Spectral Boundary Adaptation on Heart Rate Variability for Sleep and Wake Classification

Xi Long†‡, Pedro Fonseca†‡, Reinder Haakma†, Ronald M. Aarts†‡, Jerome Foussier§

†Philips Research
High Tech Campus, Prof. Holstlaan 4, 5656 AE, Eindhoven, The Netherlands
{xi.long, pedro.fonseca, reinder.haakma, ronald.m.aarts}@philips.com
‡Department of Electrical Engineering, Eindhoven University of Technology
Den Dolech 2, 5612 AZ, Eindhoven, The Netherlands
§Chair for Medical Information Technology (MedIT), RWTH Aachen University
Pauwelsstrasse 20, 52074, Germany
foussier@hia.rwth-aachen.de

Received 4 April 2013
Accepted 6 December 2013
Published 28 May 2014

A method of adapting the boundaries when extracting the spectral features from heart rate variability (HRV) for sleep and wake classification is described. HRV series can be derived from electrocardiogram (ECG) signals obtained from single-night polysomnography (PSG) recordings. Conventionally, the HRV spectral features are extracted from the spectrum of an HRV series with fixed boundaries specifying bands of very low frequency (VLF), low frequency (LF), and high frequency (HF). However, because they are fixed, they may fail to accurately reflect certain aspects of autonomic nervous activity which in turn may limit their discriminative power, e.g. in sleep and wake classification. This is in part related to the fact that the sympathetic tone (partially reflected in the LF band) and the respiratory activity (modulated in the HF band) vary over time. In order to minimize the impact of these variations, we adapt the HRV spectral boundaries using time-frequency analysis. Experiments were conducted on a data set acquired from two groups with 15 healthy and 15 insomnia subjects each. Results show that adapting the HRV spectral features significantly increased their discriminative power when classifying sleep and wake. Additionally, this method also provided a significant improvement of the overall classification performance when used in combination with other HRV non-spectral features. Furthermore, compared with the use of actigraphy, the classification performed better when combining it with the HRV features.

Keywords: Heart rate variability; sleep and wake classification; time-frequency analysis; boundary adaptation; feature extraction.

*The article is invited for the Special Issue based on the work presented at the 12th International Conference on BioInformatics and BioEngineering (BIBE 2012) that was awarded the First Prize in the student paper competition.
1. Introduction

Sleep plays an important role in human health. Night-time polysomnography (PSG) recordings, along with manually scored hypnograms, are considered the “gold standard” for objectively analyzing sleep architecture and occurrence of sleep-related problems. PSG recordings are typically recorded and analyzed in sleep laboratories, and are usually split into non-overlapping time intervals (or epochs) of 30 seconds according to the Rechtschaffen & Kales (R&K) rules.2,3

As shown in literature, monitoring heart rate variability (HRV) during bedtime is helpful in sleep staging,3,4 particularly to distinguish between rapid-eye-movement (REM) and non rapid-eye-movement (NREM).5,6 It reflects the variation, over time, of the period between consecutive heart beats. HRV is derived from the length variations of RR-intervals, i.e. time intervals between consecutive R-peaks of the QRS complex in the electrocardiogram (ECG). Spectral analysis of HRV has been widely employed in the assessment of autonomic nervous activity during bedtime.5–7

It traditionally involves the computation of the power spectral density (PSD) of an HRV series. An HRV spectrum is typically divided in three bands, namely in a very low frequency (VLF) band from 0.003 to 0.04 Hz, a low frequency (LF) band from 0.04 to 0.15 Hz, and a high frequency (HF) band between 0.15 and 0.4 Hz.8,9 These bands are then be used to compute certain properties such as the spectral power of the VLF, LF, and HF components and the power ratio of low-to-high frequency (LF/HF) components.5,10,11 In general, it has been found that the VLF spectral power is associated with long-term regulatory mechanisms, the LF spectral power is a marker of sympathetic modulation of the heart and it also reflects some parasympathetic activity when the respiratory frequency components partially fall into the LF band, the HF spectral power is related to parasympathetic activity mainly caused by respiratory sinus arrhythmia (RSA), and the LF/HF ratio is an indication of sympathetic-parasympathetic balance.9,11,12 In particular, the HRV spectrum usually contains a peak centered around the respiration frequency, located in the HF band, and another peak in the LF band which reflects, to a certain degree, sympathetic activation.8,13,14

The parameters derived from HRV PSD are often used as “features” in automatic sleep staging3 or sleep and wake classification systems.4 Previous work has used HRV spectral features with fixed boundaries for sleep and wake classification.4 This classifier exploits the fact that sympathetic tone and the respiratory activity are modulated in different frequency bands of the HRV spectrum and exhibit different properties during sleep and wake, allowing them to be distinguished.

It is known that the HRV spectrum and the dominant (or peak) frequencies of the LF and HF bands are not constant but rather vary over time according to the autonomic modulations of the heart beats.9 Hence, as long as fixed band boundaries are used to compute HRV spectral features, we might produce inaccurate estimates of cardiac autonomic activities. Since the discrimination of sleep states (or sleep and wake in our case) depends on these estimates, the classification accuracy will be
affected. To avoid this issue, we will use a feature adaption method while estimating the HRV features.

The problem of boundary adaptation has been analyzed before in other areas such as stress detection\textsuperscript{15,16} and anesthesia analysis.\textsuperscript{17} It has been suggested that the LF and HF boundaries are related to the peak frequency in the traditional LF band, called “LF peak frequency”, and the peak frequency in the traditional HF band, called “HF peak frequency”, respectively.\textsuperscript{16,17} In practice, these two peak frequencies can be estimated by determining the frequency of local maximum in the band between 0.003 and 0.15 Hz (i.e. the traditional VLF band and LF band) and in the band from 0.15 to 0.4 Hz (i.e. the traditional HF band), respectively. The working assumption here is that the peaks always fall within those two bands. By centering the new bands around these peaks instead of using fixed boundaries, we can compensate for their time-varying behavior. This should help, to some extent, reduce within- and between-subject variabilities in the way these features express sympathetic activation and respiratory activity, ultimately helping improve sleep and wake classification. Figure 1 shows an example of the mean HRV PSDs with standard errors (standard deviations) for sleep and wake states of a subject. It can be observed that, although their standard errors overlap, their mean values are not the same in different frequency ranges. This should provide an opportunity of discriminating between sleep and wake states. Figure 2 illustrates the time variation of the HRV PSD for a subject.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{HRV_PSD.pdf}
\caption{An example of the mean HRV PSD with standard errors for sleep and wake states over an entire-night’s recording of a subject.}
\end{figure}
2. Method

The proposed boundary adaptation method applied on HRV spectral features used for sleep and wake classification is described by a block diagram in Figure 3. Each block will be explained further in the following subsections.

![Block diagram of the feature adaptation method used for sleep and wake classification.](image)
2.1. Data acquisition

In total the data acquired from 30 subjects were used in our experiment. Fifteen subjects belong to healthy group and fifteen subjects are insomniacs. The insomniacs were randomly selected from a larger-sized group in order to evenly compare the classification performance between the healthy and insomnia groups, from which we ensured that the numbers of subjects are equal. A subject was considered healthy if his/her Pittsburgh Sleep Quality Index (PSQI) was less than 6, while a subject was considered insomnia based on his/her self-report. For each subject, a full PSG was recorded according to the guidelines of the American Academy of Sleep Medicine (AASM). Among the 30 subjects, the PSG recordings of fifteen insomniacs and nine healthy subjects were recorded in the Sleep Health Center, Boston, USA during 2009 (Alice 5 PSG, Philips Respironics) and of the remaining six healthy subjects in the Philips Experience Lab, Eindhoven, The Netherlands during 2010 (Vitatport 3 PSG, TEMEC). The ECG was recorded with a modified V2 Lead, sampled at 500 Hz (Boston data) and 256 Hz (Eindhoven data).

Sleep stages were manually scored on 30-s epochs by sleep experts according to the AASM guidelines as wake, REM, and each of the NREM stages (N1-N3). NREM and REM stages were further combined in a single sleep class. Each PSG recording was manually clipped to the time interval comprised between the instant when the subject turned the lights off with the intention of sleeping until the moment the lights were turned on before the subject got out of bed in the morning. The study protocol was approved by the Ethics Committee of the Center and all subjects signed an informed consent. The subject demographics including sex, age, body mass index or BMI, and sleep efficiency are summarized in Table 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean ± Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>5 males and 10 females</td>
</tr>
<tr>
<td>Age (years)</td>
<td>31.0 ± 10.4</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>24.4 ± 3.3</td>
</tr>
<tr>
<td>Sleep Efficiency (%)</td>
<td>92.3 ± 3.8</td>
</tr>
<tr>
<td>(N=15)</td>
<td></td>
</tr>
<tr>
<td>Insomnia</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>8 males and 7 females</td>
</tr>
<tr>
<td>Age (years)</td>
<td>47.4 ± 14.5</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>27.7 ± 4.5</td>
</tr>
<tr>
<td>Sleep Efficiency (%)</td>
<td>69.7 ± 14.7</td>
</tr>
<tr>
<td>(N=15)</td>
<td></td>
</tr>
</tbody>
</table>

2.2. PSD estimation

To estimate the PSDs of HRV, RR-intervals were first computed from the ECG signals. In our study, the following steps were performed to obtain an RR interval series.
A peak detector, based on the Afonso-Tompkins filter-bank algorithm,\textsuperscript{19} was used to locate the R peaks, yielding an RR-interval series.

Very short (less than 0.3 seconds) and long (more than 2 seconds) RR intervals (usually caused by ectopic heart beats, misidentification of R peaks, or badly attached electrodes during measurement) were removed.

The RR-interval series was normalized by dividing it by the mean value.

The resulting series was “re-sampled” at 4 Hz using linear interpolation.

Finally, the PSD was estimated with an autoregressive model with adaptive order automatically determined using the Akaike’s information criterion (AIC).\textsuperscript{20}

2.3. Boundary adaptation

As explained in Section 1, the use of fixed boundaries in HRV spectrum may not be appropriate to accurately represent different states of the autonomic nervous system and further to classify sleep and wake. The respiratory frequency, and therefore the corresponding peak in the HF band vary in time. Likewise, the peak corresponding to the sympathetic tone in the LF band also varies, reflecting differences in the autonomic activation during sleep. By applying a time-frequency analysis, the boundaries that define each band can be dynamically adapted so that the frequency components can be more correctly assigned to the corresponding bands. To do this, it is required to estimate the LF and HF peak frequencies, which change over time. Figure 4 illustrates, with a filled contour plot, an example of the HRV spectrum over time for a subject together with the traditional fixed frequency bands. As it can be easily seen, the dominant LF and HF peak frequencies vary over time. Moreover, it can be observed that, for some epochs, the spectral power of a frequency band spills over its neighboring bands when using the fixed boundaries. For instance, for the epochs from 140 to 150, the spectral power of the LF band also partially falls into the HF band (see Figure 4).

By adapting the boundaries of the LF and HF bands for each epoch, we can overcome the issues mentioned above. This can be achieved in the following way.

- The new high-frequency band (HF$^*$) is centered on the HF peak frequency,\textsuperscript{15,16} and has a constant bandwidth of 0.1 Hz.\textsuperscript{21} This bandwidth was chosen after analyzing the HRV PSDs of all 15 healthy subjects and empirically determining that most of the spectral power related to RSA lie within a bandwidth of 0.1 Hz. A larger bandwidth (0.25 Hz) was empirically used in other work,\textsuperscript{15,22} but we found that in some occasions it overlapped its adjacent LF band.
- The new low-frequency band (LF$^*$) is centered on the dominant frequency found in the traditional LF band, and has a bandwidth of 0.11 Hz that is similar to the traditional definition.
- The new very-low-frequency band (VLF$^*$) is defined from its traditional lower limit of 0.003 Hz up to the lower limit of the LF band.
Fig. 4. HRV spectrum versus time (30-s epoch) of a subject. The fixed boundaries of the VLF, HF, and LF bands are plotted in solid lines and the corresponding bands are indicated.

Figure 5 illustrates the adapted boundaries for the same HRV PSD shown in Figure 4. It should be noted that the LF$^*$ and HF$^*$ bands overlap in some epochs. This occurs when the LF and HF peaks are too close to each other or when there is no HF peak (often during REM sleep$^{23}$).

2.4. Feature extraction

2.4.1. HRV spectral features

After determining the bands we can extract HRV-related features for sleep and wake classification. In our study we computed the logarithm of the spectral power in the VLF$^*$, LF$^*$, and HF$^*$ bands (from here on expressed as $hrv_{vlf}$, $hrv_{lf}$, and $hrv_{hf}$) and, in addition, the ratio between the spectral powers of the LF$^*$ and the HF$^*$ bands (expressed as $hrv_{lf/hf}$). Before computing the logarithm, the power of each band was first normalized. This was achieved by dividing the power in the VLF$^*$, LF$^*$, and HF$^*$ bands by the total spectrum power.$^{10,24}$ Alternatively we could have normalized it by dividing the power in each band by the total spectrum power minus the power in the VLF$^*$ band.$^5,7$ Since we did not observe any significant differences in the final result, the first method was used.

2.4.2. Spectrum information

As mentioned, the adaptation of the new boundaries requires knowledge of information derived from the power spectrum, namely the LF and HF peak frequencies,
Fig. 5. HRV spectrum versus time (30-s epoch) of a subject. The limits of the new HF* and LF* bands are plotted in dotted and solid curves, respectively. The lower boundary of the new VLF* band (at 0.003 Hz) is plotted as a dashed line.

which must be obtained before extracting the features. The LF peak frequency can be estimated by detecting the location of the peak in the HRV spectral range from 0.003 to 0.15 Hz. The HF peak frequency can be estimated from a respiratory effort signal simultaneously recorded with the PSG data or it can be derived from the HRV series directly by searching for the peak in the range between 0.15 and 0.4 Hz. In this study, to avoid using an additional sensor modality, we used the latter approach.

2.5. Feature evaluation

A Hellinger distance metric\(^{25}\) was employed to evaluate the discriminative power (i.e. class separability) of the HRV spectral features between sleep and wake. It is estimated by computing the amount of overlap between two probability density estimates in a binary class problem, expressed as

\[
D_H(p, q) = \sqrt{1 - \sum \sqrt{p(x)q(x)}}
\]

(1)

where \(p(x)\) and \(q(x)\) are the probability density estimates of the feature values given class sleep and wake, respectively. In its most basic form, these density estimates can be computed by means of a normalized histogram with either a fixed number of bins or a specific bin size. In our study the histograms were computed with a fixed number of 100 bins. A larger Hellinger distance reflects a higher discriminative power in separating the two classes.
2.6. Sleep and wake classification

It has been demonstrated that a linear discriminant (LD-) based classifier is appropriate for the task of sleep and wake classification. Assuming that all features are normally distributed and their covariance matrices for the two classes are identical, the “linear discriminant” function is given by

\[ G_c(\vec{f}) = \frac{1}{2} (\vec{f} - \mu_c)' \Sigma^{-1} (\vec{f} - \mu_c) + \ln P(c) \]  

(2)

where \( \mu_c \) is the mean vector of the feature vector \( \vec{f} \), \( \Sigma \) is the pooled covariance matrix, and \( P(c) \) expresses the prior probability of class \( c \). In this study \( c = \text{sleep} \) as negative class or \( c = \text{wake} \) as positive class. Based on a feature vector, the epoch is assigned to one class when the computed discriminant score of this class minus that of the other class is higher than a decision making threshold \( \tau \) (here we chose \( \tau = 0 \)). For instance, an epoch is classified as \text{sleep} if \( G_{\text{sleep}}(\vec{f}) - G_{\text{wake}}(\vec{f}) > 0 \) for this epoch. Because quadratic discriminants are known to require larger sample sizes than linear discriminants and they seem to be more sensitive to possible violations of the basic assumptions of normality, the linear discriminant was used instead.

In regard to the prior probability \( P(c) \), it can be observed that the probabilities of different classes vary throughout the night. This prior probability is typically estimated during training procedure. For a given class, for example, the probability of being asleep in the middle of the night is much higher than just right after entering the bed or at the end of the night. In order to exploit these variations, instead of using a fixed prior probability we computed a time-varying prior probability for each epoch by counting the number of times that specific epoch (relative to the instant when lights were turned off) was annotated as each class. It should be pointed out that a prior probability ‘emphasis’ factor (or weight) \( \gamma (\gamma \in [0, 1]) \) is used to bias the classifier towards a pre-defined class, meaning that it can set a higher barrier of being identified to one class and at the same time a lower one to another class during decision making. Because the classes are imbalanced with much more \text{sleep} epochs than \text{wake} epochs (this will be explained later), yielding a very low prior probability of \text{wake} epochs in our study, we use this factor to “emphasize” the \text{wake} class and meanwhile “penalize” the \text{sleep} class. Therefore, the new time-varying prior probabilities after adding emphasis factor of the two classes are \( P'(\text{sleep}) = \gamma \cdot P(\text{sleep}) \) and \( P'(\text{wake}) = 1 - P'(\text{sleep}) \), where the factor \( \gamma \) of 0.79 was experimentally chosen as a proper value in the case of sleep and wake classification.

2.7. Classifier evaluation

To assess the performance of this classifier, conventional measures of sensitivity (proportion of correctly identified actual \text{wake} epochs) and specificity (proportion of correctly identified actual \text{sleep} epochs) are often used. They can be calculated as sensitivity = \( \frac{TP}{TP+FN} \) and specificity = \( \frac{TN}{TN+FP} \), with TP, TN, FP, and FN indicating the number of true positive, true negative, false positive, and false negative classifications, respectively. However, these two measures are not the most adequate
criteria for this binary classification problem. The reason is that the number of
epochs of the wake class during an entire recording lasting the whole night is natu-
really smaller than the number of epochs of the sleep class, in what is usually called
“imbalanced class distribution”. On average, the sleep and wake classes account for
respectively 92.3% and 7.7% of all epochs for the healthy subjects, and respectively
69.7% and 30.3% of all epochs in the insomnia group.

The Cohen’s Kappa coefficient of agreement\textsuperscript{31} (denoted as $\kappa$) not only provides
a better understanding of the general performance of the classifier in correctly iden-
tifying both classes, but also allows for a better interpretation of the imbalanced
problem when it is used as a criterion to optimize performance.\textsuperscript{32} Although it indi-
cates how well a classifier performs for both classes, evaluating a method with
this metric that represents a single point in the entire solution space might not be
sufficient.\textsuperscript{33} An alternative is to use a receiver operating characteristic (ROC) curve
which plots the true positive rate (i.e. sensitivity or recall) versus false positive rate
(i.e. one minus specificity) thus illustrating the classifier’s performance over the en-
tire solution space by means of varying a decision making threshold.\textsuperscript{34} However, the
ROC curve has been shown to be over-optimistic when there is a heavy imbalance
between two classes,\textsuperscript{35} for instance, sleep and wake in the healthy group. Hence, a so-
called Precision-Recall (PR) curve that plots precision versus recall is used instead,
where precision = \frac{TP}{TP + FP}, measures the positive predictive value. When comparing
different classifiers, a larger “area under the PR curve” (AUC\textsubscript{PR}) or “area under
the ROC curve” (AUC\textsubscript{ROC}) indicates a better performance. In this study, the three
metrics ($\kappa$, AUC\textsubscript{ROC} and AUC\textsubscript{PR}) were used to evaluate the performance of sleep
and wake classification with and without HRV boundary adaptation.

In addition, we combined the HRV spectral features with some other HRV (non-
spectral) features selected from the feature set used in previous work,\textsuperscript{4} including
time domain features,\textsuperscript{3,4} nonlinear measures extracted using detrended fluctuation
analysis\textsuperscript{36} and sample entropy.\textsuperscript{37} Five HRV non-spectral features were selected using
the feature selection method described in Ref. 30. This serves the purpose of
examining whether the feature adaptation method described in this paper can help
improve the classification performance when combined with other relevant features.
Note that all features were extracted from the same HRV series. Besides, we com-
pared the results with those obtained using the actigraphy feature (activity counts
over a 30-s epoch, expressed as ac), a well-known feature for sleep and wake clas-
sification.\textsuperscript{38} Finally, we also examined the classification performance by combining
the HRV features with this actigraphy feature.

3. Results

A leave-one-subject-out cross-validation (LOSOCV) procedure was conducted to
assess the discriminative power of the HRV spectral features and also to assess the
performance of our classifier. Table 2 compares the discriminative power (as mea-
sured by a Hellinger distance $D_H$) of the HRV spectral features using the traditional
Table 2. Discriminative power comparison of the HRV spectral features for healthy and insomnia groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>Feature</th>
<th>Fixed Boundaries</th>
<th>Adaptive Boundaries</th>
<th>$p$ value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>$h_{\text{L}f}$</td>
<td>0.19 ± 0.02</td>
<td>0.22 ± 0.01</td>
<td>0.0004</td>
</tr>
<tr>
<td></td>
<td>$h_{\text{L}f}$</td>
<td>0.25 ± 0.01</td>
<td>0.26 ± 0.01</td>
<td>0.0026</td>
</tr>
<tr>
<td></td>
<td>$h_{\text{H}f}$</td>
<td>0.23 ± 0.01</td>
<td>0.29 ± 0.01</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>$h_{\text{L}f/hf}$</td>
<td>0.22 ± 0.01</td>
<td>0.27 ± 0.01</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>$h_{\text{S}}$</td>
<td>0.49 ± 0.02</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Insomnia</td>
<td>$h_{\text{L}f}$</td>
<td>0.13 ± 0.01</td>
<td>0.14 ± 0.01</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>$h_{\text{L}f}$</td>
<td>0.18 ± 0.01</td>
<td>0.19 ± 0.01</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>$h_{\text{H}f}$</td>
<td>0.17 ± 0.01</td>
<td>0.21 ± 0.01</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>$h_{\text{L}f/hf}$</td>
<td>0.20 ± 0.01</td>
<td>0.21 ± 0.01</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>$h_{\text{S}}$</td>
<td>0.41 ± 0.02</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>

Note: *Significance of difference between using fixed and using adaptive boundaries was examined with a paired Wilcoxon signed-rank test.

fixed boundaries and using the adaptive boundaries for healthy and insomnia subjects. They were obtained by averaging the results computed based on training data over all iterations of the LOSOCV process.

Tables 3 and 4 summarize the classification performance obtained with and without boundary adaptation using different sets of features for the healthy and

Table 3. Classification performance (mean ± standard deviation of accuracy, sensitivity, and specificity) for healthy and insomnia groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>Feature Set</th>
<th>Accuracy (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>Actigraphy feature</td>
<td>94.8 ± 2.7</td>
<td>46.8 ± 19.6</td>
<td>99.1 ± 1.0</td>
</tr>
<tr>
<td></td>
<td>HRV spectral features (F)</td>
<td>90.3 ± 9.0</td>
<td>32.7 ± 14.1</td>
<td>95.4 ± 9.6</td>
</tr>
<tr>
<td></td>
<td>HRV spectral features (A)</td>
<td>89.3 ± 10.7</td>
<td>33.9 ± 13.8</td>
<td>94.1 ± 11.6</td>
</tr>
<tr>
<td></td>
<td>HRV features* (F)</td>
<td>89.9 ± 8.5</td>
<td>50.6 ± 13.4</td>
<td>93.3 ± 8.9</td>
</tr>
<tr>
<td></td>
<td>HRV features* (A)</td>
<td>93.1 ± 4.2</td>
<td>49.7 ± 19.2</td>
<td>96.6 ± 3.3</td>
</tr>
<tr>
<td></td>
<td>Actigraphy + HRV features* (F)</td>
<td>95.7 ± 2.1</td>
<td>56.9 ± 16.4</td>
<td>99.0 ± 0.9</td>
</tr>
<tr>
<td></td>
<td>Actigraphy + HRV features* (A)</td>
<td>95.8 ± 2.2</td>
<td>58.1 ± 18.0</td>
<td>99.1 ± 0.9</td>
</tr>
<tr>
<td>Insomnia</td>
<td>Actigraphy feature</td>
<td>79.1 ± 12.1</td>
<td>47.2 ± 17.1</td>
<td>95.3 ± 3.5</td>
</tr>
<tr>
<td></td>
<td>HRV spectral features (F)</td>
<td>65.2 ± 12.9</td>
<td>42.9 ± 20.0</td>
<td>78.6 ± 17.3</td>
</tr>
<tr>
<td></td>
<td>HRV spectral features (A)</td>
<td>69.0 ± 11.2</td>
<td>49.1 ± 16.5</td>
<td>78.5 ± 12.5</td>
</tr>
<tr>
<td></td>
<td>HRV features* (F)</td>
<td>70.1 ± 9.8</td>
<td>54.6 ± 18.4</td>
<td>80.7 ± 14.2</td>
</tr>
<tr>
<td></td>
<td>HRV features* (A)</td>
<td>72.9 ± 8.8</td>
<td>53.2 ± 15.3</td>
<td>83.0 ± 12.2</td>
</tr>
<tr>
<td></td>
<td>Actigraphy + HRV features* (F)</td>
<td>79.2 ± 10.8</td>
<td>57.4 ± 17.6</td>
<td>91.5 ± 9.0</td>
</tr>
<tr>
<td></td>
<td>Actigraphy + HRV features* (A)</td>
<td>80.6 ± 8.5</td>
<td>57.8 ± 16.9</td>
<td>92.2 ± 8.2</td>
</tr>
</tbody>
</table>

Note: F: using fixed boundaries (without adaptation) on the HRV spectral features; A: using adaptive boundaries (with adaptation) on the HRV spectral features.

*The HRV features consist of the spectral features and the non-spectral features selected from a larger feature set used in Ref. 4.
Table 4. Classification performance (mean ± standard deviation of $\kappa$ and pooled $\text{AUC}_{\text{PR}}$ and $\text{AUC}_{\text{ROC}}$) for healthy and insomnia groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>Feature Set</th>
<th>$\kappa$</th>
<th>$\text{AUC}_{\text{PR}}$</th>
<th>$\text{AUC}_{\text{ROC}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>Actigraphy feature</td>
<td>0.53 ± 0.15</td>
<td>0.67</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>HRV spectral features (F)</td>
<td>0.33 ± 0.18</td>
<td>0.30</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>HRV spectral features (A)</td>
<td>0.33 ± 0.19</td>
<td>0.36</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>HRV features† (F)</td>
<td>0.44 ± 0.25</td>
<td>0.51</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>HRV features† (A)</td>
<td>0.48 ± 0.24†</td>
<td>0.54</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>Actigraphy + HRV features* (F)</td>
<td>0.63 ± 0.10</td>
<td>0.71</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>Actigraphy + HRV features* (A)</td>
<td>0.64 ± 0.13</td>
<td>0.72</td>
<td>0.90</td>
</tr>
<tr>
<td>Insomnia</td>
<td>Actigraphy feature</td>
<td>0.45 ± 0.17</td>
<td>0.64</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>HRV spectral features (F)</td>
<td>0.20 ± 0.14</td>
<td>0.48</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>HRV spectral features (A)</td>
<td>0.25 ± 0.13†</td>
<td>0.52</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>HRV features* (F)</td>
<td>0.31 ± 0.11</td>
<td>0.56</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>HRV features* (A)</td>
<td>0.34 ± 0.12†</td>
<td>0.59</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>Actigraphy + HRV features* (F)</td>
<td>0.47 ± 0.17</td>
<td>0.70</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>Actigraphy + HRV features* (A)</td>
<td>0.50 ± 0.14†</td>
<td>0.72</td>
<td>0.81</td>
</tr>
</tbody>
</table>

*Note: F: using fixed boundaries (without adaptation) on the HRV spectral features; A: using adaptive boundaries (with adaptation) on the HRV spectral features.

†The difference between using fixed and using adaptive boundaries is significant, examined with a paired Wilcoxon signed-rank test (with $p < 0.05$).

### 4. Discussion

#### 4.1. Discriminative power

Table 2 shows that, after using the adaptation method proposed in this paper, the discriminative power of the HRV spectral features are significantly increased for the subjects in both healthy and insomnia groups (with a paired Wilcoxon signed-rank test). For comparison, the table also indicates the Hellinger distance of the actigraphy feature $\text{ac}$. Although the feature adaptation helps, to different extents, improving the discriminative power of each HRV spectral feature, it is still relatively lower than that of the actigraphy feature which addresses body motion during bedtime. As known in literature, body motion activity often happens during wake states.38,39
Spectral Boundary Adaptation on Heart Rate Variability

Fig. 6. Pooled ROC curves for sleep and wake classification using HRV spectral features, HRV features, and actigraphy feature with and without adaptation for healthy subjects (top) and insomniacs (bottom).
Fig. 7. Pooled PR curves for sleep-wake classification using HRV spectral features, HRV features, and actigraphy feature with and without adaptation for healthy subjects (top) and insomniacs (bottom).
4.2. Classification

As shown in Table 4, in general, adapting the boundaries of the HRV spectral features can improve the performance as evaluated by the three metrics. For the healthy group, it is interesting to note that the value of $\kappa$ is similar when using HRV spectral features with and without boundary adaptation. This seems to contradict the significant increase in discriminating power found with the Hellinger distance. Upon closer inspection we found that actually this occurs only for that single point in the solution space. In fact, when evaluating the performance over the entire solution space with $\text{AUC}_{\text{PR}}$ we see an increase from 0.30 to 0.36. The ROC and PR curves (plotted on Figures 6 and 7, respectively) with the use of HRV spectral features clearly show that the adapted versions are superior to the original ones, particularly in the region when recall is lower than about 0.30 or larger than about 0.60. For the insomnia group, the figures also indicate a clear improvement after adapting the HRV spectral features.

When combining the HRV spectral features with the additional HRV features indicated earlier, we see a significant increase (Wilcoxon test, $p < 0.01$) in $\kappa$ from $0.44 \pm 0.25$ (without adaptation) to $0.48 \pm 0.24$ (with adaptation) for the healthy group and from $0.31 \pm 0.11$ (without adaptation) to $0.34 \pm 0.12$ (with adaptation) for the insomnia group. The Wilcoxon significance test performed pair-wise comparison for each subject, thus indicating that boundary adaptation improved the classification performance for the majority of the subjects. Likewise, the pooled $\text{AUC}_{\text{PR}}$ and $\text{AUC}_{\text{ROC}}$ metrics increased when applying boundary adaptation. As shown in Table 4, the standard deviations of $\kappa$ are relatively large compared to the mean values, indicating large between-subject variations in the classification performance.

For comparison purposes, Tables 3 and 4 also show the classification results using the actigraphy feature $\text{ac}$. As expected, for the healthy group, it outperforms the HRV features. For the insomnia group, although the $\kappa$ value of using the HRV feature set generally is lower than using $\text{ac}$, the HRV features (in particular the adapted versions) outperforms this actigraphy feature when recall is higher than $\sim 0.55$ (see Figures 6 and 7). This indicates that the sensitivity to wake might be increased by adding these HRV features for the insomnia subjects. It also highlights the disadvantage of a metric such as $\kappa$, which only represents a single point reflecting a single solution in the space.

The classification results with the actigraphy and the HRV features are also given in the tables. Although actigraphy is adequate for sleep and wake classification, combining it with the HRV features (in particular when applying boundary adaptation on the HRV spectral features) significantly increases the classification performance measured by $\kappa$ value. The significance was confirmed with a Wilcoxon signed-rank test ($p < 0.01$).
4.3. Healthy subjects versus insomniacs

To compare between the healthy subjects and insomniacs, it makes less sense to use the pooled AUC_{PR} metric due to the difference in the ratio between the numbers of sleep and wake epochs in both groups. For instance, using a decision making rule such that all epochs are classified as wake (i.e. recall = 1), it will lead to different precision for the healthy and insomnia groups, with ∼92% and ∼70%, respectively, which only depends on their prior probabilities. Differences in class balance prevent a comparison between the area under the curves of each group. Therefore, here we used the pooled AUC_{ROC} metric instead. Figure 6 illustrates that the sleep and wake classification performances with different feature sets for the healthy subjects (top) are much better in contrast to that for the insomniacs (bottom). This confirms earlier findings, which show that discrimination between wake and sleep (especially REM sleep) is more difficult in insomniacs than in healthy subjects, when using cardiac activity\(^{40}\) or actigraphy.\(^{41}\)

4.4. Determination of adaptive boundaries

The method described in this paper shows a time-varying adaptation of the HRV spectral features that offer higher discriminative power in classifying sleep and wake states. The features are used as inputs to a sleep and wake classifier. We re-defined the spectral boundaries which are adapted to the spectrum information (related to autonomic activity) that can be obtained before feature extraction. This is because it is aimed at finding frequency bands that can more accurately capture certain aspects of physiology during sleep. For instance, the HF band should only include respiratory activity rather than sympathetic activation, which should be in the LF band. An excessively larger HF bandwidth might incorrectly include the “spillover” spectral power from sympathetic activation (see Figure 4). For this purpose, we used an HF* bandwidth of 0.1 Hz instead of the 0.25 Hz used in the traditional HF band. Alternatively, rather than using a constant HF bandwidth (0.1 Hz) in this study, it can be determined by measuring respiratory effort signals and analyzing their PSDs,\(^{16}\) but the use of an additional sensor is required.

Additionally, we observed that the LF and HF bands can overlap under different circumstances: when the peak in the LF and in the HF band are close to each other, when there is no clear peak in the HF band, or when the respiratory-frequency peak is below 0.15 Hz and therefore lies in the traditional LF band. Such overlaps (or spillovers) can be observed in Figure 4. In these situations, the overlapped part of the spectrum components will actually influence the features computed for both the LF* and the HF* bands. This may have an impact in the classification process, decreasing the accuracy of the classifier. Therefore, a more accurate method is needed for defining a threshold which separates the two bands rather than just using fixed bandwidths. This merits further investigation.

Finally, as we mentioned, the respiratory information was derived from the HRV data. Although this may not be as good an estimation as a direct measure of res-
piratory effort, it has been proven to be an available estimate of respiratory rate especially during sleep. More importantly, it does not require the use of an additional sensor to measure respiratory effort. Alternatively, the respiration rate can also be estimated from the ECG signal directly, for example by computing the changes in the “envelope” of the ECG due to the modulation induced by the respiration movements. This method will be further studied in future work.

5. Conclusion

In this study we used a method based on the time-frequency analysis of HRV spectral power to adapt HRV spectral features. It aims at providing more accurate interpretations of the sympathetic and respiratory activities in order to better discriminate between sleep and wake states. It was achieved by adapting the spectral boundaries according to the peaks found in high- and low-frequency bands of the HRV power spectral density. The adaptation improved the discriminative power of the HRV spectral features, and therefore enhanced the sleep and wake classification performance, especially after combining the adapted HRV spectral features with the other selected HRV non-spectral features. Using a linear discriminant classifier tested with leave-one-subject-out cross-validation, we achieved a significant increase on Cohen’s Kappa coefficient \( \kappa \) (from 0.44 to 0.48 for healthy subjects and from 0.31 to 0.34 for insomniacs). Furthermore, by combining these HRV features and actigraphy, we obtained a significantly increased \( \kappa \) compared with that obtained when only using actigraphy (0.64 versus 0.53 for the healthy group and 0.50 versus 0.45 for the insomnia group).

Acknowledgments

The authors would like to thank Dr. Tim Leufkens from Philips Research Eindhoven for the insightful comments and inspiring discussions.

References

X. Long et al.


34. T. Fawcett, ROC graphs: Notes and practical considerations for researchers (HP Labs, Palo Alto, CA, 2004).


