

Chapter 4

Detection of Atrial Fibrillation



Leif Sörnmo, Andrius Petrėnas and Vaidotas Marozas

4.1 Introduction

The detection of episodes of atrial fibrillation (AF) has been dealt with for more than three decades in research, and yet the challenge remains to develop a detector fully capable of handling all the problems associated with the analysis of continuous long-term ECG recordings as well as of recordings acquired by handheld devices for AF screening. Unacceptably high false alarm rates have been reported, mostly due to the presence of ectopic beats and noisy signal segments, but also due to non-AF arrhythmias manifested by rhythms patterns resembling those of AF, see, e.g., [1]. For the human reader, the following three properties are essential when detecting AF episodes:

1. the presence of a highly irregular rhythm,
2. the absence of P waves, and
3. the presence of f waves.

These properties are, to various extents, explored when developing algorithms for AF detection.

L. Sörnmo (✉)

Department of Biomedical Engineering and Center for Integrative Electrocardiology,
Lund University, Lund, Sweden
e-mail: leif.sornmo@bme.lth.se

A. Petrėnas · V. Marozas

Biomedical Engineering Institute, Kaunas University of Technology, Kaunas, Lithuania
e-mail: andrius.petrenas@ktu.lt

V. Marozas

e-mail: vaidotas.marozas@ktu.lt

Translating “highly irregular rhythm” into a detection parameter is challenging, since not much is known a priori about the features which are best suited for characterizing irregularity in AF. An abundance of detection parameters have been proposed in the literature, many of them reviewed in this chapter, and each parameter is designed to capture some specific feature of rhythm irregularity. An early study on the characterization of irregularity in AF, without also addressing the AF detection problem, posed the fundamental question whether the series of RR intervals in AF is random or deterministic [2]. The results in that study showed that the RR intervals are not entirely unpredictable, as evidenced by the nonzero correlation between the observed and the predicted RR intervals at different correlation lags. However, these findings did not apply to all patients of the analyzed data set, and, therefore, parameters related to prediction/correlation are unlikely to be good candidates for AF detection. In another study, spectral analysis demonstrated that the RR interval series during AF has a white noise-like spectrum when analyzed on a minute-by-minute scale [3].

Heart rate may be considered in AF detection as it tends to be higher in AF episodes than in sinus rhythm. Although it is obvious that heart rate alone cannot be used for detection, the power of a detection parameter describing rhythm irregularity may still be boosted by integrating information on heart rate into the definition of a parameter. Heart rate is usually characterized by the mean of the RR intervals contained in a detection window.

The detection of AF is compounded by the fact that certain arrhythmias are manifested by RR interval patterns closely resembling those observed in AF. This problem is particularly pronounced when all detection parameters describe rhythm characteristics. Hence, it is highly desirable that the detector can recognize the characteristics of confounding non-AF rhythm patterns so that the number of false alarms is minimized. Runs of ventricular premature beats (VPBs), frequent atrial premature beats (APBs), and atrial flutter, as well as bigeminy and trigeminy, are all important sources to false alarms; representative RR interval series for some of these confounding rhythms are displayed in Fig. 4.1. Another source of false alarms is inaccurate QRS detection, e.g., caused by muscle noise, motion artifacts, or large-amplitude T waves. Moreover, the risk of detecting non-AF rhythm patterns becomes increasingly higher as the detection window becomes increasingly shorter, which is required to detect short AF episodes.

When information on P waves and/or f waves is considered in AF detection, it should be paired with information on signal quality, indicating to what degree wave measurements can be trusted. Otherwise, garbage measurements may completely disrupt detection performance. Given that many clinical studies explore information derived from continuous long-term ECG recordings, often characterized by a substantial variation in noise level, information on signal quality should be an integral part of the decision-making process.

An AF episode of at least 30 s duration is considered clinically significant—a definition which was published in the ACC/AHA/ESC 2006 guidelines for management of AF patients [4], and in widespread use among clinicians. The motivation behind 30 s as minimum duration was not clearly stated, although the guidelines pointed

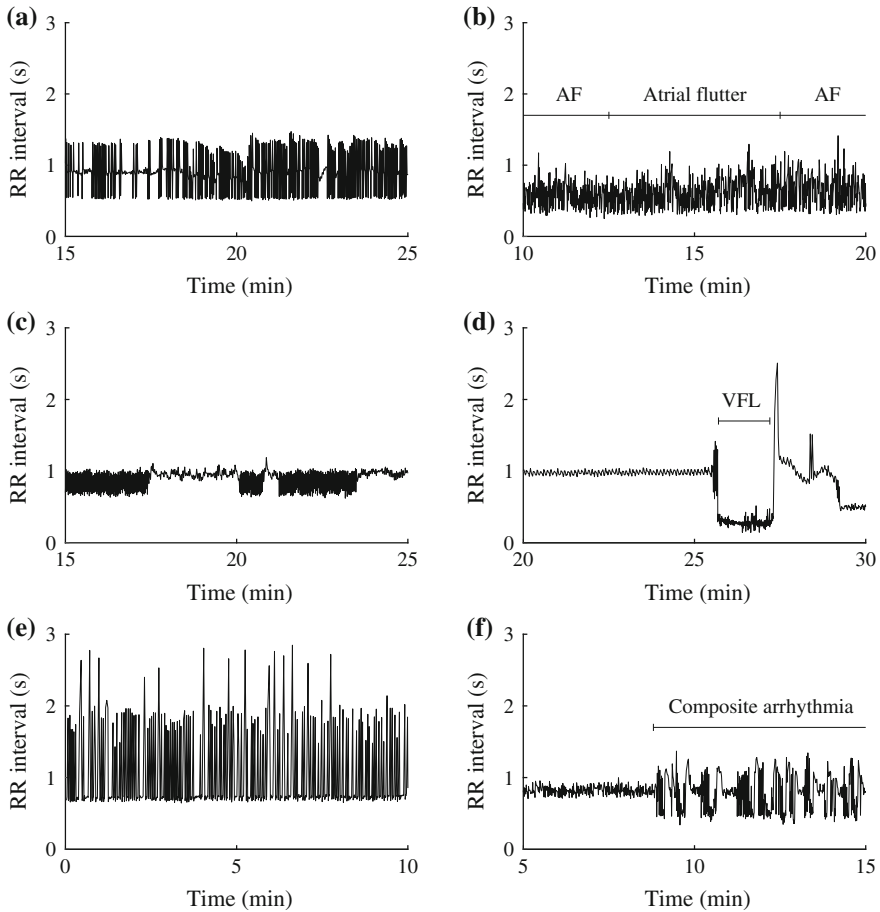


Fig. 4.1 Illustration of RR interval patterns which may confound detection of AF episodes. **a** Multiple ventricular premature beats, including bigeminy and trigeminy, **b** atrial flutter surrounded by AF, **c** second degree atrioventricular block, **d** episode of ventricular flutter (VFL), **e** sinus bradycardia, and **f** episode of a composite arrhythmia including AF, atrial flutter, atrial bigeminy, supraventricular tachycardia, atrioventricular junctional rhythm, and atrial premature beats. All examples are taken from the MIT-BIH Arrhythmia Database

out that AF episodes briefer than 30s may be relevant in “certain clinical situations involving symptomatic patients, pre-excitation or in assessing the effectiveness of therapeutic interventions.” Interestingly, the more recent guidelines published in 2014 [5] did not mention anything about minimum episode duration, whereas the 2016 guidelines [6] brought back the 30s minimum duration previously published in 2006.

In recent years, the significance of AF episodes briefer than 30s has received increasing attention in clinical research, especially concerning issues related to the

future risk of stroke and its prevention.¹ It has been suggested that such brief episodes are directly coupled to the formation of atrial thrombus, and, therefore, may be viewed as biomarkers of prolonged episodes occurring outside of the monitoring period [12–14]. When monitoring is performed during a month-long period, a patient with numerous brief episodes can very well have a higher AF burden than a patient with a few episodes which all exceed 30 s, meaning a higher thromboembolic risk for the patient with brief episodes [15], see also [16] and page 43. The concept “AF burden” is defined as the proportion of the total recording time a patient is in AF. The minimum duration of an episode which still convey clinically significant information remains to be established.

Long-term AF monitoring requires automated event detection for efficient and practical handling. Thus, the properties of the detector play a central role as they impose a lower limit on how brief an episode can be and still be detected. Most detectors described in the literature have a design that precludes the detection of episodes briefer than 30 s due to the principle adopted for detection. For example, AF detection based on RR interval histogram analysis requires a large number of RR intervals to ensure that the histogram is reasonably reliable. Indeed, some ECG-based detectors are blind to episodes briefer than two minutes, whereas, in implantable devices, a minimum episode duration of as much as six minutes has been used [17, 18]. Clinical studies reporting results on the presence of episodes briefer than 30 s have relied on commercial detectors, implementing proprietary algorithms whose detection performance have not been published [12, 13, 19, 20]. Therefore, manual review of possible AF events briefer than 30 s has been required to carry through the study [12]. Consequently, it is of substantial interest to design and evaluate AF detectors which facilitate the investigation of the clinical significance of brief episodes.

The duration of an AF episode is highly variable, extending from less than 30 s up to seven days; episodes extending beyond seven days are designated as persistent AF [4]. Similar to the problem of detecting QRS complexes, where a least informative approach is often recommended with respect to assumptions on signal properties [21], an AF detector should not involve firm assumptions on episode duration, nor on the minimum distance between two subsequent AF episodes. By merging two detected episodes, even if separated by just a few seconds, clinically relevant information could be excluded.

With the advent of handheld and smartphone-based devices for AF screening comes new possibilities to identify previously undetected AF [22–29], cf. Sects. 2.3.5 and 2.3.6, but also new challenges related to the signal quality of such patient-operated devices which, in general, is poorer than the quality associated with the clinical modalities, see Fig. 4.2 for an illustration of poor signal quality. Since handheld and smartphone-based devices are designed to record a single lead, not necessarily reflecting atrial activity, rhythm-based detection is the typical mode of operation, with information on f and P wave morphology as a bonus.

¹Paroxysmal AF manifested by episodes briefer than 30 s is sometimes referred to as *occult paroxysmal AF*, especially when asymptomatic or undetected by conventional methods [7–11].

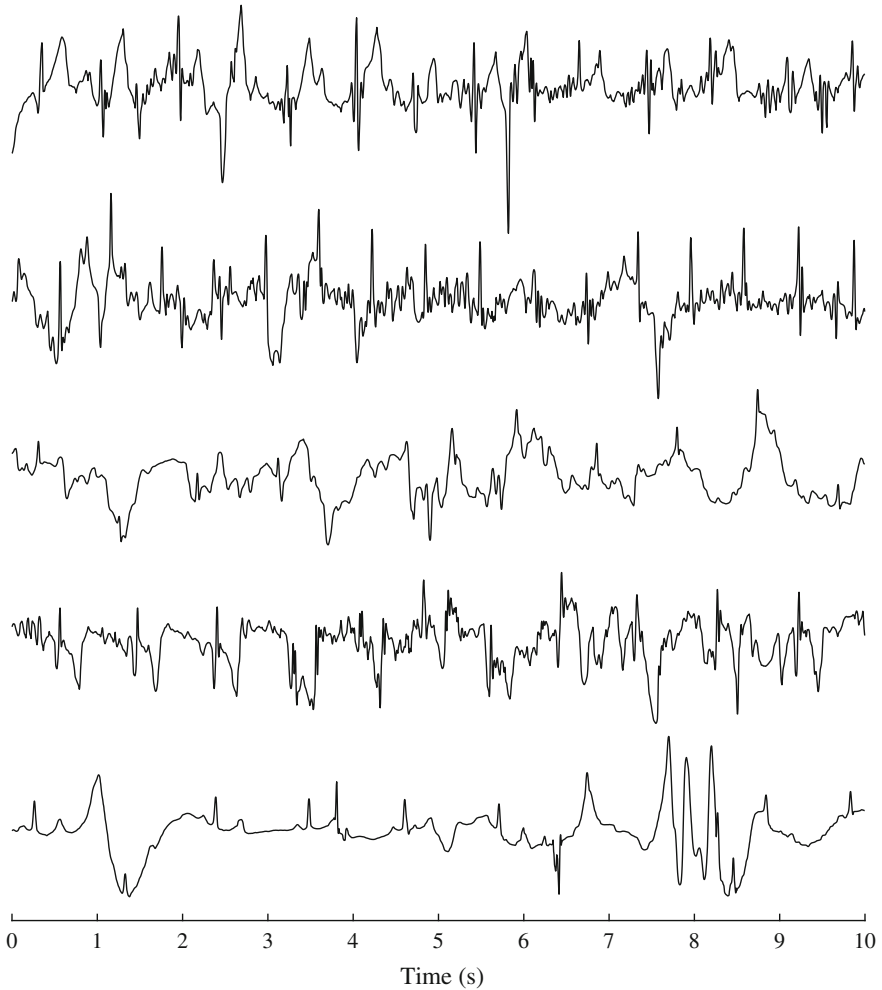


Fig. 4.2 Five examples of poor-quality ECGs recorded using a smartphone-based device. The signals are part of the database made available for the PhysioNet/Computing in Cardiology Challenge 2017 [30]

In this chapter, the main design principles used in AF detection are reviewed, either exploring rhythm information only, i.e., the RR interval series, (Sect. 4.2) or information on both rhythm and atrial wave morphology (Sect. 4.3). Aspects on detector implementation are briefly considered in Sect. 4.4, and different performance measures used in AF detection are described in Sect. 4.5. Although several reflections on performance are interspersed throughout the chapter, Sect. 4.6 has detection performance as its main theme, with a discussion on aspects which need to be considered

when evaluating performance. The chapter ends with a discussion on different types of ECG-derived information which may be explored to improve detection performance (Sect. 4.7).

4.2 Rhythm-Based AF Detection

Since reliable information on the absence/presence of P and f waves is difficult to extract at low signal-to-noise ratios (SNRs), the vast majority of AF detectors rely entirely on parameters quantifying RR interval irregularity, e.g., the degree of randomness, variability, and complexity. Another important explanation to the dominance of rhythm-based detectors is that their implementation in hardware requires far less energy than do detectors which also involve morphologic information. The RR interval series constitutes the sole input data to most detectors implemented in an implantable device, since morphologic information is difficult to extract from invasive recordings.

Over the years, detector design has been based on *ad hoc principles*, involving one or a few parameters which are fed to a simple classifier, while neither model-based statistical detection nor physiological considerations have played a significant role in the design. Nonetheless, it is obvious from the results listed in Table 4.1 that ad hoc principles have helped to push the limits of detection performance as both sensitivity and specificity have improved; for a definition of these two performance measures, see Sect. 4.5.² Still, further improvement of detector performance is warranted so that, for example, the problem of false alarms due to frequent ectopic beats, together masquerading as an AF episode, non-AF arrhythmias, or noisy signals can be adequately addressed.

Apart from using the RR interval series $x(0), \dots, x(N - 1)$ itself as detector input, the first difference,

$$\Delta x(n) = x(n) - x(n - 1), \quad n = 1, \dots, N - 1, \quad (4.1)$$

sometimes also serves as input, where N is the number of RR intervals and n is the interval index (and thus not ECG sample index). Unless the ECG recording is very short, i.e., on the order of 10–20 s, the input data is usually processed using a *sliding time window* approach in which the detection parameters are repeatedly computed as the window slides forward in time. Sliding by one RR interval at a time offers the best time resolution of episode onset and end; however, it may be necessary to take larger “slides” to reduce the amount of computations, for example, 50 intervals at a time [36].

The main principles explored for rhythm-based AF detection are described in the following. To simplify the description, detection parameters are assumed to be

²Several other detectors have been proposed besides those listed in Table 4.1. However, for various reasons, their respective performance was not evaluated on AFDB.

Table 4.1 Performance of rhythm-based AF detectors expressed in terms of sensitivity (Se) and specificity (Sp), using the MIT–BIH Atrial Fibrillation Database (AFDB) for evaluation, see Sect. 3.1. The subset AFDB₁ is identical to AFDB, except that records 4936 and 5091 are excluded for reasons of incorrect annotations. The detectors are ordered with respect to their year of publication

Method by	Year	Database	Se (%)	Sp (%)
Tateno and Glass [31]	2001	AFDB	94.4	97.2
Dash et al. [32]	2009	AFDB ₁	94.4	95.1
Lian et al. [33]	2011	AFDB	95.8	96.4
Lake and Moorman [34]	2011	AFDB	91	94
Huang et al. [35]	2011	AFDB	96.1	98.1
Shouldice et al. [36]	2012	AFDB	92	96
Lee et al. [37]	2013	AFDB ₁	98.2	97.7
Zhou et al. [38]	2014	AFDB	96.9	98.3
Asgari et al. [39]	2015	AFDB	97.0	97.1
Petrėnas et al. [40]	2015	AFDB	97.1	98.3
Zhou et al. [41]	2015	AFDB	97.4	98.4

computed in a fixed window, however, it is straightforward to replace it with a sliding window. The interested reader may want to follow up with some other rhythm-based detectors proposed over years [42–47].

4.2.1 Irregularity Parameters

Table 4.2 presents a list of parameters considered in the design of AF detectors, grouped into five categories, namely statistical dispersion, entropy, parameters based on symbolic dynamics, parameters based on the Poincaré plot, and parameters based on the time-varying coherence function. Of these categories, statistical parameters reflecting dispersion, e.g., the root mean square of successive differences, the mean of absolute successive differences, and the coefficient of variation, are the most commonly used. Some detectors base their decisions on just one parameter, combined with simple thresholding, whereas other detectors rely on a combination of parameters as input to the classifier. Certain parameters are intimately related to a statistical test, for example, the number of turning points, and, therefore, the test is described together with the parameter, instead of in Sect. 4.2.6 where different types of classifier are described.

Statistical Dispersion Parameters

The coefficient of variation (CV) of $x(n)$ has been used in AF detection [31, 48], defined by

Table 4.2 List of parameters used in rhythm-based AF detection, grouped into five different categories: statistical dispersion, entropy, symbolic dynamics, Poincaré plot-based, and time-varying coherence function

Detection parameter	Publication
Coefficient of variation	[31, 48]
Root mean square of successive differences	[32, 37]
Normalized mean of absolute successive differences	[48]
Number of turning points	[32]
Histogram-based parameters	[31, 35]
Shannon entropy	[32, 37, 38]
Sample entropy	[34, 49]
Simplified sample entropy	[40]
Symbolic dynamics	[38, 41]
Poincaré plot of $x(n)$ versus $x(n - 1)$ + bin count	[50]
Poincaré plot of $\Delta x(n)$ versus $\Delta x(n - 1)$ + bin count	[51]
Poincaré plot of $x(n)$ versus $\Delta x(n - 1)$ + bin count	[33]
Time-varying coherence function	[37]

$$P_{CV} = \frac{\sigma_x}{m_x}, \quad (4.2)$$

where m_x and σ_x denote the mean and the standard deviation of $x(n)$, respectively. The parameter P_{CV} describes dispersion but also reflects changes in heart rate since RR interval shortening, often occurring in an AF episode, is related to a smaller m_x . Using $\Delta x(n)$ instead of $x(n)$ in (4.2), the resulting mean $m_{\Delta x}$ becomes close to zero, and, therefore, to avoid division with zero, as well as to maintain the dependence on changes in heart rate, it is substituted by m_x . The performance of two single-parameter detectors, both based on P_{CV} but computed either from $x(n)$ or $\Delta x(n)$, were studied in [31]; the two detectors were found to have about the same performance.

The root mean square of successive differences (RMSSD) is defined by

$$P_{RMSSD} = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N-1} \Delta x^2(n)}. \quad (4.3)$$

Since this parameter does not reflect changes in heart rate, a heart rate dependent detection threshold can be introduced to implicitly handle such changes [32]. Accordingly, P_{RMSSD} can alternatively be interpreted as a heart rate normalized parameter applied to a fixed threshold test. Thus, the test involving a heart rate normalized P_{RMSSD} is identical to P_{CV} , with the mean and standard deviation of $\Delta x(n)$ inserted in (4.2).

Yet another dispersion parameter is the normalized mean of absolute successive differences (NMAASD) [48], defined by

$$P_{\text{NMASD}} = \frac{1}{N-1} \frac{\sum_{n=1}^{N-1} |\Delta x(n)|}{m_x}. \quad (4.4)$$

The motivation for using P_{NMASD} instead of P_{CV} , when based on $\Delta x(n)$, is unclear as the former parameter represents an approximation of the latter. Therefore, it is not surprising that the detection performance of P_{NMASD} was found to be almost the same as that of P_{CV} [48].

Thus, it may be concluded that the three dispersion parameters in (4.2)–(4.4) convey similar information. As shown below, yet another detection parameter conveys information on RR interval dispersion, though developed in the context of the Poincaré plot.

Number of Turning Points

The turning point test is a nonparametric, statistical test to determine whether the samples of a time series can be modeled by independent and identically distributed random variables. In a completely random series, any three successive samples are equally likely to occur in any of the six possible orders. In four of the orders, a turning point exists if the middle sample is a local maximum or a local minimum. Thus, the probability of a turning point in a three-sample series is $2/3$.

For a series with N samples, the number of turning points N_{TP} can be counted and compared to the expected number of turning points m_{TP} of a completely random series. If N_{TP} is too many standard deviations σ_{TP} away from m_{TP} , the series cannot be considered as completely random. Making use of the result that the mean and the standard deviation of N_{TP} are given by [52]

$$m_{\text{TP}} = \frac{2(N-2)}{3}, \quad (4.5)$$

$$\sigma_{\text{TP}} = \sqrt{\frac{16N-29}{90}}, \quad (4.6)$$

respectively, and that N_{TP} obeys an asymptotically normal distribution for a sufficiently large N , a two-sided statistical test can be used. When the number of observed turning points falls outside the 95% confidence limits, defined by $m_{\text{TP}} \pm 1.96\sigma_{\text{TP}}$, the hypothesis stating that the series is completely random can be rejected.

In AF detection, the number of observed turning points, together with other parameters, is employed for characterizing RR interval irregularity in AF [32]. Rather than using a statistical test with 95% confidence limits, the limits are determined to optimize detection performance with respect to sensitivity and specificity. When the number of turning points falls outside the optimized limits, the RR interval series is likely to exhibit periodicity, for example, due to respiratory-modulated sinus rhythm. Since it has been shown that RR intervals in AF may exhibit certain correlation [2],

the turning point test loses some of its power in detecting random RR interval series. Moreover, the turning point information is likely to cause false alarms in the presence of ectopic beats and rapid changes in rhythm, and, therefore, it is less suitable for AF detection.

Histogram-Based Parameters

Since RR interval histograms determined in sinus rhythm or AF exhibit considerable differences in shape, their shapes have been explored for AF detection. However, to make the histogram approach work, the bins must be sufficiently well-populated so that a histogram can be produced which is representative of the prevailing rhythm. This requirement implies that a large number of RR intervals has to be used for histogram construction—100 beats appears to be a minimum number [31, 35]—which, on the other hand, implies lower accuracy of the estimated onset and end times of an AF episode. If fewer and wider bins are used to allow a shorter window, the histogram becomes increasingly inadequate for discrimination between different types of cardiac rhythms. Therefore, an inherent limitation of histogram-based detection is the need of a long window, which thus precludes the detection of brief episodes.

A straightforward approach to histogram-based detection is to define a set of heuristic features which characterize the histogram, e.g., the height and the number of nonempty bins. Since a histogram in AF is usually much broader in shape than a histogram in sinus rhythm, AF is characterized by a lower height and fewer nonempty bins. If a change in heart rate occurs within the detection window, a histogram in sinus rhythm will broaden and become increasingly similar to the shape of an AF histogram. To some extent, however, this transitional problem can be avoided using the Δ RR interval histogram, since differencing not only removes slow trends present in the RR interval series, but it also makes the histogram span over a smaller range of values.

A more sophisticated approach to histogram-based detection is to compare the RR interval histogram of the detection window with a set of template histograms, stratified according to the mean RR interval length [31].³ Each template histogram is constructed from all the RR intervals contained in (nonoverlapping) windows with a mean RR interval length falling inside an interval with predefined limits, ranging, for example, from 350–399 to 1100–1149 ms in steps of 50 ms [31]. Windows whose mean length falls outside any of the predefined intervals are discarded from further analysis. The template histograms are constructed prior to detection, preferably from a huge AF database to ensure that the histograms are sufficiently representative of the underlying probability density function (PDF); the same procedure applies to Δ RR intervals.

³The relationship between histogram shape and mean heart rate has previously been investigated in noninvasive studies on atrioventricular node physiology in AF, leading to the concept of heart rate stratified histograms [53, 54].

In AF detection, the observed RR interval histogram is computed in a sliding detection window and compared to each of the template histograms [31]. For this comparison, the nonparametric Kolmogorov–Smirnov test can be used since it measures the probability of the observed RR intervals being drawn from the same population as the fixed data set, i.e., the RR intervals used for constructing the template histograms [55]. This test involves a statistic defined by the largest distance between the cumulative histogram of the observed data set and the cumulative template histogram, assessing whether the two cumulative histograms are different, see Fig. 4.3. The Kolmogorov–Smirnov test is suitable to use when two cumulative probability distributions differ in a global fashion near the center, but less suitable when the two distributions differ with respect to the number of peaks. For example, the largest distance between a bimodal and a unimodal cumulative probability distribution, both determined in AF, may not be large enough to show that the two data sets come from different populations. In such cases, the Anderson–Darling test is a better choice since it makes use of a weighted sum of the squared deviations between the two cumulative probability distributions, rather than just the largest distance at one single point [55].

Poor performance was reported when the RR series was used as input to the Kolmogorov–Smirnov test, with sensitivity and specificity of 66.3% and 99.0%, respectively [31]. Using instead Δ RR intervals as input, the sensitivity improved dramatically to 94.4%, whereas the decrease in specificity to 97.2% was relatively modest. While the authors did not provide any specific explanation to this improvement, it may be that the use of Δ RR intervals leads to better performance since the related histogram is more unimodal than that of the RR intervals, and therefore better suited for use with the Kolmogorov–Smirnov test.

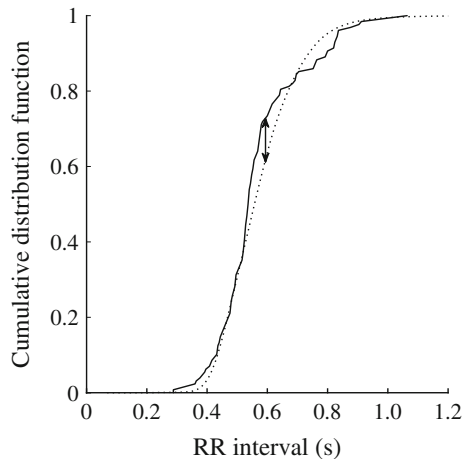


Fig. 4.3 The Kolmogorov–Smirnov test requires that the largest distance between two cumulative histograms is determined. In this example, both histograms belong to RR intervals in AF. The largest distance is marked with an arrow

The multi-template histogram approach offers the advantage of providing a much more detailed characterization of the shape of the RR interval distribution than does the single template histogram in which all RR intervals are merged. On the other hand, it is well-known that the shape of RR interval histograms exhibits considerable intra- as well as inter-patient variability, and unimodal as well as bimodal shapes are often observed in AF [56–58]. Consequently, an AF detector relying on a set of template histograms is likely to perform less satisfactory when these types of variability are pronounced.

Another approach to histogram-based AF detection is to compare two Δ RR interval histograms determined from the first and the last part of the detection window [35], thus replacing the above-mentioned comparison to template histograms. The sum of the squared difference between the corresponding bin counts of the two histograms is used as a detection parameter: this difference remains small as long as the same rhythm persists, but increases when a transition from sinus rhythm to AF occurs, or vice versa. Since the information carried by the squared difference turned out to be insufficient for achieving satisfactory detection performance, the number of nonempty bins, the height of the histogram, and the standard deviation of the Δ RR intervals were also used as detection parameters to improve discrimination between sinus rhythm and AF.

Shannon Entropy

The Shannon entropy quantifies the uncertainty (unpredictability) of the information content of a “message” such as the RR interval series [59]. In statistical terms, the entropy increases as the PDF becomes increasingly uniform, and decreases when the PDF becomes increasingly concentrated around a certain value. In other words, large entropy indicates low predictability of the information content, and vice versa. The Shannon entropy (ShEn) is defined by

$$I_{\text{ShEn}} = - \sum_{i=1}^B p(x_i) \log_2(p(x_i)), \quad (4.7)$$

where the message is synonymous to the outcome of a random variable x assuming B different values, i.e., (x_1, \dots, x_B) ; the probability of each value is given by $p(x_i)$. Since I_{ShEn} ranges from 0 to $\log_2(B)$, the right hand side of (4.7) is sometimes normalized with $\log_2(B)$ to facilitate interpretation. In practice, the probability $p(x_i)$ is estimated from the message itself, usually by computing the histogram. The probability of the i -th bin is estimated by

$$\hat{p}(x_i) = \frac{N(i)}{N}, \quad (4.8)$$

where $N(i)$ denotes the count of the i -th bin.

The Shannon entropy I_{ShEn} has been considered in AF detection since it is typically much larger in AF than in sinus rhythm [32]. The computation of I_{ShEn} is based on a modified RR interval series in which the longest and the shortest RR intervals are first removed to reduce the influence of outlier values. The histogram is constructed from the remaining RR intervals, with the bins equally spaced over an interval defined by the shortest and longest RR intervals of the modified series. The authors concluded that at least 16 bins should be used to obtain I_{ShEn} with reasonable accuracy.

It has been found that I_{ShEn} is associated with a degradation in performance at higher heart rates, i.e., from about 90 beats per minute (bpm) and higher [49]. This finding can be explained by noting that the probability distribution $\hat{p}(x_i)$ becomes increasingly narrower as the heart rate increases, illustrated by the following example where the variation in heart rate, set to 5 bpm, is identical at different heart rates. For a heart rate of 60 bpm, the RR intervals corresponding to 55 and 65 bpm have the lengths 1090 and 923 ms, respectively, and thus the difference in length is 167 ms. On the other hand, for a heart rate of 120 bpm, the RR intervals corresponding to 115 and 125 bpm have the lengths 521 and 480 ms, respectively, i.e., the difference in length has shrunk to 41 ms. Since I_{ShEn} is computed from the RR intervals, and not from the instantaneous heart rate, it is obvious that the power of I_{ShEn} to discriminate AF from sinus rhythm becomes increasingly worse as the heart rate becomes increasingly higher.

Rather than computing I_{ShEn} directly from the RR interval series, the ΔRR interval series can be mapped to a symbolic series, defined by an alphabet, containing only 10 symbols, which is used for computation of I_{ShEn} [38]. The mapping function quantizes the changes present in the RR interval series by relating the changes to a “reference RR series” resulting from lowpass filtering of the RR interval series. The quantization grid is dynamic in the sense that it is defined by the properties of another, even more lowpass filtered version of the RR interval series; linear, time-invariant lowpass filters are employed, where the lowpass filters are obtained by ad hoc design. The results suggested that the use of symbolic dynamics provides a path to better performance, probably explained by the quantization operation which helps to improve the separation between normal beats and beats in AF when described by I_{ShEn} .

In a subsequent study, bearing considerable resemblance to the one in [38], the authors delved further into the use of symbolic series and Shannon entropy [41]. The main difference between the two detectors is that the instantaneous heart rate is employed, rather than the RR interval series, for generating a symbol series, using a quantization grid with fixed steps. While the authors do not provide any explanation to why the instantaneous heart rate leads to slightly better detection performance, this result seems plausible since the above-mentioned limitation, i.e., when I_{ShEn} is computed from the RR interval series at different heart rates [49], is then sidestepped.

Sample Entropy

While the Shannon entropy is based on the probability of a certain RR interval length to occur, the sample entropy (SampEn) reflects self-similarity of a signal, and therefore used as a measure of signal complexity [60, 61]. The sample entropy is defined as the negative natural logarithm of the conditional probability of a signal repeating itself for m samples within the tolerance r will also repeat itself for $m + 1$ samples, where self-matches are excluded [60],

$$I_{\text{SampEn}} = -\ln \left(\frac{B(m+1, r)}{B(m, r)} \right), \quad (4.9)$$

where $B(m, r)$ is the probability of pairs of sequences which match for m samples. A small value of I_{SampEn} indicates that the signal repeats itself and therefore is regular, whereas a large value indicates a complex (irregular) signal. In terms of AF detection, this means that a transition from sinus rhythm to AF is manifested by a considerably increase in I_{SampEn} , and vice versa.

To estimate the probability $B(m, r)$, the RR interval series $x(0), \dots, x(N-1)$ is first divided into m -length subsequences, described by the vectors

$$\mathbf{x}(i) = \begin{bmatrix} x(i) \\ \vdots \\ x(i+m-1) \end{bmatrix}, \quad i = 0, \dots, N-m-1. \quad (4.10)$$

Similarity between two subsequences, beginning at i and j , respectively, is measured by the maximum norm, defined by

$$\|\mathbf{x}(i) - \mathbf{x}(j)\|_{\infty} = \max_{k=0, \dots, m-1} |x(i+k) - x(j+k)|, \quad i, j = 0, \dots, N-m-1. \quad (4.11)$$

Two subsequences are considered similar when $\|\mathbf{x}(i) - \mathbf{x}(j)\|_{\infty}$ is within a fixed tolerance r . Accordingly, the average number of similar subsequences is given by

$$\hat{B}_i(m, r) = \frac{1}{N-m-1} \sum_{j=0, j \neq i}^{N-m-1} H(r - \|\mathbf{x}(i) - \mathbf{x}(j)\|_{\infty}), \quad (4.12)$$

where self-matches are excluded. The maximum number of similar subsequences is equal to $N - m - 1$. The Heaviside step function $H(z)$ is defined by

$$H(z) = \begin{cases} 1, & z \geq 0, \\ 0, & z < 0. \end{cases} \quad (4.13)$$

The probability of two m -length subsequences being similar is estimated by

$$\begin{aligned}\hat{B}(m, r) &= \frac{1}{N-m} \sum_{i=0}^{N-m-1} \hat{B}_i(m, r) \\ &= \frac{1}{(N-m)(N-m-1)} \sum_{i=0}^{N-m-1} \sum_{j=0, j \neq i}^{N-m-1} H(r - \|\mathbf{x}(i) - \mathbf{x}(j)\|_\infty).\end{aligned}\quad (4.14)$$

Since an estimate of $B(m+1, r)$ is required before I_{SampEn} can be computed, (4.11)–(4.14) are also evaluated for $m+1$.

When computing I_{SampEn} in a short window, required for detection of brief AF episodes, the likelihood that none of the few subsequences match is high, especially for a small r . Accordingly, the denominator $\hat{B}(m, r)$ in (4.9) may become zero, leading to that I_{SampEn} is undefined. In order to address this problem, the probabilities in (4.9) can be converted to probability densities by division of the volume of the matching regions [62],

$$-\ln\left(\frac{B(m+1, r)}{(2r)^{m+1}}\right) + \ln\left(\frac{B(m, r)}{(2r)^m}\right) = -\ln\left(\frac{B(m+1, r)}{B(m, r)}\right) + \ln(2r).\quad (4.15)$$

This conversion, serving as a normalization, allows direct comparison of sample entropies computed for different values of r . As a result, the standard approach to selecting r , i.e., an r taken as a fraction of the standard deviation of the input data [60], may be replaced by an approach in which r is data-dependent. The operating value of r is then determined by incrementing r until $B(m, r)$ becomes nonzero; in AF analysis 30 ms has been used as initial value of r , after which r is incremented in steps of 5 ms.

Based on statistical analysis of different RR interval series in AF, it has been observed that the mean RR interval length \bar{m}_x provides predictive information on AF independently of I_{SampEn} [34]. In AF detection, this observation can be accounted for by simply subtracting the logarithm of \bar{m}_x from the expression on the right hand side of (4.15), leading to a new entropy measure, labeled the *coefficient of sample entropy* (CSampEn) and defined by [34]

$$I_{\text{CSampEn}} = I_{\text{SampEn}} + \ln(2r) - \ln(\bar{m}_x).\quad (4.16)$$

The inclusion of \bar{m}_x implies that I_{CSampEn} , as desired, increases in AF when the heart rate is usually higher, whereas it decreases in sinus rhythm when the heart rate is usually lower.

Before I_{CSampEn} can be computed, the subsequence length m needs to be determined. Use of the shortest possible subsequence, i.e., $m=1$, may be motivated by the observation that the autocorrelation function of RR intervals in AF is essentially zero for nonzero lags [3]. Another, more straightforward motivation is that better detection performance is obtained for $m=1$ than for a larger m [34]; for additional aspects on the choice of m and r , see Sect. 6.4.4.

It should be pointed out that I_{SampEn} was preceded chronologically by the *approximate entropy* I_{ApEn} [63], defined in the same way as I_{SampEn} except that self-matches are included in (4.12). However, it has been shown that I_{ApEn} is biased, heavily dependent on the number of samples N , and lacks relative consistency [60], and therefore less used than I_{SampEn} .

A variation on I_{SampEn} is the *fuzzy entropy* where the Heaviside function $H(z)$ in (4.13) is replaced by a function which fuzzifies the samples and thereby avoids that similarity of subsequences is treated as either/or [64]. The use of fuzzy entropy has found its way into the analysis of heart rate variability [65] and f wave characterization [66], whereas it remains to be shown whether it can provide better performance in AF detection.

Probability of Pairs of Matching RR Interval Subsequences

A simpler approach to entropy-based AF detection is to only consider the probability $B(m, r)$, forming part of the definition of I_{SampEn} in (4.9) [40, 67]. This approach is advantageous from an implementation viewpoint since $B(m + 1, r)$ is not needed, and neither the ratio of probabilities nor the natural logarithm have to be computed. Another advantage is that the problem of an undefined I_{SampEn} is circumvented. In this approach, the maximum norm in (4.12) is replaced with the Euclidean norm between two m -length subsequences. The following expression is used in place of $\hat{B}(m, r)$ [67],

$$\hat{C}(m, r) = \frac{2}{(N - m)(N - m - 1)} \sum_{i=0}^{N-m-1} \sum_{j=i+1}^{N-m} H(r - \|\mathbf{x}(i) - \mathbf{x}(j)\|), \quad (4.17)$$

where the Euclidean norm is denoted $\|\cdot\|$ and the normalization factor is given by the maximum value of the double sum. The estimator $\hat{C}(m, r)$ differs from $\hat{B}(m, r)$ with respect to the difference between $\mathbf{x}(i)$ and $\mathbf{x}(j)$ which is only counted once in $\hat{C}(m, r)$; self-matches are avoided in both estimators.⁴

An AF detector based on $B(m = 1, r)$ has been proposed in [40], offering the additional implementation advantages that neither the maximization in (4.11) nor the Euclidean distance in (4.17) need to be performed. The probability of two RR intervals differing less than r is estimated by

$$\hat{B}(m = 1, r) = \frac{2}{(N - 1)(N - 2)} \sum_{i=0}^{N-2} \sum_{j=i+1}^{N-1} H(r - |x(i) - x(j)|). \quad (4.18)$$

Before application of a detection threshold, the probability $\hat{B}(m = 1, r)$ is divided by an estimate of the mean length of the RR intervals contained in the detection window

⁴It should be noted that $\hat{C}(m, r)$ constitutes an essential part of the *correlation dimension*, a measure introduced to describe the dimensionality of the space occupied by a set of random samples [68].

to emphasize that AF is usually accompanied by a higher heart rate [40]. Thus, the resulting detection parameter, denoted the *simplified sample entropy* (SSampEn), is defined by

$$I_{\text{SSampEn}} = \frac{\hat{B}(m = 1, r)}{\bar{m}_x}, \quad (4.19)$$

where \bar{m}_x is obtained from exponential averaging of the RR intervals, excluding the RR intervals related to ectopic beats which previously have been flagged by a simple algorithm, see Sect. 4.2.5. The ratio in (4.19) bears considerable resemblance to the coefficient of variation in (4.2), since the numerator is a dispersion measure (though thresholded and therefore not changing in the same continuous way as does the standard deviation in (4.2)) and the denominator is given by the mean of the RR intervals.

Another possible approach to accounting for information on heart rate in (4.18) is to replace the fixed tolerance r with a tolerance defined as a function of the heart rate in the detection window, i.e., $r \rightarrow r(\bar{m}_x)$. If a fixed r is still preferred, it can, as already mentioned, be taken as a fraction of the standard deviation determined from a huge data set [60].

4.2.2 Poincaré-Based Parameters

The scatter plot of successive pairs of RR intervals ($x(n), x(n + 1)$), known as the Poincaré plot, is a simple technique for characterizing different types of cardiac rhythms. This type of plot was introduced for analyzing nonlinear aspects of heart rate variability, constructed from a series of RR intervals spanning over a long time period, i.e., up to several days [69–72]. The Poincaré plot has also served as the guiding design principle when developing AF detectors, but then a much shorter time period determined by the detection window is subject to analysis, i.e., typically ranging from 60 to 120 s. Since the Poincaré plot constructed from the RR intervals in AF is much more scattered than the plot constructed from normal sinus rhythm and atrial or ventricular ectopic rhythms, illustrated in Fig. 4.4, the challenge to be addressed is one of translating the scattering observed in AF to a set of detection parameters. The following two approaches have been pursued:

1. parameters reflecting the density of points in different regions of the Poincaré plot [33, 51, 73], and
2. parameters providing a geometrical characterization of the points in the Poincaré plot [50].

In addition to relying on ($x(n), x(n + 1)$) as the basis for producing a Poincaré plot, these two approaches can alternatively rely on ($\Delta x(n), \Delta x(n + 1)$) or ($x(n), \Delta x(n)$) which also convey information on beat-to-beat irregularity in the RR interval series.⁵

⁵Yet another approach proposed for characterizing the Poincaré plot is the complex correlation measure, quantifying the point-to-point (temporal) variation of the RR series [74], see also [75].

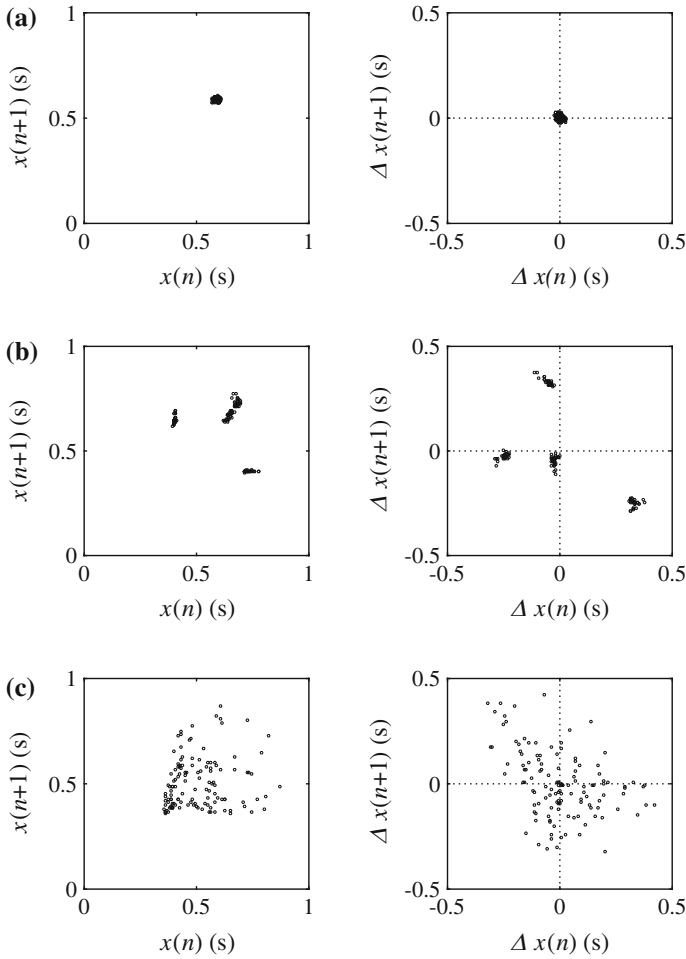


Fig. 4.4 Poincaré plots defined by $(x(n), x(n + 1))$ and $(\Delta x(n), \Delta x(n + 1))$ (left and right column, respectively), resulting from **a** normal sinus rhythm, **b** sinus rhythm with ectopic beats, and **c** AF. All plots are based on 128 RR intervals

The first AF detector to explore the point density of a Poincaré plot was defined by $(\Delta x(n), \Delta x(n + 1))$ [51, 76]. Hence, the proposed analysis is not confined to just the first quadrant, as is the case for $(x(n), x(n + 1))$, but covers all four quadrants since $\Delta x(n)$ can assume both positive and negative values. The quadrants are divided into a square grid, where the cells are treated as bins of a two-dimensional histogram; the bin size is a design parameter which should be set to a small value, e.g., 25 ms.

However, this measure has not received any attention in AF detection, probably because it is better suited for discriminating between ectopic rhythms and normal sinus rhythm than between AF and normal sinus rhythm.

Moreover, the Poincaré plot is divided into different regions defined so that their respective populations of points correlate with different rhythms, as manifested by the pattern of the three successive RR intervals required for computing $\Delta x(n)$ and $\Delta x(n + 1)$, see Fig. 4.5. First, the total number of bins populated by at least one point (“nonzero bins”) is computed for all regions, excluding a circular region enclosing origo which is populated by points related to normal sinus rhythm. Then, the total number of bins is corrected by not only subtracting the number of points in region 0, but also a number reflecting the presence of APBs; APBs tend to cluster in certain regions since they are often accompanied by a compensatory pause. An AF episode is detected whenever the corrected total number of bins exceeds a predefined threshold, provided that the number of points reflecting the presence of atrial tachycardia falls below another predefined threshold. The presence of atrial tachycardia is determined by a heuristic combination of the number of points found in different regions relevant to this particular arrhythmia, see Fig. 4.5; for a detailed description of the algorithm, see [51, 76].

Using the Poincaré plot defined by $(x(n), \Delta x(n))$, a much simpler approach to AF detection has been proposed in [33], particularly well-suited for use in implantable loop recorders. This approach was later applied to AF detection in polysomnographic recordings [73]. In the plot, the first and the fourth quadrants are analyzed since $\Delta x(n)$ can assume both positive and negative values. Again, the two quadrants are divided into a square grid with cells treated as bins. All bins with at least one point are counted, and an AF episode is detected whenever the total count exceeds a predefined, fixed threshold. Obviously, many more bins will be nonzero for an irregular rhythm such as AF than for normal sinus rhythm. In contrast to [51], this approach does not require that the Poincaré plot is divided into different regions, thereby simplifying detector implementation. The count of nonzero bins defines the detection parameter P_{NZPP} .

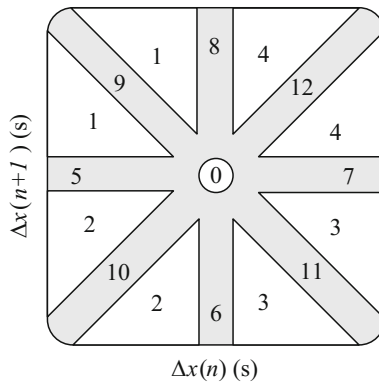


Fig. 4.5 Definition of regions in a Poincaré plot defined by $(\Delta x(n), \Delta x(n + 1))$ [51]. Normal sinus rhythm usually populates the circular, origo-centered region 0, whereas AF populates all regions except region 0. Atrial tachycardia usually populates regions 6, 7, 9, and 11, whereas atrial and ventricular premature beats usually populate regions 1–4

As already pointed out, histogram-based detectors suffer from the disadvantage of requiring a large number of RR intervals to achieve adequate performance, especially when a two-dimensional histogram is analyzed. Therefore, it is not surprising that a 2 min detection window is recommended to ensure that the different regions of the Poincaré plot (which may be viewed as a counterpart to histogram bins) are reasonably well-populated [51]. When neither histogram shape nor population size are of importance, a much shorter detection window may be employed, e.g., 64 beats, without having to trade much in performance [33]. The introduction of regions offer, on the other hand, a means to detect other rhythms than AF, e.g., atrial flutter or APBs. It should be pointed out that the relative advantage of using a Poincaré plot defined either by $(x(n), \Delta x(n))$ or $(\Delta x(n), \Delta x(n+1))$, rather than by $(x(n), x(n+1))$, remains to be established.

The second approach to Poincaré-based AF detection involves parameters providing a geometrical characterization of how the points $(x(n), x(n+1))$ populate the plot [50]; see also [77] where some of the original ideas appeared. As will be obvious from the following, detection parameters involving distances in the Poincaré plot are related to the statistical dispersion measures described earlier. Accordingly, the main merit of the Poincaré plot seems to be its use as a conceptual tool for designing parameters, while the plot itself does not provide much novel information. In contrast to the Poincaré-based detector proposed in [51], the geometrical parameters do not treat any particular region of the Poincaré plot as more likely to be populated when AF is present, but simply quantifies certain type of dispersion of the RR interval series.

In normal sinus rhythm, the points of the Poincaré plot are typically dispersed around the line of identity, i.e., $x(n) = x(n+1)$, forming a cluster whose shape resembles an ellipse. One of the axes of the ellipse, usually the major axis, has the same orientation as the line of identity, whereas the other axis is perpendicular. The dispersion of points along these two axes is quantified by first performing a 45° rotation of $(x(n), x(n+1))$, defined by

$$\begin{bmatrix} y(n+1) \\ y(n) \end{bmatrix} = \begin{bmatrix} \sin \frac{\pi}{4} & \cos \frac{\pi}{4} \\ \cos \frac{\pi}{4} & -\sin \frac{\pi}{4} \end{bmatrix} \begin{bmatrix} x(n+1) \\ x(n) \end{bmatrix}, \quad n = 0, \dots, N-2, \quad (4.20)$$

where $y(n)$ lies on the axis perpendicular to the line of identity. Then, the standard deviations $\sigma_{y,0}$ and $\sigma_{y,1}$ of $y(n)$ and $y(n+1)$, respectively, describe the shape of the cluster. The standard deviations are defined by

$$\sigma_{y,j} = \sqrt{\frac{1}{N-1} \sum_{n=0}^{N-2} (y(n+j) - \bar{m}_y)^2}, \quad j = 0, 1, \quad (4.21)$$

where \bar{m}_y denotes the mean value of $y(n)$. In a broader sense, $\sigma_{y,0}$ may be interpreted as a parameter characterizing the short-term variability of the RR intervals, whereas $\sigma_{y,1}$ characterizes long-term variability [71, 78].

On the other hand, the point distribution in AF differs significantly from that in normal sinus rhythm, implying that the assumption of a cluster with elliptic shape loses its meaning. Still, $\sigma_{y,0}$ has been employed as a detection parameter to quantify short-term variability [50], see also [79], but not $\sigma_{y,1}$ since it reflects a much coarser time scale than does $\sigma_{y,0}$. The transformation in (4.20) implies that successive RR intervals should be differenced,

$$y(n) = \frac{1}{\sqrt{2}}(x(n+1) - x(n)) = \frac{\Delta x(n)}{\sqrt{2}}, \quad (4.22)$$

and, therefore, the mean value of $y(n)$ is close to zero. Hence, the standard deviation $\sigma_{y,0}$ is well-approximated by

$$\sigma_{y,0} \approx \sqrt{\frac{1}{2(N-1)} \sum_{n=1}^{N-1} \Delta x^2(n)}, \quad (4.23)$$

which describes the dispersion of points around the diagonal line in the Poincaré plot. It is evident that $\sigma_{y,0}$, apart from different normalization factors, is identical to P_{RMSSD} in (4.3) and employed in [32] but then without any reference to the Poincaré plot. When distance measures are used for characterizing the plot $(x(n), x(n+1))$, the differenced RR interval series $\Delta x(n)$ is a quantity appearing naturally.

The idea of fitting an ellipse to the Poincaré plot stems from the analysis of long-term ECG data. When adapting this idea to AF detection, the resulting plot must be based on much fewer RR intervals (i.e., only those inside the detection window), leading to that the shape of the Poincaré plot becomes dot-like rather than ellipse-like, see Fig. 4.4. Still, the ellipse-inspired analysis of RR intervals has been considered for AF detection.

Another geometrical detection parameter inspired by the Poincaré plot is based on the Euclidean distance between two successive points $(x(n), x(n+1))$ and $(x(n+1), x(n+2))$, describing the local rate of change in the RR interval series [50]. This parameter, denoted σ_c , is defined as the mean of all Euclidean distances contained in the detection window,

$$\sigma_c = \frac{1}{N-2} \sum_{n=1}^{N-2} \sqrt{\Delta x^2(n) + \Delta x^2(n+1)}, \quad (4.24)$$

$$= \frac{1}{N-2} \sum_{n=1}^{N-2} \sqrt{\sum_{k=0}^1 \Delta x^2(n+k)}, \quad (4.25)$$

which, similar to $\sigma_{y,0}$, represents a measure of RR interval dispersion. Before use in AF detection, both $\sigma_{y,0}$ and σ_c have been “normalized” by the mean RR interval

length \bar{m}_x ,⁶ exemplified by

$$\sigma'_{y,0} = \frac{\sigma_{y,0}}{\bar{m}_x}. \quad (4.26)$$

Thus, similar to the coefficient of sample entropy in (4.16) and the simplified and heart rate modified sample entropy in (4.19), the parameters $\sigma'_{y,0}$ and σ'_c are designed so that an increase in heart rate contributes to improved detection performance.

4.2.3 Time-Varying Coherence Function

A linear systems approach to AF detection is provided by exploring the difference in spectral coherence of the RR intervals in two adjacent windows: the spectral coherence remains high as long as normal sinus rhythm is present in both windows, whereas it changes rather abruptly at the time when an AF episode either begins or ends. This approach was proposed in [37], benefitting from previously presented results on how to estimate the time-varying coherence function (TVCF) from the time-varying transfer functions obtained from the samples of two adjacent windows [80].

Assuming that the data in the two windows are viewed as the input and output signals of a linear system, denoted $x(n)$ and $y(n)$, respectively, the time-varying coherence function is defined by

$$C_{xy}(\omega, n) = \frac{|S_{xy}(\omega, n)|^2}{S_x(\omega, n)S_y(\omega, n)}, \quad (4.27)$$

where $S_{xy}(\omega, n)$ is the time-varying cross-spectrum between $x(n)$ and $y(n)$, and $S_x(\omega, n)$ and $S_y(\omega, n)$ are the time-varying spectra of $x(n)$ and $y(n)$, respectively. Conversely, when $y(n)$ is viewed as the input signal and $x(n)$ as the output signal, the time-varying coherence function is defined by

$$C_{yx}(\omega, n) = \frac{|S_{yx}(\omega, n)|^2}{S_x(\omega, n)S_y(\omega, n)}. \quad (4.28)$$

Accounting for the fact that the time-varying coherence function can be computed both forwards and backwards, an overall TVCF can be defined by

$$C^2(\omega, n) = C_{xy}(\omega, n)C_{yx}(\omega, n). \quad (4.29)$$

Introducing the two time-varying transfer functions characterizing the linear system when either $x(n)$ or $y(n)$ is the input signal,

⁶In [50], the computation of \bar{m}_x includes all RR intervals in the window except the first and last RR intervals, i.e., $x(0)$ and $x(N-1)$; however, the interpretation of \bar{m}_x is similar to that otherwise used in this chapter.

$$H_{x \rightarrow y}(\omega, n) = \frac{S_{xy}(\omega, n)}{S_x(\omega, n)}, \quad (4.30)$$

$$H_{y \rightarrow x}(\omega, n) = \frac{S_{yx}(\omega, n)}{S_y(\omega, n)}, \quad (4.31)$$

the overall TVCF can be expressed as [80]

$$C^2(\omega, n) = |H_{x \rightarrow y}(\omega, n)H_{y \rightarrow x}(\omega, n)|^2. \quad (4.32)$$

The two filters $H_{x \rightarrow y}(\omega, n)$ and $H_{y \rightarrow x}(\omega, n)$ can be determined using a model-based approach in which the samples of the two windows are assumed to be characterized by an autoregressive moving average (ARMA) model. This approach is preferred over a spectrogram-based approach due to its better frequency resolution, provided that the ARMA model is adequate for the analyzed data. Both the model parameters and the model order are determined using an optimization technique developed especially for the identification of time-varying linear systems [81]. Results have demonstrated that the model order estimate depends on the length of the detection window: longer windows require higher model orders.

Figure 4.6 illustrates one of the essential properties of $C^2(\omega, n)$, namely that the variation across the frequency axis is almost nonexistent in normal sinus rhythm,

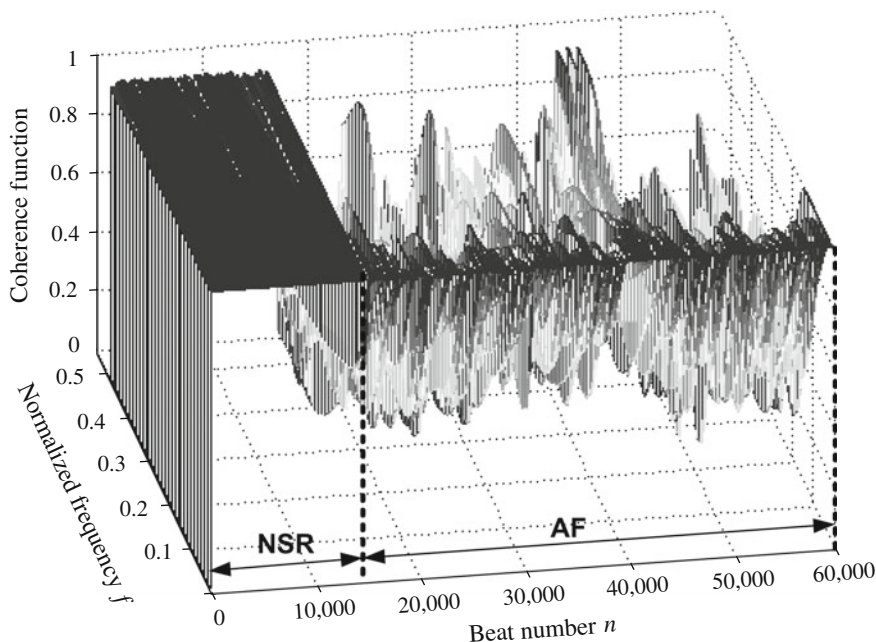


Fig. 4.6 Time-varying coherence function $C^2(\omega, n)$ computed from of an RR interval series containing a transition from normal sinus rhythm (NSR) to AF ($\omega = 2\pi f$). Both detection windows contain 128 beats, and slide with 128 beats at a time. (Reprinted from [37] with permission)

whereas the variation increases at the onset of the AF episode—an increase which becomes more pronounced at higher frequencies. Based on this observation, the variance of $C^2(\omega, n)$ is computed across the frequency axis for each beat n , and used as detection parameter.

4.2.4 Parameter Time Series Exemplified

For an 80-min ambulatory ECG recording with two AF episodes and several runs of ectopic beats, the time series of different detection parameters are displayed in Fig. 4.7. The series are computed using a 128-beat sliding detection window, except for $I_{SSampEn}$ which is computed using an 8-beat window [40]; the window slides one beat at a time.

A number of observations can be made from Fig. 4.7, first and foremost that normal sinus rhythm and AF episodes are easily distinguished in all series. Another observation is that the impact of the runs of ectopic beats, for example, those occurring before the second AF episode, differ quite considerably between the series: while the impact is small for $I_{SSampEn}$, it is quite substantial for P_{CV} and P_{NMASD} since the ectopic beats are manifested by parameter values which actually exceed those belonging to the AF episodes. Thus, to reduce the number of false alarms, techniques for handling the influence of ectopic beats need to be implemented, see Sect. 4.2.5. Yet another observation to be made from Fig. 4.7 is that I_{ShEn} has more pronounced “background” fluctuations in normal sinus rhythm than the other detection parameters.

4.2.5 Ectopic Beat Handling

An important aspect to address in rhythm-based AF detection is the presence of ectopic beats, often abundant in numbers. The inclusion of a processing block excluding or flagging RR intervals related to VPBs and APBs can, as already pointed out, considerably improve the specificity of a detector. At the same time, ectopic beat handling must not alter the RR intervals which form an AF episode so that the sensitivity is lowered.

In many detectors, no explicit strategy is implemented for handling ectopic beats, but the parameters characterizing rhythm irregularity are fed directly to the classifier, see, e.g., [34, 36, 46–48]. When the ΔRR interval histogram constitutes the basis for detection, rhythms with frequent VPBs are sometimes falsely detected as AF when the Kolmogorov–Smirnov test is involved [31]. The source of the problem is the compensatory pause which accompanies most types of VPB, leading to a negative ΔRR interval immediately followed by a positive. Consequently, the histogram bears resemblance to a histogram determined in AF. It has been noted that the cumulative RR interval histogram determined from rhythms with frequent VPBs exhibits a “prominent shoulder” at around 400–600 ms, while the AF histogram usually does

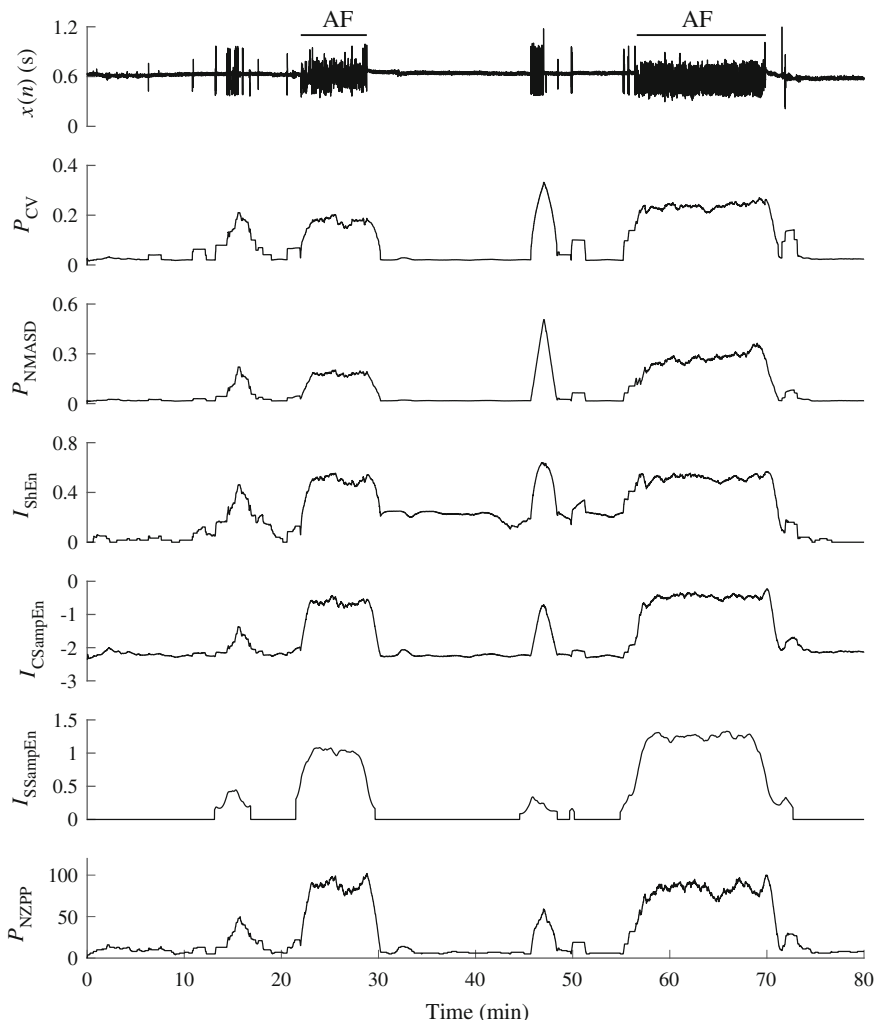


Fig. 4.7 An RR interval series $x(n)$ (top diagram) and related series of different detection parameters: coefficient of variation P_{CV} , normalized mean of absolute successive differences P_{NMAASD} , Shannon entropy I_{ShEn} (16 bins, $[0.2, 1.7]$ s, step 0.1 s), coefficient of sample entropy $I_{CSampEn}$ ($m = 1$, $r = 0.03$ s), simplified sample entropy $I_{SSampEn}$ ($r = 0.03$ s), and number of nonzero bins in the Poincaré plot P_{NZPP} (bin size 25 ms)

not [31]. Preliminary results showed that the number of VPB-related false alarms can be reduced by introducing a test on the height and width of a potential shoulder; however, no details have been provided on how to implement a test for identifying a prominent shoulder.

When the Poincaré plot is the starting point for computing a detection parameter, the bin population pattern may be considered for singling out ectopic beats. For

example, bigeminy is manifested by clustered points populating just a few bins [33, 51], whereas AF is manifested by points which are much more scattered. When the Poincaré plot is defined by $(x(n), x(n + 1))$, changes in heart rate within the detection window smears the clustered points related to ectopic beats, which in turn increases the number of false alarms; this problem is likely to be less pronounced when the plot is defined by $(\Delta x(n), \Delta x(n + 1))$.

One of the first rhythm-based detectors to involve handling of ectopic beat was described in [32], see also [37], embracing three different ratio series defined by successive RR intervals. In order to eliminate a VPB, preceded by a short RR interval $x(n)$ and followed by a compensatory pause $x(n + 1)$, the following three conditions need to be fulfilled for $x(n)$ and $x(n + 1)$ to be excluded from the RR interval series:

$$\frac{x(n)}{x(n - 1)} < \gamma_1, \quad (4.33)$$

$$\frac{x(n + 1)}{x(n)} > \gamma_{99}, \quad (4.34)$$

$$\frac{x(n + 1)}{x(n + 2)} > \gamma_{25}. \quad (4.35)$$

The thresholds γ_1 , γ_{25} , and γ_{99} denote the 1st, 25th, and 99th percentiles, respectively, of the RR interval ratio histogram of the current detection window. Obviously, these percentiles are increasingly difficult to determine with sufficient reliability as the window becomes shorter. The application of the conditions in (4.33)–(4.35) is illustrated in Fig. 4.8a and b for an RR interval series containing bigeminy and ectopic beats, and then followed by an AF episode. The ectopic beats are eliminated in the thinned output series, whereas the episode of bigeminy is characterized by much flattened RR intervals and reduced irregularity of AF.

Median filtering may be used to eliminate occasional ectopic beats from the RR interval series, while preserving the sharp changes that typically characterize the onset and end of an AF episode. Such filters have been implemented with lengths ranging from 3 [40] to 17 [38], where longer median filters offer better elimination of ectopic beats, but increases the risk of missed brief AF episodes. Therefore, bearing in mind the growing interest in detection of brief episodes, short median filters are to be preferred. Figure 4.8c and d illustrate how the RR interval series is altered when using 3- and 17-point median filters, respectively. The ectopic beats are eliminated in the filtered output, but the episode of bigeminy is largely unaltered and the irregularity of AF is much reduced, especially for the 17-point filter.

In addition to eliminating ectopic beats with median filtering, a set of ad hoc tests, similar to those in (4.33)–(4.35), have been suggested which are also based on the series of ratios of successive RR intervals [35]. The sequence of RR interval ratios is determined for common non-AF arrhythmias, e.g., bi- and trigeminy, and used to build a database with template patterns. The sequence of ratios inside the detection window is correlated to all the template patterns, and the presence of AF is ruled out whenever a sufficient number of correlation matches are found. In this approach,

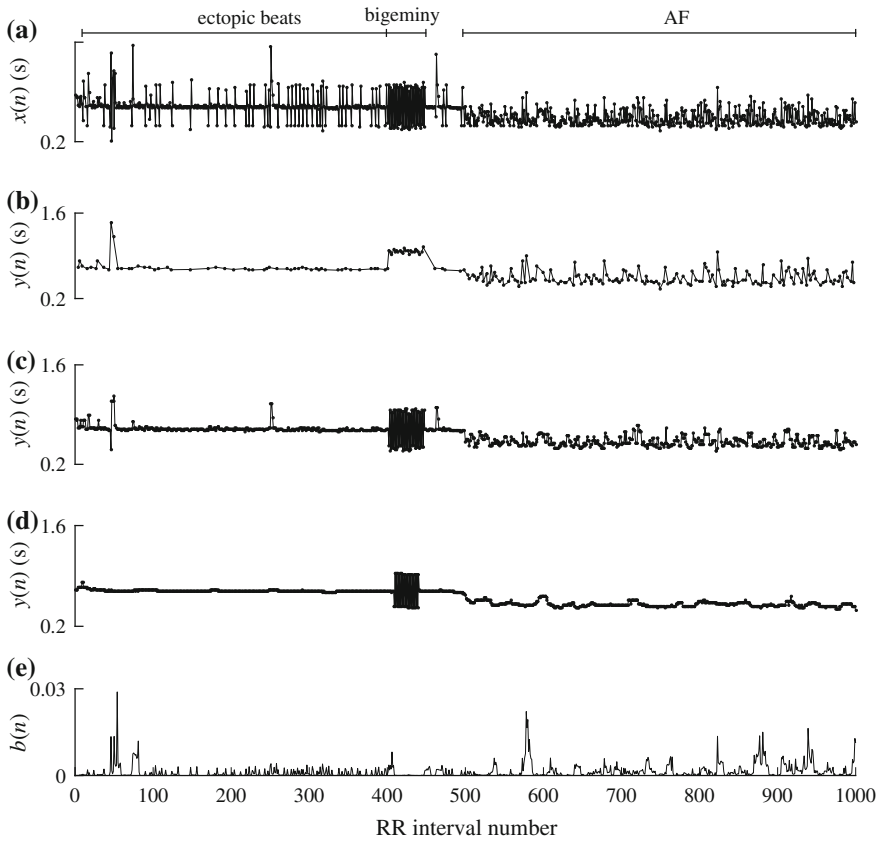


Fig. 4.8 **a** An RR interval series $x(n)$ containing ectopic beats and bigeminy, followed by an AF episode with onset at about interval #500. **b** The output $y(n)$ when applying the three conditions in (4.33)–(4.35) to $x(n)$. **c** The output $y(n)$ from 3-point median filtering and **d** 17-point median filtering. **e** The function $b(n)$ in (4.36), whose only purpose is to flag when bigeminy is present, is computed for $M = 8$; this function does not replace $x(n)$. It should be noted that the output samples in (b) are thinned in time compared to $x(n)$, whereas no thinning is introduced in (c)–(e)

ectopic beat handling is part of the classifier, since no processed RR interval sequence results. Several thresholds need to be set before the tests can be applied—settings whose influence on performance remain to be established.

A simple flag function has been proposed to indicate whether the observed rhythm is likely to be in AF, defined by [40]

$$b(n) = \left(\frac{\sum_{m=0}^{M-1} x_m(n-m)}{\sum_{m=0}^{M-1} x(n-m)} - 1 \right)^2, \quad n = M, \dots, N-1, \quad (4.36)$$

where n is the end time of the sliding detection window, M is an even-valued integer, and $x_m(n)$ is the output of a three-point median filter. For regular rhythms as well as for bigeminy, the ratio in (4.36) is approximately equal to 1 since $x_m(n)$ and $x(n)$ resemble each other; thus, $b(n)$ is approximately equal to 0. On the other hand, in AF, the variability in $x_m(n)$ is lower than that in $x(n)$ due to the median filtering, and, as a consequence, $b(n)$ increases to indicate AF presence. The squaring operation in (4.36) is introduced to improve the differentiation of AF from non-AF rhythms. In contrast to the criteria in (4.33)–(4.35), resulting in the exclusion of RR intervals, the purpose of $b(n)$ is to serve as a weighting function suitable for use in signal fusion. Figure 4.8e illustrates the behavior of $b(n)$ in the presence of an episode of bigeminy, being flagged by values close to zero.

Given that rhythm-based AF detection is the predominant mode of operation in mHealth monitoring devices and implantable loop recorders, further development of techniques for better handling ectopic beats is warranted.

4.2.6 Classification

The most common approach to designing a classifier is to simply apply one or several threshold tests to the parameters (“features”) selected for AF detection. Information on RR interval irregularity is often condensed into one single feature, see, e.g., [31, 33, 34, 38, 40], but as many as nine features, with nine accompanying threshold tests, have also been considered [35]. The threshold values can be determined by optimizing a suitable performance measure, e.g., the area under the receiver operating characteristic (ROC) (Sect. 4.5), with respect to the features of interest using a training data set. The optimized thresholds are then used to evaluate performance on a test data set. Alternatively, the determination of a threshold may be based on some underlying statistical assumptions associated with the feature [31].

When the classifier involves many features, the question arises whether a feature conveys unique information or correlates with the other features. If correlated, which is often the case, the features can be decorrelated using principal component analysis (PCA) so that only the most relevant features are retained, i.e., the dimensionality of the feature vector is reduced. It is well-known that low-dimensional feature vectors generalize better to data not presented during training, thereby leading to more robust detection performance [82]. Another obvious advantage is that fewer features imply less computations. Although feature selection has been considered in AF detection, then involving an improved version of the sequential forward floating selection algorithm [46], this approach has yet to find its way into AF detection on a broader scale.

A simple approach to understanding the relevance of individual features in multi-feature threshold testing is to establish their relative contribution to detection performance, for example, by determining the performance with and without a test involving a certain feature. Such an insight may help to render the detector structure more effective, of particular importance when the detector is aimed at implementation in a low power device. In rhythm-based AF detection, no study has yet reported on the significance of individual tests, whereas one study has presented results on rhythm and morphology based detection, demonstrating that rhythm irregularity plays a more significant role in detection [83].

Another, even simpler, approach to understanding the relevance of a feature is to determine the histograms of the feature for RR intervals observed in either AF or non-AF rhythms, using some suitable database [38, 40, 41, 49]. Then, the extent by which these two histograms overlap serves as a preliminary indication of the feature's discriminatory power. The histograms of different parameters, previously described in this chapter, are presented in Fig. 4.9. Using AFDB, the parameters are computed from the RR intervals contained in a sliding 128-beat window, except $I_{SSampEn}$ which is computed in a sliding 8-beat window. Visual inspection of Fig. 4.9 shows that the least histogram overlap is exhibited by $I_{SSampEn}$, and therefore this parameter is particularly well-suited for AF detection. Interestingly, the simple-structured feature P_{NZPP} , defined by the number of nonzero bins in the Poincaré plot, is also associated with a small overlap. On the other hand, I_{ShEn} is associated with the largest overlap, thus questioning its suitability for use in AF detection. When the Shannon entropy is computed from a symbolic sequence, determined either from the RR intervals or the instantaneous heart rate, the histogram overlap has been found to decrease, see [38, 41].

In addition to using a traditional classifier defined by a set of threshold tests, pattern classification techniques have been investigated for AF detection, including support vector machines (SVMs) [39, 50, 84] and linear discriminant analysis (LDA) [36]; the former technique has the advantage of offering better flexibility as the decision boundaries can be nonlinear [85]. In these studies, the dimension of the feature vector ranges from 2 to as large as 24. It should be noted that LDA-based classification requires many more computations for training than does simple threshold testing, as the sample mean vector and the covariance matrix for both non-AF and AF data are needed to compute the discriminant function. For SVM, only two design parameters need to be set, both related to the degree with which misclassifications should be penalized [50, 84].

From Table 4.1, it is evident that detection based on a single threshold test offers performance superior to detection based on a classifier incorporating multi-threshold tests or an SVM. For example, the single-test detector in [40] performs better than does the detector using an SVM [39]. At a first glance, this result may stand out as unexpected as an SVM offers so much more freedom with respect to the location of the decision boundaries, and therefore an SVM should perform better. A possible explanation to this result may be that the SVM does not generalize well from training to testing when a small or nonrepresentative training set has been used. A more likely

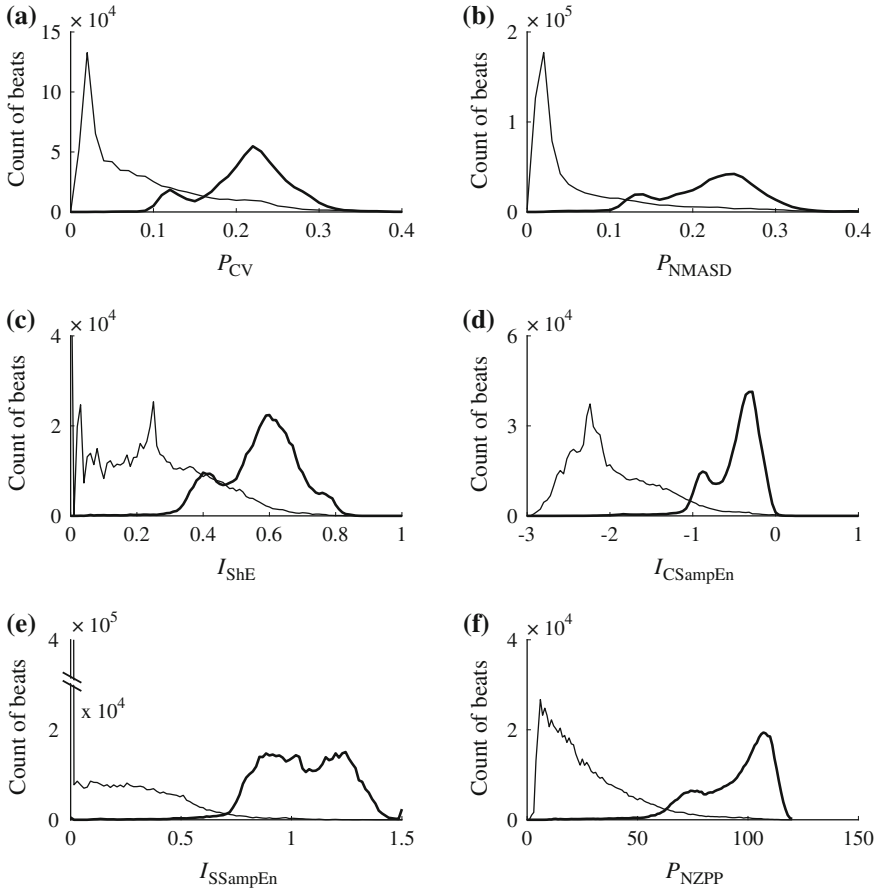


Fig. 4.9 Histograms for six different detection parameters, determined either in AF (thick line) or non-AF (thin line). **a** Coefficient of variation P_{CV} , **b** normalized mean of absolute successive differences P_{NMASD} , **c** Shannon entropy I_{ShE} , **d** coefficient of sample entropy $I_{CSampEn}$, **e** simplified sample entropy $I_{SSampEn}$, and **f** number of nonzero bins in the Poincaré plot P_{NZPP} . The values used to compute the parameter time series displayed in Fig. 4.7 were also used in this figure

explanation, though unrelated to the SVM, is that less powerful features were used, leading to inadequate handling of non-AF rhythms.

Detectors involving machine learning techniques have yet to demonstrate performance exceeding that of classical threshold-based AF detection. However, this relation may very well change in the future since databases for training are continuously growing—a change which implies time-consuming and meticulous work by expert cardiologists to ensure that the databases are adequately annotated.

None of the above-mentioned approaches to classification offer built-in immunity to non-AF rhythms such as bi- and trigeminy, frequent APBs and VPBs, supraventricular tachycardia, and atrioventricular junctional rhythms, and, therefore, ectopic beat

handling prior to classification will have significant repercussions on performance. This aspect is illustrated by considering the performance of the detector in [40] when implemented with and without such handling. In that detector, the fusion of I_{SSampEn} , computed using a sliding 8-beat window, and $b(n)$, indicating the likelihood of AF presence, results in a parameter which is subjected to simple thresholding. Using the MIT-BIH Normal Sinus Rhythm Database (NSRDB), containing several occurrences of bigeminy, cf. Sect. 3.1, the incorporation of $b(n)$ in the detector leads to a dramatic improvement in performance since the specificity increases from 93.2 to 98.6%, whereas the sensitivity remains essentially the same (this is a previously unpublished result).

4.3 Rhythm and Morphology Based AF Detection

Although AF is accompanied by changes in both rhythm and atrial wave morphology, rhythm-based detection continues to be the preferred mode of operation since the RR intervals can be determined much more reliably in noisy signals than information on atrial activity [38, 86]. Since rhythm-based detectors tend to produce false alarms in sinus rhythms with ectopic beats, complete atrioventricular block, as well as in patients with prescribed ventricular rate-controlling medication, it is natural to also analyze whether P waves are absent and/or f waves are present so that the false alarm rate can be reduced. Thus, information on atrial wave morphology needs to be included in the decision process, illustrated by the block diagram of an AF detector in Fig. 4.10a. While the performance of rhythm-based AF detectors is not critically dependent on the lead selected for signal processing, lead selection is crucial when morphologic information is involved since f waves have much lower amplitude in leads positioned farther away from the atria; such lead-dependence is less pronounced for P wave amplitude.

Only a handful of AF detectors have been designed in which information on both rhythm and atrial wave morphology are subject to analysis. The performance reported in the literature must be regarded as rather disappointing since, indeed, none of the detectors achieve performance superior to that of a well-performing rhythm-based detector, see Table 4.3. This result may be explained by the use of detector structures not accounting for the fact that the noise level usually changes over time. As a consequence, measurements characterizing atrial activity are not always reliable, but may actually contribute to worsen the performance rather than to improve it [87]. Hence, an important guiding design principle is to account for the prevailing noise level in the detector structure, implying that information on atrial activity becomes less influential when decisions are made at higher noise levels, and vice versa. Ultimately, when the noise level exceeds a certain threshold, the detector structure should simplify to one based on only the RR interval series, cf. Sect. 4.2. Pursuing the design of a detector accounting for noise calls for the development of a noise level estimator. The noise-dependent mode of operation of an AF detector is described by the block diagram in Fig. 4.10b.

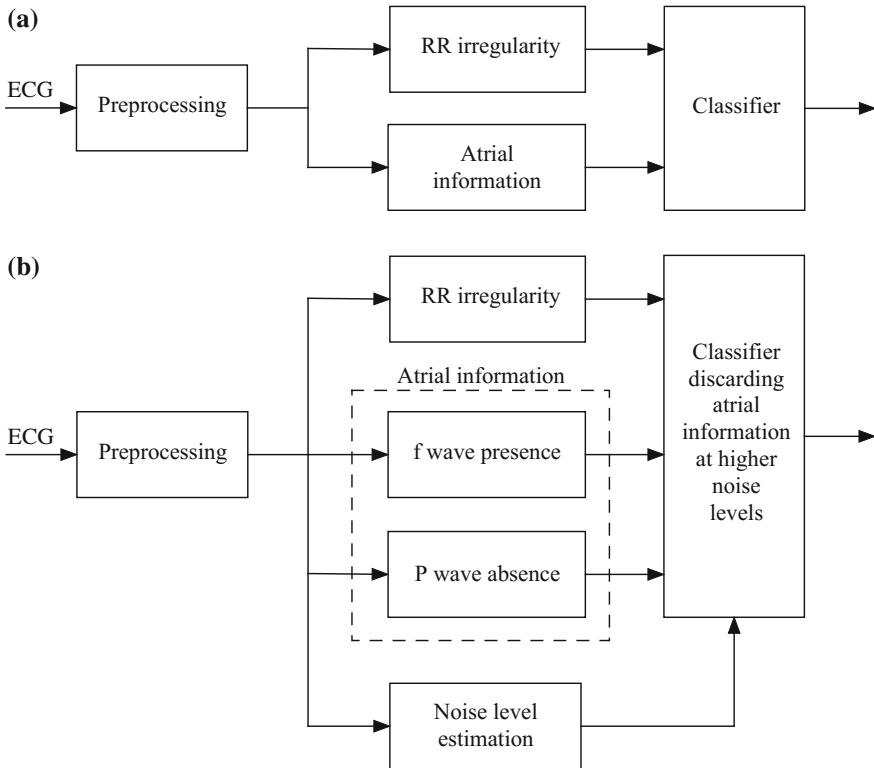


Fig. 4.10 General structure of AF detectors described in the literature. **a** Block diagram of a detector exploring atrial morphology independently of the prevailing noise level. **b** Block diagram of a detector whose classifier is designed to increasingly discard information on atrial wave morphology as the noise level increases

This section provides an overview of the building blocks required for processing information on atrial wave morphology, as well as for estimating the noise level. Some detectors explore information on either P waves or f waves, while others explore both types of waves.

4.3.1 P Wave Detection Information

The problem of P wave detection/delineation has been thoroughly treated in the literature, with emphasis on automated interpretation of diagnostic ECGs where highly accurate measurements of P wave amplitude and duration are of critical importance [91–93]. The prediction of patients prone to AF based on P wave morphology represents another, more immediate application where accurate measurements are

Table 4.3 The performance of five detectors based on both rhythm and morphology, together with the performance figures of rhythm-based detection already presented in Table 4.1. The subset AFDB₁ is defined in Table 4.1, AFDB₂ is identical to AFDB, except that records 00735 and 03665 are excluded since they do not include ECG signals, only RR interval information, AFDB₃ contains only 20 of the 25 records since five records do not have sufficient sinus rhythm data for training, and AFDB₄ excludes a huge number of unspecified non-AF segments to balance the sizes of AF and non-AF records. The difficulties associated with comparing detection performance are considered in Sect. 4.6, applying especially to the best-performing detector

Method by	Year	Database	Se (%)	Sp (%)
Dash et al. [32]	2009	AFDB ₁	94.4	95.1
Lian et al. [33]	2011	AFDB	95.8	96.4
Lake and Moorman [34]	2011	AFDB	91	94
Huang et al. [35]	2011	AFDB	96.1	98.1
Shouldice et al. [36]	2012	AFDB	92	96
Lee et al. [37]	2013	AFDB ₁	98.2	97.7
Zhou et al. [38]	2014	AFDB	96.9	98.3
Asgari et al. [39]	2015	AFDB	97.0	97.1
Petrėnas et al. [40]	2015	AFDB	97.1	98.3
Zhou et al. [41]	2015	AFDB	97.4	98.4
Babaeizadeh et al. [87]	2009	AFDB ₂	93	98
Carvalho et al. [83]	2012	AFDB ₂	93.8	96.1
Ladavich and Ghoraani [88]	2015	AFDB ₃	98.1	91.7
Ródenas et al. [89]	2015	AFDB ₂	96.5	94.2
Xia et al. [90]	2018	AFDB ₄	98.3	98.2

essential [94–96]. In AF detection, however, the demands on accuracy are more relaxed since the absence of P waves can be established without first having to estimate P wave onset and end.

A straightforward approach to determining whether P waves are absent is to use a measure reflecting morphologic similarity between the samples in two consecutive “PR intervals”, with the correlation coefficient and the mean square difference as examples of such a measure [87]. In sinus rhythm, P wave morphology is usually stable from one beat to the next, and, therefore, such a measure would indicate a high degree of similarity. In AF, on the other hand, P waves are replaced with f waves which are unsynchronized with the QRS complexes, and, consequently, the degree of similarity between two PR intervals is much lower. Once pairwise comparison has been performed for all beats in the detection window, the average of the resulting similarity measurements can be compared to a threshold to determine whether P waves are absent.

In a related approach, the samples of the PR interval are correlated to the samples of a fixed P wave template [83]. The template is determined by averaging all annotated P waves of a huge annotated database [97, 98]; further considerations on template-based P wave detection can be found in [99]. By analyzing the sequence of correlation

coefficients determined from all beats in the detection window, a P wave is detected whenever the correlation coefficient exceeds a fixed threshold. P waves are considered absent when the P wave occurrence ratio, defined as the number of detected P waves to the total number of beats in the window, falls below another fixed threshold.

Rather than quantifying P wave absence directly in the ECG signal, as is usually the case, it can be quantified in a signal resulting from PQRST cancellation of the ECG, thus composed of PQRST-related residuals in normal sinus rhythm and f waves in AF [100]. In this approach, the term “P wave absence” has a different meaning since the input signal no longer contains P waves; however, the term is still useful since an “imaginary” PR interval can be analyzed. It has been shown that PQRST cancellation can be accomplished by means of an echo state network which offers the advantage of handling substantial variation in normal beat morphology as well as the presence of ectopic beats [101]; for a description of the echo state network, see Sect. 5.5.3. In the canceled signal, all possible pairwise combinations of the PR intervals are considered in the detection window, not just the pairs defined by consecutive PR intervals as in [87]. The squared error is computed for pairs of PR intervals, and then averaged over all possible combinations to produce a measure of P wave absence. The PR interval has a fixed location relative to the fiducial point of the QRS complex, with its onset and end preceding the fiducial point by 250 and 50 ms, respectively.

A radically different approach to AF detection is to completely leave out all rhythm information and only explore whether P waves are absent [88, 89].⁷ The main motivation for pursuing this approach is that rhythm information may not be discriminative enough to reliably detect AF in patients on rate-controlled medication or with pacemaker, where rhythm irregularity is reduced. It is obvious from Table 4.3 that these two detectors have performance inferior to the best-performing rhythm-based detectors.

As many as nine features have been employed for describing different P wave properties: six features describing P wave amplitude in contiguous 20 ms intervals, and three features describing variance, skewness, and kurtosis of the samples in the PR interval, located, as above, at a fixed distance from the QRS fiducial point [88]. In contrast to the three above-mentioned approaches, which all produce a simple scalar parameter for determining P wave absence, this approach is considerably more complicated as a training phase is required for each patient before AF detection can take place. This phase involves a Gaussian mixture model whose model parameters have to be determined from a half hour long ECG segment containing sinus rhythm; each P wave is represented by the nine-dimensional feature vector. In the testing phase, the Mahalanobi distance between the features of the candidate P wave and the features of the patient-specific P wave model is computed, indicating P wave absence when the distance is sufficiently large.⁸

⁷Strictly speaking, this type of detector does not explore both rhythm and morphology. However, since information on P wave absence is still required, the detector is described in this section.

⁸The idea of studying the deviation from “normality”, i.e., whether a P wave is absent, is closely related to the concept of novelty detection [102, 103].

The entropy of different scales, *wavelet entropy*, constitutes a set of features explored in AF detection [89]. In order to compute the wavelet entropy, the samples in the TQ interval are first subject to wavelet decomposition [21], resulting in the wavelet coefficients $w_{i,k}$, where i and k denote scale and time, respectively. Then, the relative energy E_i is computed for each scale,

$$E_i = \frac{\sum_{k=0}^{K_i-1} w_{i,k}^2}{\sum_{l=1}^J \sum_{k=0}^{K_l-1} w_{l,k}^2}, \quad i = 1, \dots, J, \quad (4.37)$$

where J denotes the number of scales, and K_l denotes the length of $w_{l,k}$ at scale l . The wavelet entropy is obtained as the Shannon entropy of E_i , cf. (4.7), except that the probabilities $p(x_i)$ are replaced by the relative energies E_i , which, by definition, sum to 1. Statistical analysis of AFDB showed that TQ intervals with P waves were associated with significantly lower wavelet entropies than TQ intervals with f waves. This finding is due to that the relative energy is much more concentrated to one scale for P waves than for f waves.

The variability of the length of the PR interval may serve as an indirect measure of P wave absence [87]. Obviously, this length can only be determined when a P wave is present, requiring that the onset of the P wave and the onset of the QRS complex have first been determined. While PR interval variability is undefined in AF, a surrogate measure may be used in which the onset of an f wave is treated as the onset of a P wave, leading to a PR interval variability which is much larger in AF than in normal sinus rhythm. Considering the imprecise definition of PR interval variability in AF, it is doubtful whether this measure is sufficiently powerful for AF detection.

The above-mentioned techniques for determining P wave absence vary quite substantially in complexity, ranging from simple similarity measures to advanced, statistical modeling of P waves. When a similarity measure is computed between the samples of two PR intervals, e.g., the correlation coefficient or the mean square difference, no particular polarity or morphology of the P wave is favored. This is an important advantage when the objective is to quantify a rather unspecific concept such as ‘‘P wave absence.’’ On the other hand, a template-based similarity measure can be expected to perform less well in rhythms with varying P wave morphology, but also for morphologies which are approximately orthogonal (in mathematical terms) to the template, i.e., the correlation coefficient is approximately zero although a P wave is present. Statistical modeling of P wave properties offers more degrees of freedom than the template-based approach, however, such modeling also requires training in each patient on lengthy data which have to be recorded in sinus rhythm; such data is not always available.

It is obvious that information on P wave absence becomes increasingly unreliable as the noise level increases, eventually reaching a ‘‘breakdown’’ level that differs from one technique to another depending on the robustness of the design. For example,

a template-based approach is likely more robust to noise than an approach where P wave onset needs to be determined. In addition, information on P wave absence is more reliable when extracted from more than one lead: by analyzing two leads instead of one, the specificity of a P wave based AF detector has been shown to increase from 91.7 to 94.6%, while the sensitivity remained essentially the same [88].

4.3.2 *f* Wave Detection Information

The sparse use of *f* wave information in AF detection is due to the difficulty to reliably characterize low amplitude *f* waves in the presence of noise, as well as to reliably determine *f* wave characteristics from the TQ interval. Not only is it challenging to determine the endpoint of the T wave in AF, but the TQ interval becomes increasingly shorter as the heart rate increases. Eventually, the TQ interval may have shrunk to such an extent that the *f* waves are completely concealed by ventricular activity, thus precluding further analysis. This problem can, however, be addressed by means of *f* wave extraction—a signal processing operation which is thoroughly reviewed in Chap. 5. While *f* wave extraction facilitates AF detection, it also increases the complexity of the detector structure so that it may no longer be feasible to implement in a battery-powered device.

Basal time domain information on *f* wave presence can be obtained by counting the number of *f* waves in the TQ interval, with *f* waves considered present whenever the count exceeds one, otherwise absent [104]. The width of a signal fluctuation must exceed a certain threshold to be counted as an *f* wave; in [104], *f* wave width is defined as the time elapsed between two level crossings. In order to avoid that noise fluctuations are counted, the amplitude of a fluctuation must exceed an adaptive threshold related to both the amplitude of the TQ interval and the peak amplitude of the T wave. Another means to combat false counts of *f* waves is to first bandpass filter the observed signal so that baseline wander and noise of muscular origin are reduced. However, even with such filtering, it is well-known that *f* wave analysis relying on level crossing patterns remains vulnerable to noise since the spectral content of filtered muscle noise overlaps with that of *f* waves [21]. The consequences of a vanishing TQ interval at higher heart rates, i.e., a count of zero *f* waves, was not addressed in [104].

Spectral characterization is another approach to determining *f* wave presence, assuming that *f* wave extraction is first performed so that all samples in the detection window are suitable for spectral analysis, not just samples in the TQ interval [83, 100]. Since the spectral peak corresponding to the *f* wave repetition rate (dominant AF frequency, DAF) is typically the largest, parameters describing signal bandwidth have been proposed as a measure of *f* wave presence. Figure 4.11a illustrates the spectrum of an extracted *f* wave signal. In this example, the DAF, located at 6 Hz, is the main spectral feature, but important information may also be conveyed by

the second and third harmonics, see Sect. 6.3.2. In general, two or more harmonics are more likely to be present in patients with paroxysmal AF than in patients with permanent AF.

The *normalized spectral concentration* is defined by [100], see also [105, 106],

$$F_{\text{SC}} = \int_{\Omega_a} P'_d(\omega) d\omega, \quad (4.38)$$

where $P'_d(\omega)$ denotes the normalized power spectrum of the extracted f wave signal $\hat{d}(n)$, defined by

$$P'_d(\omega) = \frac{1}{\sigma_d^2} P_d(\omega), \quad (4.39)$$

and σ_d^2 the variance of $\hat{d}(n)$. The integration interval Ω_a is centered around the dominant spectral peak located within the interval $[\omega_{a,0}, \omega_{a,1}]$, usually chosen to be [4, 12] Hz. When f waves are present, the spectral concentration is closer to 1, whereas it is closer to 0 when sinus rhythm is present. The power spectrum $P_d(\omega)$ may be estimated using a nonparametric technique, e.g., Welch's method, or a parametric technique, e.g., Burg's method [107].

Spectral entropy is another parameter used for determining f wave presence [83], defined by

$$F_{\text{SE}} = - \int_{\Omega_a} P'_d(\omega) \ln(P'_d(\omega)) d\omega. \quad (4.40)$$

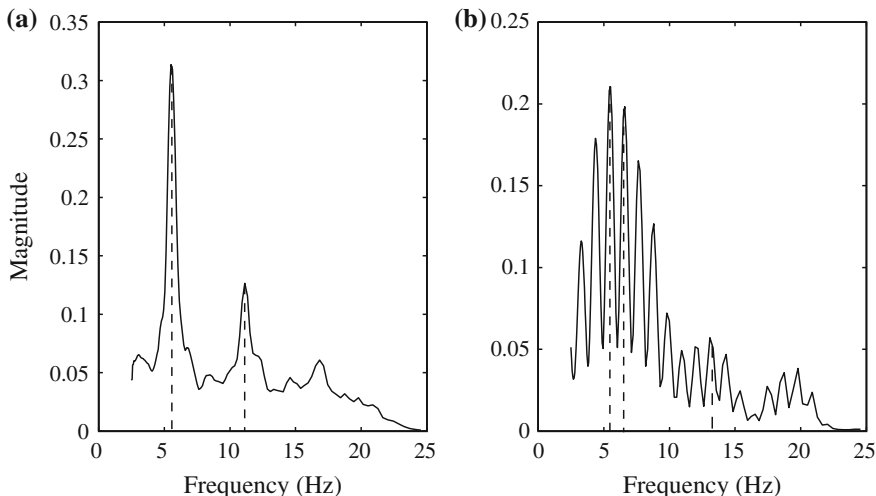


Fig. 4.11 The power spectrum of **a** an extracted f wave signal, and **b** a QRST-cancelled signal observed in sinus rhythm. The two largest spectral peaks are indicated with vertical lines

As the bandwidth of $P'_d(\omega)$ becomes increasingly narrower in Ω_a , and thus more likely to reflect AF, the spectral entropy becomes increasingly smaller.

Unlike F_{SE} , the *Kullberg–Leibler divergence*, also known as *relative spectral entropy*, accounts for the similarity between $P'_d(\omega)$ and a template power spectrum $P'_t(\omega)$ [83], defined by

$$F_{KL} = \int_{\Omega_a} P'_d(\omega) \ln \left(\frac{P'_d(\omega)}{P'_t(\omega)} \right) d\omega. \quad (4.41)$$

Ideally, the template power spectrum $P'_t(\omega)$ should be determined so that it is representative of f waves for all patients, e.g., by computing a gross power spectrum from a huge database with high quality f waves. However, not only varies the DAF substantially from patient to patient, but so does f wave morphology. As a result, the practical utility of a template power spectrum is limited, and the information on f wave presence conveyed by F_{KL} can hardly be viewed as representative. In [83], $P'_t(\omega)$ was determined from AFDB and used, in combination with F_{SE} , to decide whether f waves are present. The dominant peak of $P'_t(\omega)$ was found to be located at about 2 Hz, which is far below the expected range of the DAF.

Though not developed specifically for determining f wave presence in AF detection, a set of simple threshold tests have been proposed for judging whether the structure of $P'_d(\omega)$ relates to AF [108]. The tests involve the following ad hoc spectral parameters:

- The SNR, where “signal” is defined as the mean of the magnitudes of the first and second harmonics, and “noise” as the magnitude halfway between the two harmonics.
- The deviation of the second largest peak in $P'_d(\omega)$ from the expected position of the second harmonic, aiming at excluding signal segments with a “ringing” spectrum, e.g., due to P waves occurring at slow rates.
- The ratio between the magnitudes of the second largest and the largest peak in $P'_d(\omega)$, detecting when the second harmonic is too large.
- The squared error between the spectrum of the sliding window and an exponentially averaged spectrum based on past signal segments not containing muscle noise or residuals due to poor f wave extraction.

The spectrum in Fig. 4.11a fulfills the above four tests to be considered an AF spectrum, whereas the spectrum in Fig. 4.11b does not; the test outcome is correct in both cases.

The additional value of including information on atrial wave morphology in AF detection is illustrated in Fig. 4.12, where ECGs with either several APBs or respiratory sinus arrhythmia are analyzed. Using the fuzzy logic detector in [100] which processes information on P wave absence and f wave presence, none of the two non-AF rhythms are detected as AF, whereas both are falsely detected as AF when the coefficient of sample entropy $I_{CSampEn}$ of the RR intervals is used as detection parameter [34]. The decision functions of the two detectors are displayed in Fig. 4.12.

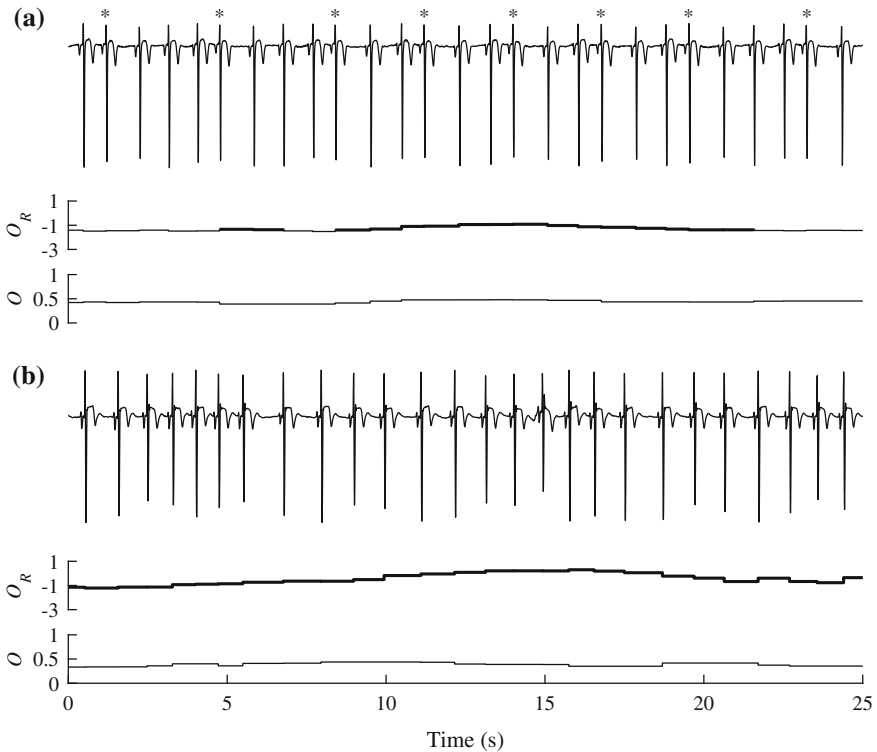


Fig. 4.12 Non-AF arrhythmias causing false alarms in rhythm-based detection, but not in rhythm and morphology based detection: **a** Frequent atrial premature beats (marked with “*”), and **b** respiratory sinus arrhythmia. Atrial fibrillation is detected (thicker line) whenever the decision function, denoted O_R for rhythm-based detection [34] and O for rhythm and morphology based detection [100], exceeds the detection threshold

4.3.3 Noise Level Estimation

Although an AF detector must operate at highly varying noise levels, remarkably little attention has been paid to the problem of how to adjusting detector operation relative to such variation. Rather, the observed ECG signal is processed in the same way, irrespective of the prevailing noise level [83, 87, 88]. One explanation to this structural omission may be related to the challenge of how to integrate noise information into the classifier so that information on atrial wave morphology becomes increasingly discarded as the noise level increases, see Fig. 4.10. Another, more fundamental explanation may be related to the development of the noise level estimator itself, which should be designed so that the estimate actually reflects the noise level, but not the cardiac activity.

One of the very few AF detectors operating in a noise-dependent mode was proposed in [100]. In that detector, the extracted f wave signal $\hat{d}(n)$, produced by an echo state network, serves as the starting point for estimating the noise level. The estimator is defined by the root mean square value $R_{\hat{d}}$ of $\hat{d}(n)$, weighted by a ratio of spectral entropies:

$$\hat{N}_{\text{WRMS}} = R_{\hat{d}} \cdot \frac{\int_{\Omega_n} P_{\hat{d}}(\omega) \log_2 P_{\hat{d}}(\omega) d\omega}{\int_{\Omega_a} P_{\hat{d}}(\omega) \log_2 P_{\hat{d}}(\omega) d\omega}. \tag{4.42}$$

The numerator is computed in a spectral band dominated by noise, defined by $\Omega_n \in [\omega_{n,0}, \omega_{n,1}]$, and the denominator in a spectral band dominated by f waves, cf. (4.38). The definitions of spectral entropy in (4.40) and (4.42) differ with respect to the logarithm—a difference with little importance from a practical viewpoint. The estimator \hat{N}_{WRMS} produces smaller values when $P_{\hat{d}}(\omega)$ reflects the presence of f waves, but larger values when muscle noise and motion artifacts are present. Figure 4.13 illustrates the estimation of noise level, demonstrating that the estimate tracks the changes in noise level during the last 15 s, while it remains uninfluenced by the f waves of the first AF episode.

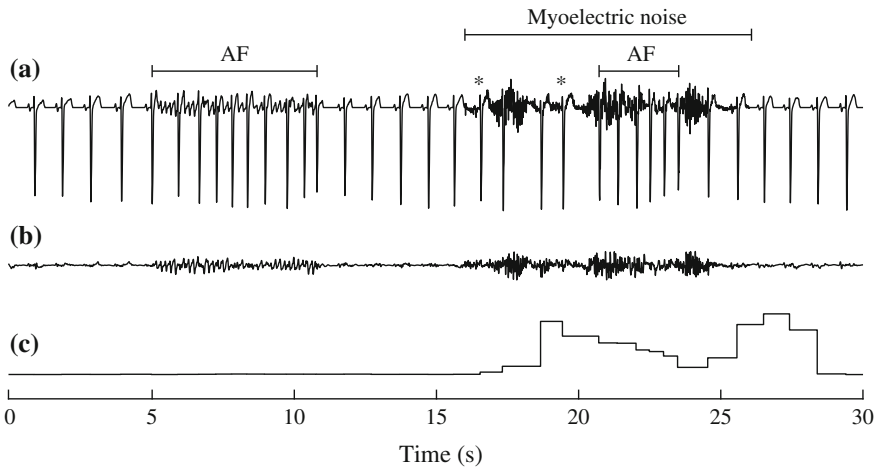


Fig. 4.13 Noise level estimation based on (4.42). **a** The first 15 s of the signal are noise-free, then followed by a 10 s burst of myoelectric noise. The second AF episode is preceded by two atrial premature beats (marked with “*”). **b** f wave signal extracted using an echo state network. **c** The noise level estimate \hat{N} , defined in (4.42), is delayed due to that it is computed in a sliding 5-beat window

The wavelet entropy of the samples in the TQ interval can, in addition to quantifying P wave absence (Sect. 4.3.1), be used as a noise level estimator. While the energy of P waves is mostly confined to one scale, the noise energy is more evenly distributed across the different scales, implying that noise is associated with higher wavelet entropy than P waves. It should be emphasized that the wavelet entropy measures signal organization, and, therefore, contrary to the estimator in (4.42), not proportional to noise level.

If the purpose of the noise level estimator is instead to provide information on whether the RR interval sequence can be reliably analyzed for AF detection, other approaches to noise level estimation may be considered [109–115]. For example, the noise level can be associated with the differences in output from two different QRS detectors, where one is tuned to be more sensitive to noise than the other; large differences in QRS detection then represents an indirect measure of a high noise level [111]. Thus, this type of signal quality index does not have to be integrated into the classifier of the AF detector, but can be treated as independent information indicating whether the samples in the detection window should be processed [114]. Given that signal quality assessment is essential for f wave characterization, it is further considered in Sect. 6.5.

4.3.4 *Ectopic Beat Handling*

Detectors which process information on both rhythm and morphology offer indirect handling of ectopic beats, either through the analysis of P wave absence [87, 88] or the analysis of P wave absence in combination with f wave presence [100]. None of these detectors implement any of the techniques for ectopic beat handling previously described in Sect. 4.2.5 for rhythm-based AF detection. When detection is confined to analysis of P wave absence, the number of false detections due to frequent APBs can be considerably reduced since an APB is preceded by a P wave, on condition that the detector can cope with P wave morphologies that differ from the dominant morphology in normal sinus rhythm [100]. A complication arises, however, when APB prematurity is so pronounced that the P wave is hidden in the preceding T wave, thereby increasing the risk of falsely detecting frequent APBs as AF. In addition, frequent VPBs increase the risk of false detections since VPBs are not preceded by a P wave. Despite these complications, detectors using information on both P wave absence and f wave presence are likely to perform better in ectopic rhythms than would a rhythm-based detector.

If AF detection is implemented in a system for automated ECG analysis, whether for resting or continuous long-term recordings, classification of beat morphology is a built-in functionality which may be utilized for excluding segments with VPBs before AF detection is performed. Such exclusion can also be based on beat classification

performed jointly with AF detection [83]. Alternatively, the output of the built-in beat morphology classifier can be used to augment the feature vector created for AF detection.⁹

4.3.5 Classification

The considerations concerning classification in rhythm-based AF detection earlier discussed in Sect. 4.2.6 are equally valid for detection based on both rhythm and atrial wave morphology. With morphologic information included in the feature vector, the noise level should also be included so that the reliability of the parameters describing P wave absence and f wave presence can be assessed by the classifier. However, such an approach has not yet permeated the design of detectors, but classifiers are rather trained on data with considerable variation in noise level, with the objective to produce a fixed classifier suitable for use on data with both low and high noise levels.

One of the very first rhythm and morphology based detectors was described in [87], where the decisions were based on a feature vector composed of one rhythm parameter (the transition probability matrix of a stationary first-order Markov process [42]) and two P wave related parameters (P wave similarity and PR interval variability), see Sect. 4.3.1. A regression decision tree technique was considered for classification, implemented as a series of simple threshold tests, without involving any assumptions on the statistical distribution of the features.

In order to classify more accurately the nine P wave amplitude features described in Sect. 4.3.1, a multivariate mixture model was introduced in [88]. In this model, the features are characterized by a PDF defined as a sum of Gaussians, where each Gaussian is defined by its mean vector and covariance matrix. The model parameters, as well as the number of Gaussians in the sum, are determined by the expectation-maximization algorithm, requiring that a patient-specific training phase is first performed [85]. Once the statistical model has been identified, the likelihood of P wave absence is evaluated for each beat in the detection window. Based on the combined likelihood for all beats in the window, a decision is taken whether an AF episode is present.

The first detector architecture to offer joint processing of features describing rhythm irregularity, P wave absence, as well as f wave presence, was proposed in [83, 121]. A feedforward artificial neural network (ANN) was used as classifier, trained on a subset of records from AFDB.

A comparison of the performance figures listed in Table 4.3 is unfortunately not straightforward since both sensitivity and specificity differ from detector to detector. Nonetheless, the performance figures clearly indicate that detectors based on both

⁹While classification of beat morphology is not reviewed here, it deserves to be mentioned that this classification problem has received, and continues to receive, considerable attention in the literature, see, e.g., [116–120].

rhythm and morphology do not offer performance superior to that of rhythm-based detectors. In fact, the much earlier presented rhythm-based detector in [31] offers better performance than does the detector in [83], where account is made of both P wave and f wave information. This, rather disappointing result may be explained by the use of decision boundaries not adjusted in relation to the prevailing noise level. Interestingly, the authors of [83, 87–89] all point out noise as an important source to performance degradation of their respective detectors, although none of the detectors were designed to account for noise.

In fact, few of the above-mentioned detectors have a structure which lends itself to the handling of noise information. For example, it is unclear how an ANN-based classifier trained on signals with low noise levels generalizes to signals with higher levels. This observation is likely to apply also to classifiers based on a regression decision tree or a Gaussian mixture model.¹⁰

The first AF detector to account for information on noise level was proposed in [100], having a structure which agrees with that displayed in Fig. 4.10b. The information fed to the classifier consists of four different parameters, describing rhythm irregularity, P wave absence, f wave presence, and noise level as defined by (4.42). The classification is based on a Mamdani-type fuzzy logic in which the four input parameter values are mapped by a membership function to indicate the degree of belonging to a certain fuzzy set. For the parameters describing rhythm irregularity, P wave absence, f wave presence, the fuzzy sets relate to sinus rhythm and AF, whereas the fuzzy set relates to low level and high level for the noise parameter. The fuzzified parameter values are then combined using a set of fuzzy if–then rules, producing an output between 0 and 1 reflecting the likelihood that the detection window contains AF. With simplicity as the guiding star, the fuzzy rules are defined such that more weight is assigned to rhythm irregularity, and less weight to P wave absence and f wave presence, when the noise level is high, and vice versa when the noise level is low [100]. An AF episode is detected whenever the output exceeds a fixed threshold, which, for the example presented in Fig. 4.14 as well as for the overall detector evaluation, was simply set to 0.5.

An important advantage with the fuzzy logic classifier is that no training phase is required. On the other hand, the membership functions and fuzzy rules are defined by a large number of parameters which need to be set to reflect basic knowledge on AF. It should be noted that detector in [100] has not been subject to performance evaluation on AFDB since the method for f wave extraction requires a reference lead with negligible atrial waves which is not available in all recordings of that database.

Another approach to noise-dependent classification is to simply exclude beats whose noise level exceeds a certain fixed threshold [89]. The noise threshold is chosen so that the agreement with manual annotation of noisy beats is optimized. In noisy ECG segments, detector operation is suspended as information on P wave

¹⁰Neither is noise level taken into account in classification of beat morphology, even though it is well-known that certain beats are difficult to cluster due to excessive noise. This problem was indirectly addressed in [120], where an elegant technique based on switching Kalman filters was proposed for detecting “strange” beat morphologies falling outside the well-established clusters.

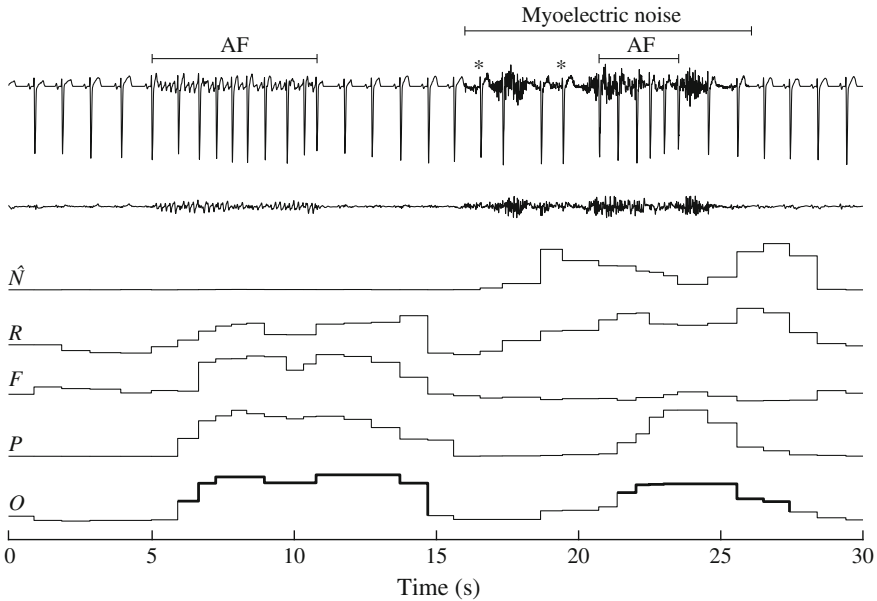


Fig. 4.14 Rhythm and morphology based AF detection using a fuzzy logic classifier. The example in Fig. 4.13 is here extended to also include trends on rhythm irregularity (R), f wave presence (F), and P wave absence (P)

absence cannot be determined. This property stands in contrast to the detector in [100] which continues to operate at higher noise levels, but then “resorting” to information on rhythm irregularity.

It should be noted that the most recent rhythm and morphology based detector listed in Table 4.3 offers slightly better performance than do any of the other detectors. This detector is based on a deep convolution neural network whose input is either the short-term Fourier transform (STFT) or the stationary wavelet transform of consecutive 5 s segments of the ECG signal, i.e., the input signal contains both atrial and ventricular activity [90]. Thus, the design of the detector is not driven by physiology—none of the three properties mentioned in the beginning of this chapter are taken into consideration—but emphasis is given to general ECG properties as well as nonphysiological aspects such as whether color or greyscale should be used to represent the STFT. While this approach to AF detection has potential, the performance figures must be called into question for reasons related to the use of a subset of AFDB in combination with tenfold cross-validation, further discussed in Sect. 4.6.

4.4 Implementation Aspects

When AF detection is to be implemented in a battery-powered, portable device, aspects such as computationally efficient algorithms and minimized memory usage are essential to ensure so that the device can operate continuously over an extended period of time. These requirements become even more crucial when AF detection is to be implemented in an implantable device, for example, a loop recorder. However, details on detector implementation are sparse in the literature, and those which have been published apply to rhythm-based detection where the input data, i.e., the RR series, has a very low rate, thus requiring few computations. On the other hand, for detectors exploring both rhythm and morphology, the input data rate is dramatically higher since the analysis of atrial wave morphology requires that the original ECG samples are available.

Thus, the amount of computations differs vastly between AF detectors, ranging from the simple rhythm-based detector using bin counts of the RR-based Poincaré plot to make decisions [33] to the detector using an echo state network for f wave extraction and fuzzy logic for decision-making [100]. The former detector can be implemented without multiplications, whereas the latter detector requires a huge amount of floating point multiplications as well as much memory to implement the different processing steps. Detailed information on the required amount of computations and memory is lacking for most detectors, with the exception of the rhythm-based detector exploring the combination of symbolic dynamics and the Shannon entropy as detection principle [41]. The computational complexity is analyzed by determining the number of arithmetic operations, shifts, and conditional expressions required per RR interval. Another, much more sweeping approach is to determine the time required by the central processing unit (CPU) and the amount of memory consumed during AF detection [122]. However, figures on CPU time and memory consumption are heavily system-dependent, and, therefore, it is difficult to make a fair comparison to the figures reported in other studies.

Hardware implementation of an invasive AF detector not only must consider requirements on computational complexity, but also energy dissipation when operating in idle and active mode. Idle energy is dominated by the leakage drawn by the memory retaining data, and active energy is minimized by reducing computational complexity. For a rhythm-based AF detector, with its low input data rate, minimization of computational complexity may, in fact, turn out to be less of a concern than minimization of required memory.

The rhythm-based detector in [32], using the number of turning points N_{TP} , the root mean square of successive differences P_{RMSSD} , and the Shannon entropy I_{ShEn} as parameters for characterizing the RR interval series, has been implemented in hardware, resulting in a fabricated application-specific integrated circuit (ASIC) optimized for ultra-low voltage operation [123]. The main reason for choosing the detector in [32] for implementation was that no storage of data was required for online training. It was demonstrated that the three parameters can be efficiently implemented thanks to that resource sharing of arithmetic units reduces the require-

ments of memory capacity, and that time multiplexing efficiently implements the arithmetic operations required to evaluate the conditions in (4.33)–(4.35) to remove VPBs. A potential AF episode is detected when all three threshold tests are fulfilled, each test involving one parameter. Rather than computing all three parameters first, only the parameter with the lowest cost from an energy consumption perspective is computed and tested. If the test is not fulfilled, the computation of the other parameters is unnecessary, and so on; N_{TP} was found to be the parameter with the lowest cost. The results suggested that the energy required to operate the detector for several years is well within what is provided by the battery of an implantable device [123].

4.5 Performance Measures

The predominant approach to quantifying detection performance is to compare the labels of the detected beats to those of the annotated beats contained in the database—the labels being either AF or non-AF. Such a comparison results in the following four counts,

N_{TP} = #beats in AF correctly detected as AF (true positive),

N_{TN} = #beats in non-AF correctly detected as non-AF (true negative),

N_{FP} = #beats in non-AF falsely detected as AF (false positive),

N_{FN} = #beats in AF falsely detected as non-AF (false negative),

which are required for computing the two most commonly used performance measures,

$$\text{Sensitivity} = \frac{N_{TP}}{N_{TP} + N_{FN}}, \quad (4.43)$$

$$\text{Specificity} = \frac{N_{TN}}{N_{FP} + N_{TN}}. \quad (4.44)$$

Performance is often studied by displaying sensitivity versus (1–specificity) for different values of a detection threshold, resulting in the ROC [34, 37]. From this curve, the threshold value achieving the desired trade-off between sensitivity and specificity can be chosen. The ROC is sometimes condensed into an overall, scalar measure defined as the area under the curve (AUC), where an area of 1 represents perfect performance and an area of 0.5 random performance. The AUC is considered a robust performance measure because all possible detection thresholds are involved. In AF detection, certain parameter values have been determined by maximizing the AUC [39, 41, 51, 88].

In addition, the following measures have been employed to describe detection performance:

$$\text{Positive predictive value} = \frac{N_{TP}}{N_{TP} + N_{FP}}, \quad (4.45)$$

$$\text{Detection accuracy} = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{FN} + N_{FP} + N_{TN}}. \quad (4.46)$$

It should be noted that detection accuracy should only be used when the two classes AF and non-AF have approximately the same size. Otherwise, Matthews correlation coefficient may be a better choice to evaluate the performance of binary classifiers such as the ones used in AF detection [124, 125].

Sensitivity and specificity based on the counts from a beat-to-beat comparison obviously convey important information on detection performance; however, these two measures also suffer from the disadvantage of not reflecting the episodic nature of paroxysmal AF. This is illustrated by the following scenario where an ECG recording is assumed to contain two AF episodes, one hour-long and another just 10-beat-long. The detector correctly identifies the long episode, but misses the brief one—a likely scenario given that the window length of most AF detectors precludes the detection of a 10-beat episode. The change in sensitivity due to a missed, brief episode is negligible, and illustrates that performance measures based on a beat-to-beat comparison tend to gloss over when brief episodes are missed. Accordingly, valuable clinical information may be lost. A similar glossing takes place in situations when numerous brief episodes are falsely detected, although the corresponding ROC still indicates almost perfect performance; this drawback is illustrated by the example in Fig. 4.15.

A kindred solution would be to replace the beat-to-beat comparison with an episode-to-episode comparison. Such a replacement will, however, raise a number of questions which need to be resolved: What is the meaning of “true negative” in episode-based detection? To what extent must the detected episode overlap with the annotated episode to be treated as a correct detection? Should a minimum duration be imposed on a detected episode to avoid that single beats, falsely labeled as AF beats, are counted as AF episodes?

Inspired by the work in [126] on performance measures appropriate for evaluating the detection of transient ischemia in long-term ECG recordings, these questions have been discussed in the context of AF detection [36]. Since an episode of non-AF beats has little meaning, the number of true negatives N_{TN} is undefined, and, therefore, only sensitivity and positive predictive value can be computed, requiring that the following, redefined counts are determined:

N_{TP} = #AF episodes correctly detected as AF episodes (true positive),

N_{FP} = #non-AF episodes falsely detected as AF episodes (false positive),

N_{FN} = #AF episodes falsely detected as non-AF episodes (false negative).

An episode is judged as correctly detected if it overlaps the annotated episode with at least 50%, otherwise the episode is labeled non-AF [33]. While the minimum duration of a detected episode not necessarily has to be stated, it is indirectly determined by

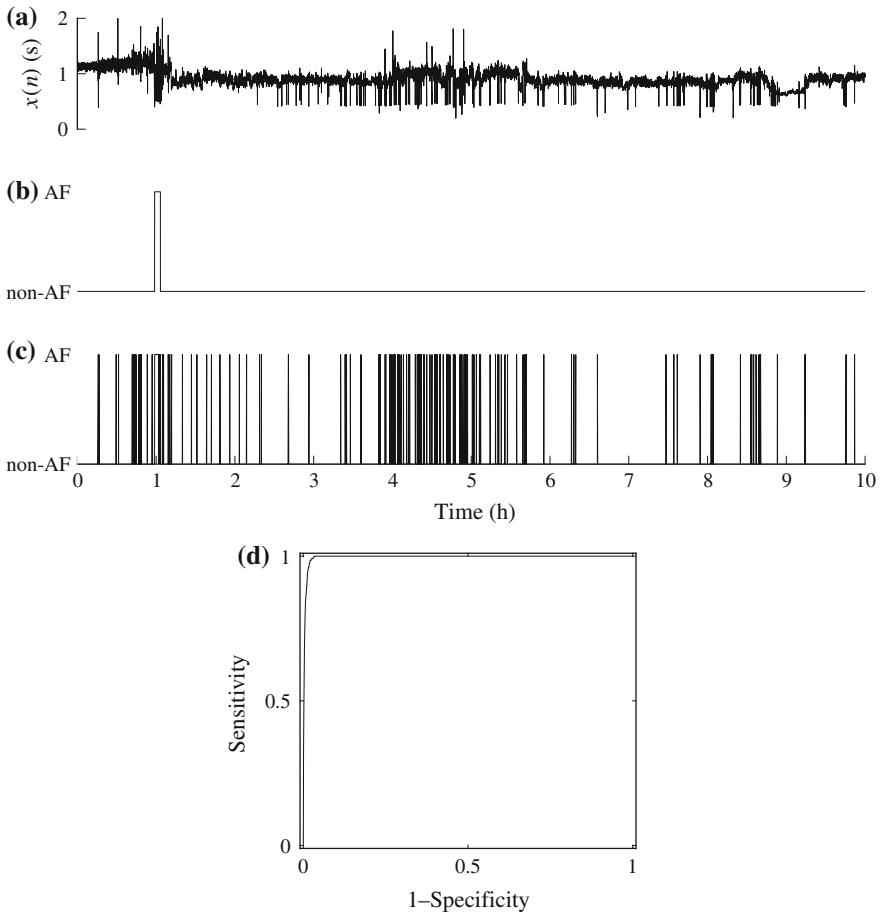


Fig. 4.15 **a** An RR interval series $x(n)$ and **b** AF episode annotation. **c** Output from a detector based on the coefficient of sample entropy (computed in a 12-beat window), and **d** related ROC. The detector correctly identifies the single AF episode, but also produces numerous false detections due to the presence of ectopic beats. Still, the corresponding ROC indicates that almost perfect detection performance is achieved

the choice of window length. For a 100-beat window, the beat-based sensitivity of 0.92, reported in [36] and listed in Table 4.1, dropped to 0.71 when episode-based sensitivity was considered instead. This drop in sensitivity illustrates that the use of a 100-beat window precludes the detection of brief episodes.

Episode-based performance measures have not yet gained a foothold in the literature on AF detection, although such measures provide information which is complementary to beat-based measures. The popularity of beat-based measures may be due to their ease of computation, but also to the many ECG applications where beat-based performance measures have become well-established. However, neither beat-based

nor episode-based measures provide information on the detectability of episodes with varying lengths.

The delay between the annotated onset of the episode and the onset produced by the detector represents another type of performance measure which has received attention in the literature [32, 35, 39, 89]. From an algorithmic viewpoint, the time delay introduced by the detector needs to be established to make a comparison with episode onset/end annotations meaningful. From a clinical viewpoint, however, a short time delay is of subordinate importance to the above-mentioned performance measures, since very few ECG applications call for immediate action after the initiation of an episode.

4.6 Detection Performance

4.6.1 ECG Databases

Detection performance is commonly evaluated on one or several publicly available, annotated databases of long-term ECG recordings, where AFDB holds the position as the most popular database. While the availability of public databases certainly facilitates the comparison of performance, conclusions drawn from the performance figures presented in Table 4.3, or the tables presented in e.g., [86, 89, 127], should be made with caution for a number of reasons. Since both specificity and sensitivity differ from one detector to another, performance is not easily compared. Better, though not perfect, is to first compute the ROC for each detector, and then determine the sensitivity at a fixed specificity, or vice versa, which leads to a more relevant comparison.

Another complicating factor is that detection performance is not always established from the analysis of the entire AFDB, but from a subset of varying size. In some studies, records 4936 and 5091 were omitted, since the annotations were deemed to be incorrect (AFDB₁) [32, 37]. While rhythm-based detectors can analyze all 25 records of the AFDB, only 23 records can be analyzed by detectors based on rhythm and morphology since two records lack the original ECG signals (AFDB₂). Yet another complicating factor is that certain detectors require a minimum length of normal sinus rhythm to fulfill detector training, in one case leading to the exclusion of as many as 5 out of the 25 records (AFDB₃) [88]. Moreover, in detector training, it is highly desirable to analyze data sets containing AF and non-AF segments which are balanced in size. A straightforward approach to handling the fact that AFDB contains about 80% more non-AF segments than AF segments is therefore to discard the excess amount of non-AF segments (AFDB₄) [90]; unfortunately, the non-AF data set cannot be reproduced in other studies since the segments were randomly excluded. However, such a drastic exclusion of data precludes any meaningful comparison of detection performance—a fact which should be kept in mind when assessing the results in Table 4.3.

From a comparative perspective, the picture becomes even more complicated when performance is evaluated on tiny subsets of RR interval series [67] or beats [127], excerpted from the records in AFDB. Such data excerption not only tends to exaggerate performance figures due to inclusion of better-than-average data quality, but the reproduction of results is not possible due to the lack of detail on what data were actually excerpted.

Other public databases have been analyzed to provide a more complete description of detection performance, notably NSRDB, MIT-BIH Arrhythmia Database (MITDB), and Long-Term AF Database (LTAfDB) [98], see Sect. 3.1. Since NSRDB contains no significant arrhythmias, it can only provide information on specificity, e.g., [32, 33, 35, 37, 41, 51]. The MITDB contains several types of arrhythmia, including AF and atrial flutter, and may be used to evaluate both specificity and sensitivity [32, 33, 37, 41]; however, as pointed out in Sect. 3.1, MITDB contains relatively few AF episodes, and, consequently, performance figures describing episode detection are not representative. The LTAfDB, containing many more and much longer ECG recordings than AFDB, is well-suited for performance evaluation, though not very often used [41].

Some studies involve proprietary ECG databases, acquired to strengthen the results obtained on public databases [37], or used for classifier training [100]. Another reason for acquiring a database is that public databases do not always account for the signal characteristics pertinent to the application of interest.

4.6.2 *Training and Evaluation*

Widely different approaches have been considered for classifier training and performance evaluation—an observation illustrated by the way different data sets are handled by the detectors listed in Table 4.3. In some studies, either a proprietary database or LTAfDB were used for training, accompanied by a performance evaluation on AFDB [38, 40, 41, 87]. Such an approach is preferred since it avoids that the same patients are used for both training and evaluation. In other studies, no information is provided on the data set used for training [33, 37], whereas AFDB or some other databases is used for evaluation.

With respect to training, AFDB has been used to determine optimal detection thresholds [32, 34], or to select an optimal set of features for classification [36], accompanied by performance evaluation on other databases. Although the results from evaluation are the important ones in these studies, the positively biased results obtained from training on AFDB were also reported. Later on, these results have been included in comparisons of detector performance [38, 88, 89, 128], although the figures are not fully representative. This observation applies even more to the results reported in [35], where AFDB was used for both detector development and evaluation.

In an effort to reduce positive bias, AFDB can be partitioned into different subsets, one for training and another for performance evaluation. The subsets have

been formed either by random selection of non-AF/AF segments, division into disjoint subsets of equal size for use in stratified twofold cross-validation [39], or tenfold cross-validation [90]. A small subset of AFDB was used in [83] for training, whereas the entire AFDB was used for evaluation. One of the subsets was used for training in [39] and the other for evaluation, followed by reverse use of the two subsets; the results from the two evaluations were then averaged to yield the overall performance. It is highly questionable whether the performance figures of cross-validation on AFDB can be compared to those obtained for a detector which have been trained on a separate database, especially when considering that AFDB only contains 25 patients [129].

The above-mentioned approaches to training and evaluation are population-based, however, patient-based training may be pursued as well [88]. For each patient in AFDB, detector training was based on the initial part of the ECG record, whereas evaluation was based on the remaining part. However, before training, all beats with “irregularities” were excluded from the training data set using manual review, introducing positive bias in the results. In addition, the practical use of the detector is limited since good-quality signals are not always available for training, nor is manual review prior to AF analysis feasible in clinical routine.

Based on the above considerations, it is evident that a comparison of detection performance is seriously challenged by the presence of positive bias. Independent data sets for training and evaluation should ideally be analyzed, however, not uncommonly, the same patient is part of both data sets. Therefore, as already pointed out, caution should be exercised when comparing detection performance, e.g., with respect to sensitivity and specificity as in Table 4.3.

It deserves to be noted that AF detectors using adaptive filtering for f wave extraction, such as the one in [100], cannot be trained and evaluated on AFDB since none of the two leads is appropriate for use as a reference lead, i.e., none of the leads contains negligible atrial activity. This problem may be addressed using a proprietary multi-lead database for training, and simulated multi-lead signals for performance evaluation [100].

4.6.3 Simulated ECG Signals

Performance evaluation is typically based on real ECG signals annotated with respect to the onset and end of AF episodes, whereas simulated ECG signals are rarely used. This stands in contrast to the evaluation of f wave extraction performance, where simulated signals are frequently used—the main reason being that manual annotations are irrelevant in f wave extraction. Nonetheless, simulated ECG signals have a place in AF detection since certain properties of clinical or technical significance, e.g., atrial ectopy, episode duration, and noise level, can be easily controlled in such signals, whereas public databases may not allow adequate investigation of these properties.

Detection accuracy has been investigated on simulated signals with different noise levels, both with and without the presence of APBs [100]. The results put spotlight on the importance of proper handling of APBs, and show that detection based on both rhythm and morphology provides much higher accuracy than does rhythm-based detection in the presence of APBs, especially at lower noise levels where P wave absence and f wave presence can be reliably estimated. A similar relationship exists between detection accuracy and episode duration, i.e., detection based on both rhythm and morphology provides much higher accuracy in finding brief episodes of varying duration than does rhythm-based detection (5, 10, 20, and 30-beat duration were investigated).

Simulated ECG signals can also serve as a means to establish the SNR below which AF detector operation no longer is recommended. In one of the few studies to address this issue, simulated muscle noise was added to real ECGs, contained in LTAFDB, at different SNRs [114]. The noisy ECG signals were then used to evaluate the influence of noise on QRS detection, as well as on rhythm-based AF detection. The results suggested an essentially linear reduction in AF detection accuracy with respect to SNR when expressed in terms of decibels. The evaluation of performance in noise is even more important for AF detectors analyzing both rhythm and morphology.

4.6.4 Brief AF Episodes

Despite the clinical interest in occult PAF and related risk of future stroke, little attention has been paid to the detection of brief AF episodes. Although AFDB contains a few brief episodes, it is completely dominated by long episodes, cf. Fig. 3.1b, so that missed brief episodes have little influence on beat-based performance measures. Interestingly, some studies report the number of missed brief episodes: 30 out of the 254 episodes in AFDB₁ were missed by the detector in [32], all missed episodes having a duration less than 75 beats. In another study [35], 32 out of the 299 episodes in AFDB were missed, the main reason again being missed brief episodes (durations from 4 to 62 beats).

Indirect evaluation of performance with respect to brief episodes can be accomplished by analyzing the influence of different lengths of the detection window on performance. The window length imposes a minimum duration on AF episode detectability. While the exact relationship between window length and episode duration depends on the detection principle used, an episode with a duration of about half the window length or shorter will, in general, be missed, illustrated in Fig. 4.16. The choice of window length is a trade-off: a shorter window facilitates the detection of brief AF episodes, whereas a longer window implies more reliable parameter estimates (assuming that the window contains the same rhythm), but also a larger amount of computations.

Over the years, the trend has been to design detectors with increasingly shorter windows, primarily motivated by the wish to reduce the time to decision [32, 89] and the amount of computations [84]. The recommended window length in rhythm-

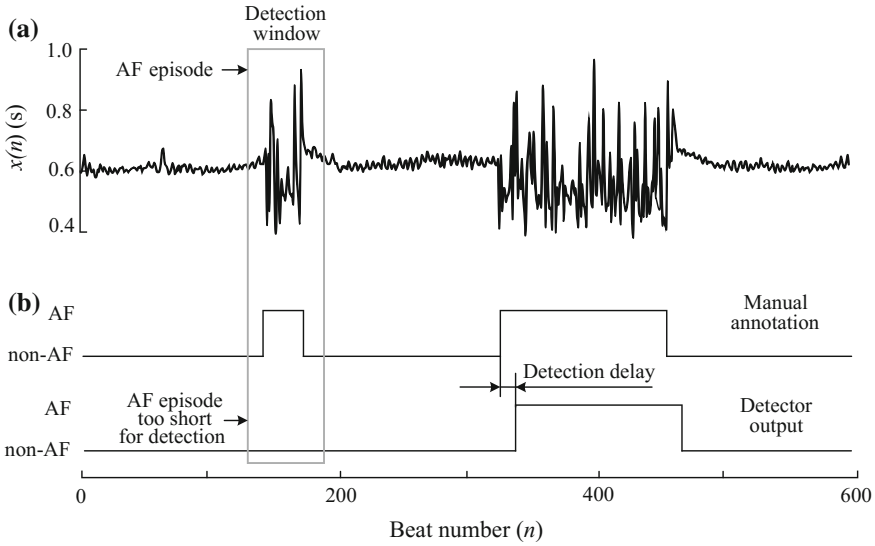


Fig. 4.16 **a** RR series from record 4043 of AFDB containing a brief AF episode (20 beats) and a longer AF episode. The sliding detection window, displayed as a box, is too wide to allow detection of the first episode. **b** Annotation of the RR interval series and detector output. The delay in detecting the second AF episode is indicated. (Reprinted from [32] with permission)

based detection has decreased from 180 s in 1992 [130] to just 8 beats in 2015 [40], whereas, for rhythm and morphology based detectors, even shorter window lengths has been considered, i.e., 5 beats [100].

The degradation in performance when using a short window is well-illustrated by the detector based on the time-varying coherence function [37], briefly described in Sect. 4.2.3. Using a 128-beat window, sensitivity of 98.2% and specificity of 97.7% were obtained on AFDB₁, see Table 4.1. Using instead a 32-beat window, the sensitivity and specificity dropped to 96.7% and 96.1%, respectively. Despite the degradation in performance, the authors concluded that a shorter window is still of interest, since it will likely provide a more accurate description of AF burden. For the simple-structured detector exploring the distribution of the Poincaré point population [33], the use of a 128-beat window resulted in a sensitivity of 95.9% and a specificity of 95.4%, dropping to 94.4% and 92.6%, respectively, for a 32-beat window.

Alternatively, direct evaluation of performance can be accomplished by means of simulated ECG signals in paroxysmal AF, where episode duration is controlled by a set of model parameters [40]. The direct approach to evaluation is illustrated in Fig. 4.17, where detection accuracy is presented as a function of median episode duration, denoted T_E , for two different AF detectors. The simulated signals are produced by the model described in Sect. 3.3, and constructed from either synthetic or real components. Figure 4.17 underlines not only the expected result that shorter episodes imply decreased detection accuracy, but it also demonstrates that a detector

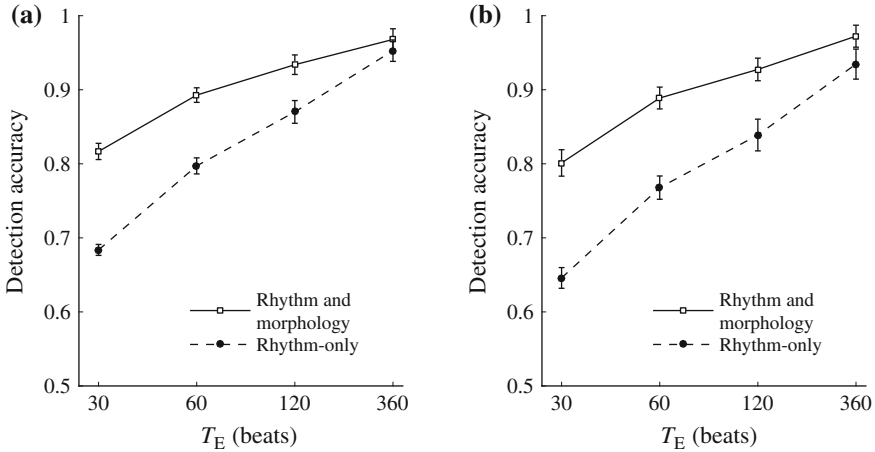


Fig. 4.17 Detection accuracy as a function of median episode duration T_E when the ECG signals are generated using **a** synthetic and **b** real components. The noise level is set to $20\ \mu\text{V}$ RMS. The rhythm-based detector is described in [40], and the detector based on both rhythm and morphology in [100]

based on rhythm and morphology performs better than a detector based on rhythm-only; the difference in performance increases as T_E becomes increasingly shorter. Comparing Fig. 4.17a and b, it is obvious that detection accuracy is essentially independent of whether synthetic or real components are used to produce the simulated ECG. However, as T_E becomes increasingly shorter, the difference in performance between the detector based on rhythm and morphology and the detector based on rhythm-only becomes increasingly larger for real components than for synthetic components. This drop in performance is likely explained by the pathological rhythms present in the database from which the RR interval series were extracted.

4.7 Additional Detector Information

Certain ECG signal properties have been explored for the purpose of predicting either the onset or the end of an AF episode. Similar to heart rate, the properties are not of immediate importance to detector design, but may be integrated in the detector, for example, using a threshold whose level is adjusted in relation to the proneness with which a transition occurs from sinus rhythm to AF, or vice versa. Whether such integration improves detection performance remains to be demonstrated. Considering that more than 90% of all AF episodes are triggered by APBs [131–135], successful prediction of AF onset can be accomplished with a simple test on whether the rate of APBs, not followed by a regular RR interval, increases. This test is combined

with other tests on runs of atrial bigeminy/trigeminy and the duration of short runs of paroxysmal atrial tachycardia [136].

Another approach to predicting the onset of paroxysmal AF is to analyze changes in heart rate variability (HRV) which may precede an AF episode. Indeed, in many patients, AF onset is immediately preceded by a significant reduction of the ratio between the low and the high frequency HRV components [137–139], a pattern which is not detectable after spontaneous recovery of sinus rhythm [140]. Alternatively, changes in HRV may be characterized by entropy, with results suggesting that AF onset is preceded by reduced complexity of the RR intervals [141], see also [142]. In yet another approach, AF onset could be predicted by combining spectral, bispectral, and nonlinear features, using a machine learning technique for classification of the preceding HRV pattern [143]. For the above-mentioned studies on APB- and HRV-based prediction, a 30-min segment immediately preceding AF onset is usually considered for evaluating prediction performance.

Different P wave properties have been explored for predicting AF. For example, changes in P wave morphology due to abnormal interatrial conduction are observed in patients bound to develop AF [144, 145], prolongation of the maximum P wave duration may predict recurrent AF [146–148], as well as shortening of the minimum P wave duration [147, 149]. Moreover, changes in the dynamics of P wave morphology may predict AF onset [96, 148]. However, changes in P wave properties occur over a much longer time frame than changes associated with APB- and HRV-based prediction: the former type of changes occurs over weeks to months, whereas the latter over minutes. Hence, information on P wave related changes are less useful in AF detection.

The prediction of AF termination takes its starting point in the analysis of f wave properties, and typically requires that the ventricular activity has been cancelled before prediction can take place. Among the properties explored, the DAF has been found to exhibit gradual slowing just before termination of paroxysmal or persistent AF [150, 151]. Results from studying the wavelet entropy of f waves, employed as a measure reflecting unpredictability in time as well as frequency, suggest that f waves are characterized by decreasing entropy as the termination is approaching [152].

Information on physical activity will most likely play a role in AF detection in the quest to reduce the number of falsely detected episodes, especially since accelerometers are nowadays standard implementation in ECG devices. Although it remains to be demonstrated that AF detectors analyzing both bioelectrical and physical information offer better performance, a preliminary study shows that the number of falsely classified arrhythmias due to noise and artifacts can be considerably reduced when accelerometer information is taken into account [153]. The potential of accelerometer information in AF detection is further supported by results showing that AF episodes can be detected from accelerometers attached to the chest [154], or from an electromechanical vibration sensor attached to a bed mattress [155], without involving the analysis of the ECG.

References

1. A. Haeberlin, L. Roten, M. Schilling, F. Scarcia, T. Niederhauser, R. Vogel, J. Fuhrer, H. Tanner, Software-based detection of atrial fibrillation in long-term ECGs. *Heart Rhythm* **11**, 933–938 (2014)
2. K.M. Stein, J. Walden, N. Lippman, B.B. Lerman, Ventricular response in atrial fibrillation: random or deterministic? *Am. J. Physiol.* **277**, H452–458 (1999)
3. J. Hayano, F. Yamasaki, S. Sakata, A. Okada, S. Mukai, T. Fujinami, Spectral characteristics of ventricular response to atrial fibrillation. *Am. J. Physiol.* **273**, H2811–2816 (1997)
4. V. Fuster, L.E. Rydén, D.S. Cannom, H.J. Crijns, A.B. Curtis et al., ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines and the European Society of Cardiology Committee for Practice Guidelines developed in collaboration with the European Heart Rhythm Association and the Heart Rhythm Society Heart Association Task Force on practice guidelines and the European Society of Cardiology Committee for Practice Guidelines developed in collaboration with the European Heart Rhythm Association and the Heart Rhythm Society. *Europace* **8**, 651–745 (2006)
5. C.T. January, L.S. Wann, J.S. Alpert, H. Calkins, J.E. Cigarroa et al., 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Heart Rhythm Society. *Circulation* **130**, 2071–2104 (2014)
6. P. Kirchhof, S. Benussi, D. Kotecha, A. Ahlsson, D. Atar, B. Casadei, M. Castella, H.C. Diener, H. Heidbuchel, J. Hendriks, G. Hindricks, A.S. Manolis, J. Oldgren, B.A. Popescu, U. Schotten, B. Van Putte, P. Vardas, S. Agewall, J. Camm, G. Baron Esquivias, W. Budts, S. Cacerj, F. Casselman, A. Coca, R. De Caterina, S. Deftereos, D. Dobrev, J.M. Ferro, G. Filippatos, D. Fitzsimons, B. Gorenek, M. Guenoun, S.H. Hohnloser, P. Kolh, G.Y. Lip, A. Manolis, J. McMurray, P. Ponikowski, R. Rosenhek, F. Ruschitzka, I. Savelieva, S. Sharma, P. Suwalski, J.L. Tamargo, C.J. Taylor, I.C. Van Gelder, A.A. Voors, S. Windecker, J.L. Zamorano, K. Zeppenfeld, 2016 ESC guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Eur. Heart J.* **37**, 2893–2962 (2016)
7. R.C.S. Seet, P.A. Friedman, A.A. Rabinstein, Prolonged rhythm monitoring for the detection of occult paroxysmal atrial fibrillation in ischemic stroke of unknown cause. *Circulation* **124**, 477–486 (2011)
8. J.W. Keach, S.M. Bradley, M.P. Turakhia, T.M. Maddox, Early detection of occult atrial fibrillation and stroke prevention. *Heart* **101**, 1097–1102 (2015)
9. J.G. Andrade, T. Field, P. Khairy, Detection of occult atrial fibrillation in patients with embolic stroke of uncertain source: a work in progress. *Front. Physiol.* **1**, 1–9 (2015)
10. D.J. Miller, K. Shah, S. Modi, A. Mahajan, S. Zahoor, M. Affan, The evolution and application of cardiac monitoring for occult atrial fibrillation in cryptogenic stroke and TIA. *Curr. Treat. Options Neurol.* **18**, 17 (2016)
11. J.O. Cerasuolo, L.E. Cipriano, L.A. Sposato, The complexity of atrial fibrillation newly diagnosed after ischemic stroke and transient ischemic attack: advances and uncertainties. *Curr. Opin. Neurol.* **30**, 28–37 (2017)
12. A.H. Tayal, M. Tian, K.M. Kelly, S.C. Jones, D.G. Wright, D. Singh, J. Jarouse, J. Brillman, S. Murali, R. Gupta, Atrial fibrillation detected by mobile cardiac outpatient telemetry in cryptogenic TIA or stroke. *Neurology* **71**, 1696–1701 (2008)
13. A.A. Rabinstein, J.E. Fugate, J. Mandrekar, J.D. Burns, R.C. Seet, S.A. Dupont, T.J. Kauffman, S.J. Asirvatham, P.A. Friedman, Paroxysmal atrial fibrillation in cryptogenic stroke: a case control study. *J. Stroke Cerebrovascular Dis.* **22**, 1405–1411 (2013)
14. A.H. Abdul-Rahim, K.R. Lees, Paroxysmal atrial fibrillation after ischemic stroke: how should we hunt for it? *Expert Rev. Cardiovasc. Ther.* **11**, 485–494 (2013)
15. C.G. Favilla, E. Ingala, J. Jara, E. Fessler, B. Cucchiara, S.R. Messé, M.T. Mullen, A. Prasad, J. Siegler, M.D. Hutchinson, S.E. Kasner, Predictors of finding occult atrial fibrillation after cryptogenic stroke. *Stroke* **46**, 1210–1215 (2015)

16. E.I. Charitos, U. Stierle, P.D. Ziegler, M. Baldewig, D.R. Robinson, H. Sievers, T. Hanke, A comprehensive evaluation of rhythm monitoring strategies for the detection of atrial fibrillation recurrence: insights from 647 continuously monitored patients and implications for monitoring after therapeutic interventions. *Circulation* **126**, 806–814 (2012)
17. T. Etgen, M. Hochreiter, M. Mundel, T. Freudenberger, Insertable cardiac event recorder in detection of atrial fibrillation after cryptogenic stroke: an audit report. *Stroke* **44**, 2007–2009 (2013)
18. J. Reiffel, A. Verma, J.L. Halperin, B. Gersh, S. Tombul, J. Carrithers, L. Sherfese, P. Kowey, Rationale and design of REVEAL AF: a prospective study of previously undiagnosed atrial fibrillation as documented by an insertable cardiac monitor in high-risk patients. *Am. Heart J.* **167**, 22–27 (2014)
19. A.C. Flint, N.M. Banki, X. Ren, V.A. Rao, A.S. Go, Detection of paroxysmal atrial fibrillation by 30-day event monitoring in cryptogenic ischemic stroke: The stroke and monitoring for PAF in real time (SMART) registry. *Stroke* **43**, 2788–2790 (2012)
20. S.B. Silverman, Paroxysmal atrial fibrillation: Novel strategies for monitoring and implications for treatment in stroke. *Curr. Treat. Options Cardio. Med.* **18**, 1–13 (2016)
21. L. Sörnmo, P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications* (Elsevier (Academic Press), Amsterdam, 2005)
22. N. Lowres, L. Neubeck, J. Redfern, S.B. Freedman, Screening to identify unknown atrial fibrillation. A systematic review. *Thromb. Haemost.* **110**, 213–222 (2013)
23. B. Vaes, S. Stalpaert, K. Tavernier, B. Thael, D. Lapeire, W. Mullens, J. Degryse, The diagnostic accuracy of the MyDiagnostick to detect atrial fibrillation in primary care. *BMC Fam. Pract.* **15**, 113 (2014)
24. F. Kaasenbrood, M.H.F.H. Rutten, L.J. Gerhards, A.W. Hoes, R.G. Tieleman, Yield of screening for atrial fibrillation in primary care with a hand-held, single-lead electrocardiogram device during influenza vaccination. *Europace* **18**, 1514–1520 (2016)
25. L. Desteghe, Z. Raymaekers, M. Lutin, J. Vijgen, D. Dilling-Boer, P. Koopman, J. Schurmans, P. Vanduyhoven, P. Dendale, H. Heidbuchel, Performance of handheld electrocardiogram devices to detect atrial fibrillation in a cardiology and geriatric ward setting. *Europace* **19**, 29–39 (2017)
26. E. Svennberg, J. Engdahl, F. Al-Khalili, L. Friberg, V. Frykman, M. Rosenqvist, Mass screening for untreated atrial fibrillation: the STROKESTOP study. *Circulation* **131**, 2176–2184 (2015)
27. E. Svennberg, M. Stridh, J. Engdahl, F. Al-Khalili, L. Friberg, V. Frykman, M. Rosenqvist, Safe automatic one-lead electrocardiogram analysis in screening for atrial fibrillation. *Europace* **19**, 1449–1453 (2016)
28. S.R. Steinhubl, R.R. Mehta, G.S. Ebner, M.M. Ballesteros, J. Waalen, G. Steinberg, P. Van Crocker, Jr., E. Felicione, C. T. Carter, S. Edmonds, J. P. Honcz, G. D. Miralles, D. Talantov, T. C. Sarich, E. J. Topol, Rationale and design of a home-based trial using wearable sensors to detect asymptomatic atrial fibrillation in a targeted population: the mHealth screening to prevent strokes (mSToPS) trial. *Am. Heart J.* **175**, 77–85 (2016)
29. M.P. Turakhia, D.W. Kaiser, Transforming the care of atrial fibrillation with mobile health. *J. Interv. Card. Electrophysiol.* **47**, 45–50 (2016)
30. G. D. Clifford, C. Liu, B. Moody, L.-W. H. Lehman, I. Silva, Q. Li, A. Johnson, and R. G. Mark, “AF classification from a short single lead ECG recording: the PhysioNet Computing in Cardiology Challenge 2017, in *Proceedings of Computing in Cardiology*, vol. 44 (2017)
31. K. Tateno, L. Glass, Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and deltaRR intervals. *Med. Biol. Eng. Comput.* **39**, 664–671 (2001)
32. S. Dash, K.H. Chon, S. Lu, E.A. Raeder, Automatic real time detection of atrial fibrillation. *Ann. Biomed. Eng.* **37**, 1701–1709 (2009)
33. J. Lian, L. Wang, D. Muessig, A simple method to detect atrial fibrillation using RR intervals. *Am. J. Cardiol.* **107**, 1494–1497 (2011)

34. D. E. Lake, J.R. Moorman, Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices. *Am. J. Physiol. (Heart Circ. Physiol.)* **300**: H319–H325 (2011)
35. C. Huang, S. Ye, H. Chen, D. Li, F. He, Y. Tu, A novel method for detection of the transition between atrial fibrillation and sinus rhythm. *IEEE Trans. Biomed. Eng.* **58**, 1113–1119 (2011)
36. R.B. Shouldice, C. Heneghan, P. de Chazal, Automatic detection of paroxysmal atrial fibrillation, in *Atrial fibrillation – basic research and clinical applications*. (J. Choi, ed.), chap. 7, pp. 125–146, InTech (2012)
37. J. Lee, Y. Nam, D.D. McManus, K.H. Chon, Time-varying coherence function for atrial fibrillation detection. *IEEE Trans. Biomed. Eng.* **60**, 2783–2793 (2013)
38. X. Zhou, H. Ding, B. Ung, E. Pickwell-MacPherson, Y. Zhang, Automatic online detection of atrial fibrillation based on symbolic dynamics and shannon entropy. *Biomed. Eng. Online* **13**, 18 (2014)
39. S. Asgari, A. Mehrmi, M. Moussavi, Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine. *Comput. Biol. Med.* **60**, 132–142 (2015)
40. A. Petrėnas, V. Marozas, L. Sörnmo, Low-complexity detection of atrial fibrillation in continuous long-term monitoring. *Comput. Biol. Med.* **65**, 184–191 (2015)
41. X. Zhou, H. Ding, W. Wu, Y. Zhang, A real-time atrial fibrillation detection algorithm based on the instantaneous state of heart rate. *PLoS ONE* **10**, e0136544 (2015)
42. G.B. Moody, R.G. Mark, A new method for detecting atrial fibrillation using R-R intervals. in *Proceedings of Computers in Cardiology* vol. 10, pp. 227–230 (1983)
43. S. Cerutti, L.T. Mainardi, A. Porta, A.M. Bianchi, Analysis of the dynamics of RR interval series for the detection of atrial fibrillation episodes, in *Proceedings of Computers in Cardiology*, vol. 24, pp. 7–80 (1997)
44. S. Shkurovich, A.V. Sahakian, S. Swiryn, Detection of atrial activity from high-voltage leads of implantable ventricular defibrillators using a cancellation technique. *IEEE Trans. Biomed. Eng.* **45**, 229–234 (1998)
45. D. Duverney, J.M. Gaspoz, V. Pichot, F. Roche, R. Brion, A. Antoniadis, J.C. Barthelemy, High accuracy of automatic detection of atrial fibrillation using wavelet transform of heart rate intervals. *Pacing Clin. Electrophysiol.* **25**, 457–462 (2002)
46. F. Yaghouby, A. Ayatollahi, R. Bahramali, M. Yaghouby, A.H. Alavi, Towards automatic detection of atrial fibrillation: a hybrid computational approach. *Comput. Biol. Med.* **40**, 919–930 (2010)
47. C.-T. Lin, K.-C. Chang, C.-L. Lin, C.-C. Chiang, S.-W. Lu, S.-S. Chang, B.-S. Lin, H.-Y. Liang, R.-J. Chen, Y.-T. Lee, L.-W. Ko, An intelligent telecardiology system using a wearable and wireless ECG to detect atrial fibrillation. *IEEE Trans. Info. Tech. Biomed.* **14**, 726–733 (2010)
48. P. Langley, M. Dewhurst, L.D. Marco, P. Adams, F. Dewhurst, J. Mwita, R. Walker, A. Murray, Accuracy of algorithms for detection of atrial fibrillation from short duration beat interval recordings. *Med. Eng. Phys.* **34**, 1441–1447 (2012)
49. J. Lee, B. Reyes, D. McManus, O. Mathias, K. Chon, Atrial fibrillation detection using an iPhone 4S. *IEEE Trans. Biomed. Eng.* **60**, 203–206 (2013)
50. J. Park, S. Lee, M. Jeon, Atrial fibrillation detection by heart rate variability in Poincaré plot. *Biomed. Eng. Online* **8**, 1–12 (2009)
51. S. Sarkar, D. Ritscher, R. Mehra, A detector for a chronic implantable atrial tachyarrhythmia monitor. *IEEE Trans. Biomed. Eng.* **55**, 1219–1224 (2008)
52. M.S. Kendall, A. Stuart, J.K. Ord, *The Advanced Theory of Statistics*, vol. 3, 4th edn. (High Wycombe: Charles Griffin, 1983)
53. S.B. Olsson, N. Cai, M. Dohnal, K.K. Talwar, Noninvasive support for and characterization of multiple intranodal pathways in patients with mitral valve disease and atrial fibrillation. *Eur. Heart J.* **7**, 320–333 (1986)
54. N. Cai, M. Dohnal, S.B. Olsson, Methodological aspects of the use of heart rate stratified RR interval histograms in the analysis of atrioventricular conduction during atrial fibrillation. *Cardiovasc. Res.* **21**, 455–462 (1987)

55. J. Dickinson Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*, 5th edn. (Chapman and Hall/CRC, 2010)
56. J. Tebbenjohanns, B. Schumacher, T. Korte, M. Niehaus, D. Pfeiffer, Bimodal RR interval distribution in chronic atrial fibrillation: impact of dual atrioventricular nodal physiology on long-term rate control after catheter ablation of the posterior atrionodal input. *J. Cardiovasc. Electrophysiol.* **11**, 497–503 (2000)
57. S. Rokas, S. Gaitanidou, S. Chatzidou, C. Pamboucas, D. Achtipis, S. Stamatelopoulos, Atrioventricular node modification in patients with chronic atrial fibrillation: role of morphology of RR interval variation. *Circulation* **103**, 2942–2948 (2001)
58. V.D.A. Corino, F. Sandberg, L.T. Mainardi, L. Sörnmo, An atrioventricular node model for analysis of the ventricular response during atrial fibrillation. *IEEE Trans. Biomed. Eng.* **58**, 3386–3395 (2011)
59. C.E. Shannon, A mathematical theory of communication. *Bell Sys. Tech. J.* **27**, 379–423 (1948)
60. S.J. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol.* **278**, H2039–H2049 (2000)
61. D.E. Lake, J.S. Richman, M.P. Griffin, J.R. Moorman, Sample entropy analysis of neonatal heart rate variability. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **283**, R789–R797 (2002)
62. D.E. Lake, Renyi entropy measures of heart rate Gaussianity. *IEEE Trans. Biomed. Eng.* **53**, 21–27 (2006)
63. M. S. Pincus, A.L. Goldberger, Physiological time-series analysis: what does regularity quantify? *Am. J. Physiol.* 266 (Heart Circ. Physiol.) **35**: H1643–H1656 (1994)
64. W. Chen, Z. Wang, H. Xie, W. Yu, Characterization of surface EMG signal based on fuzzy entropy. *IEEE Trans. Neural Sys. Rehab. Eng.* **15**, 266–272 (2007)
65. A. Avolio, Heart rate variability and stroke: strange attractors with loss of complexity. *J. Hypertension* **31**, 1529–1531 (2013)
66. M. Julián, R. Alcaraz, J.J. Rieta, Comparative assessment of nonlinear metrics to quantify organization-related events in surface electrocardiograms of atrial fibrillation. *Comput. Biol. Med.* **48**, 66–76 (2014)
67. L. Hong-wei, S. Ying, L. Min, L. Pi-ding, Z. Zheng, A probability density function method for detecting atrial fibrillation using R-R intervals. *Med. Eng. Phys.* **31**, 116–123 (2009)
68. P. Grassberger, I. Procaccia, Characterization of strange attractors. *Phys. Rev. Lett.* **50**, 346–349 (1983)
69. T. Anan, K. Sunagawa, H. Araki, M. Nakamura, Arrhythmia analysis by successive RR plotting. *J. Electrocardiol.* **23**, 243–248 (1990)
70. P.W. Kamen, H. Krum, A.M. Tonkin, Poincaré plot of heart rate variability allows quantitative display of parasympathetic nervous activity in humans. *Clin. Sci. (Lond.)* **91**, 201–208 (1996)
71. M. Brennan, M. Palaniswami, P. Kamen, Poincaré plot interpretation using a physiological model of HRV based on a network of oscillators. *Am. J. Physiol. Heart Circ. Physiol.* **283**, H1873–H1886 (2002)
72. M. Malik, Standard measurements of heart rate variability, in *Dynamic electrocardiography* ed. by M. Malik, A.J. Camm, chap. 2, (Wiley–Blackwell, New York, 2004), pp. 13–21
73. K. Monahan, Y. Song, K. Loparo, R. Mehra, F.E. Harrell Jr., S. Redline, Automated detection of atrial fibrillation from the electrocardiogram channel of polysomnograms. *Sleep Breath* **20**, 515–522 (2015)
74. C.K. Karmakar, A.H. Khandoker, J. Gubbi, M. Palaniswami, Complex correlation measure: a novel descriptor for Poincaré plot. *BioMed. Eng. Online* **8**, 37–48 (2009)
75. L. Zhang, T. Guo, B. Xi, Y. Fan, K. Wang, J. Bi, Y. Wang, Automatic recognition of cardiac arrhythmias based on the geometric patterns of Poincaré plots. *Physiol. Meas.* **36**, 283–301 (2015)
76. R. Mehra, J. Gillberg, P. Ziegler, S. Sarkar, Algorithms for atrial tachyarrhythmia detection for long-term monitoring with implantable devices, in *Understanding atrial fibrillation: the signal processing contribution* ed. by L.T. Mainardi, L.Sörnmo, S. Cerutti, chap. 8 (Morgan & Claypool, San Francisco, 2008), pp. 175–214

77. H. Käsmacher, S. Wiese, M. Lahl, Monitoring the complexity of ventricular response in atrial fibrillation. *Discrete Dynamics Nature Soc.* **4**, 63–75 (2000)
78. M. Brennan, M. Palaniswami, P. Kamen, Do existing measures of Poincaré plot geometry reflect nonlinear features of heart rate variability? *IEEE Trans. Biomed. Eng.* **48**, 1342–1347 (2001)
79. R.A. Thuringham, An electrocardiogram marker to detect paroxysmal atrial fibrillation. *J. Electrocardiol.* **40**, 344–347 (2007)
80. H. Zhao, S. Lu, R. Zou, K. Ju, K.H. Chon, Estimation of time-varying coherence function using time-varying transfer functions. *Ann. Biomed. Eng.* **33**, 1582–1594 (2005)
81. R. Zou, H. Wang, K.H. Chon, A robust time-varying identification algorithm using basis functions. *Ann. Biomed. Eng.* **31**, 840–853 (2003)
82. F. van der Heijden, R.P.W. Duin, D. de Ridder, D.M.J. Tax, *Classification, Parameter Estimation and State Estimation—An Engineering Approach using Matlab* (Wiley, New York, 2005)
83. P. Carvalho, J. Henriques, R. Couceiro, M. Harris, M. Antunes, J. Habetha, Model-based atrial fibrillation detection, in *ECG signal processing, classification and interpretation* ed. by A. Gacek, W. Pedrycz (Springer London, 2012), pp. 99–133
84. R. Colloca, A.E.W. Johnson, L. Mainardi, G.D. Clifford, A support vector machine approach for reliable detection of atrial fibrillation events, in *Proceedings of Computing in Cardiology*, vol. 40, pp. 1047–1050 (2013)
85. R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, 2nd edn. (Wiley–Interscience, New York, 2001)
86. N. Larburu, T. Lopetegi, I. Romero, Comparative study of algorithms for atrial fibrillation detection, in *Proceedings of Computing in Cardiology*, vol. 38, pp. 265–268 (2011)
87. S. Babaeizadeh, R.E. Gregg, E.D. Helfenbein, J.M. Lindauer, S.H. Zhou, Improvements in atrial fibrillation detection for real-time monitoring. *J. Electrocardiol.* **42**, 522–526 (2009)
88. S. Ladavich, B. Ghoraani, Rate-independent detection of atrial fibrillation by statistical modeling of atrial activity. *Biomed. Signal Process. Control* **18**, 274–281 (2015)
89. J. Ródenas, M. García, R. Alcaraz, J.J. Rieta, Wavelet entropy automatically detects episodes of atrial fibrillation from single-lead electrocardiograms. *Entropy* **17**, 6179–6199 (2015)
90. Y. Xia, N. Wulan, K. Wang, H. Zhang, Detecting atrial fibrillation by deep convolutional neural networks. *Comput. Biol. Med.* **93**, 84–92 (2018)
91. P. Laguna, R. Jané, P. Caminal, Automatic detection of wave boundaries in multilead ECG signals: validation with the CSE database. *Comput. Biomed. Res.* **27**, 45–60 (1994)
92. J.P. Martínez, R. Almeida, S. Olmos, A.P. Rocha, P. Laguna, A wavelet-based ECG delineator: evaluation on standard databases. *IEEE Trans. Biomed. Eng.* **51**, 570–581 (2004)
93. J. Dumont, A. Hernández, G. Carraut, Improving ECG beats delineation with an evolutionary optimization process. *IEEE Trans. Biomed. Eng.* **57**, 607–615 (2010)
94. L. Clavier, J.-M. Boucher, R. Lepage, J.-J. Blanc, J.-C. Cornily, Automatic P-wave analysis of patients prone to atrial fibrillation. *Med. Biol. Eng. Comput.* **40**, 63–71 (2002)
95. F. Censi, G. Calcagnini, C. Ricci, R.P. Ricci, M. Santini, A. Grammatico, P. Bartolini, P-wave morphology assessment by a Gaussian functions-based model in atrial fibrillation patients. *IEEE Trans. Biomed. Eng.* **54**, 663–671 (2007)
96. A. Martínez, D. Abásolo, R. Alcaraz, J.J. Rieta, Alteration of the P-wave non-linear dynamics near the onset of paroxysmal atrial fibrillation. *Med. Eng. Phys.* **37**, 692–697 (2015)
97. P. Laguna, R. G. Mark, A. L. Goldberger, and G. B. Moody, “A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG, in *Proceedings of Computers in Cardiology*, Vol. 23, pp. 673–676 (1997)
98. A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **101**, E215–220 (2000)
99. I. Dotsinsky, Atrial wave detection algorithm for discovery of some rhythm abnormalities. *Physiol. Meas* **28**, 595–610 (2007)

100. A. Petrėnas, L. Sörnmo, A. Lukoševičius, V. Marozas, Detection of occult paroxysmal atrial fibrillation. *Med. Biol. Eng. Comput.* **53**, 287–297 (2015)
101. A. Petrėnas, V. Marozas, L. Sörnmo, A. Lukoševičius, An echo state neural network for QRST cancellation during atrial fibrillation. *IEEE Trans. Biomed. Eng.* **59**, 2950–2957 (2012)
102. S. Roberts, L. Tarassenko, A probabilistic resource allocating network for novelty detection. *Neural Comput.* **6**, 270–284 (1994)
103. M. Markou, S. Singh, Novelty detection: a review. *Signal Process.* **83**, 2481–2497 (2003)
104. X. Du, N. Rao, M. Qian, D. Liu, J. Li, W. Feng, L. Yin, X. Chen, A novel method for real-time atrial fibrillation detection in electrocardiograms using multiple parameters. *Ann. Noninvasive Electrocardiol.* **19**, 217–225 (2014)
105. F. Castells, J.J. Rieta, J. Millet, V. Zarzoso, Spatiotemporal blind source separation approach to atrial activity estimation in atrial tachyarrhythmias. *IEEE Trans. Biomed. Eng.* **52**, 258–267 (2005)
106. R. Llinares, J. Igual, J. Miró-Borrás, A fixed point algorithm for extracting the atrial activity in the frequency domain. *Comput. Biol. Med.* **40**, 943–949 (2010)
107. S.M. Kay, *Modern Spectral Estimation Theory and Application* (Prentice-Hall, New Jersey, 1999)
108. M. Stridh, A. Bollmann, S.B. Olsson, L. Sörnmo, Time-frequency analysis of atrial tachyarrhythmias: detection and feature extraction. *IEEE Eng. Med. Biol. Mag.* **25**, 31–39 (2006)
109. Q. Li, R.G. Mark, G.D. Clifford, Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. *Physiol. Meas.* **29**, 15–32 (2008)
110. G.D. Clifford, D. Clifton, Wireless technology in disease management and medicine. *Ann. Rev. Med.* **63**, 479–92 (2012)
111. J. Behar, J. Oster, Q. Li, G.D. Clifford, ECG signal quality during arrhythmia and its application to false alarm reduction. *IEEE Trans. Biomed. Eng.* **60**, 1660–1666 (2013)
112. A.E.W. Johnson, J. Behar, F. Andreotti, G.D. Clifford, J. Oster, Multimodal heart beat detection using signal quality indices. *Physiol. Meas.* **36**, 1665–1677 (2015)
113. C. Orphanidou, T. Bonnici, P. Charlton, D. Clifton, D. Vallance, L. Tarassenko, Signal-quality indices for the electrocardiogram and photoplethysmogram: derivation and applications to wireless monitoring. *IEEE J. Biomed. Health Inform.* **19**, 832–838 (2015)
114. J. Oster, G.D. Clifford, Impact of the presence of noise on RR interval-based atrial fibrillation detection. *J. Electrocardiol.* **48**, 947–951 (2015)
115. N. Gambarotta, F. Aletti, G. Baselli, M. Ferrario, A review of methods for the signal quality assessment to improve reliability of heart rate and blood pressures derived parameters. *Med. Biol. Eng. Comput.* **54**, 1025–1035 (2016)
116. S.H. Rappaport, L. Gillick, G.B. Moody, R.G. Mark, QRS morphology classification: quantitative evaluation of different strategies, in *Proceedings of Computers in Cardiology*, Vol. 9, pp. 33–38 (1982)
117. M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, L. Sörnmo, Clustering ECG complexes using Hermite functions and self-organizing maps. *IEEE Trans. Biomed. Eng.* **47**, 838–848 (2000)
118. P. de Chazal, M. O’Dwyer, R.B. Reilly, Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Trans. Biomed. Eng.* **51**, 1196–1206 (2004)
119. M. Llamedo, J.P. Martínez, Heartbeat classification using feature selection driven by database generalization criteria. *IEEE Trans. Biomed. Eng.* **58**, 616–625 (2011)
120. J. Oster, J. Behar, O. Sayadi, S. Nemati, A.E.W. Johnson, G.D. Clifford, Semisupervised ECG ventricular beat classification with novelty detection based on switching Kalman filters. *IEEE Trans. Biomed. Eng.* **62**, 2125–2134 (2015)
121. R. Couceiro, P. Carvalho, J. Henriques, M. Antunes, M. Harris, J. Habertha, Detection of atrial fibrillation using model-based ECG analysis, in *Proceedings of International Conference on Pattern Recognition (ICPR)*, vol. 19, pp. 1–5 (2008)
122. T. Jeon, B. Kim, M. Jeon, B.-G. Lee, Implementation of a portable device for real-time ECG signal analysis. *Biomed. Eng. Online* **13**, 1–13 (2014)

123. O. Andersson, K.H. Chon, L. Sörnmo, J. Neves Rodrigues, A 290 mV sub- V_T ASIC for real-time atrial fibrillation detection. *IEEE Trans. Biomed. Circ. Syst.* **9**: 377–386 (2015)
124. B. Matthews, Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta* **405**, 442–451 (1975)
125. P. Baldi, S. Brunak, Y. Chauvin, C.A. Andersen, H. Nielsen, Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics* **16**, 412–424 (2000)
126. F. Jager, G.B. Moody, A. Taddei, R. Mark, Performance measures for algorithms to detect transient ischemic ST segment changes, in *Proceedings of Computers in Cardiology*, vol. 18, pp. 369–372 (1991)
127. R.J. Martis, U.R. Acharya, H. Prasad, C.K. Chua, C.M. Lim, Automated detection of atrial fibrillation using Bayesian paradigm. *Knowledge-Based Syst.* **54**, 269–275 (2013)
128. M. García, J. Ródenas, R. Alcaraz, J.J. Rieta, Application of the relative wavelet energy to heart rate independent detection of atrial fibrillation. *Comput. Meth. Progr. Biomed.* **131**, 157–168 (2016)
129. A. Isaksson, M. Wallman, H. Göransson, M.G. Gustafsson, Cross-validation and bootstrapping are unreliable in small sample classification. *Pattern Recogn. Letter* **29**, 1960–1965 (2008)
130. J. Slocum, A.V. Sahakian, S. Swiryn, Diagnosis of atrial fibrillation from surface electrocardiograms based on computer-detected atrial activity. *J. Electrocardiol.* **25**, 1–8 (1992)
131. M. Haïssaguerre, P. Jaïs, D.C. Shah, A. Takahashi, M. Hocini, G. Quiniou, S. Garrigue, A. Le Mouroux, P. Le Métayer, J. Clémenty, Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins. *N. Engl. J. Med.* **339**, 659–666 (1998)
132. S.-A. Chen, M.-H. Hsieh, C.-T. Tai, C.-F. Tsai, V.S. Prakash, W.-C. Yu, T.-L. Hsu, Y.-A. Ding, M.-S. Chang, Initiation of atrial fibrillation by ectopic beats originating from the pulmonary veins: electrophysiological characteristics, pharmacological responses, and effects of radiofrequency ablation. *Circulation* **100**, 1879–1886 (1999)
133. C. Kolb, S. Nürnberger, G. Ndrepepa, B. Zrenner, A. Schömgig, C. Schmitt, Modes of initiation of paroxysmal atrial fibrillation from analysis of spontaneously occurring episodes using a 12-lead Holter monitoring system. *Am. J. Cardiol.* **88**, 853–857 (2001)
134. D. Wallmann, D. Tüller, K. Wustmann, P. Meier, J. Isenegger, M. Arnold, H.P. Mattle, E. Delacrétaç, Frequent atrial premature beats predict paroxysmal atrial fibrillation in stroke patients. An opportunity for a new diagnostic strategy. *Stroke* **38**, 2292–2294 (2007)
135. K.T. Nguyen, E. Vittinghoff, T.A. Dewland, M.C. Mandyam, P.K. Stein, E.Z. Soliman, S.R. Heckbert, G.M. Marcus, Electrocardiographic predictors of incident atrial fibrillation. *Am. J. Cardiol.* **118**, 714–719 (2016)
136. T. Thong, J. McNames, M. Aboy, B. Goldstein, Prediction of paroxysmal atrial fibrillation by analysis of atrial premature complexes. *IEEE Trans. Biomed. Eng.* **4**, 561–569 (2004)
137. J.L. Huang, W.C. Wen, W.L. Lee, M.S. Chang, S.A. Chen, Changes of autonomic tone before the onset of paroxysmal atrial fibrillation. *Int. J. Cardiol.* **66**, 275–283 (1998)
138. M. Bettoni, M. Zimmermann, Autonomic tone variations before the onset of paroxysmal atrial fibrillation. *Circulation* **105**, 2753–2759 (2002)
139. C. Gallo, P.P. Bocchino, M. Magnano, L. Gaido, D. Zema, A. Battaglia, M. Anselmino, F. Gaita, Autonomic tone activity before the onset of atrial fibrillation. *J. Cardiovasc. Electro-physiol.* **28**, 304–314 (2017)
140. F. Lombardi, D. Tarricone, F. Tundo, F. Colombo, S. Belletti, C. Fiorentini, Autonomic nervous system and paroxysmal atrial fibrillation: a study based on the analysis of RR interval changes before, during and after paroxysmal atrial fibrillation. *Eur. Heart J.* **25**, 1242–1248 (2004)
141. D.-G. Shin, C.-S. Yoo, S.-H. Yi, J.-H. Bae, Y.-J. Kim, J.-S. Park, G.-R. Hong, Prediction of paroxysmal atrial fibrillation using nonlinear analysis of the R-R interval dynamics before the spontaneous onset of atrial fibrillation. *Circ. J.* **70**, 94–99 (2006)
142. Y. Chesnokov, Complexity and spectral analysis of the heart rate variability dynamics for distant prediction of paroxysmal atrial fibrillation with artificial intelligence methods. *Artif. Intell. Med.* **43**, 151–165 (2008)

143. M. Mohebbi, H. Ghasseman, Prediction of paroxysmal atrial fibrillation based on non-linear analysis and spectrum and bispectrum features of the heart rate variability signal. *Comput. Meth. Progr. Biomed.* **5**, 40–49 (2012)
144. F. Holmqvist, P.G. Platonov, J. Carlson, W. Zareba, A.J. Moss, Altered interatrial conduction detected in MADIT II patients bound to develop atrial fibrillation. *Ann. Noninvasive Electrocardiol.* **14**:268–275 (2009)
145. F. Holmqvist, P.G. Platonov, S. McNitt, S. Polonsky, J. Carlson, W. Zareba, A.J. Moss, Abnormal P-wave morphology is a predictor of atrial fibrillation development and cardiac death in MADIT II patients. *Ann. Noninvasive Electrocardiol.* **15**: 63–72 (2010)
146. H. Gonna, M.M. Gallagher, X.H. Guo, Y.G. Yap, K. Hnatkova, A.J. Camm, P-wave abnormality predicts recurrence of atrial fibrillation after electrical cardioversion: a prospective study. *Ann. Noninvasive Electrocardiol.* **19**, 57–62 (2014)
147. J.B. Nielsen, J.T. Kühn, A. Pietersen, C. Graff, B. Lind, J.J. Struijk, M.S. Olesen, M.F. Sinner, T.N. Bachmann, S. Haunsø, B.G. Nordestgaard, P.T. Ellinor, J.H. Svendsen, K.F. Kofoed, L. Køber, A.G. Holst, P-wave duration and the risk of atrial fibrillation: results from the copenhagen ECG study. *Heart Rhythm* **12**, 1887–1895 (2015)
148. G. Conte, A. Luca, S. Yazdani, M.L. Caputo, F. Regol, T. Moccetti, L. Kappenberger, J.-M. Vesin, A. Auricchio, Usefulness of P-wave duration and morphologic variability to identify patients prone to paroxysmal atrial fibrillation. *Am. J. Cardiol.* **119**, 275–279 (2017)
149. I.C.Y. Chang, E. Austin, B. Krishnan, D.G. Benditt, C.N. Quay, L.H. Ling, L.Y. Chen, Shorter minimum P-wave duration is associated with paroxysmal lone atrial fibrillation. *J. Electrocardiol.* **47**, 106–112 (2014)
150. F. Nilsson, M. Stridh, A. Bollmann, L. Sörnmo, Predicting spontaneous termination of atrial fibrillation using the surface ECG. *Med. Eng. Phys.* **26**, 802–808 (2006)
151. P. G. Platonov, V.D.A. Corino, M. Seifert, F. Holmqvist, L. Sörnmo, Atrial fibrillatory rate in the clinical context: natural course and prediction of intervention outcome. *Europace* **16**: iv110–iv119 (2014)
152. R. Alcaraz, J.J. Rieta, Application of wavelet entropy to predict atrial fibrillation progression from the surface ECG. *Comput. Math. Meth. Med.* **13**, 1–9 (2012)
153. T. Tanantong, E. Nantajeewarawat, S. Thiemjarus, False alarm reduction in BSN-based cardiac monitoring using signal quality and activity type information. *Sensors* **15**, 3952–3974 (2015)
154. T. Koivisto, M. Pänkäälä, T. Hurnanen, T. Vasankari, T. Kiviniemi, A. Saraste, J. Airaksinen, Automatic detection of atrial fibrillation using MEMS accelerometer. in *Proceedings of Computing in Cardiology*, vol. 42, pp. 829–832 (2015)
155. C. Brueser, J. Diesel, M.D.H. Zink, S. Winter, P. Schauer, S. Leonhardt, Automatic detection of atrial fibrillation in cardiac vibration signals. *IEEE J. Biomed. Health Inform.* **17**, 162–171 (2013)