A NEW METHOD FOR DETECTING ATRIAL FIBRILLATION USING R-R INTERVALS

George B. Moody and Roger G. Mark

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

Summary

Markov process models of atrial fibrillation (AF) have been used as a basis for AF detection with limited success. The Markov modelling technique is equivalent to calculating the arithmetic mean of a set of scores which reflect the relative likelihood of observing pairs of R-R intervals in AF vs. making the same observations outside of AF. Beginning with such a set of scores, we show that other techniques of averaging result in improved detector performance. Further, we show that models which deal with a continuum of R-R intervals can be used to obtain performance superior to what appears possible using finite-state-machine (FSM) models. We compare the performance of several such detectors on an AF database composed of 260 hours of two-lead Holter recordings selected for the presence of intermittent atrial fibrillation.

The problem of AF detection

The detection of atrial fibrillation by an ECG analysis program is motivated not only by the clinical significance of the rhythm per se, but also by the desire to improve the accuracy of ventricular arrhythmia detection, since many of the commonly used heuristics for PVC detection fail in the presence of AF.

Atrial fibrillation and flutter are characterized by chaotic electrical activity in the atria, which may contract at up to 600 bpm in humans. Frequently, the atrial activity is apparent in the surface ECG as atrial flutter wavelets. At other times, it may be impossible to observe any direct evidence of atrial activity. In such cases, it may be possible to diagnose AF on the basis of a sudden disappearance of regularly-occurring P-waves.

The atrial impulses are conducted sporadically through the AV node, resulting in a highly irregular ventricular response in atrial fibrillation, which may become regular during atrial flutter. This irregular ventricular response produces an R-R interval sequence which bears a signature of AF easily recognizable to the trained observer. Because of the difficulty of detecting atrial activity in the surface ECG, the most robust techniques for automated AF detection depend on making inferences from the R-R interval sequence. In figure 1, R-R intervals for 15 minutes of ECG, including a seven-minute AF episode, are shown. During the AF episode, a dramatically different pattern of R-R intervals is clearly visible.

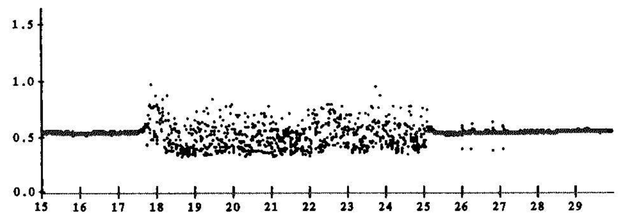

Figure 1. The pattern of R-R intervals during an AF episode. The episode begins at 17:45 and ends at 25:05. Between minutes 26 and 27, several SVPBs are visible.

Markov process models of the R-R interval sequence

An attractive approach for AF detection is due to Gersch[1], who proposed modelling the R-R interval sequence as a three-state Markov process. Each interval is characterized as representative of one of the three states {S,R,L} by classifying it as short, regular, or long. In the present work, intervals were called short if they did not exceed 85% of the mean interval, long if they exceeded 115% of the mean, and regular otherwise. The mean interval is determined recursively by the relation

$$rrmean(i) = 0.75 \cdot rrmean(i-1) + 0.25 \cdot rr(i)$$

for all observed R-R intervals rr(i) which do not exceed 1.5 seconds. We have shown previously[2] that R-R interval predictors of this form are adequate for judging prematurity.

Algra[3] used morphologic information in a 6-state Markov process, in which each of the three states based on R-R intervals is divided into two states based on a "width" measure which tends to separate PVCs from other beats. In the present work, we simply ignored intervals bounded by PVCs; this technique loses some information (for example, regular intervals which follow PVCs are rare in NSR, and relatively common in AF) but is quite adequate to permit good performance.

Using a suitable database, one can compile statistics of transitions between states, and derive transition probability matrices for various rhythms. We selected 12 half-hour records (the "learning set") from the MIT/BIH Arrhythmia Database[4] for this purpose, of which 2 contain only AF, 5 contain AF and normal sinus rhythm, and 5 contain a variety of other rhythms considered likely to confuse an AF detector. Results are summarized in table 1.

		(AF) from				(other rhythms) from		
		S	R	L		S	R	L
to	S	351	734	303	S	141	301	246
	R	723	4828	1351	R	142	12668	575
	L	330	992	431	L	404	236	375

$$S_{ij} = \begin{matrix} -0.075 & -1.460 & 0.346 \\ -0.806 & 0.256 & -0.304 \\ 0.828 & -1.926 & 0.426 \end{matrix}$$

Table 1. RR interval transition statistics for MIT/BIH database tapes 201, 202, 203, 207, 209, 210, 213, 219, 220, 221, 222, and 223 (the "learning set" used in the present work). S_{ij} is the S (score) matrix described in the text.

Assume that the R-R interval sequence

$$T = \{ t_1, t_2, \dots, t_n \} \qquad (t_i \text{ in } \{S,R,L\})$$

is controlled by a stationary first-order Markov process characterized by the transition probability matrix

$$p_{i,j,R} = P(t_i \mid t_j, R) \qquad (R \text{ in } \{AF, other\})$$

The conditional probability for observation of T given process R is (by the Markov hypothesis)

$$P(T \mid R) = \prod_{i=1}^{n-1} p_{i+1,i,R}$$

By choosing R among the available rhythm models such that $P(T|R)$ is maximum, we have the maximum likelihood procedure for choosing the most probable rhythm.

A useful computational efficiency may be achieved by storing the logarithms of the transition matrices and adding the appropriate elements rather than multiplying the elements of the original matrices. Furthermore, if the goal is to make a binary decision (e.g., is AF present or not?), we may calculate a single transition matrix for all non-AF rhythms, and divide its elements by those of the AF transition matrix. We can now define the matrix

$$S_{ij} = k \cdot \log (p_{ij,\ other} / p_{ij,\ AF})$$

the negative elements of which have the property that the transitions they represent are relatively more likely to occur in AF than otherwise.

The Markov process model suggests that a sequence of n intervals can be classified by simply adding n-1 appropriately chosen elements of S, and declaring AF if the sum is negative. We shall refer to the sum and analogous quantities as AF predictors. If desired, a bias

$$S_0 = k \cdot \log [P(other) / P(AF)]$$

based on the a priori probabilities of AF and other rhythms may be added to the predictor. It may also be desirable to incorporate hysteresis into the decision process, since the non-stationary statistics of AF can cause substantial short-term fluctuations in the predictor.

The remaining degree of freedom is the choice of n, which is a compromise between the expected increase in accuracy from basing the decision on a larger context, and the requirement for reasonably prompt response to rhythm changes, which may vary depending on the application. For real-time monitors, a choice of n = 20 works reasonably well.

Extensions to the Markov process model

The Markov process model approach outlined above is equivalent to calculating the arithmetic mean of appropriately chosen elements of the matrix S. The elements of S may be considered as scores which are most negative for those transitions which are most characteristic of AF. The inclusion of a priori probabilities of AF and other rhythms is equivalent to resetting the reference level of the AF predictor. The need for hysteresis in the decision process may be described in terms of the "noise" in the signal represented by the time sequence of scores as a function of time.

Collectively, these observations led us to the conclusion that the AF detection problem can be reduced to the problem of processing the sequence of elements of S, now treated as a function of time, using standard signal-processing techniques.

Two classic signal processing problems are present here. The first is quantization error, which in this context corresponds to erroneous assignment of an interval to a state. A useful technique for reducing the effects of quantization error is interpolation; in this case, we may consider the S matrix as a set of nine samples of a continuous function S' of the normalized current and previous R-R intervals (see figure 2).

The second problem is noise in the signal, a problem addressed by filtering, of which taking the arithmetic mean (as we have seen is implied by the basic Markov process model) is a special case. Although motivated originally by an interest in improving efficiency, we found that simple first-order filters of the form

$$V(t) = k \cdot v_{obs} + (1 - k) \cdot V(t-1)$$

are markedly superior to arithmetic averaging for noise removal in this context. The constant k should be rather small; we found that k = 1/64 works well.

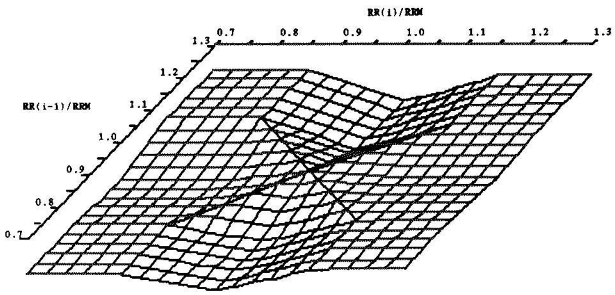

Figure 2. The function S', constructed by interpolating between the elements of the matrix S. Negative values (low points) indicate regions in the pattern space more likely to be observed in AF than elsewhere.

On the same data set which was used to develop the S matrix, we tested the basic Markov process model against variations which employed a first-order filter, interpolation, or both. Receiver operating curves were generated for the four detectors (see figure 3). It is apparent that first-order filters have the effect of attenuating the tail of the "other rhythm" distribution, while interpolation has the same effect on the tail of the AF distribution. These results suggest that noise in the predictor signal due to short-term irregularities outside of AF is largely removed by filtering, while in AF quantization error is a more significant problem, the effects of which are mitigated by interpolation.

R-R predictor arrays for AF detection

Observing that R-R intervals in AF, unlike those in most other rhythms, are unpredictable, an alternative to Markov process modelling is to consider the mean error of a well-designed R-R interval predictor as an AF predictor. An elegant method which treats each of the N previous intervals as predictors of the current interval is due to Schluter[5]. The mean errors of each predictor are calculated using a first-order filter. The predictor with the lowest mean error is designated the "best predicting interval". If the mean error of the best predicting interval exceeds a threshold, AF is declared.

The AF database

Although the MIT/BIH Arrhythmia Database makes evaluation of an AF detector straightforward, such evaluations may be reasonably criticized on the basis of the relatively small sample size, especially since the same database was used for development of the detector. For this reason we sought to develop a much larger database, annotated only with respect to AF onsets and terminations, which could form a testbed for AF detector evaluation.

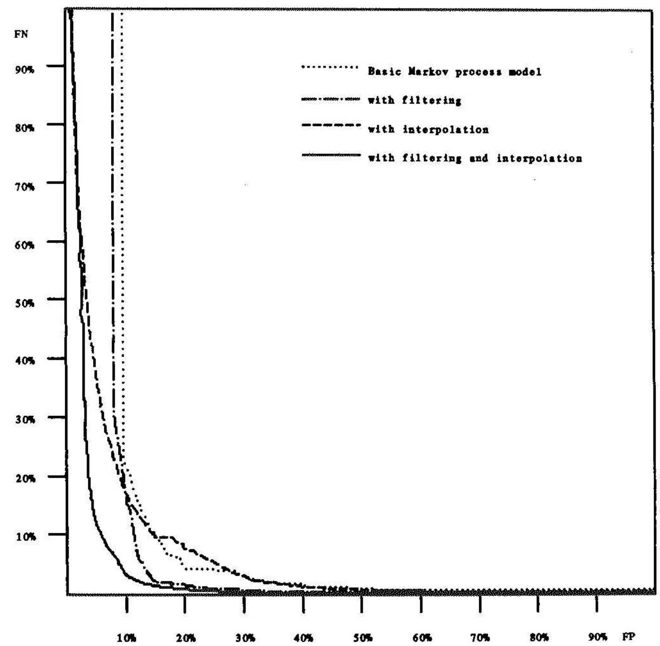

Figure 3. Receiver operating curves (ROCs) for the basic Markov process model and three variations. Based on these curves, decision boundaries were chosen to produce the lowest total error rate. Below are shown AF sensitivity (Se) and positive predictivity (+P) based on these choices for the four methods shown above and for a fifth method described in the text.

Detector	AF Se	AF +P
Basic Markov model	90.03%	80.14%
with first-order filter	96.15%	82.34%
with interpolation	88.22%	81.31%
with filtering and interpolation	96.09%	86.79%
R-R predictor array	95.44%	80.52%

From the library of over 8000 24-hour Holter recordings collected by the Arrhythmia Laboratory of Beth Israel Hospital, we digitized 26 ten-hour excerpts of tapes on which episodes of paroxsysmal atrial fibrillation had been diagnosed. Each digitized excerpt was interpreted by an experimental arrhythmia analysis program[6] which recorded beat annotations and R-R intervals in disk files. The R-R intervals were then plotted as in figure 1 and examined visually for evidence of rhythm change. For each possible rhythm change, a short excerpt of the two-channel digitized ECG, with time reference marks and the program's beat annotations, was plotted on a chart recording. Rhythm changes were manually noted on the paper strips to an accuracy of one R-R interval, and recorded in disk files. The contents of the AF database are summarized in table 2.

Rhythm	Hours	Episodes	Beats
Atrial fibrillation	94.99	319	532,276
Atrial flutter	1.98	40	16,563
Other rhythms	163.03	293	733,550
Total	260.00	652	1,282,369

Table 2. AF database profile.

Evaluation of detectors on the AF database

Three of the detectors shown in figure 3 were evaluated on the AF database using thresholds chosen to minimize total errors on the learning set. Results are summarized in table 3. Hysteresis based on additional experience with the learning set was incorporated into all detectors. PVC-bounded intervals (as determined from the annotations produced by the ECG analysis program) were excluded from the inputs of detectors 1, 2a, and 3.

It is clear that the basic Markov process model, with thresholds set for operation near the "knee" of its ROC on the learning set, is operating at some distance from that point on the AF database. Although its sensitivity is excellent, there is an unacceptably high false positive rate. These observations suggest that this detector is tuned to its learning set. The results for detectors 2a and 2b demonstrate that the addition of filtering and interpolation to the basic Markov model results in a robust detector. Detector 2b is identical to 2a, but PVC-bounded intervals are not removed from the input. There is some degradation as a consequence, but performance remains adequate. Results for the R-R predictor array approach show significant degradation with respect to sensitivity on the learning set, suggesting that this detector may lack generality.

Detector	AF Se	AF +P
1. Basic Markov model	99.59%	65.97%
2a. (1) + filtering + interpolation	93.58%	85.92%
2b. (2a) + PVC-bounded intervals	90.65%	82.38%
3. RR predictor array	75.79%	91.93%

Table 3. AF database evaluation results (260 hours)

Conclusions

We have shown that the Markov process model for AF detection is equivalent to determining the arithmetic mean of a series of scores based on the R-R interval sequence. Application of a first-order filter to the same series of scores has the effect of reducing false positive detections of AF, while the use of interpolation between elements of the discrete score matrix has the effect of reducing false negatives. Performance of this strategy improves if PVC-bounded intervals are excluded from consideration, but remains acceptable if they are not. Finally, in contrast to the other techniques examined, the performance of the new method does not change significantly when tested on a large database, suggesting it is statistically well-behaved.

Acknowledgements

The authors are grateful to P. Devlin, who contributed programs for R-R interval plotting and for digitization of Holter tapes; to E. Moody, who annotated a significant portion of the AF database, and provided many helpful suggestions; and to P. Albrecht, who wrote the rhythm comparison program which was used to derive AF database performance measurements.

References

1. Gersch, W., Eddy, P., and Dong, E., Jr. Cardiac arrhythmia classification: a heart-beat interval - Markov chain approach. Comp. Biomed. Res. 4:385-392. 1970.

2. Chernoff, D., Lee, T., Moody, G., and Mark, R. Evaluation of R-R interval predictors using an annotated ECG database. Computers in Cardiology 8:359-362. Long Beach, California: IEEE Computer Society. 1981.

3. Algra, A., Vinke, R., and Zeelenberg, C. Prescanning 24 hour ECG recordings selecting normal sinus rhythms using the theory of Markov chains. Computers in Cardiology 9:49-54. 1982.

4. Mark, R., Schluter, P., Moody, G., Devlin, P., and Chernoff, D. An annotated ECG 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.

5. Schluter, P. The design and evaluation of a bedside cardiac arrhythmia monitor. Ph.D. thesis, Massachusetts Institute of Technology, Dept. of Electrical Engineering, Cambridge, Mass. 1981.

6. Moody, G., and Mark, R. Development and evaluation of a two-channel ECG analysis program. Computers in Cardiology 9:39-44. 1982.