High Accuracy of Automatic Detection of Atrial Fibrillation Using Wavelet Transform of Heart Rate Intervals

DAVID DUVERNEY, JEAN-MICHEL GASPOZ, VINCENT PICHOT, FRÉDÉRIC ROCHE, RICHARD BRION, ANESTIS ANTONIADIS, and JEAN-CLAUDE BARTHÉLÉMY

From the *Service d’Exploration Fonctionnelle Cardio-Réspiratoire, Laboratoire de Physiologie, Hôpital Universitaire Nord, Saint-Etienne, France, †Clinique de Médecine II et Division de Cardiologie, Département de Médecine Interne, Hôpital Universitaire, Genève, Suisse, ‡Service de Cardiologie, Hôpital D’Instruction des Armées, Desgenettes, 69275 Lyon, and the §Laboratoire de Modélisation et de calcul, LMC-IMAG, Université Joseph Fourier, Tour IRMA, Grenoble, France

DUVERNEY, D., ET AL.: High Accuracy of Automatic Detection of Atrial Fibrillation Using Wavelet Transform of Heart Rate Intervals. Permanent and paroxysmal AF is a risk factor for the occurrence and the recurrence of stroke, which can occur as its first manifestation. However, its automatic identification is still unsatisfactory. In this study, a new mathematical approach was evaluated to automate AF identification. A derivation set of 30 24-hour Holter recordings, 15 with chronic AF (CAF) and 15 with sinus rhythm (SR), allowed the authors to establish specific RR variability characteristics using wavelet and fractal analysis. Then, a validation set of 50 subjects was studied using these criteria. 18 with CAF, 16 with SR, and 15 with paroxysmal AF (PAF); and each QRS was classified as true or false sinus or AF beat. In the SR group, specificity reached 99.9%; in the CAF group, sensitivity reached 99.2%; in the PAF group, sensitivity reached 96.1%, and specificity 92.9%. However, classification on a patient basis provided a sensitivity of 100%. This new approach showed a high sensitivity and a high specificity for automatic AF detection, and could be used in screening for AF in large populations at risk. (PACE 2002; 25[Pt. 1]:457–462)

Introduction

In the United States and Europe, one of four cerebrovascular strokes (i.e., 75,000 strokes per year) is associated with atrial fibrillation (AF).1 The incidence of AF increases, between the 65–74 and the 75–84 age groups, from 1.76 to 4.27 for 100-person-years, respectively, in men, and from 1.01 to 2.16, respectively, in women.2 A prevalence of AF as high as 83% was described in a population aged ≥85 years.3 The presence of AF is a strong predictor of the occurrence and recurrence of stroke and of increased mortality after a stroke on the short- and long-term,4,5 and the risk associated with it is increased with age.6–8 According to the cohorts studied, the absolute risk of stroke attributed to AF varied from 3% to 67%,9,10 while the relative risk of stroke ranged from 1.0 to 6.9.11 Interestingly, the absolute risk of a silent clinical stroke was in the same range.12–16 Importantly, the type of AF, permanent or paroxysmal (PAF), did not statistically modify the risk of stroke occurrence.17

All these figures make the tracking of AF a major health priority.20 Several approaches have been considered: intracardiac recordings at the atrial level21–24 and at the conduction system level,25 averaged atrial activity of esophageal signals,26,27 and ambulatory electrocardiograph (ECG). Such recordings have been shown to be the best suited to detect rhythm abnormalities because of their easiness of use and of their unique ability to detect frequent paroxysmal forms.

Several attempts have been made to create efficient methods for the recognition of AF in Holter recordings. The simple Lorentz plot allowed recognition of scattering specific to AF.28 Other methods were intended to improve visual ECG reading.29 Neural networks have been applied, analyzing RR intervals30,31 and baseline fluctua-
tions. However, in Holter systems or event recorders, the identification of AF still has to be improved, published sensitivities ranging from 82.4% to 96.6% and specificities from 92% to 92.3%. Due to its ability to precisely characterize at any time RR variability, the new methods of time frequency analysis of RR intervals associated to a fractal classification could prove to be a powerful tool in AF detection. Thus, the accuracy of AF recognition on standard Holter recordings was investigated using that combination of mathematical heart rate variability (HRV) analysis.

Methods

Population

Two sets of subjects were included in the study and were randomized into two groups. First a derivation set of 30 subjects, 15 with a sinus rhythm (SR), and 15 with chronic AF (CAF) established a threshold criteria.

A validation set of 50 subjects was then analyzed using the previously established thresholds: 19 suffering from CAF, 15 from PAF, and 16 control patients with normal SR.

All subjects were recorded during 24 consecutive hours. The 24-hour Holter recordings were directly digitized at a rate of 100 Hz on a recorder (DuoHolter, Novacor, Paris, France).

Recordings

The 24-hour Holter recordings, recorded using three standard leads, were analyzed and AF episodes were first assessed by two independent cardiologists using full interactive review and, if necessary, edited. Thus, for each recording, a list of consecutive properly labeled RR intervals was obtained for further mathematical processing.

Mathematical Methods

RR Intervals were obtained from the Holter system, each QRS being given a number to synchronize the clinical and the mathematical identification of AF episodes.

Globally, the identification of AF required the use of a cascade of two sequential complementary analyses of RR intervals, the first one (discrete wavelet transform [DWT]) identifying periods of high HRV coefficient, the second one (fractal analysis) classifying these high variability periods into physiological (SR) or pathological (AF) rhythms.

Thus, the HRV level was first established all along the recording, using DWT of heart rate intervals, to ensure a precise time localization of the beginning and of the end of each high variability event (Fig. 1). DWT, a discretion of the wavelet transform, analyzes a signal by a family of functions ψjk, created by successive dilatation and translation of a mother wavelet ψ, where

\[ \psi_{jk}(x) = 2^{j/2} \psi(2^j x - k) \]

with j,k ψ Z. The transform is written:

\[ W_f(j,k) = <f, \psi_{jk}> \]

where \(<, >\) denotes scalar products in L^2(R). For a signal composed of 2^j values, this transform leads to series of 2^j-1 coefficients written c_j(k), giving the contribution to signal projection at position 2^j k with scale 2^j. Squaring the obtained coefficients provides the power of the transformed signal for each time and for each scale index. Different time areas are each characterized by an homogeneous power level computed from the highest scales of their DWT (low j). In this study, a quadratic spline of order 3 wavelet transform was used.

As SR can present with HRV indexes as high as those found in AF, a second step was needed. That second step analyzed another HRV property of AF using a fractal analysis on high variability periods identified during the first step, which classified these periods as SR or AF rhythm. When calculating power spectral density (PSD) using Fourier transform, a normal SR has a general trend.

Figure 1. Heart rate variability of a paroxysmal episode of atrial fibrillation analyzed using wavelet transform. The importance of the variability, at each level of frequency, is represented by the amplitude of the coefficients (i.e., the amplitude of the vertical lines). One can easily recognize the onset and the disappearance of the high variability period in the central part of the figure with, at the higher scales, a precision of two RR intervals for the beginning and the end of the episode. However, it has to be noted that high variability coefficients are not related to atrial fibrillation only, as sinus rhythm can also reach such variability. Thus this procedure represents only the first step.
in 1/fβ, thus showing a single linear downsloping pattern when plotted in a log-log plot (Fig. 2, left). Conversely, the PSD of AF presents with two different slopes: one in a high frequency band (10^-1-10^-2 Hz) and the other in a low frequency band (10^-3-10^-4 Hz); the slope in the high frequency band is higher (β ≈ 0) than those obtained for SR (β ≈ 1) (Fig. 2, right). To model the 1/fβ power law spectrum, the concept of fractional Brownian motion (FBM) was applied. This non-stationary autosimilar process has been widely used as a model for biological processes, and particularly HRV. The autosimilarity property is characterized by the Hurst exponent H, which is related to β by the equation β = 2H + 1. The increment process of FBM can also be used to estimate the Hurst exponent of the FBM process by means of DWT with

\[ \text{var}(d_j(n)) = 2(2H-1)j^{-\beta} \sigma^2 (2 - 2^{2H-1}) \]

where \( d_j \) are details coefficients at scale j. It leads to the following equation:

\[ \log_{10}(\text{var}(d_j(n))) = (2H-1)j + f(H,\alpha) \]

Each of the previously identified high HRV areas was individually reprocessed accordingly, to calculate the associated Hurst exponent corresponding to the FBM, using the HRV of that particular area as the increment process. As the difference between SR and AF lay on the highest frequencies, the Hurst exponent was computed only with the five highest scales.

Two thresholds were associated with these two steps. The first threshold was, thus, a level of variability, set to 6.10^-3 second, while the second threshold, set to 0.7, was a value of the Hurst exponent.

Each threshold was computed using the derivation set to obtain the combination giving the best separation between SR and AF, and was then applied to the validation set.

Software Tools

The signal processing was performed using MATLAB (The Mathworks Inc., Natick, MA, USA) and the statistical analysis with Statview 4.5 (Abacus Concepts, Berkeley, CA, USA) on a Power Macintosh. Results are presented as mean ± SD (m ± SD).

Results

Derivation Set

There was a total of 1,334,296 beats in the SR group (heart rate 64.5 ± 8.2), and the mean recording duration was 23.0 ± 3.6 hours. There was a total of 1,131,696 beats in the CAF group (heart rate 62.9 ± 9.4), and the mean recording duration was 20.0 ± 3.8 hours.

Specificity of AF detection based on QRS

![Figure 2. Power spectral density of heart rate variability calculated using Fourier transform, and plotted on a log-log plot. Sinus rhythm (SR) (left) bears a single linear downsloping pattern, while atrial fibrillation (AF) (right) presents with two different slopes. The Hurst coefficient is also the slope of the high frequency band. This second step, applied to HRV periods identified during the first step, allowed to separate appropriately SR and AF rhythms.](image-url)
complexes reached 99.9%, since there were 223 false AF QRS in the SR group (Table I).

Sensitivity of AF detection based on QRS complexes reached 99.7%, since there were 15 short episodes of 227 ± 208 beats interpreted as SR rhythm beats (Table I).

Validation Set

SR Group

There was a total of 1,083,537 SR beats in the SR (heart rate 73.3 ± 10.2) and the mean recording duration was 15.4 ± 1.1 hours.

Standard interpretation of 24-hour ECG-recordings of the 16 SR subjects did not reveal any cardiac arrhythmia.

Two AF episodes were falsely detected in the same subject (duration of 154 and 293 beats, respectively; heart rate of 81.5 ± 14.8 and 76.0 ± 8.5 beats/min, respectively). Thus, specificity of AF detection based on QRS complexes reached 99.9% (1,083,090 true-negative QRS vs 447 false-positive ones [Table I]).

CAF Group

There was a total of 2,064,197 AF beats in the CAF group (heart rate 96.3 ± 27.6) and the mean recording duration was 18.8 ± 4.6 hours. The number of QRS complexes was 108,642 ± 26,510 beats per patient.

A total of 40 episodes were falsely identified as SR in seven patients (heart rate of 82.5 ± 9.5, duration of 399 ± 372 beats). The longest undetected period reached 2,075 beats and the shortest 120 beats. Sensitivity for AF detection reached 99.2% based on QRS complexes (2,048,179 true-positive QRS vs 16,018 false-negative ones [Table I]).

One patient accounted for 24 of these 40 episodes, while the other ones occurred in six other patients. These false-negative episodes were due to atrial flutter and to atrial tachycardia in, re-
spectively, 34 and 1 of the 40 cases. In five of the seven patients, recognition of AF was delayed at the beginning of the recording; however, the delay in AF recognition was limited to 178.4 ± 70 beats.

PAF Group

In the PAF group (recording duration 19.9 ± 3.0 hours), there was 857,157 SR and 687,086 AF beats; AF beats were distributed in 36 episodes (Fig. 3) with a mean QRS number of 19,085 ± 24,238 (range 52–94,990 beats) and a mean duration of 2.5 ± 3.4 hours (range 34 s to 15 hours 36 min). During these episodes, the heart rate was 127.2 ± 26.5 beats/min (RR range 3.65–0.37 s).

Based on QRS identification, sensitivity reached 96.1%, as 26,804 AF QRS over 687,086 were misclassified as SR; specificity reached 92.6%, as 63,565 SR QRS over 857,157 were misclassified as AF beats (Table I).

As all the patients of the group with PAF were recognized as having AF, sensitivity, based on patients, was 100%.

In that group, there was false and excessive detection of AF episodes: 10 short of the 36 episodes remained undetected (length 260 beats, HR 123.2 ± 28.1 beats/min); there were 37 false detections as 1,718 ± 2,423 beats (range 147–11, 491).

For detected episodes, there were some delays or advances in the recognition of episodes at their beginning or at their end, and dropouts during the episodes. At the beginning of the episodes, detection was delayed in eight episodes by 91 ± 95 beats (range 2–255); detection was premature in 23 episodes by 166 ± 621 beats (range 2–2,994). At the end of the episodes, two were prematurely interrupted (333 and 3,486 beats); 21 were abnor-

Table I.

<table>
<thead>
<tr>
<th>Accuracy of QRS Classification Accuracy</th>
<th>AF Beats</th>
<th>SR Beats</th>
<th>True SR Beats</th>
<th>False AF Beats</th>
<th>True AF Beats</th>
<th>False SR Beats</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Derivation Set</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR group</td>
<td>—</td>
<td>1,334,296</td>
<td>1,334,073</td>
<td>223</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>99.9</td>
</tr>
<tr>
<td>CAF group</td>
<td>1,131,696</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1,128,291</td>
<td>3,405</td>
<td>99.7</td>
<td>—</td>
</tr>
<tr>
<td>Validation Set</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR group</td>
<td>—</td>
<td>1,083,537</td>
<td>1,083,090</td>
<td>447</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>99.9</td>
</tr>
<tr>
<td>CAF group</td>
<td>2,064,197</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>2,048,179</td>
<td>16,018</td>
<td>99.2</td>
<td>—</td>
</tr>
<tr>
<td>PAF group</td>
<td>687,086</td>
<td>857,157</td>
<td>793,592</td>
<td>63,565</td>
<td>660,284</td>
<td>26,804</td>
<td>96.1</td>
<td>92.6</td>
</tr>
</tbody>
</table>

Sensitivity and specificity of the validation group were higher for permanent sinus rhythm (SR) or chronic atrial fibrillation (CAF) than for paroxysmal atrial fibrillation (PAF) episodes.

AF = atrial fibrillation.

April 2002, Part I  PACE, Vol. 25, No. 4
AUTOMATIC AF DETECTION ACCURACY

Figure 3. Duration distribution of the 36 episodes of atrial fibrillation in the paroxysmal atrial fibrillation group. The mean ± SD duration was 2.5 ± 34 hours (range 34–57,298 s). The fourth episode (*, duration 126 s) remained undetected.

nally prolonged with a delay of 60 ± 103 beats (range 10–429).

False interruptions were observed in 15 cases during PAF episodes, of which 13 occurred in the same patient, with a duration of 1,484 ± 151 beats; the length of the last three, in two different patients, were 781 and 2,001 beats, respectively.

These false-negative episodes were due to atrial flutter or atrial tachycardia in, respectively, 5 and 8 of the 14 cases.

Discussion

When applied to a validation group, the authors’ algorithm associating wavelet transform and fractal analysis provided a sensitivity of 99.2% (CAF group) for the detection of AF QRS complexes, and a specificity of 99.9% (SR group). Its accuracy was slightly less in the PAF group, with a sensitivity of 96.1% and a specificity of 92.6%. However, when based on patients, sensitivity was 100% in the CAF and in the PAF group.

Intraarrhythmic recordings have often been used to detect AF, and analyses of the atrial electric activity at the esophageal level99 with sensitivity and specificity ranges of 100–95.5% and 95–100%, respectively. These recordings were not used, as the goal of the study was AF recognition based on noninvasive standard Holter recordings.

For the purpose of automatic detection, Holter recordings have been analyzed by neural networks,10 which reached a sensitivity of 92% and a specificity of 92.3% in QRS recognition. Using a neural network fed by RR interval information and morphological analysis of the baseline tracing, Cubanski et al.,11 working on fixed groups of ten consecutive QRS, found a sensitivity of 82.4% and a specificity of 96.6%. Thus, the sensitivity and specificity of AF detection by the methods used in the present study favorably compare with prior studies.

Study Limitations

There are some limitations. First, the evaluation of the two parameters, the variability index, and the Hurst exponent needs at least 64 consecutive beats to be established, which limits the power of detection of short episodes. Second, the use of HRV analysis alone is sensitive to arrhythmias that can alter HRV, notably supraventricular extrasystoles, supraventricular tachycardia, and other kinds of arrhythmias. Thus, false-positive detection can be increased due to the presence of numerous isolated or grouped supraventricular arrhythmias, inducing important nonphysiological HRV. As a matter of fact, false-positive detection occurred in 0.04% of the QRS of the SR group, and in 0.0% of the QRS of the PAF subjects. The false-positive QRS could be correctly reclassified by analyzing ECG tracings simultaneously recorded.

On the contrary, false-negative detection can occur in atrial arrhythmias, like paroxysmal supraventricular tachycardia or atrial flutter, due to the relative regularity of such rhythms. In clinical settings, a false-negative detection is more problematic than a false-positive one, since automatic detection does not keep the tracing for the user, which precludes any visual correction. In this study, false-negative episode detection was frequent with a unique PAF episode not exceeding 260 QRS. Improvement of the algorithm could further decrease its false-negative detection rate, however, at the cost of a decrease in specificity.

Conclusion

As more and more strokes are being recognized as the first manifestation of AF, the authors’ algorithm appears well suited to automatically screen for AF in large populations. This is of particular importance in the elderly, in whom AF, paroxysmal and chronic, is frequent and represents a major risk of stroke. In that view, implemented on already available long-term event recorders bearing automatic ECG analysis, the algorithm could help to face an important health priority by improving screening for AF and providing preventive measures before dramatic complications.

Acknowledgements: The authors thank Marjoly Victoire and Gisèle Guillot for their participation in the project.
References


