B. Moody and R. G. Mark, "Development and evaluation of a 2-lead ECG analysis program," in *Computers in Cardiology*. IEEE Computer Society Press, 1982, pp. 39-44.
- [6] P. Laguna, R. Jané, and P. Caminal, "Automatic detection of wave boundaries in multilead ECG signals: Validation with the CSE database," *Comput. Biomed. Resear.*, vol. 27, no. 1, pp. 45-60, February 1994.

Address for correspondence:
 Pablo Laguna. Centro Politécnico Superior (Univ. Zaragoza)
 Maria de Luna 3, 50015 Zaragoza, Spain.
 e-mail: laguna@posta.unizar.es
 Web: <http://apollo.cps.unizar.es/deps/GTC>