A NOISE STRESS TEST FOR ARRHYTHMIA DETECTORS

George B. Moody, Warren K. Muldrow, and Roger G. Mark

Massachusetts Institute of Technology, Cambridge, Mass., and
Beth Israel Hospital, Boston, Mass., USA

Summary

Noise tolerance is an important aspect of arrhythmia detector performance which is difficult to measure using conventional techniques. By adding noise in known amounts to annotated EKG recordings, the noise-handling capabilities of an arrhythmia detector may be stressed, and one can observe how detector performance measurements vary as functions of noise level and noise type. We have developed a database of noise recordings, and describe their use in automated noise stress tests of arrhythmia detectors.

Background

Noise is the principal source of error in well-designed arrhythmia detectors. Developers of arrhythmia detectors need a tool which can allow them to observe the effects of noise on their systems, and which can aid in optimizing algorithm decisions based on signal quality. Evaluators need a tool which can help them to assess detector noise tolerance in a quantitative, reproducible way. Clinicians can make better use of arrhythmia detectors if they are aware of the limits of acceptable signal quality.

A beat-by-beat evaluation of a detector using an annotated digital database is the best available method of measuring performance. Existing databases, because of their relatively small size, do not exhibit the wide variety of noise in the range of contexts in which it may be encountered in clinical practice.

The concept of the noise stress test

The "noise stress test" is a quantitative, reproducible technique for assessing the performance of arrhythmia detectors in the presence of noise and artifact. It is a beat-by-beat evaluation using existing annotated ECG recordings[1,2] and conventional performance measures[3]. Realistic noise of the types encountered in ECG recording is added to clean ECGs, and the noisy ECGs thus produced are analyzed by the detector under test. By varying the type and level of noise, detector performance may be measured as a function of noise characteristics. The noise stress test is designed to be strictly reproducible so that the effects of changes in a detector may be assessed, and so that comparisons between detectors may be made.

The idea of adding noise to clean ECGs to test the noise tolerance of ECG analysis programs was proposed independently by two of the present authors[4] and by the developers of the Hannover diagnostic ECG analysis program (Alraun, Zywietz, et al.[5]). The Hannover group examined the effects of noise on the accuracy of their program's measurements of time-domain ECG features. The noise sources were artificial, and included band-limited white noise, line frequency noise, low frequency sine and sawtooth waveforms, and step function-like artifacts. The ECG signals were constructed by concatenating identical, essentially noise-free, beats at fixed R-R intervals.

The noise database

Our goal was to provide a highly realistic simulation of real-world noisy ECG recordings. We chose to classify noise according to its source rather than by its frequency-domain characteristics, into the following categories:

1. Baseline wander is the familiar low-frequency signal usually caused by motion of the subject or the leads.
2. Electrode motion artifact is often the most difficult to handle since it can closely mimic elements of the ECG signal. It is usually the result of intermittent mechanical forces acting on the electrodes.
3. Muscle noise (EMG) has a spectrum which overlaps that of the ECG, but which extends to higher frequencies.
4. Power line interference occurs at multiples of the mains frequency (50 or 60 Hz).

In the present work, we ignored power line interference since it is easily removed using a digital filter, and presents no problem to the detectors which were available for testing.

Noise recordings may be obtained from digital or analog simulators, from noisy ECG recordings from which the ECG has been removed, or by recording signals from electrodes placed in such a way that the ECG is not visible. Each method has drawbacks: one cannot prove that all significant characteristics of real noise are reproduced in a simulation, since noise-generating mechanisms are not fully understood; it may be difficult to remove all ECG information from a noisy ECG recording; and recordings from specially-placed electrodes record noise from different positions than those which are used in typical ECG recordings. We adopted the last approach, which, because of its simplicity, seemed most likely to work well.

0276-6574/84/0000/0381$01.00 © 1984 IEEE

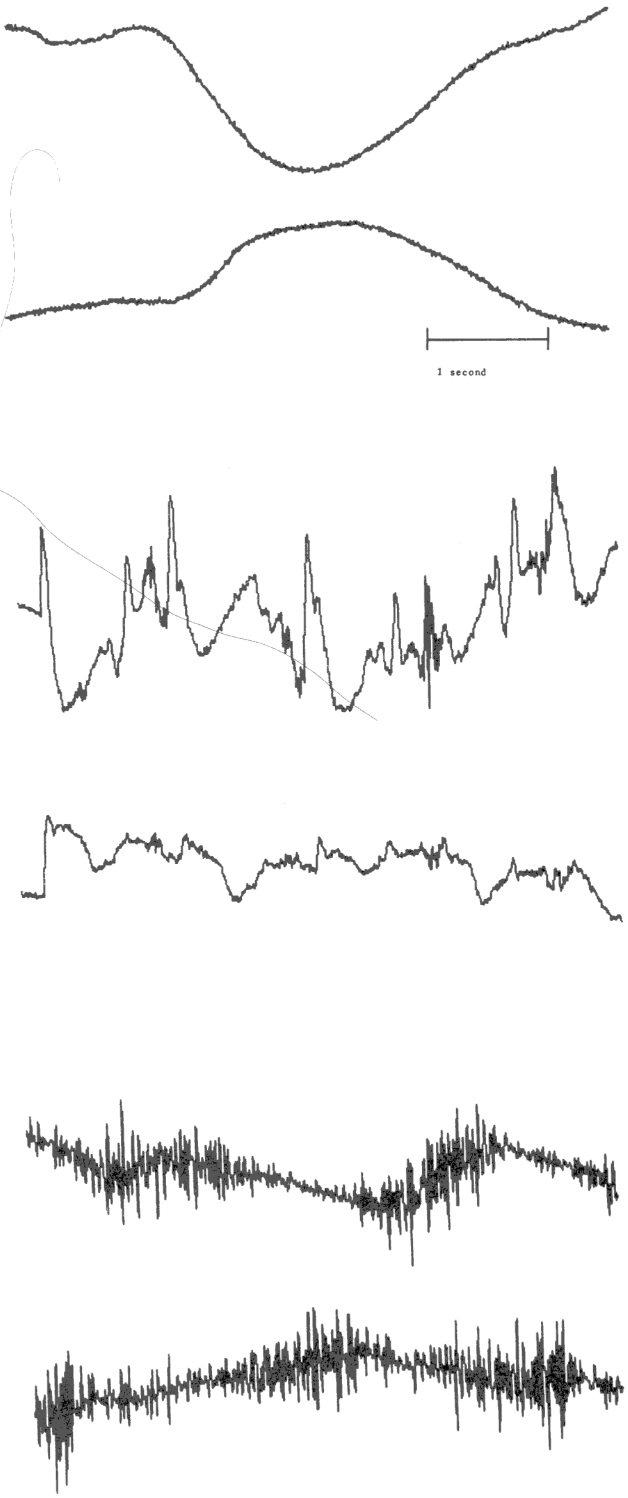


Figure 1. Five-second excerpts from two-channel noise recordings, showing baseline wander (top), electrode motion artifact (center), and muscle noise (bottom).

We recorded approximately 25 hours of noise using Avionics model 445 Holter recorders, with standard cables and electrodes. Volunteers wore electrodes on their thighs and arms with lead axes chosen to eliminate ECG signals. Two channels of noise were recorded simultaneously; though correlation between channels is generally low, it is non-zero, and must be maintained if the test is to be realistic. The analog noise recordings were digitized at 250 Hz per channel using the same equipment and techniques used for digitization of the ECG databases. By visual inspection, the data were separated into the three categories of baseline wander, electrode motion artifact, and muscle noise. Half-hour segments of each category were prepared by concatenating segments of similar noise type, avoiding discontinuities in amplitude.

Figure 1 illustrates each of the three types of noise. All three excerpts include a sizeable low-frequency component which is characteristic of baseline wander. In practice, we found it nearly impossible to record electrode motion artifact or muscle noise without baseline wander.

Test protocol

Figure 2 illustrates the software components of the noise stress test and the data flow between them. The process of combining the ECG and noise is performed under the control of a program which follows a list of instructions describing how the noise level is to be varied during the course of the test. The ECG analysis program reads the noisy ECG and generates a file of beat annotations, which are compared against those supplied with the database.

The results of a noise stress test will be dependent on the characteristics of the ECG as well as the noise. The intrinsic difficulty of tapes varies; PVCs which are very similar to normals may be readily identifiable if the signal is clean, for example, but may not be distinguishable in a noisy

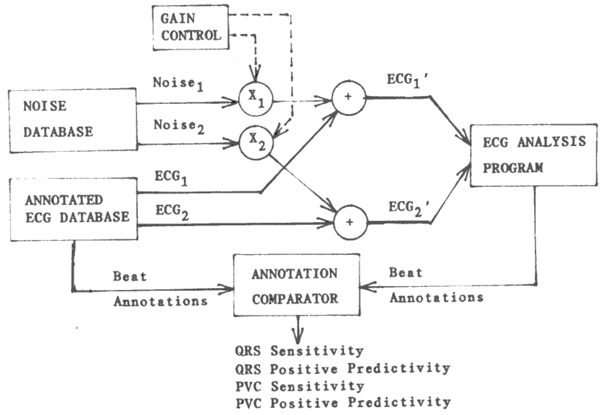


Figure 2. The noise stress test. Gains X_1 and X_2 are chosen to produce equal noise/signal ratios in ECG$_1$' and ECG$_2$' (see text).

context. Artifacts may resemble beats to a greater or lesser degree, since QRS morphologies vary greatly from tape to tape. If noise or sudden changes in signal amplitude are present in the original ECG tape, an accurate calibration of the test will be difficult to achieve.

To reduce variability in results due to the choice of ECG tape, we used essentially noise-free tapes, and normalized the noise levels by the ECG amplitude. Tapes from the AHA database are generally less noisy than those in the MIT/BIH database; we chose one tape from each available class, where possible choosing clean recordings with significant numbers of PVCs.

Normalization of the noise level requires calculation of the signal and noise amplitudes before beginning the test. Assuming that the ECG tape is noise-free, the signal amplitude may be obtained by measuring the mean peak-to-peak amplitude of the normal QRS complexes, excluding outliers. This calculation implicitly defines the QRS complexes as the signal. Since a substantial low-frequency component is present in all of the noise recordings, an overall RMS amplitude measurement would overestimate the amplitude of the noise in the frequency range of interest. For this reason, we measured the RMS amplitude of each 10-second segment of noise (after removing the local DC component) and calculated the noise amplitude as the mean of these measurements. Based on the calculated noise and signal amplitudes, the noise gain may be set to produce any desired noise/signal ratio.

Each test consisted of 10 to 20 runs of a single tape, using noise of a single category, and investigated the effects of varying the noise/signal ratio between 0 and 2. Each run consisted of a 5-minute learning period, a 15-minute stress period, and a 15-minute recovery period. Noise was added during the stress period only, using fixed gain settings calculated to produce equal noise/signal ratios on each channel.

Using an automated beat-by-beat comparison process which we have described elsewhere[3], four standard performance measures (QRS and PVC sensitivity and positive predictivity) were derived for the stress period of each run. Performance curves may be drawn based on the measures derived from a complete set of runs. The plots on the next page illustrate the performance of two independently developed arrhythmia analysis programs. Program A performs well at relatively low noise/signal ratios but is heavily penalized for "shutting down" at high noise levels, usually because of ADC saturation. Program B avoids ADC saturation and shutdown, and performs better at high noise levels.

From these results, it is clear that baseline wander and muscle noise can be tolerated by both detectors at much higher levels than electrode motion artifact. Figure 3 illustrates typical signal quality in the critical range of electrode motion noise/signal ratios between 0.2 and 0.6, where performance begins to deteriorate. At the upper end of this range, visual interpretation becomes quite difficult.

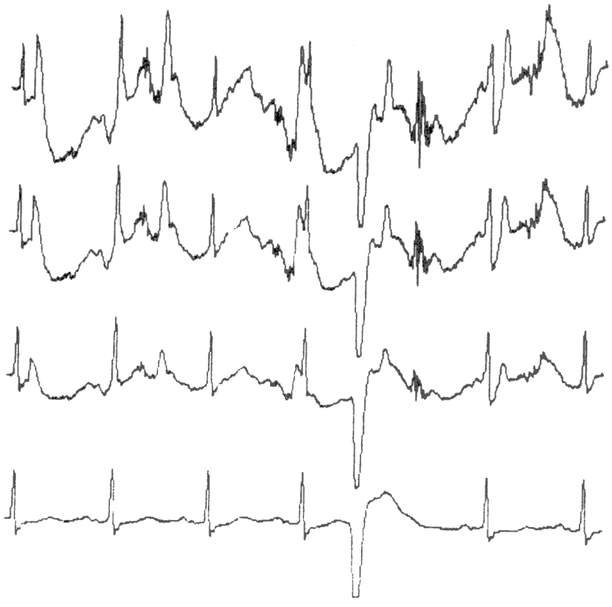
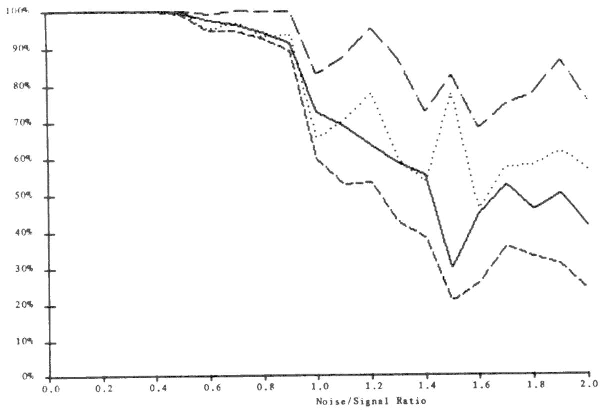


Figure 3. Five seconds of ECG, to which has been added the electrode motion artifact shown in figure 1. From top to bottom, the noise/signal ratios are 0.6, 0.4, 0.2, and 0 (no noise added).

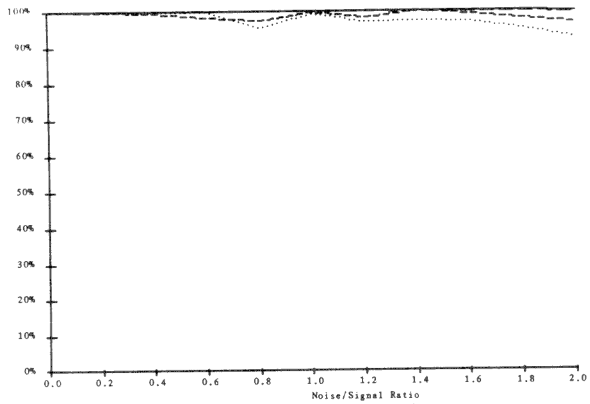
Variations and extensions

Performance curves such as those shown below contain more information than is needed to make a comparison between two detectors, or two versions of the same detector, with respect to noise tolerance. A useful concept is that of a critical performance threshold, which describes the lowest noise/signal ratio at which performance becomes unacceptable. The noise tolerance of a detector may then be characterized in terms of mean critical performance thresholds for each noise type.

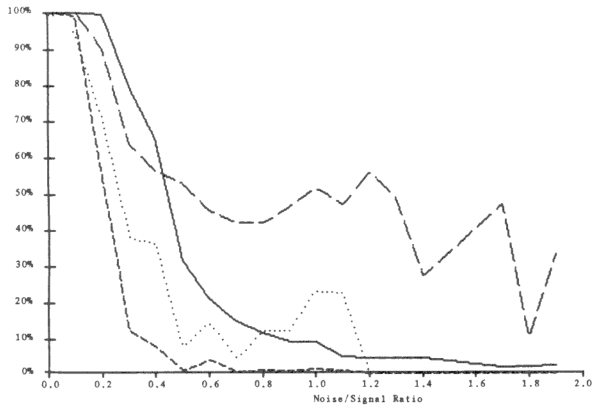
If periods of shutdown are recorded by the detector and excluded from the beat-by-beat comparisons, it becomes possible to evaluate the efficacy of the shutdown algorithm. Ideally one would prefer a detector which cannot be made to perform unacceptably; based on the noise stress test, the developer can optimize decision thresholds for shutdown. If (as we found for the detectors we tested) some aspects of performance deteriorate at lower noise/signal ratios than others, a "structured shutdown" might be appropriate. For example, as noise/signal ratio increases, program B begins to make PVC detection errors, then begins to make false QRS detections, and finally begins to miss beats. This suggests that a suitable shutdown strategy may involve disabling (or raising the thresholds for) ectopic beat alarms first, then high rate alarms, and finally all alarms.



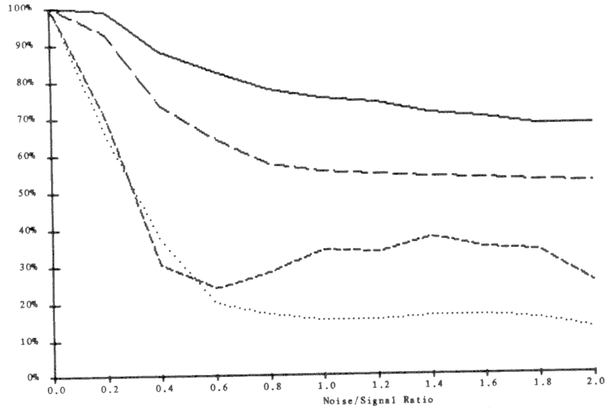
Program A (Tape 6009 with baseline wander)



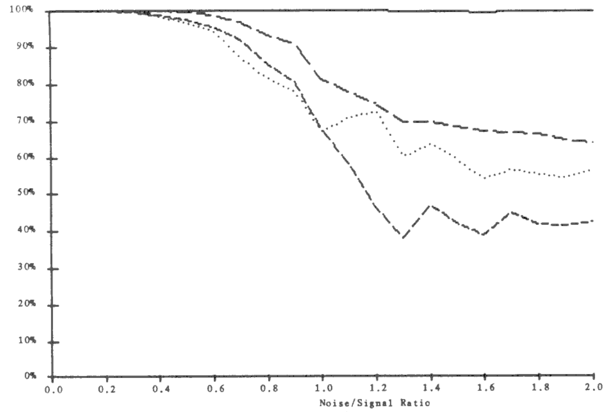
Program B (Tape 6009 with baseline wander)



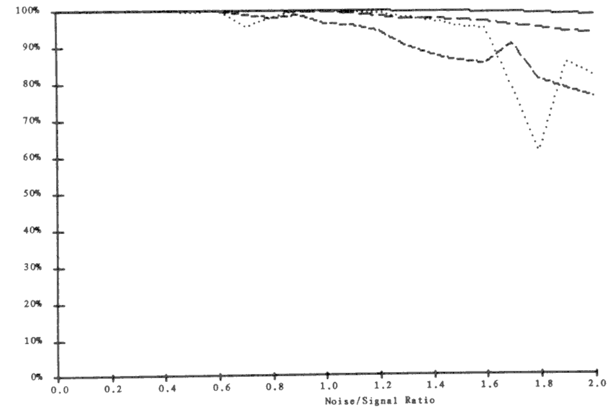
Program A (Tape 6009 with electrode motion artifact)



Program B (Tape 6009 with electrode motion artifact)



Program A (Tape 6009 with muscle noise)



Program B (Tape 6009 with muscle noise)

——— QRS Sensitivity
— — — QRS Positive Predictivity
- - - - - - PVC Sensitivity
.... PVC Positive Predictivity

References

1. Mark, R., Schluter, P., Moody, G., Devlin, P., and Chernoff, D. An annotated database for evaluating arrhythmia detectors. In Frontiers of Engineering in Health Care, pp. 205-210. Proc. 4th Annual Conf. IEEE Engineering in Medicine and Biology Society. 1982.

2. Hermes, R. and Oliver, G. Use of the American Heart Association database. In Ambulatory Electrocardiographic Recording, pp. 165-181. Eds. Wenger, N., Mock, M., and Ringqvist, R. Yearbook Med. Pub. 1980.

3. Schluter, P., Mark, R., Moody, G., Olson, W., Peterson, S. Performance measures for arrhythmia detectors. Computers in Cardiology 7:267-270. Long Beach, California: IEEE Computer Society. 1980.

4. Moody, G. and Mark, R. How can we predict real-world performance of an arrhythmia detector? Computers in Cardiology 10:71-76. 1983.

5. Alraun, W., Zywietz, C., Borovsky, D., Willems, J. Methods for noise testing of ECG analysis programs. Computers in Cardiology 10:253-256. 1983.