This report provides the arrhythmia performance of the ST/AR (ST and Arrhythmia) algorithm. For a description of the algorithm, see the Arrhythmia Monitoring ST/AR Algorithm Application Note 453564115631.

Introduction

The performance of the algorithm used to detect and classify arrhythmias is fundamental to the effectiveness of computerized arrhythmia monitoring. It is important that the algorithm enable the system to alert the clinical staff to true arrhythmia events without unnecessarily distracting them from their duties. Such an algorithm will allow the monitoring system to generate accurate cumulative data that will be useful in supporting therapeutic decisions.

Procedures exist to facilitate the unbiased comparison of competing algorithms against identical data. A group of industry and user representatives under the auspices of the Association for the Advancement of Medical Instrumentation (AAMI) have published their recommendations for algorithm testing and reporting - Recommended Practice Document ANSI/AAMI EC57.

This application note summarizes the results of the PVC Performance Validation Test, Atrial Fibrillation detection test and the Ventricular Fibrillation Detection Test.
Reference Databases for Ventricular Arrhythmia Testing

Databases are available to the public for the testing of arrhythmia detection algorithms. These databases consist of records of patient ECG waveforms, together with a set of annotation files in which each beat has been labeled by expert cardiologists. Two of the databases that have been used to assess the performance of the ST/AR arrhythmia algorithm are as follows:

- The American Heart Association (AHA) database, distributed by the Emergency Care Research Institute (ECRI), consists of data from 80 patients, two of whom are paced.
- The Massachusetts Institute of Technology and Beth Israel Hospital (MIT-BIH) database consists of data from 48 patients, four of whom are paced.

The AHA and MIT-BIH databases were developed specifically for the purpose of evaluating ventricular arrhythmia detection algorithms. Both of these databases have two simultaneous channels of ECG.

Even though both the AHA and MIT-BIH databases contain some paced patients, data from paced patients (in digital form) do not retain sufficient pace pulse signal quality to allow for reliable detection of pace pulses. Therefore, the current recommended practice document recommends that the paced patients not be included in the evaluation.

QRS and PVC Performance Validation Test

The QRS and PVC Performance Validation Test quantifies the performance of the arrhythmia algorithm in QRS detection and PVC classification.

Performance Measures

The performance measures quantify the extent to which the algorithm results agree with the expert annotation of a given patient record. The annotation or labeling of each beat in the reference database by clinical experts is called the “true” classification of that beat.

QRS Detection Sensitivity is the percentage of true QRSs that the algorithm classifies as QRSs. It is calculated as the number of QRSs the algorithm correctly classifies, divided by the number of true QRSs, and then converted to a percentage.

A consistently high value for this performance measure indicates that the system is able to detect a high percentage of the QRS complexes.

QRS Detection Positive Predictivity is the percentage of beats that the algorithm classifies as QRSs that are true QRSs. It is calculated as the number of QRSs the algorithm correctly classifies, divided by the total number of beats that the algorithm classifies as QRSs, and then converted to a percentage.

A consistently high value for this performance measure indicates that the beats designated as QRSs by the algorithm are truly QRSs a high percentage of the time.

PVC Detection Sensitivity is the percentage of true PVCs that the algorithm classifies as PVCs. It is calculated as the number of PVCs the algorithm correctly classifies, divided by the number of true PVCs, and then converted to a percentage.

If the algorithm detects the PVCs of a majority of patients with a high level of sensitivity, one can expect the system to detect PVCs with a high degree of reliability. A sensitivity of 100%, that is, correctly identifying all PVCs, is the ultimate goal. Actual performance will tend to fall short of that goal.

PVC Detection Positive Predictivity is the percentage of beats that the algorithm classifies as PVCs that are true PVCs. It is calculated as the number of PVCs the algorithm correctly classifies, divided by the total number of beats that the algorithm classifies as PVCs, and then converted to a percentage.

A consistently high value for this performance measure indicates that the beats designated as PVCs by the algorithm are truly PVCs a high percentage of the time.

PVC Detection False Positive Rate is the percentage of beats that the algorithm incorrectly labels as PVCs relative to all non-PVC beats. It is calculated as the number of non-PVCs the algorithm incorrectly classifies as PVCs, divided by the number of non-PVC beats, and then converted to a percentage.

This measure is an indication of the most visible annoyance to the clinical staff. If the false positive rate is high, the arrhythmia monitoring system tends to generate a large number of false alarms. If a large percentage of patients have a low false positive rate, the system maintains a relatively high level of credibility.

It is important that an arrhythmia algorithm have:

- High QRS Detection Sensitivity
- High QRS Detection Positive Predictivity
- High PVC Detection Sensitivity
- High PVC Detection Positive Predictivity
- A low PVC Detection False Positive Rate

These five parameters together show how well an algorithm correctly detects QRS complexes and PVCs.

AAMI recommends that when comparing performance results and evaluating arrhythmia algorithm performance, it is important that all patient records in the database (except those of paced patients) be included.
Performance Results
The following results pertain to the ST/AR arrhythmia algorithm. ST/AR can perform multi-lead or single-lead analysis. This setting depends on your unit’s default setting or can be changed on an individual basis. In certain cases, such as the M3/M4 Monitor only single-lead analysis is available. In the following document both multi-lead and single-lead performance will be described.

Multi-Lead Results Summary
The following table summarizes the performance results of the multi-lead analysis.

Table 1: Multi-Lead Performance Summary

<table>
<thead>
<tr>
<th>Average Performance Measures</th>
<th>AHA</th>
<th>MIT-BIH</th>
</tr>
</thead>
<tbody>
<tr>
<td>QRS detection sensitivity (%)</td>
<td>99.80</td>
<td>99.66</td>
</tr>
<tr>
<td>QRS Detection Positive Predictivity (%)</td>
<td>99.87</td>
<td>99.86</td>
</tr>
<tr>
<td>PVC Detection Sensitivity (%)</td>
<td>95.96</td>
<td>94.25</td>
</tr>
<tr>
<td>PVC Detection Positive Predictivity (%)</td>
<td>98.34</td>
<td>96.38</td>
</tr>
<tr>
<td>PVC Detection False Positive Rate (%)</td>
<td>0.16</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Single-Lead Results Summary
The following table summarizes the performance results of the QRS and PVC performance validation test when using the best single lead for analysis.

Table 2: Single-Lead Performance Summary

<table>
<thead>
<tr>
<th>Average Performance Measure</th>
<th>AHA</th>
<th>MIT-BIH</th>
</tr>
</thead>
<tbody>
<tr>
<td>QRS detection sensitivity (%)</td>
<td>99.65</td>
<td>99.12</td>
</tr>
<tr>
<td>QRS Detection Positive Predictivity (%)</td>
<td>99.88</td>
<td>99.90</td>
</tr>
<tr>
<td>PVC Detection Sensitivity (%)</td>
<td>94.20</td>
<td>94.75</td>
</tr>
<tr>
<td>PVC Detection Positive Predictivity (%)</td>
<td>98.56</td>
<td>96.39</td>
</tr>
<tr>
<td>PVC Detection False Positive Rate (%)</td>
<td>0.14</td>
<td>0.27</td>
</tr>
</tbody>
</table>

These results show that the algorithm performs well on both of the reference databases whether using single-lead or multi-lead analysis.

The MIT-BIH database presents more difficult situations in which to detect PVCs because the second channel of this database is very noisy; there is a low signal-to-noise ratio and, in many cases, low signal amplitude.

The high values for PVC Detection Sensitivity indicate that the algorithm detects PVCs with a high degree of reliability. The consistently high values for PVC Detection Positive Predictivity indicate that the beats designated as PVCs by the algorithm are truly PVCs a high percentage of the time.

The PVC Detection False Positive Rates of less than 0.3% indicate that for every 1000 beats, the algorithm incorrectly classifies less than 3 PVCs.

Ventricular Fibrillation Detection Test
The Ventricular Fibrillation Detection Test quantifies the performance of the arrhythmia algorithm in detection of ventricular fibrillation (VF).

Performance Measures
The following three performance measures together describe algorithm performance in ventricular fibrillation detection:

Relative VF Detection Time is measured as the length of time from the beginning of a VF episode to the time that the arrhythmia system generates a Ventricular Fibrillation/Ventricular Tachycardia red alarm. In some instances, a negative alarm time results when the arrhythmia algorithm detects alarms before the annotated episode beginning time. In these cases, the alarms are triggered by the ventricular tachycardia which immediately preceded the onset of fibrillation or flutter. These are considered to be early detections of life-threatening episodes, not “false-positive” errors.

VF Detection Sensitivity is the percentage of true ventricular fibrillation episodes that the algorithm detects as ventricular fibrillation.

VF False Positive Detection is the number of times that the algorithm incorrectly detects ventricular fibrillation.

Results Summary
The following two histograms summarizes the VF detection time results of the VF detection test for multi-lead and for single-lead analysis. There are a total of 15 VF episodes in the AHA and MIT-BIH databases. With single-lead analysis each channel is treated separately resulting in 30 VF episodes. The horizontal axis in the histogram represents a range of relative VF detection times from 45 seconds before the onset of VF to 25 seconds after the onset of the episode. The vertical axis represents the number of VF episodes. The solid bars

ST/AR Arrhythmia Performance
indicate the number of VF episodes for which the relative VF detection time was in the range spanned by that bar.

Figure 1 Multi-Lead Algorithm Relative VF Detection Time AHA and MIT-BIH Databases

In terms of VF detection sensitivity and the AHA and MIT-BIH databases, a Ventricular Fibrillation or Ventricular Tachycardia red alarm is always generated by the arrhythmia system immediately preceding or within a few seconds of the VF start time. In terms of VF false positive detection, the algorithm detected only two false VF episodes across both databases. Both of these alarms occurred on ventricular tachycardia episodes with high heart rate. These are considered an appropriate announcement of life-threatening events, not false-positive errors.

Ventricular fibrillation and flutter is a chaotic and unpredictable event. There is no existing database that can be used to ensure that an arrhythmia algorithm will achieve 100% detection accuracy in all clinical episodes. However, the ST/AR arrhythmia algorithm has demonstrated 100% sensitivity in generating red alarms during life-threatening episodes of ventricular fibrillation using the AHA and MIT-BIH databases.

Atrial Fibrillation Detection Test

The atrial fibrillation detection test quantifies the performance of the arrhythmia algorithm in detecting atrial fibrillation (AF).

Reference Database

For performance evaluation, several databases, which were not used during the algorithm development, were used for the testing. The dataset consists of a total of 234 cardiologists annotated adult patient-records with 75 records contain AF episodes. The dataset includes three publicly available databases, including the MIT AF database with 23 patient-records, the AHA testing database with 75 patient-records, and the long-term ST database with 86 patient-records. The dataset also includes four proprietary databases collected from 4 different hospitals. These proprietary databases are labeled as Pafdb A, B, C, and D. The number of AF, AFL, and AF/FL episodes and their total duration are summarized in Table 3 below for all the individual databases.
Performance Measures

The performance measures quantify the extent to which the algorithm results agree with the cardiologists’ annotated AF episodes. The performance measures used to measure AF detection accuracy include the following four measures as described in the Recommended Practice Document ANSI/AAMI EC57: for reporting AF detection performance.

**AF Duration Detection Sensitivity** is calculated as the duration of overlapping between the algorithm-marked episode and the cardiologist-annotated episode divided by the duration of the cardiologist-annotated episode, and then converted to a percentage. This performance measure indicates how well the algorithm is able to marked the true AF episode.

**AF Duration Detection Positive Predictivity** is calculated as the duration of overlapping between the algorithm-marked episode and the cardiologist-annotated episode divided by the duration of the algorithm-marked episode, and then converted to a percentage. This performance measure indicates what percentage of the algorithm-marked episode duration is actually correct.

**AF Episode Detection Sensitivity** is the percentage of true AF episodes that the algorithm detected as AF. It is calculated as the number of AF episodes the algorithm correctly detected divided by the number of cardiologist-annotated episodes, and converted to a percentage.

**AF Episode Detection Positive Predictivity** is the percentage of the AF episodes that the algorithm detected as AF that are true AF. It is calculated as the number of AF episodes the algorithm correctly detected divided by the total number of AF episodes that the algorithm detected as AF, and then converted to a percentage.

**Results Summary**

For the final performance calculation, episodes less than 1 minute in duration and all atrial flutter episodes are excluded. The following table summarizes the performance results of the AF detection algorithm using the performance measures described above. These performance evaluation results show that the AF detection algorithm has achieved a high level of accuracy in detecting AF episode over a large patient population.

Table 3: Summary of AF Testing Database (not used in development)

<table>
<thead>
<tr>
<th>Database Name</th>
<th>Total Number of Records</th>
<th>Records with AF Epi</th>
<th>AF Epi #</th>
<th>D(sec)</th>
<th>AFL Epi #</th>
<th>D(sec)</th>
<th>AF/FL Epi #</th>
<th>D(sec)</th>
<th>Non AF #</th>
<th>D(sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>afdb</td>
<td>23</td>
<td>23</td>
<td>8</td>
<td>0</td>
<td>291</td>
<td>336232</td>
<td>14</td>
<td>5877</td>
<td>0</td>
<td>501290</td>
</tr>
<tr>
<td>ahatest</td>
<td>75</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>41665</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>10800</td>
</tr>
<tr>
<td>ltstdb</td>
<td>86</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>89730</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2147484</td>
</tr>
<tr>
<td>Pafdb A</td>
<td>10</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>24</td>
<td>37490</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>7200</td>
</tr>
<tr>
<td>Pafdb B</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>34638</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pafdb C</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>60008</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pafdb D</td>
<td>34</td>
<td>31</td>
<td>0</td>
<td>2</td>
<td>31</td>
<td>190120</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>14294</td>
</tr>
<tr>
<td>Total</td>
<td>234</td>
<td>75</td>
<td>8</td>
<td>4</td>
<td>358</td>
<td>789883</td>
<td>14</td>
<td>5877</td>
<td>12</td>
<td>32294</td>
</tr>
<tr>
<td>Total Hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>948.74</td>
</tr>
</tbody>
</table>

Abbreviations: AF = Atrial Fibrillation, AFL = Atrial Flutter, AF/FL = Combination of AF and AFL, NonAF = Rhythm is not AF or AFL, Epi = Episode, D(sec) = Duration of episode in seconds
## Table 4: AF Detection Performance Summary

<table>
<thead>
<tr>
<th>Database Record</th>
<th>AF Duration Detection</th>
<th>AF Episode Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Positive Predictivity</td>
</tr>
<tr>
<td>234 Records (All AF testing database)</td>
<td>94%</td>
<td>97%</td>
</tr>
<tr>
<td>232 Records (Excluding 2 sinus arrhythmia records)</td>
<td>94%</td>
<td>98%</td>
</tr>
<tr>
<td>23 records (MIT AF Database)</td>
<td>94%</td>
<td>98%</td>
</tr>
</tbody>
</table>

## Conclusion

The testing procedures recommended by AAMI using the reference databases provide a way to compare the performance of the arrhythmia monitoring algorithms. From the results presented in this application note, the excellent performance of the ST/AR arrhythmia algorithm is evident. This ability is basic to the effective monitoring of arrhythmia events in the clinical setting.
Philips Healthcare is part of
Royal Philips Electronics

Interested?
Would you like to know more about our imaginative products? Please do not hesitate to contact us. We would be glad to hear from you.

On the web
www.philips.com/healthcare

Via e-mail
healthcare@philips.com

By fax
+31 40 27 64 887

By mail
Philips Healthcare
Global Information Center
P.O Box 1286
5602 BG Eindhoven
The Netherlands

By phone
Asia
Tel: +852 2821 5888

Europe, Middle East, Africa
Tel: +49 7031 463 2254

Latin America
Tel: +55 11 2125 0764

North America
Tel: +1 425 487 7000
800 285 5585 (toll free, US only)