

# A Graphical Evaluation Tool to Utilize ECG Data Without Reference Annotation

Yu-He Zhang, Saeed Babaeizadeh

Advanced Algorithm Research Center, Philips Healthcare, Andover, MA, USA

## Abstract

*Without reference annotation, statistical metrics such as sensitivity and positive predictive value (PPV) cannot be calculated. Annotating a large ECG database may not be feasible, hence, the interest in developing an evaluation tool that does not require reference annotation. We developed a tool for evaluating key performance attributes (KPA) including arrhythmia detection, heart rate, ST value, and noise tolerance. The tool has three layers of KPA graphics. The top layer includes interactive distribution graphs of the KPA values for aggregated results for the entire database. From this top layer the user can select an individual record to launch interactive trending graphs that display the KPA values, or their discrepancies, for a time span on that particular record. From this second layer the user can identify any KPA value of interest (e.g., a specific arrhythmia label) to view the underlying ECG waveform. Navigating through these three layers, the user is able to quickly confirm the validity of KPA reported by the algorithm. We modified the noise tolerance of an exercise ECG arrhythmia algorithm. Then used this tool to visually verify the resulting improvement on the Telemetric and Holter ECG Warehouse (THEW) stress database E-OTH-12-0927-015. We confirmed the visual verification of improvement by manually annotating a small subset of records in this database.*

## 1. Introduction

Manual evaluation of ECG records is a slow and tedious process. It needs skills with many hours of training and years of experience. Computerized algorithms have been developed to imitate human expertise which significantly reduced burden and improved efficiency of the process. Before being approved for clinical use, algorithms will have to be validated using datasets with reference annotation that are obtained either using some other modality or through an agreement among multiple human experts. Software tools that are developed to evaluate an algorithm usually provide statistical performance of the algorithm on clusters of ECG records. To investigate an

individual record, a different software tool is typically needed to visualize the ECG and provide accessories for morphological analysis. To make use of both sets of statistical and visualization tools, users may have to frequently switch between the tools that would add overhead to the process.

It is known that an algorithm's performance decreases when the noise level is increased. Furthermore, different types of noise impact the algorithm performance differently. ECG records with high level of added noise are not easy to annotate even for human experts, which is one reason that large exercise stress ECG databases with expert annotation are hard to find. Regardless of the rigorous development and validation, no ECG algorithm will ever have perfect accuracy under varying conditions in the field. The final interpretation of the ECG may always be done by expert clinicians.

In addition to monitoring and diagnostic software, tools are in increased demand for other tasks. For example, in a hospital with large daily flow of ECG reports and handful clinicians, software tools can be used to pre-screen the records to reduce the burden of manual screening.

We developed a combination suite of algorithm and data tools for quick visual evaluation of key performance attributes (KPA) such as arrhythmia detection, heart rate (HR), ST value, and noise tolerance. Using this toolbox one can quickly go from the KPA distributions on the entire database to the ECG waveform of any particular KPA values for an individual record.

## 2. Materials & Methods

### 2.1. Database

The algorithms under evaluation are for arrhythmia and ischemia detection in cardiac stress test. These algorithms have been independently verified and validated with annotated ECGs from MIT-BIH Arrhythmia database [1], AHA Arrhythmia database, and European Society of Cardiology (ESC) ST-T database. The ECGs in those databases have relatively low level of noise.

To evaluate the algorithms under stressed noisy conditions, we used Telemetric and Holter ECG Warehouse (THEW) stress database E-OTH-12-0927-015 [2]. There are 927 records with durations mostly between 11 and 30 minutes. Those records are not annotated by human experts for beat labels, arrhythmia, or ST values.

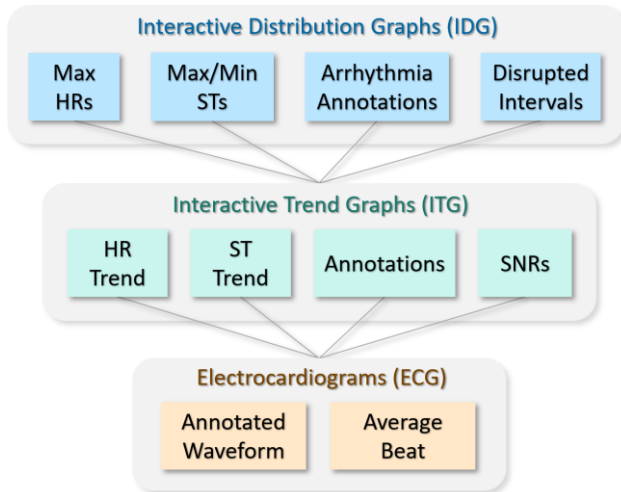


Figure 1. The 3 layers of the hierarchical architecture.

## 2.2. Architecture

We developed a suite of tools for quick visual inspection of KPAs such as arrhythmia detection, HR, ST value, and noise tolerance. The suite is architected with 3

layers of graphics as shown in Figure 1. The top layer is the Interactive Distribution Graphs (IDG) that can launch 1 of the 4 KPA distribution graphs, max HR, max/min ST, arrhythmia annotations, and disrupted intervals, for the entire database. The second layer is the Interactive Trending Graphs (ITG) for each lead that contains trending curves of HR and ST, and scatter plots of arrhythmia annotations and Signal-to-Noise Ratio (SNR). The third layer is the waveform (ECG) at the KPA of interest.

### 2.2.1. Interactive Distribution Graphs

The IDG is a set of graphs for the KPAs over the entire database, including maximum heart rate distribution and maximum ST elevation/depression distribution for each of the 12 leads, arrhythmia annotations, disrupted intervals (the impact of stress/noise on algorithms), as shown in Figure 2. Each mark on the max HR and max/min ST distributions in IDGs represents one particular record. In the arrhythmia annotations and disrupted intervals IDGs, each horizontal line represents a record with KPAs color coded at their corresponding times. The records are sorted with the highest counts or durations starting from the bottom. Based on the KPA distributions in the IDGs for the

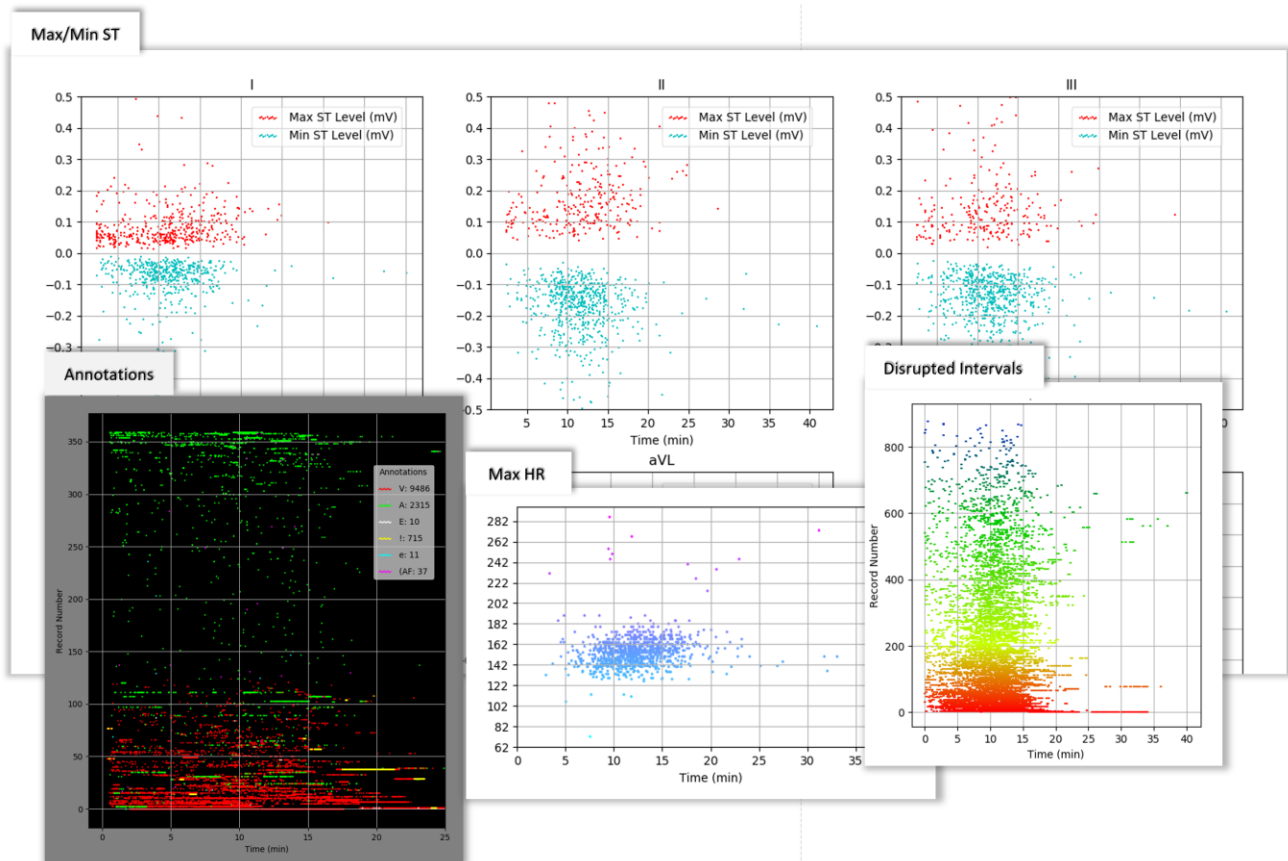


Figure 2. The 4 IDGs for all records in the database: 1) Max/Min STs for every lead; 2) Max HRs; 3) Arrhythmia Annotations; 4) Disrupted Intervals.

927 THEW records, as shown in Figure 2, the stress test appeared to reach the most intense stage around the 11<sup>th</sup> to 13<sup>th</sup> minute window. There was a clear statistical correlation between the noise levels, the heart rates, and the ST levels for reaching their highest deviations from the resting baseline around the 11<sup>th</sup> to 13<sup>th</sup> minute window as shown in Figure 2. The arrhythmia distribution, on the other hand, does not show a significant correlation with the stress level.

### 2.2.2. Interactive Trending Graphs

The ITG consists of trending KPAs for each of the leads, including trending curves of ST levels and heart rates, and scatter plots of arrhythmia annotations and SNR for each lead, as shown in part of Figure 3 for lead V5 in record PR01\_190204\_4. The HR trend shows that the stress exercise progressed to the peak at around 15<sup>th</sup> minute with max HR of 170 BPM and ST depression of around 0.24 mV at around 13.5<sup>th</sup> minute. It also shows 9 PVCs and 3 APCs detected by the algorithm during the stress test.

### 2.2.3. ECG

A segment of ECG waveform with annotations generated by the algorithm is instantaneously prompted and refreshed by pressing and gliding over the KPAs of interest in the ITG, as shown in part of Figure 3. Every beat of the ECG waveform is supposed to be annotated by the algorithm. A missing annotation usually indicates the algorithm failure to detect a beat, mainly due to noise. The instant ECG segments directly verify the accuracy of those corresponding KPAs, therefore the validity of the output of the

algorithm. An average QRS template is also displayed as inlet for ST value inspection.

Many automated algorithms use only few leads for arrhythmia analysis although more leads may be available. These “arrhythmia leads” are usually selected to be orthogonal to each other to maximize the detectable physiological information. These leads, however, may not necessarily be the cleanest leads throughout the test. But with ITGs and ECGs available for all leads, 12 for each of the THEW records that we used, user can choose any lead for better judgment, as it may be cleaner or have more prominent morphological profile.

### 2.3. Workflow

User can start launching from any of the 4 IDGs (max

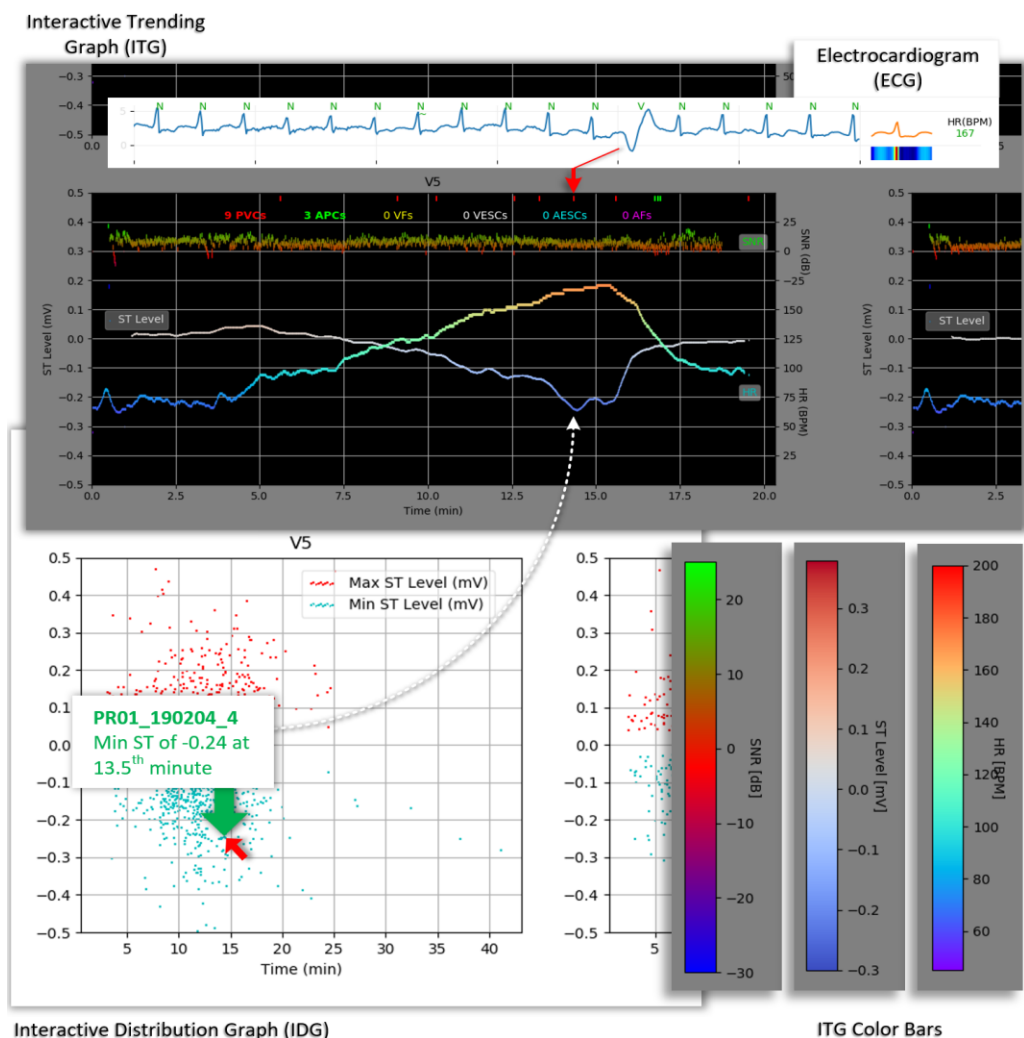


Figure 3. The workflow: 1. Launch Max/Min ST IDG. 2. Move cursor to point of interest (POI) for V5 at minimum ST of -0.24 at 13.5<sup>th</sup> minute, which prompts up record PR01\_190204\_4. 3. Clicking POI on IDG launches ITG for the record. 4. Move cursor to POI on ITG at a PVC annotation at around 13.5<sup>th</sup> minute. 5. Clicking POI on ITG launches a segment of ECG with annotations and average QRS template. Moving pressed cursor refreshes the ECG continuously.

HR, max/min ST levels, arrhythmia annotations, disrupted intervals) or directly from an ITG of any particular record. As an example, we start from the max/min ST IDG. We see from the ST IDG that the distribution of max ST depression demonstrated that leads II, III, aVF, V4, V5, V6 were more likely to reveal prominent ST depression. This is in agreement with the THEW database clinical records which identified the number of records with ST depression as following: I: 3, II: 260, III: 258, aVR: 0, aVL: 2, aVF: 260, V1: 0, V2: 5, V3: 22, V4: 322, V5: 337, V6: 336. When the user navigates across the ST IDGs, the record information will pop up for the record where the cursor is landed on. As shown in Figure 3, the cursor landed on record PR01\_190204\_4 at around 13.5<sup>th</sup> minute with a ST depression of 0.24 for lead V5. Clicking on this KPA launches the ITG for the record. The ITG shows the trending progress of the ST levels, HRs, detected arrhythmia beats, and SNR for the signal quality. Now user can click any KPA of interest, such as the lowest ST, the highest HR, the better signal quality, or a particular arrhythmia beat label, to launch the ECG waveform. As indicated in Figure 3, we choose the PVC annotation at around 13.5<sup>th</sup> minute as an example. The ECG waveform clearly shows that a PVC beat was correctly identified by the algorithm. In addition, the morphology of the average QRS template on the right side of the ECG waveform confirms the ST depression.

## 2.4. Algorithm Comparison

The tools can be used to compare different algorithms side by side for the accuracy of declared KPA values, such as the arrhythmia annotations. Figure 4 shows a cropped portion of the ITG for record PR01\_281003\_5 with 2 rows of annotations generated by 2 different versions of an algorithm, with version A at the top row and version B at the bottom. The version A was developed with a goal to improve the noise tolerance under stressed conditions. As we can see side-by-side, the 2 versions generated different labels for some beats. User can promptly investigate the differences by launching and gliding through the ECG waveforms at the spot of the KPA discrepancies. The beat annotations for algorithm versions A and B are displayed

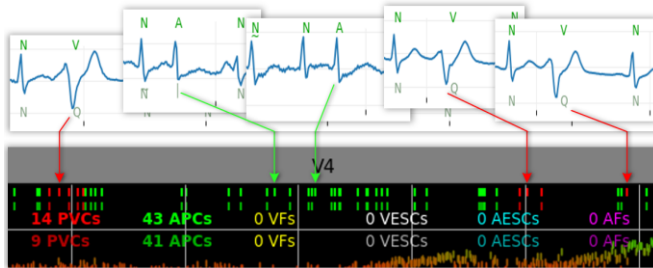


Figure 4. Part of ITG and ECGs for arrhythmia annotation comparison of 2 versions of an algorithm for record PR01\_281003\_5.

above and below the ECG waveforms, respectively. Looking at the ECG waveform, one can verify that the algorithm version A was able to identify the 3 PVCs and the 2 APCs that were missed by version B.

## 3. Results and Discussion

As shown in section 2, the tools provided a fast path to navigate from the entire database down to the waveform morphology for a particular record without having to switch between platforms. The KPAs displayed at the higher layer graphs can be investigated in the lower layer graphs through just few mouse clicks. At the lowest layer, the ECG waveform provides the visual evidence of the morphological details. We modified an exercise ECG arrhythmia algorithm with the goal of improving the noise tolerance. To get quantitative assessment of the tool's reliability, we first visually verified the resulting improvement on the THEW stress database E-OTH-12-0927-015 (n=927), specifically on the accuracy of the arrhythmia annotations. Then, to confirm the visual verification, we manually annotated only the 20 records in this database for which the algorithms had reported the most PVCs. We verified that the corresponding positive predictive value for these 20 records had indeed increased from 70% to 90%, while the improved algorithm was still able to maintain 95% sensitivity. This improvement was mainly due to higher tolerance of motion artifact.

## 4. Conclusion

Our graphical tool can be used to accurately evaluate the performance of algorithms on an unannotated database with a fraction of effort needed for annotating the entire database. It could also be a handy clinical research tool for quick evaluation of a live growing database where reference annotation is not yet obtained.

## References

- [1] Moody GB, Mark RG. The impact of the MIT-BIH Arrhythmia Database. *IEEE Eng in Med and Biol* 20(3):45-50 (May-June 2001). (PMID: 11446209).
- [2] Couderec JP. The telemetric and holter ECG warehouse initiative (THEW): a data repository for the design, implementation and validation of ECG-related technologies. *Conf Proc IEEE Eng Med Biol Soc* 2010; 2010. p. 6252-5.

Address for correspondence:

Yu-He Zhang  
3000 Minuteman Road, Andover, MA 01810  
yu-he.zhang@philips.com