Atrial fibrillation detection using a novel cardiac ambulatory monitor based on photo-plethysmography at the wrist

Citation for published version (APA):

DOI:
10.1161/JAHA.118.009351

Document status and date:
Published: 01/08/2018

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

Take down policy
If you believe that this document breaches copyright please contact us:
openaccess@tue.nl
providing details. We will immediately remove access to the work pending the investigation of your claim.
Atrial Fibrillation Detection Using a Novel Cardiac Ambulatory Monitor Based on Photo-Plethysmography at the Wrist

Alberto G. Bonomi, PhD; Fons Schipper, MSc; Linda M. Eerikäinen, MSc; Jenny Margarito, MSc; Ralph van Dinther, PhD; Guido Muesch, MSc; Helma M. de Morree, PhD; Ronald M. Aarts, PhD; Saeed Babaeizadeh, PhD; David D. McManus, MD; Lukas R.C. Dekker, MD

Background—Long-term continuous cardiac monitoring would aid in the early diagnosis and management of atrial fibrillation (AF). This study examined the accuracy of a novel approach for AF detection using photo-plethysmography signals measured from a wrist-based wearable device.

Methods and Results—ECG and contemporaneous pulse data were collected from 2 cohorts of AF patients: AF patients (n=20) undergoing electrical cardioversion (ECV) and AF patients (n=40) that were prescribed for 24 hours ECG Holter in outpatient settings (HOL). Photo-plethysmography and acceleration data were collected at the wrist and processed to determine the inter-pulse interval and discard inter-pulse intervals in presence of motion artifacts. A Markov model was deployed to assess the probability of AF given irregular pattern in inter-pulse interval sequences. The AF detection algorithm was evaluated against clinical rhythm annotations of AF based on ECG interpretation. Photo-plethysmography recordings from apparently healthy volunteers (n=120) were used to establish the false positive AF detection rate of the algorithm. A total of 42 and 855 hours (AF: 21 and 323 hours) of photo-plethysmography data were recorded in the ECV and HOL cohorts, respectively. AF was detected with >96% accuracy (ECV, sensitivity=97%; HOL, sensitivity=93%; both with specificity=100%). Because of motion artifacts, the algorithm did not provide AF classification for 44±16% of the monitoring period in the HOL group. In healthy controls, the algorithm demonstrated a <0.2% false positive AF detection rate.

Conclusions—A novel AF detection algorithm using pulse data from a wrist-wearable device can accurately discriminate rhythm irregularities caused by AF from normal rhythm. (J Am Heart Assoc. 2018;7:e009351. DOI: 10.1161/JAHA.118.009351.)

Key Words: arrhythmia (heart rhythm disorders) • cardioversion • screening • self-management
cardiac monitoring strategies (i.e., implantable cardiac loop recorders) are costly and/or burdensome. Unobtrusive and inexpensive solutions to detect AF would greatly contribute to prevention of AF-related complications, including stroke and heart failure.17

In this study, we sought to investigate the performance of a novel AF detection strategy using pulse data from a wrist-based photo-plethysmography (PPG) recording device. This technology offers a low-cost, unobtrusive, and convenient solution for long-term patient monitoring. The objective was 2-fold: (1) to determine the performance of the algorithm for AF detection in a population of patients with known AF before and after electrical cardioversion and (2) to assess the AF detection accuracy among patients with AF as well as healthy adults.

Methods

Study Sample

A clinical trial was conducted to collect wrist PPG and ECG reference data in 2 different cohorts of patients: (1) 20 AF patients before and after elective electrical cardioversion (ECV) and (2) 40 patients prescribed to undergo 24/48-hour ECG Holter registration in daily living conditions (HOL) who had been diagnosed with paroxysmal or persistent AF. Patients were excluded from participation if they showed presence of any pathological cardiac arrhythmia different from AF. The study was approved by the medical ethical committee of MEC-U (Medical Research Ethics Committees United) in The Netherlands. All participants provided written informed consent. The study was registered with number: NL53827.100.15 at www.toetsingonline.nl. Data, analytical methods, and study materials will not be made available to other researchers for purposes of reproducing the results or replicating the procedure.

Protocol

Pre- and post-cardioversion

Patients admitted for electrical cardioversion for AF were approached for informed consent. Consenting participants were asked to wear a wrist-based PPG measurement device (CM3 Generation-3, Connected Sensing; Philips, Eindhoven, The Netherlands) and a single-lead ECG apparatus (Actiwave Cardio, CamNtech Ltd, Cambridge, UK) before sedation, while lying in bed before the cardioversion procedure. Approximately 60 minutes of pulse and ECG data were acquired before and after the electrical cardioversion from all participants. Before delivery of the electric shock, the ECG measurement device was detached from the patient chest and reattached when the doctor confirmed that the procedure was finished. Cardioversion was successful in restoring normal sinus rhythm (NSR) in all patients. This protocol allowed gathering data during both AF and normal sinus rhythm under controlled, resting conditions.

Holter test

Patients referred for Holter monitoring as part of their usual clinical care were approached from a single center. After obtaining informed consent and after placement of a 12-lead ECG Holter monitor (H12+; Mortara, Milwaukee, WI, USA), the wrist-wearable PPG monitor was positioned on the participant’s non-dominant arm. Participants were asked to wear the PPG monitor at home for the duration of the Holter monitoring period. When the recording period was complete, participants were asked to return the Holter and the wrist-wearable PPG device to the hospital.

Additional Data Sets

Two other data sets were used in this study only for validation purposes of the AF detection algorithm. An overview of all the data sets available for analysis is presented in Table 1.

Healthy control data set

A data set of wrist PPG waveforms recorded in daily living conditions was used to qualitatively assess the likelihood of false positive AF detection in a cohort of apparently healthy subjects. Volunteers were healthy adults recruited from 5 different observational studies. As part of the inclusion criteria, trial participants had to report no history of cardiac disease, including AF. These trials were approved by the Philips internal committee for biomedical experiments, and all participants provided written informed consent that allowed
retrospective use of the collected data for the purpose of this study. Participants were monitored using the wrist-based wearable device over a minimum of 7 hours and up to 4 weeks.

### MIT-BIH AF data set

The MIT-BIH AF (Massachusetts Institute of Technology-Beth Israel Hospital Atrial Fibrillation Database) data set available at PhysioNet consists of 25 long-term ECG recordings of ≈10 hours in duration of patients suffering from paroxysmal AF (www.physionet.org). This data set is commonly used for validation of ECG-based AF detection algorithms because it has a variety of AF cases. In our study, this data set was used to evaluate the ability of the AF detection algorithm to process heartbeat sequences and detect transient episodes of AF. One record (5091) was excluded from our study because AF was only present for less than 1 minute in duration.

### Apparatus

**Wrist-wearable PPG and acceleration sensors**

Cardiac rhythm and body motion were measured using a novel ambulatory monitoring modality embodied in a wrist-wearable device (Figure 1) equipped with the Cardio and Motion Monitoring Module (CM3 Generation-3), which included a PPG sensor and an accelerometer. The PPG sensor used a reflective modality, and the light source was formed by 2 green light LEDs. Sampling frequency of the raw data was 128 Hz, and the dynamic range of the accelerometer was ±8G.

### Table 1. Overview of the Observational Trials and Data Sets Analyzed in This Study

<table>
<thead>
<tr>
<th>Trial / Data Set</th>
<th>Inclusion Criteria</th>
<th>Dropout (No. of Patients)</th>
<th>Recording Duration</th>
<th>Measurement Condition</th>
<th>Types of Data</th>
<th>Reference</th>
<th>Analytical Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECV (n=20)</td>
<td>AF patients admitted for electrical cardioversion</td>
<td>Atrial flutter (2)</td>
<td>1 h before and 1 h after cardioversion</td>
<td>In-hospital bed rest</td>
<td>Wrist PPG ECG</td>
<td>Visual inspection of single-lead ECG</td>
<td>Fine-tuning (50%) and holdout testing (50%) of the AF detection algorithm</td>
</tr>
<tr>
<td>HOL (n=40)</td>
<td>Patients diagnosed with paroxysmal or persistent AF prescribed for ECG Holter</td>
<td>Atrial flutter (5) Atrial tachycardia (1)</td>
<td>24 to 48 h</td>
<td>Daily life</td>
<td>Wrist PPG ECG</td>
<td>Visual inspection of 12-lead ECG</td>
<td>Fine-tuning (50%) and holdout testing (50%) of the AF detection algorithm</td>
</tr>
<tr>
<td>Healthy Control I to V (n=120)</td>
<td>None</td>
<td>None</td>
<td>Daily life</td>
<td>Wrist PPG Not available</td>
<td>None</td>
<td>Visual inspection of single-lead ECG</td>
<td>Assess the likelihood of false-positive AF detection</td>
</tr>
<tr>
<td>Trial I (n=18)</td>
<td>18 to 65 y, M/F healthy adults</td>
<td>None</td>
<td>3 times 24 h</td>
<td>Daily life</td>
<td>Wrist PPG</td>
<td>Not available</td>
<td>Fine-tuning (50%) and holdout testing (50%) of the AF detection algorithm</td>
</tr>
<tr>
<td>Trial II (n=17)</td>
<td>40 to 65 y, M/F healthy adults</td>
<td>None</td>
<td>2 times 8 to 10 h</td>
<td>Bedtime monitoring</td>
<td>Wrist PPG</td>
<td>Not available</td>
<td>Fine-tuning (50%) and holdout testing (50%) of the AF detection algorithm</td>
</tr>
<tr>
<td>Trial III (n=25)</td>
<td>18 to 65 y, M/F healthy adults</td>
<td>None</td>
<td>24 h</td>
<td>Daily life</td>
<td>Wrist PPG</td>
<td>Not available</td>
<td>Fine-tuning (50%) and holdout testing (50%) of the AF detection algorithm</td>
</tr>
<tr>
<td>Trial IV (n=46)</td>
<td>18 to 65 y, M/F healthy adults working in clinical environment</td>
<td>None</td>
<td>3 times 24 h</td>
<td>Daily life (home and work shift)</td>
<td>Wrist PPG</td>
<td>Not available</td>
<td>Fine-tuning (50%) and holdout testing (50%) of the AF detection algorithm</td>
</tr>
<tr>
<td>Trial V (n=14)</td>
<td>18 to 65 y, M/F professional truck drivers</td>
<td>None</td>
<td>5 to 7 d repeated up to 3 times</td>
<td>Daily life (home and on-road shift)</td>
<td>Wrist PPG</td>
<td>Not available</td>
<td>Fine-tuning (50%) and holdout testing (50%) of the AF detection algorithm</td>
</tr>
<tr>
<td>MIT-BIH AF (n=25)</td>
<td>Patients with AF mostly paroxysmal</td>
<td>AF event &lt;1 min (1)</td>
<td>10 h</td>
<td>Daily life</td>
<td>ECG</td>
<td>Visual inspection of ECG</td>
<td>Accuracy assessment for transient AF events</td>
</tr>
</tbody>
</table>

AF indicates atrial fibrillation; ECV, electrical cardioversion; HOL, Holter cohort; MIT-BIH AF, Massachusetts Institute of Technology-Beth Israel Hospital Atrial Fibrillation Database; M/F, male/female; PPG, photo-plethysmography.
activities of daily living was provided in previous publications.\textsuperscript{19–21} The PPG sensor was able to sense changes in the reflected light from the skin caused by blood volume variations directly related to the cardiac systolic and diastolic cycle.

**Data Processing**

**Inter-pulse intervals**

The PPG waveform acquired at the wrist of patients was processed to: (1) determine the temporal location of fiducial points which corresponded to each heartbeat, according to a previously published method,\textsuperscript{22} and (2) calculate the time interval between successive heart beats, here called the inter-pulse interval (IPI). The wrist acceleration signal was used to compute the amount of motion associated with each IPI. In case the motion intensity associated with a certain IPI exceeded a predefined threshold, such IPI was labeled as invalid and discarded from the analysis.\textsuperscript{23}

**AF detection algorithm**

A first-order 11-states Markov model was used to calculate probability of AF given a sequence of IPI. Such an algorithm, inspired by the work of Moody and Mark,\textsuperscript{24} has been previously trained and validated for AF detection on ECG data from Holter recordings.\textsuperscript{25} Output from the Markov model was an AF probability reflecting the relative likelihood of observing a sequence of pulse intervals during AF versus making the same observation outside AF. In our study, the Markov model was adapted in order to process IPI sequences from PPG for AF detection. Such adaptation focused on fine-tuning the probability threshold applied to the Markov model output to identify episodes of AF (see Data S1 for details). Invalid IPIs, such as those obtained in the presence of considerable wrist motion, were not used to assess and update the AF probability. An epoch-by-epoch approach was introduced to determine presence or absence of AF. Epoch length was set to 60 seconds with 50% overlap rate.
(30-second sliding interval). If the sum of valid IPIs within the epoch exceeded 30 seconds, such an epoch was considered of sufficient quality to generate a classification of arrhythmia. Median AF probability associated with each valid IPI within an epoch was compared with a threshold and used to identify presence or absence of AF. When the quality of the epoch was insufficient (eg, the sum of valid IPIs <30 seconds), the epoch was classified as an “unknown rhythm.”

**ECG reference**

ECG was acquired to generate a reference standard for presence or absence of AF. Actiwave Cardio was used in the ECV cohort to gather ECG data at 1024 Hz, and the raw waveform was visually inspected to annotate each detected heartbeat as AF, NSR, premature atrial contraction (PAC), premature ventricular contraction (PVC), or unknown/artefacts. The 12-lead Holter recorder used in the HOL cohort provided as output the temporal location of each detected beat (R-peak), which was used to determine inter-beat intervals (IBI), and the associated classification according to proprietary rhythm classification software (Veritas; Mortara, Milwaukee, WI). An experienced study technician reviewed the ECG recordings and identified AF or other arrhythmias. Rhythm type annotations were processed to generate a ground truth of AF events. The same epoch-by-epoch approach introduced to process the output of the Markov model was applied to the rhythm labels of each IBI after time synchronization with PPG-derived IPIs. AF was defined as any 60-second epoch in which ≥50% of the IBIs carried an AF label. Otherwise, the epoch was labeled as non-AF. The overall data processing method is schematically represented in Figure 1.

**Statistical Analysis**

**Training and testing**

Data from the ECV and HOL cohorts were split in 2 parts (a training and test set) of approximately the same size and including the same amount of AF data. Patients were randomly assigned to either the training or test set. The training data set was used to fine-tune the AF probability threshold used to classify AF from the output of the Markov model; the test data set was used to determine the algorithm accuracy on a hold-out group of patients.

**Literature algorithms**

Several algorithms based on features describing IBI variability and entropy have been proposed in the literature to detect AF. In our study, we selected the normalized absolute deviation (NAdEv) and normalized absolute difference (NADiff), which are normalized variability measures, and the coefficient of the sample entropy (COSen) features to process IPI sequences and compare their AF detection accuracy against the Markov model. Algorithms based on such features were selected because they were extensively validated using large data sets of ambulatory monitoring data (>70 Holter records) as well as the MIT-BIH AF data set, and the optimal threshold to use for AF detection was disclosed in previous studies. In addition, a multiparametric algorithm was developed by averaging a surrogate probability of AF derived from normalized absolute deviation, normalized absolute difference, and coefficient of the sample entropy. The AF probability \( p^\text{AF}(f) \) for each feature \( f \) was determined by applying a logistic function to the normalized difference between the feature value \( f(i) \), where \( i \) indicates the current epoch and the threshold \( \text{Th}^\text{AF}(f) \) for AF detection (as disclosed in the literature; Equation 1).

\[
p^\text{AF}(i) = \frac{1}{1 + e^{-kx_f(i)}} \tag{1}
\]

\[
x_f(i) = \frac{f(i) - \text{Th}^\text{AF}(f)}{\text{max}_f - \text{min}_f} \tag{2}
\]

Normalization was based on the feature range \((\text{max}_f - \text{min}_f)\) obtained from the ECV and HOL training data set (see Equation 2), and \( k \) indicated the steepness of the logistic function curve which was set to 3.5. AF was detected when the average combined probability exceeded 0.5.

**Accuracy and coverage estimation**

AF classification was evaluated by comparing the algorithm output with the ground truth of rhythm annotations. Epoch-by-epoch comparison led to define accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and coverage of the classification method. Coverage was defined as the percentage of monitoring time during which a classification of either AF or non-AF could be made. Episode and duration accuracy was also determined according to the algorithm specified by the Association for the Advancement of Medical Instrumentation standards for ambulatory ECG analyzers. All episodes <1 minute were excluded from episode and duration statistics. Results are reported in 2 ways: gross and average statistics. The gross method calculates the statistics by aggregating all records into one. The average method reports the mean of statistics derived for each individual record. While processing the healthy control data set, given the absence of a rhythm annotation ground truth, the records were all considered as corresponding to non-AF. Data for not-normally distributed variables are represented as median and range; otherwise, mean±SD was used.

**Results**

Among prospectively enrolled study participants, 6 were excluded from the HOL group and 2 from the ECV group.
because of having rhythms other than AF, including atrial flutter or atrial tachycardia. Characteristics of study participants are presented in Table 2. ECG and wrist PPG data were collected for an average of 42 hours (21 hours of AF recordings) in the ECV and 855 hours (323 hours of AF recordings) in the HOL cohort, respectively. An example of the collected PPG data during AF and NSR is shown in Figure 2.

Among participants undergoing cardioversion, 3.5% (113 beats [2–756]) of the beats (median [min–max]) were annotated as PACs and 0% (0 beats [0–1527]) as PVCs. In the HOL trial, 14 participants were in persistent AF throughout the recording period, whereas 22 had NSR. Among participants in NSR, 0.11% (94 beats [5–24 229]) and 0.02% (24 beats [0–7 148]) were PACs and PVCs, respectively. Anthropometric characteristics of healthy volunteers recruited in 5 observational trials used to create the healthy control data set are presented in Table 3. In this data set, a total of around 1 year of wrist PPG data (8815 hours) was available for analysis.

AF probability threshold of the Markov model was established from the receiver operating characteristics (ROC) plot by searching in the region with >95% sensitivity which threshold conferred the largest combined specificity measured as the average value between the ECV and HOL training set. Using this threshold, AF detection accuracy was 96.3/6.8% in the ECV training data set (sensitivity = 97.3/3.7%; specificity = 94.3/13.6%) and 99.5/1.7% in the HOL training data set (sensitivity = 97.7/4.2%; specificity = 99.5/1.7%). Table 4 summarizes the accuracy obtained on the hold-out test data set for the ECV and HOL cohorts. The Markov model offered a larger ROC area under the curve as compared with conventional algorithms (NADev, NADiff, and COSen) and the multiparametric model on both the ECV and HOL test data sets (Figure 3). As a result of motion artefacts in the HOL data set, coverage for AF detection reached 56/16% (35–95%) during the home monitoring period. According to the Association for the Advancement of Medical Instrumentation standards, the accuracy of the AF detection algorithm showed a

<table>
<thead>
<tr>
<th>Table 2. Patients’ Characteristics Included in the ECV and HOL Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ECV (N=18)</strong></td>
</tr>
<tr>
<td><strong>Baseline characteristics</strong></td>
</tr>
<tr>
<td>N (%) of males</td>
</tr>
<tr>
<td>Age, y</td>
</tr>
<tr>
<td>Weight, kg</td>
</tr>
<tr>
<td>Height, m</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
</tr>
<tr>
<td>Monitoring, h</td>
</tr>
<tr>
<td><strong>Comorbidities, n (%)</strong></td>
</tr>
<tr>
<td>Hypertension</td>
</tr>
<tr>
<td>Hyperlipidemia</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td>Coronary artery diseases</td>
</tr>
<tr>
<td>Heart failure</td>
</tr>
<tr>
<td>Valve disease</td>
</tr>
<tr>
<td>Stroke</td>
</tr>
<tr>
<td><strong>Medication, n (%)</strong></td>
</tr>
<tr>
<td>Beta-blocker</td>
</tr>
<tr>
<td>Calcium-channel blockers</td>
</tr>
<tr>
<td>Statin</td>
</tr>
<tr>
<td>Antiarrhythmic</td>
</tr>
<tr>
<td>Digoxin</td>
</tr>
<tr>
<td>Oral anticoagulant</td>
</tr>
</tbody>
</table>

(Min–Max) indicates range of values; BMI, body mass index; ECV, cohort of patients undergoing elective electrical cardioversion; HOL, cohort of patients prescribed for a Holter test; M, mean value; SD, standard deviation.
large sensitivity and positive predictive value on both episode and duration statistics (Table 4). To validate our algorithm, we also examined the performance of the AF detection Markov model using ECG data from the MIT-BIH AF data set. The algorithm demonstrated 94% accuracy in terms of average and gross statistical performance (Table 5).

Among study participants free from arrhythmia, the Markov model algorithm very rarely detected pulse irregularities attributable to AF. Only 811 epochs were labeled as AF and those occurred in only 25 of 120 subjects (21%). Overall, this would correspond to an estimated <0.2% false positive rate in daily living conditions (6.7 hours of AF/367-day monitoring period). Other published algorithms and the multiparametric model showed a much larger number of epochs labeled as AF (from approximately 2 up to 50 times more false positives) on the same data set of healthy controls (Figure 4).

**Discussion**

We demonstrated that a novel algorithm analyzing pulse data from a wrist-based wearable device can accurately detect pulse irregularities attributable to AF in both hospital conditions as well as in daily life. The AF detection algorithm showed superior performance compared with other published algorithms designed to measure variability and entropy of heartbeat sequences. Furthermore, the specificity of the Markov model, and its ability to minimize false positive AF detection (0.2%), was qualitatively verified on a large data set of apparently healthy subjects. As reported in previous studies examining pulse recordings from wrist-based wearables, motion artefacts were common and frequently corrupted PPG waveforms. Because the AF detection algorithm was designed to analyze only uncorrupted pulse recordings, AF detection coverage was, on average, approximately half (56%) of the recording time among participants wearing the wrist-based device in the home setting.

Previous studies presented the accuracy of various techniques to unobtrusively record PPG and identify occurrence of AF from pulse irregularities. Such different techniques could be categorized into solutions aimed at either instantaneous or continuous patient assessment. Applications using data provided by the camera sensor and lamp of a smartphone have been extensively validated to record pulse rate variability at the fingertip and identify AF during a spot check.29,30 Continuous PPG measurements have been also investigated in very recent studies using earlobe-, finger-, and wrist-wearable sensors.31–33 Accuracy of these solutions has been tested mostly under controlled hospital conditions, and AF detection was mainly derived from information carried by variability and entropy measures of IPI sequences. We recently showed that PPG measurements in hospital conditions are substantially different from those carried out in ambulatory settings. Statistical features of IPI sequences trained to classify AF in hospitalized patients did not show

![Figure 2.](https://ahajournals.org/doi/10.1161/JAHA.118.009351)
**Table 3. Subjects Characteristics of Volunteers Included in Healthy Control Data Set**

<table>
<thead>
<tr>
<th>Trial</th>
<th>No.</th>
<th>M (%)</th>
<th>Age, y</th>
<th>Weight, kg</th>
<th>Height, m</th>
<th>BMI, kg/m²</th>
<th>Monitoring, Hours</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>18</td>
<td>10 (56)</td>
<td>32.9±8.5</td>
<td>69.4±12.0</td>
<td>1.75±0.11</td>
<td>22.6±2.0</td>
<td>68.6±8.9</td>
<td>Healthy adults monitored continuously for 3 days</td>
</tr>
<tr>
<td>II</td>
<td>17</td>
<td>9 (53)</td>
<td>49.7±10.1</td>
<td>81.3±10.9</td>
<td>1.78±0.07</td>
<td>25.7±2.8</td>
<td>13.2±4.2</td>
<td>Healthy adults monitored during sleep for 1 or 2 nights</td>
</tr>
<tr>
<td>III</td>
<td>24</td>
<td>11 (46)</td>
<td>39.4±14.7</td>
<td>74.8±11.5</td>
<td>1.75±0.11</td>
<td>24.6±4.5</td>
<td>23.5±1.4</td>
<td>Healthy adults monitored for 1 day</td>
</tr>
<tr>
<td>IV</td>
<td>46</td>
<td>6 (13)</td>
<td>38.6±10.7</td>
<td>66.9±14.2</td>
<td>1.65±0.08</td>
<td>24.7±4.7</td>
<td>78.3±21.4</td>
<td>Clinical professionals monitored for 3 days</td>
</tr>
<tr>
<td>V</td>
<td>15</td>
<td>15 (100)</td>
<td>49.7±7.3</td>
<td>93.1±13.5</td>
<td>1.81±0.07</td>
<td>28.5±4.3</td>
<td>216.3±121.3</td>
<td>Truck drivers monitored for 5 to 7 days and repeated up to 3 times</td>
</tr>
<tr>
<td>ALL</td>
<td>120</td>
<td>51 (43)</td>
<td>40.7±12.2</td>
<td>73.7±15.2</td>
<td>1.72±0.11</td>
<td>24.9±4.3</td>
<td>68.8±69.0</td>
<td></td>
</tr>
</tbody>
</table>

No. number of subjects; data are expressed as mean±SD (Min–Max); ALL, statistics of the combined data sets including the 5 different observational trials (I–V). BMI indicates body mass index.

**Clinical Relevance**

The gold standard for AF diagnosis remains the ECG.9,12 Yet, a commercially available smartwatch. 34 This illustrates the same performance for AF detection using signals recorded outside of controlled hospital environments. 22 The gold standard for AF diagnosis remains the ECG.9,12 Yet, noninvasive ECG monitoring devices available today, such as pulse photoplethysmography (PPG), represent a more-reliable way to detect AF than processing sequences of average heart rate from a smartwatch. 14 This represents a more-reliable way to detect AF than processing sequences of average heart rate from a smartwatch. 14 Our AF detection algorithm consistently and accurately discriminated AF from normal rhythm in both controlled hospital as well as home conditions. The Markov model showed a detrimentally large number of false positives as compared to the Markov model. Fine-tuning their decision threshold for AF detection using PPG data, and perhaps combining and expanding the number of features in a single detection algorithm, could improve the specificity.

AF detection algorithms, when applied to the healthy control data set, showed a detrimentally large number of false positives compared with the MIT-BIH AF data set, showed a detrimentally large number of false positives as compared to the Markov model. Fine-tuning their decision threshold for AF detection using PPG data, and perhaps combining and expanding the number of features in a single detection algorithm, could improve the specificity.
Ambulatory rhythm monitoring with PPG may also be used to better classify a patient’s condition into paroxysmal, persistent, or permanent AF. Clinical distinction between AF categories is needed to personalize treatment and select appropriate therapies for patients. However, data from implantable devices show a large discrepancy between clinical classification of AF patients and the actual AF burden or temporal persistence.37 Charitos et al37 observed that in a 1-year follow-up, only 47% of paroxysmal patients actually suffered from short and self-terminating episodes of AF. Likewise, barely 33% of patients clinically classified as persistent AF showed arrhythmic events >7 consecutive days. This suggests that a wrist-wearable PPG sensor could remove the discordance between clinical assessment of AF and actual AF persistence, and perhaps lead to better identification of patients who would likely respond to the appropriate therapy.

Objective assessment of AF burden has also been stressed as important to confer the correct stroke risk to AF patients. Indeed, a clear trend of increasing stroke risk from paroxysmal to persistent to permanent AF has been suggested from the ACTIVE-A (Atrial Fibrillation Clopidogrel Trial with Irbesartan for Prevention of Vascular Events) and AVERROES (Apixaban Versus Acetylsalicylic Acid [ASA] to Prevent Stroke in Atrial Fibrillation Patients Who Have Failed or Are Unsuitable for Vitamin K Antagonist Treatment) trials.40 Yet, clinical guidelines for stroke prevention do not take the AF burden into account to gauge stroke risk in patients, in part because of the lack of data from long-term rhythm monitors.41 In the near future, convenient and broadly applicable solutions for measuring AF burden using PPG may pave the way to generate fresh insights into relative stroke risk among patients with brief episodes of AF compared with individuals with longer episodes or greater burden of AF.

Clinical relevance of transient, paroxysmal AF detected during acute conditions, such as infection, pneumonia, acute myocardial infarction, or surgery (so-called secondary AF), is unknown. However, recent studies42 suggest that secondary AF may have clinical significance43 despite the commonly held notion that secondary AF is caused by a reversible precipitant.41 Evidence indicates that around half of patients suffering from this type of transient AF would manifest episodes of arrhythmia in a 5-year period.42 In this patient population, a sensitive and convenient method to detect pulse irregularities attributable to AF over a long-term period may help facilitate risk stratification for stroke and health deterioration.

Limitations

A limitation of this study was the lack of transient and short AF episodes in daily living conditions. The HOL cohort only included 14 participants with persistent AF (100% of recording windows). Thus, we could not determine the
impact of loss of monitoring coverage observed during the
day on detection accuracy of intermittent AF. Yet, the
sensitivity of the presented Markov model for detection of
transient AF was validated using the MIT-BIH AF data set,
albeit from ECG data. Future studies are needed to
objectively evaluate the clinical implications of low monitor-
ing coverage for AF detection despite the long-term
monitoring modality offered by the technology. A second
limitation was the absence of a rhythm annotation reference
for the PPG recordings used to test the false positive rate

**Table 5. Accuracy of the Markov Model Designed to Detect Pulse Irregularities Attributable to AF and of Literature Algorithms According to MIT-BIH AF Data Set**

<table>
<thead>
<tr>
<th>Algorithm/Feature</th>
<th>Accuracy % (CI)</th>
<th>Sensitivity % (CI)</th>
<th>Specificity % (CI)</th>
<th>PPV % (CI)</th>
<th>NPV % (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Markov model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>94±12 (89–98)</td>
<td>84±21 (77–90)</td>
<td>98±5 (96–100)</td>
<td>90±21 (80–96)</td>
<td>92±14 (87–97)</td>
</tr>
<tr>
<td>Gross</td>
<td>94</td>
<td>86</td>
<td>99</td>
<td>98</td>
<td>91</td>
</tr>
<tr>
<td><strong>NADev</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>94±6 (92–96)</td>
<td>95±6 (93–98)</td>
<td>89±17 (82–95)</td>
<td>67±36 (52–80)</td>
<td>97±6 (95–99)</td>
</tr>
<tr>
<td>Gross</td>
<td>94</td>
<td>98</td>
<td>92</td>
<td>89</td>
<td>99</td>
</tr>
<tr>
<td><strong>NADiff</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>92±13 (86–97)</td>
<td>92±16 (85–97)</td>
<td>92±15 (87–97)</td>
<td>80±30 (69–93)</td>
<td>95±11 (91–99)</td>
</tr>
<tr>
<td>Gross</td>
<td>92</td>
<td>91</td>
<td>93</td>
<td>89</td>
<td>94</td>
</tr>
<tr>
<td><strong>COSen</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>91±9 (85–94)</td>
<td>89±10 (85–92)</td>
<td>88±16 (82–94)</td>
<td>66±35 (48–78)</td>
<td>91±16 (86–97)</td>
</tr>
<tr>
<td>Gross</td>
<td>91</td>
<td>91</td>
<td>90</td>
<td>86</td>
<td>94</td>
</tr>
<tr>
<td><strong>Multiparametric</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>95±8 (91–97)</td>
<td>96±7 (94–99)</td>
<td>91±15 (85–97)</td>
<td>73±31 (57–84)</td>
<td>98±4 (96–100)</td>
</tr>
<tr>
<td>Gross</td>
<td>95</td>
<td>98</td>
<td>93</td>
<td>90</td>
<td>98</td>
</tr>
</tbody>
</table>

Average, indicates accuracy described by mean ± SD of the value across patients; Gross, accuracy determined by aggregating results from each patient into one; AF indicates atrial fibrillation; CI, 95% confidence interval of the mean as obtained by bootstrapping and resampling of the data set (100 iterations); NPV, negative predictive value; PPV, positive predictive value. Average indicates accuracy described by meanSD of the value across patients. Any AF event <1 minute was removed from the analysis, and 1 record (5091) was discarded because AF was only present for <1 minute.
Figure 4. Histogram indicating the number of subjects for each categorical amount of atrial fibrillation (AF) epochs as output by the AF detection algorithm (Markov model), literature algorithms (NADev, NADiff, and COSen) and a multiparametric algorithm (combining NADev, NADiff, and COSen) with the control data set of apparently healthy subjects. COSen indicates coefficients of the sample entropy; NADev, normalized absolute deviation; NADiff, normalized absolute difference.
the AF detection algorithm in the healthy control data set. In this study, we assumed that volunteers recruited in the different observational studies were not suffering from AF. It is possible that these participants experienced brief episodes of undetected AF given that they were not wearing contemporaneous Holter monitors. Another important limitation of this study was that the diagnostic accuracy of the Markov model for AF discrimination is unproven in populations affected by other cardiac rhythm disorders, such as atrial flutter or tachycardia. Among the few dropouts of the HOL cohort suffering from continuous atrial flutter, we noticed that AF was rarely detected in some patients (2 patients, specificity: 96.4% and 92.1% with coverage >50%), but often in others (3 patients, specificity: 7.5%, 17.6%, and 2.8% with coverage >50%). In addition, the Markov model detected AF only occasionally in the dropout patient from the HOL trial suffering from continuous atrial tachycardia (specificity, 78.8% with coverage >50%). Follow-up studies are needed to collect more data from large cohorts of free-living individuals affected by paroxysmal AF and other cardiac rhythm abnormalities to confirm the algorithm’s performance for AF detection.

Conclusion
A novel algorithm analyzing pulse data from a wrist-wearable device can accurately detect pulse irregularities associated with AF. Our findings suggest that wrist-based wearables capable of pulse monitoring can be used as long-term monitors for AF and thereby enhance detection of paroxysmal AF and assessment of AF burden among patients affected by this dysrhythmia.

Acknowledgments
The authors thank N. Sturkenboom (MD), L. Verborg (MD), R. Eerdekens (MD), L. van den Heuvel, the personnel of the Cardiostim and Holter departments, and the Holter analysts of the Catharina Hospital for their help and contributions in the data collection.

Sources of Funding
This research was performed within the framework of the strategic joint research program on Data Science between TU/e, Catharina Hospital, and Philips Electronics Nederland B.V.

Disclosures
Bonomi, Margarito, van Dinther, Muesch, de Morree, and Aarts are employed by Philips Research. Schipper and Babaeizadeh are employed by Philips Healthcare.


MARKOV MODEL TO DETECT ATRIAL FIBRILLATION FROM PULSE INTERVALS

This section describes the details of the Markov model used to determine the likelihood of AF from a sequence of inter-pulse-intervals (IPI). The IPI sequence is modeled as an 11-states first-order Markov process in which each observed IPI is associated to one of 11 different possible states. This association is done firstly by determining the ratio between the observed IPI and the average IPI in the sequence, and secondly by classifying such ratio into 11 categories of values ranging from 0.75 to 1.25 at steps of 0.05. This corresponds to identify to which extent the observed IPI is shorter, similar, or larger than average. In accordance with the work of Moody and Mark\textsuperscript{1}, the average IPI ($\overline{IPI}$) was determined recursively as follows:

$$\overline{IPI}(i) = 0.75 \times \overline{IPI}(i - 1) + 0.25 \times IPI(i)$$

ECG recordings were used in a previous study\textsuperscript{2} as training data to define the statistics of transition from one state to another and derive transition probability matrixes in either presence or in absence of AF. Assuming a sequence $X$ of $n$ states:

$$X = \{x_1, x_2, x_3, \ldots x_n\} \quad (x_i \in \{s_{75}, s_{80}, s_{85}, \ldots s_{125}\})$$

where $s$ indicates one of the 11-states of the model, the transition probability matrix in presence of AF or absence of AF (not\textunderscore AF) can be defined as:

$$p_{j,k,AF} = P(x_j \mid x_k, AF)$$
\[ p_{j,k,\text{not}_{-}AF} = P(x_j \mid x_k, \text{not}_{-}AF) \]

Where \( x_j \) and \( x_k \) are two consecutive states in the process and \( p_{j,k,R} \) their relative frequency during a certain heart rhythm \( R \) (\( AF \) or \( \text{not}_{-}AF \)). From this we can derive the conditional probability of \( X \) given the process \( R \) (either \( AF \) or \( \text{not}_{-}AF \)):

\[
P(X \mid R) = \prod_{i=1}^{n-1} p_{i+1,i,R}
\]

The rhythm \( R \) that maximizes \( P \) indicates the maximum likelihood procedure for the observed sequence of states \( X \). As suggested by Moody and Mark\(^1\), the logarithm of the transition matrix for \( \text{not}_{-}AF \) was divided by that of \( AF \) and used to determine the Markov score \( (S_i) \) iteratively by adding each element in the resulting matrix according to the transitions between the \( n \) observed states \( X \).

\[
S_i = \sum_{i=1}^{n-1} \log \left( \frac{p_{i+1,i,\text{not}_{-}AF}}{p_{i+1,i,AF}} \right)
\]

In our work \( n \) was kept equal to the entire length of the IPI sequence. As a final step the Markov score was processed using a previously published averaging filter\(^1\). The resulting Markov score \( (S_i) \) tended to decrease below zero in presence of \( AF \) and such value was compared to a threshold to make a binary decision between presence and absence of \( AF \). The Markov model which was designed and trained using ECG data was optimized in our study to process IPI sequences from PPG for \( AF \) detection. Such optimization focused solely on fine-tuning the probability threshold applied to the Markov score \( (S_i) \) to identify episodes of \( AF \).
Supplemental References:
