The Long–Term ST Database: A Research Resource for Algorithm Development and Physiologic Studies of Transient Myocardial Ischemia

F Jager[1,2], A Taddei[3], M Emdin[3], G Antolič[4], R Dorn[1], GB Moody[2,5], B Glavić[1], A Smrdel[1], M Varanini[3], M Zabukovec[1], S Bordigiago[3], C Marchesi[3], RG Mark[2,5]

[1]Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia
[2]Harvard–M.I.T. Division of Health Sciences and Technology, Cambridge, MA, USA
[3]CNR Institute of Clinical Physiology, Pisa, Italy
[4]Department of Cardiology, University Medical Center, Ljubljana, Slovenia
[5]Cardiology Division, Beth Israel Deaconess Medical Center, Boston, MA, USA

Abstract

We present the Long Term ST Database, a collection of eighty 24-hour two and three lead ECG records from ambulatory subjects with transient ST segment abnormalities. The database provides a comprehensive standard research resource for quantitatively assessing the performance of automated detectors of transient ischemia, and for supporting basic research into the mechanisms and dynamics of transient ischemia. Records of the database contain annotated significant transient ischemic ST episodes, non-ischemic ST episodes caused by heart-rate related changes, non-ischemic ST events due to axis shifts or QRS conduction changes, and individual QRS and rhythm annotations, all made by human experts.

1. Introduction

Development of the *Long Term ST Database (LTST DB)* started in January 1995 under the joint project between the FCIS (Ljubljana) and MIT (Boston) sponsored by the U.S.-Slovenian Science and Technology Joint Fund Secretariat which yielded in a database of eleven 24-hour annotated ambulatory records. In 1997, Medtronic, Inc. agreed to sponsor further development of the database. At that time, the research groups from ICP (Pisa), BIDMC (Boston) and UMC (Ljubljana) joined to the project. In 1999, Zymed, Inc. agreed to additionaly sponsor the project in order to add a set of three lead ambulatory records. Goals of the LTST DB are to represent the wide variety of “real-world” data with many examples of significant non-ischemic ST events to gain development of improved algorithms to differentiate ischemic ST episodes from non-ischemic ST events, to include variety of temporal patterns in ST level on long run to enable researchers to study lengthly physiologic mechanisms and dynamics of ischemia, to include variety of rhythm abnormalities to enable studies on their correlations with ischemic ST episodes, and to permit more reliable prediction of clinical performance from the first-order performance statistics due to long duration records. In our previous report, we explained aspects and reasons for development of the LTST DB, and reported current status of protocols and annotating of transient ST events [1]. As the pool of collected and annotated records has grown, the characteristics of transient ST events became more clear. It was necessary to enhance definitions of transient ST events as well as annotation protocol to accurately cover all possible characteristics and origins of ST events. The aim of this paper is to report the final design and construction of the database. We present the selection criteria for records, improved annotation protocol with definitions of significant transient ST events, annotating procedure, improved tools to identify, classify and semiautomaticaly annotate ST events, and present the content of the records of the database.

2. The LTST database

The LTST DB contains 24-hour ambulatory records selected from Holter recordings obtained in routine clinical practice settings in Europe and in the United States between 1994 and 2000. Contributions were made from the BIDMC, ICP, Brigham and Women’s Hospital (Boston), and Duke University Medical Center (Durham). The records were collected to model real-world clinical conditions as far as possible with or without known coronary artery disease while containing significant number of ischemic and non-ischemic ST events. Analog records were made using standard AECG recorders. Since AECG recorders preserves frequency content in the signals typically up to 30 Hz, or to 45 Hz in best cases, we digitized the records at 128 or 250 samples per second per channel depending on the scanning system (Marquette MARS, ICR7200, Oxford Medilog 4-24, REMCO LP 103, ZYMED) with

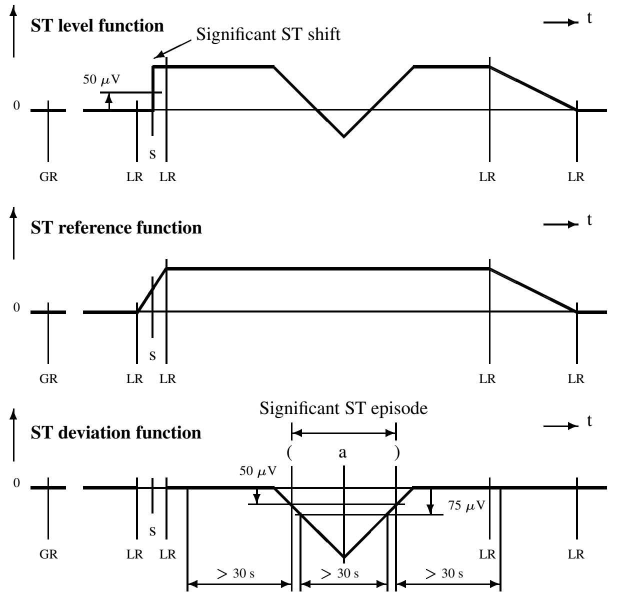

Figure 1. Representation of annotation protocol, definition of significant ST shift and significant ST episode, and schematic representation of annotating. ST deviation function is obtained as change in ST level function from which the ST reference function is subtracted. For the legend refer to table 1.

the resolution of 12 bits. After resampling the records to unique sampling frequency of 250 samples per second and adjusting amplitude scale to 200 ADC units per mV, the records were preprocessed [1]. Trends of the derived features, original ECG data and clinical informations formed the basis for selecting the records. Each selected record contains significant transient ST segment episodes corresponding to known ischemia (ischemic ST episodes), significant non-ischemic heart-rate related transient ST episodes, significant non-ischemic ST events due to axis shifts (postural changes), or significant non-ischemic ST events due to changes in QRS conduction. Some of the records contain arrhythmias such as atrial and ventricular ectopy, atrial fibrillation, and/or atrio-ventricular and intraventricular conduction defects. Other records were selected to include examples of baseline ST displacement resulting from conditions such as hypertension, ventricular dyskinesia, and effects on medications. We also included a number of 24-hour records with proven acute myocardial ischemia such as effort, resting, unstable, mixed, or Prinzmetal's angina originally recorded at the ICP from which the 2-hour excerpts in the European Society of Cardiology ST-T Database (ESC DB) were obtained. The LTST DB contains *sixty-five* 24-hour two lead and *fifteen* three lead ambulatory records with significant ST events annotated by human experts. The records also provide individual QRS and rhythm annotations made by an expert Holter scanning technician using a Marquette MARS system. Each record is accompanied with detailed and compact clinical summary including: age and sex; Holter report; report on other clinical investigations (if performed) such as ventriculography, ECG stress test, thallium positron emission tomography, echocardiography, stress echo, coronary angiography; relevant information on patient conditions what includes history, medications, history of arterial hypertension, previous myocardial infarction, left ventricular function, cardiomyopathy, valve disease, electrolyte disturbances, hypercapnia, intraventricular conduction block, baseline ECG, final diagnosis; and technical information about the record including recorder, leads, date and starting time of recording.

During development of the LTST DB, it became obvious that non-ischemic ST events such as sudden axis shifts, slow changes in QRS axis, QRS conduction changes and slow drifts of ST level in general appear without regularity, and are thus forming mixtures of non-ischemic ST events. For these reasons, the expert annotators established improved annotation protocol. ST events were defined and annotated independently in each channel. Initially, the protocol requires manual identification of the isoelectric and J points simultaneously in all ECG leads throughout the records. The basis for annotating ST events in each ECG lead was the *ST level function* (see figure 1), which was defined as change of *ST segment amplitude* over time measured 80 ms after the J point, or 60 ms after if heart rate exceeds 120 bpm. ST segment amplitude measurements and positions of the isoelectric and J point were obtained on time-averaged (16 seconds) heart beats derived for each "clean" beat which passed the preprocessing phase. ST level function typically varies widely in amplitude due to drifts, position changes, changes in conduction, intermittent QRS conduction changes, heart rate changes, and ischemia. Since non-ischemic ST events could also cause significant (> 50 μV) shifts in ST level function, the annotating cardiologists manually *tracked* the ST segment level to eliminate these non-ischemic ST changes. The resultant *ST reference function* (defined as piecewise linear function between the knot points as annotated throughout the record by local-reference annotations in the ST level function) approximates the ST reference level and was after that subtracted from the original ST level function to form the *ST deviation function*. To successfully annotate ST events, annotators considered ST level and ST deviation functions, original ECG signals, time series of QRS complex and ST segment KL coefficients, and clinical information about the patient (final diagnosis, other investigations, patient history). Annotators also established few characteristic categories relating to time-domain ST segment and QRS complex morphology describing their temporal change, origin or nature:

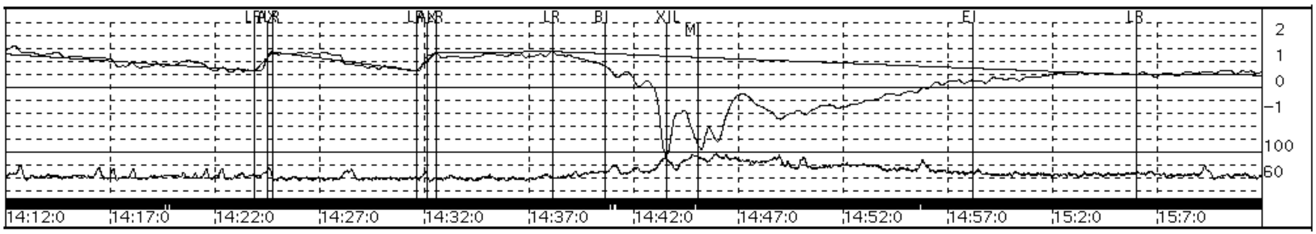

Figure 2. Example of tracking the ST segment level in the lead 0 of patient sb0047. The ST level function and ST reference function (piecewise linear) are shown higher (resolution 100 μV per unit). Instantaneous heart rate function is lower ([bpm]). Two significant ST shifts due to shifts in the electrical axis of the heart (AX) are annotated at approximate times 14:24:30 and 14:32:0. They were bounded and tracked out by local references (LR). A significant ischemic ST episode (BI, XIL, EI) is also bounded by two local references.

• *Change in QRS complex morphology* due to slow or sudden (postural - axis shifts) changes in the cardiac electrical axis characterized by change in the Q, R or S wave amplitude, or due to right or left bundle branch block, or other slow or sudden intraventricular conduction defects characterized by bizarre and wider QRS complex.

• *Non-ischemic change in ST segment morphology* due to simultaneous slow or sudden changes in the cardiac electrical axis, due to simultaneous slow or sudden conduction defects, or due to slow drifts. Drifts are characterized by slow and persistent non-ischemic change in ST segment slope and shape within longer period, and may or may not be accompanied by change in heart rate. Category of drift includes heart-rate related diurnal changes and effects of medication on repolarization.

• *Non-ischemic heart-rate related change in ST segment morphology* characterized by change in heart rate and by non-ischemic change in ST segment morphology within shorter period (J-point depression with positive slope, parallel shift of ST segment compared to reference or basal ST segment).

• *Ischemic change in ST segment morphology* characterized by ischemic change in ST segment morphology (flattening, down sloping, changing shape, scooping, elevation) and may or may not be accompanied by change in heart rate.

• *Noisy ST interval* characterized by consecutive ST segments which cannot be evaluated by annotators because of noises.

During establishing the ST reference level, the annotators first set simultaneously for each ECG lead the *global reference* annotation (for the annotation codes see table 1) representing the first stable five-minute interval after the beginning of the record without significant variation in the ST level (basal interval). Annotators tracked the ST segment level by a sequence of *local references*. They tracked everything but ischemic and heart-rate related ST episodes, and changes due to noises. Individual ST episodes of both types, or salvos or sequences of these episodes, were preceded and ended by a local reference. The ST segment level was tracked in the cases of drift, or in the cases of non-ischemic change in ST segment morphology which had to be accompanied by simultaneous change in QRS complex morphology and also evident in the time course of QRS complex KL coefficients. Changes of ST level function tracked were significant ($> 50\mu$V) or not. Any significant sudden-step change of ST level function which was accompanied by simultaneous sudden-step change in QRS complex morphology was bounded by a local reference before and after the step change and was annotated as *significant axis shift* or *significant conduction change* according to its nature. Figure 2 shows an example of tracking the ST segment level when two significant ST shifts and significant ST episode are present. Other ST events were annotated in the ST deviation function. In order to be annotated, transient ST episode had to be significant satisfying following criteria:

• An episode begins when the magnitude of the ST deviation function first exceeds 50μV;

• The deviation must reach 75μV or more throughout a continuous interval of at least 30 seconds;

• The episode ends when the deviation becomes smaller than 50μV, provided that it does not exceed 50μv in the following 30 seconds;

Significant ST episodes accompanied by non-ischemic heart-rate related change in ST segment morphology were annotated as *significant heart-rate related ST episodes*, while those accompanied by ischemic change in ST segment morphology as *significant ischemic ST episodes*. Sometimes significant axis shifts or conduction changes appeared within significant ST episodes. In these cases, they were not tracked out, but annotated within the episodes. Sometimes significant ST episodes appeared due to noisy ST intervals as bumps of ST deviation function. Those shorter were annotated as *noisy events* at their extrema, while those longer as *unreadable intervals*. Longer intervals with all heart beats rejected during preprocessing as noisy were annotated as unreadable intervals as well.

Annotating procedure consisted from three phases. The first phase consisted from deriving ST level functions on the basis of manually adjusted positions of the isoelectric level and J point in time-averaged (16 seconds) heart beats which corresponded to frequent manually set "dummy" annotations along the records [1]. These positions estimated correct positions of the isoelectric and J point for the corresponding heart beats in the centers of averaging windows. The positions of the isoelectric level and J point for the rest of clean heart beats were calculated by interpolating the two positions at dummy annotations along the records. In the second phase, annotators set local references to estimate ST reference functions and set annotations indicating significant ST shifts. During the third phase, the ST reference functions were reviewed/corrected, significant ST episodes were annotated automatically in the ST deviation function, and after that manually verified/corrected. Annotators reached the agreement on the annotations during joint meetings.

Identification and classification of transient ST episodes was accomplished by expert cardiologists using SEMIA (semiautomatic), version 3.0 (mainly developed by the fifth author), a special purpose graphic event-driven user interface and signal-processing tool designed specially for this project. It provides detailed insights into the data at multiple resolutions, examines data at any point, presents the real ECG waveforms and trends of ST and QRS functions, and supports manual and automatic annotation of the records by cardiologists at different sites interacting via the Internet and without paper tracings. New features of version 3.0, as compared to previous version [1], are following: up to three ECG channels, user selects "lead", "data" and "KL" windows to form his/her own interface, dynamic menus not permitting an annotator to go out of the required protocol and stay consistent, several fast modes simplifying repeatable operations, memorizing operations in the mouse buttons, and enhanced menus regarding automatic deletion and changing attributes of annotations.

The LTST DB record files are in the WFDB format and contain detailed clinical information for the subjects, waveform data, true QRS annotations, and ST annotations. Each record is composed by header file **s**N*xxxx*.**hea** (where N is 2 or 3 describing the number of ECG signals and *xxxx* is the record number), signal file **s**N*xxxx*.**dat**, ARISTOTLE's QRS annotation file **s**N*xxxx*.**ari**, QRS annotation atruth file **s**N*xxxx*.**atr**, ST annotation atruth file, **s**N*xxxx*.**sta**, and ST segment measurements atruth file **s**N*xxxx*.**16a**. Header files describe the format of the signal files and contain technical information about the records, comments of expert annotators, and detailed and compact clinical summaries for the subjects. ARISTOTLE's QRS annotation files contain automatically derived QRS annotations and heart-beat fiducial points which were used during preprocessing and annotating the records. QRS annotation atruth files contain individual QRS and rhythm annotations made by expert Holter scanning technician. ST annotation atruth files contain ST annotations (see table 1), while ST measurements atruth file contain measurements obtained on

Code	Meaning
GRST n	*Global reference*
LRST $n \pm llll$	*Local reference*
s[cc] st n	*Significant ST shift*
([rt] st $n \pm dddd$	*Beginning of significant ST episode*
a[rt] st $n \pm dddd$	*Extrema of significant ST episode*
[rt] st $n \pm dddd$)	*End of significant ST episode*
noi $n \pm dddd$	*Noise*
(urd n	*Beginning of unreadable interval*
urd n)	*End of unreadable interval*

Table 1. ST annotation codes used for the LTST DB. [cc]: type of ST shift (none: axis shift, cc: conduction change), [rt]: type of ST episode (none: ischemic, rt: heart-rate related), n: lead number (0, 1 or 2), $llll$: ST level in μV, $dddd$: ST deviation in μV.

average heart beats. These measurements were attached back to individual heart beats in the centers of averaging windows. An annotation corresponds to each clean beat and contains: ST amplitude measurements at the points J+80(60)ms, J+0ms, J+20ms, J+40ms, J+60ms, J+80ms, J+100ms, and J+120ms; positions of the isoelectric level and J point relative to the ARISTOTLE's fiducial point; and the number of heart beats left and right to the center beat included into the average beat.

3. Discussion and conclusions

Database development was complex, resource intensive, and time consuming. Semiautomatic interactive graphic tools were critical. They allowed us to work paperless and facilitated international cooperation via Internet. Project benefited from the expertise, resources, and experiences of the research groups. Samples of the database will be available on *http://www.physionet.org*, while the database will be published on CD-ROMs and will be released in the spring of 2001.

Acknowledgements

This work was sponsored by Medtronic, Inc., Minneapolis, MN, USA, and by Zymed, Inc., Camarillo, CA, USA.

References

[1] Jager F, Moody GB, Taddei A, Antolič G, Emdin M, Smrdel A, Glavić B, Marchesi C, Mark RG. A long-term ST database for development and evaluation of ischemia detectors. In Computers in Cardiology 1998. Piscataway: IEEE Society Press, 1998; 301–304.

Address for correspondence:

Franc Jager
Faculty of Computer and Information Science
Tržaška 25 / 1001 Ljubljana / Slovenia
Tel./Fax: +386-61-1768-362/1264-647
franc@manca.fri.uni-lj.si