Christian Heinze*, Constantin Hütterer, Thomas Schnupp, Gustavo Lenis and Martin Golz

Drowsiness discrimination in an overnight driving simulation on the basis of RR and QT intervals

Abstract: We examined if ECG-based features are discriminative towards drowsiness. Twenty-five volunteers (19–32 years) completed 7×40 minutes of monotonous overnight driving simulation, designed to induce drowsiness. ECG ($512\,s^{-1}$) was recorded continuously; subjective ratings of drowsiness on the Karolinska sleepiness scale (KSS) were polled every five minutes. ECG recordings were divided into 5-min segments, each associated with the mean of one self- and two observer-KSS ratings. Those mean KSS values were binarized to obtain two classes *not drowsy* and *drowsy*. The Q-, R- and T-waves in the recordings were detected; R-peak positions were manually reviewed; the Q- and T-detection method was tested against the manual annotations of Physionet's QT database. Power spectral densities of RR intervals (RR-PSD) and quantiles of the empirical distribution of heart-rate corrected QTc intervals were estimated. Support-vector machines and random-holdout cross-validation were used for the estimation of the classification error. Using either RR-PSD or QTc features yielded mean test errors of 79.3 ± 0.3 % and 82.7 ± 0.5 %, respectively. Merging RR and QTc features improved the accuracy to 88.3 ± 0.2 %. QTc intervals of the class *drowsy* were generally prolonged compared to *not drowsy*. Our findings indicate that the inclusion of QT intervals contribute to the discrimination of driver sleepiness.

Keywords: RR interval, QT interval, Karolinska sleepiness scale,driving simulation, Lomb-Scargle periodogram, support vector machine

https://doi.org/10.1515/cdbme-2017-0117

1 Introduction

The analysis of EEG and EOG provides, to date, the best results for the detection of driver drowsiness, therefore it's an established standard in this field of research. It is, however, an impractical approach for the implementation of driver monitoring. The ECG, in comparison, can be easily recorded, has an amplitude that is several magnitudes larger, and is less afflicted by noise. Analyses of RR intervals have established that, under increasing sleepiness, heart rate variability is generally raised due to increasing parasympathetic autonomic activity. Such statements, however, can only be made on a larger time scale, as it has been shown, for example, that ECG features cannot reliably estimate the probability of microsleeps within a time frame of 30 to 120 seconds [1].

Recently, we have conducted a study involving an overnight driving simulation that was designed to induce a high level of drowsiness. In this work, we investigated the relationship between RR-interval variability and driver sleepiness and, particularly, whether the inclusion of QT interval information contributes to the discrimination of subjective sleepiness.

2 Material

Twenty-five volunteers (12 female, 13 male, 19–32 years old) participated in the study. All subject had been accustomed to the laboratory's driving simulator, which consists of a real-car body placed inside a dark, isolated room. The displayed driving scene was a country road at night, without any road intersections or other cars present. The driving was thus reduced to a simple lane- and speed-holding task, which was designed to induce a high level of monotony and drowsiness.

Before the experiment, subjects had slept for a minimum of seven hours at night. They were instructed to get up not later than 09.00 and to spend the rest of the day without naps.

***Corresponding author: Christian Heinze:** University of Applied Sciences Schmalkalden, Germany, e-mail: c.heinze@hs-sm.de
Constantin Hütterer, Thomas Schnupp, Martin Golz: University of Applied Sciences Schmalkalden, Germany, e-mail: c.huetterer@stud.fh-sm.de, thomas.schnupp@mailbox.org, m.golz@hs-sm.de
Gustavo Lenis: Karlsruhe institute of technology, Germany, e-mail: gustavo.lenis@icloud.com

Subjects arrived at the laboratory at 23.00 and were fitted for a polygraphic recording, including EEG, EOG and ECG (one bipolar channel, 512 s^{-1}). The experiment protocol began at 01.00 and consisted of seven repeated blocks of one hour length. In each block, the subject performed the driving task for 40 minutes. Every five minutes, the subject was prompted to rate his sleepiness on the Karolinska sleepiness scale (KSS, nine levels ranging between 1: *extremly alert* and 9: *very sleepy, fighting sleep*). Right after, two observers outside the simulator room submitted their rating of the subject's sleepiness as well. After 40 minutes, the subject left the simulator and performed a series of psychomotor tests, which took no more than 15 minutes total. The next block began at the following hour, and so on. The experiment ended at 08.00.

At the start of the experiment at 01.00, a subject had been awake for at least 16 hours, which corresponds to the waking portion of a normal day. The experiment procedure itself prevented sleep. The combination of extended wakefulness, partial sleep deprivation and task monotony was designed to provoke high sleepiness and microsleep during driving. Subjects were constantly monitored via IR-cameras; any occuring microsleeps, manifested as prolonged eye lid closure, rolling eye movements, head nods, etc., were logged by the observers. Reviews of microsleep annotations and other analyses are still in progress. Therefore, EEG and EOG data, the psychomotor tests and mircosleeps are not presented in this paper.

3 Methods

3.1 R and QT annotation

The ECG recording of each 40-min driving period was divided into eight 5-min segments. R-peaks in the ECG were automatically annotated using a method described in [2]. False-positive or missed peaks R-peak were manually corrected, and unprocessible noisy signal portions were flagged. RR-interval time series were then established from successive R-peaks. To detect the beginning of Q- and the end of T-waves (abbreviated as Q and T from here on) and hence to determine QT intervals of the ECG, we considered two annotation methods: (A) a toolchain that locates waveforms by generating and matching templates of each waveform, which was developed at the Karlsruhe Institute of Technology [1], and (B) a method that was awarded in the Physionet QT-interval-measurement challenge in 2006 [3], which consecutively applies continuous and discrete wavelet transform in order to separate P- and T-waves, QRS-complexes, and noise [4]. To assess the reliability of both methods, we tested them against Physionet's QT database [5], which consists of 105 records with 15 min of a two-channel ECG (250 s^{-1}). In each record, averagely 34 heartbeats were fully annotated by experts. Only fully annotated and normal heartbeats[1] were considered for comparing both methods.

Table 1: Root mean squared error (RMS), median of absolute deviations (MAD), and sensitivity (Sen) of Q, T, and QT detection for methods A, B, and both methods combined.

methods		A	B	A&B
Q	RMS	23,3 ms	20,4 ms	15,9 ms
	MAD	17,1 ms	11,8 ms	10,4 ms
	Sen	99,8 %	99,0 %	100 %
T	RMS	67,4 ms	57,5 ms	49,6 ms
	MAD	51,4 ms	26,6 ms	31,2 ms
	Sen	96,2 %	86,5 %	98,3 %
QT	RMS	62,9 ms	68,3 ms	51,3 ms
	MAD	46,4 ms	30,3 ms	31,6 ms

As a first result, method B detected Q and T more precisely, but less sensitively than method A. Curiously, though, the QT intervals determined by method A deviated from the true (expert) QT intervals with a smaller error than those determined by method B. It appeared that the respective locations of Q and T determined by method B tended to deviate from their true locations in opposite directions, which added up to a higher QT-interval error in average. For improvement, we averaged each Q and T location between the two methods. As a result, error and sensitivity of Q, T, and QT detection were improved by the combined method, in comparison to the single methods (see table 1). Finally, QT-interval time series were constructed from matching Q and T; QT-intervals were corrected for heart rate with Bazett's formula $\mathrm{QTc} = \mathrm{QT}/\sqrt{\mathrm{RR}}$.

3.2 Feature extraction

We estimated power spectral densities (PSD) of both RR- and QTc-interval series for each 5-min segment. The Fourier transform, which is mostly used for PSD estimation, assumes that a time series X_j has been sampled at equidistant times t_j. Fourier transforms of data sampled at unequally spaced times usually introduces low-frequency noise into the spectral estimate and diminishes the amplitude of true spectral peaks. Data such as RR-interval series are unequally spaced by nature, therefore interpolation and equidistant resampling is often applied to them. This, however, may produce false spectral peaks [6] and also makes the spectral estimation dependent

[1] A total of 2559 heartbeats met this criteria.

on the choice of the interpolation method. Least-squares spectral estimation methods like the Lomb-Scargle periodogram [7,8] overcome this problem and can be applied directly to unequally spaced time series X_j:

$$P_x(\omega) = \frac{1}{2}\left(\frac{\left[\sum_j X_j \cos\omega(t_j-\tau)\right]^2}{\sum_j \cos^2\omega(t_j-\tau)} + \frac{\left[\sum_j X_j \cos\omega(t_j-\tau)\right]^2}{\sum_j \cos^2\omega(t_j-\tau)}\right),$$

where τ is a time delay and is defined by

$$\tau = \frac{1}{2\omega}\tan^{-1}\left(\frac{\sum_j \sin(2\omega t_j)}{\sum_j \cos(2\omega t_j)}\right).$$

The maximum of P_x occurs at period ω where the sum of squares of a sinusoidal fit to X_j becomes minimal. Time delay τ is chosen such that the sinusoids are mutually orthogonal at sample times t_j, it thus adjusts for phase shifts that may be caused by unequally spaced t_j. In case of equidistantly sampled data, P_x has the same statistical distribution as the Fourier transform periodogram [8].

The maximum frequencies for PSD estimation were restricted to 0.4 Hz for RR- and to 0.22 Hz for QTc-interval series, respectively. In heart rate variability analysis, the PSD of RR intervals conventionally is integrated into the three bands very-low frequency (0–0.04 Hz), low frequency (0.04–0.15 Hz) and high frequency (0.15–0.4 Hz). In previous investigations, however, we found that finer grained RR-PSD bands yield better results when they are used as features for discrimination [9]. Therefore, we integrated the PSD of RR and QTc intervals in every segment into 0.02-Hz wide bands between 0 Hz and maximum frequency. We also estimated the 5th, 15th, …, and 95th percentile over the empirical distribution of RR and QTc intervals in every segment in order to characterize the intervals' probabilities.

3.3 Classification and validation

A class label was assigned to each segment that was determined by the average of the subject's and the observers' KSS ratings within it. The 40th and 60th percentile of the distribution of all mean KSS values was 7.0 (corresponds to *sleepy, but no difficulty to remain awake*) and 8.0 (*sleepy, some effort to keep alert*), respectively, so every segment with mean KSS ≤ 7.0 or ≥ 8.0 was labelled *not drowsy* (c_+) or *drowsy* (c_-), respectively. 650 and 571 segments were labelled with c_+ and c_-, respectively, while 207 others had mean KSS > 7.0 and < 8.0, and were therefore discarded.

We used support vector machines (SVM) [10] to discriminate the features of each segment according to their class membership. A SVM separates a labelled input data set $(\boldsymbol{x}_1|y_1), \ldots, (\boldsymbol{x}_n|y_n)$ by finding a hyperplane $\boldsymbol{w}_o \cdot \boldsymbol{x} + b_o = 0$ with a maximal margin between the classes. This optimal hyperplane is solely determined by the support vectors $\boldsymbol{x}_+$ and $\boldsymbol{x}_-$ (with $y_+ = c_+$ and $y_- = c_-$, respectively), that is, those input vectors of different classes with minimal distance to each other. If the training data cannot be separated without error, then a penalty term $\sum_{i=1}^{n} \xi_i$, with slack variables $\xi_i \geq 0$ keeping track of the number of errors, has to be used. This term restricts the range of the Lagrange multipliers α_i that are needed to solve the hyperplane optimazation problem: $0 \leq \alpha_i \leq C, i = 1, \ldots, n$. In order to support nonlinear separation functions, the SVM can be extended by kernel functions $K(\boldsymbol{u}, \boldsymbol{v})$ that map inputs $\boldsymbol{u}$ to vectors $\boldsymbol{v}$ into a higher-dimensional space. We used a radial-basis kernel function $K(\boldsymbol{u}, \boldsymbol{v}) = \exp(-\gamma\|\boldsymbol{u} - \boldsymbol{v}\|^2)$. The regularization parameter C and the kernel width γ are free parameters that have to be determined empirically by minimizing the classification error in a cross-validation scheme. In order to estimate the classification error or, conversely, the accuracy, we multiple random sub-sam¬pling: input data was randomly partitioned into a 90-% training set and a 10-% test set, which was repeated 50 times. The small test size, which come at the prize of more repetitions, is motivated by keeping the estimation bias small.

Table 2: Regularization parameter C, kernel width γ, training and test set accuracy for the discrimination of four feature set combinations.

Feature sets	C	γ	Training	Test
1: QTc perc.	1.4	-1.5	81.5 ± 0.0 %	79.3 ± 0.3 %
2: RR-PSD	0	-1.5	89.4 ± 0.0 %	82.7 ± 0.5 %
3: 1 + 2	0.6	-1.5	98.7 ± 0.0 %	86.7 ± 0.5 %
4: 3 + RR perc.	0.3	-1.5	98.7 ± 0.0 %	88.3 ± 0.2 %

4 Results

As described above, we considered PSD and percentiles of RR and QTc intervals, respectively, within 5-min segments for classification. We preliminarly tested all 15 possible combinations of those four feature sets. QTc-PSD yielded the worst test set accuracy. It also impaired any combination it was part of, and was therefore excluded from further analysis. Table 2 shows the training and test set accuracies from the final SVM-classification for the four best combinations, with RR-PSD + RR- + QTc-percentiles yielding the best test accuracy. Figures 1 and 2 compare the features' manifestations between classes. They show that RR-PSD and QTc percentiles are generally higher for the not drowsy condition.

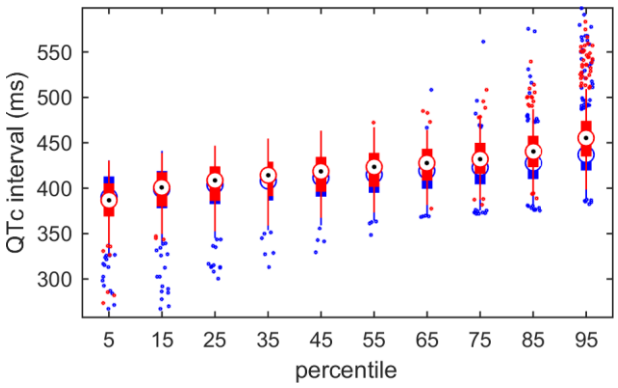

Figure 1: Distribution difference of QTc intervals between classes. Red: drowsy, blue: not drowsy.

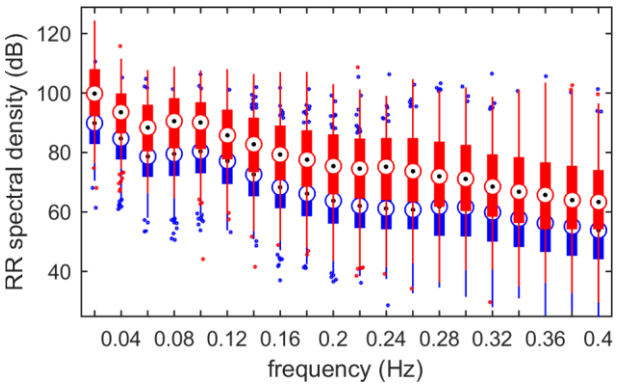

Figure 2: Distribution difference of RR spectral densities between classes (red: drowsy, blue: not drowsy).

Acknowledgment: We thank Professor Olaf Dössel and his research team at the Karlsruhe Institute of Technology for their ECG analysis software.

5 Discussion

This work followed up on an investigation of the same study [11]. There, KSS-rated drowsiness was similarly classified using different combinations of spectral and nonlinear RR-interval features; the best test set accuracy amounted to 79.7 ± 0.2 %. By including information derived from QTc intervals into a comparable analysis, we could improve on these results by about eight percent. This improvement cannot be attributed to a mere increase of feature space dimensionality: in both investigations, there were combinations with less features than the winning combinations that yielded inferior results.

Others have examined the statistical characteristics of various ECG intervals in the vicinity of microsleep events, which were recorded in an older, but similar study conducted in our driving simulator [1]. They found that epochs centered around microsleeps had significantly prolonged QTc intervals, in comparsion to epochs without microsleeps. Along this line, our work indicates as well that QT elongation might be caused by extreme drowsiness. It was noteworthy that spectral features of QTc-interval series provided no discriminative information for the problem at hand. Only time-invariant attributes like distribution quantiles contributed to an improved result. This might be an indication that time-dependent QT variability analysis is generally not useful.

To our knowledge, the best driver drowsiness detection is provided by EEG and EOG data annotated with microsleep events; after proper training and optimization, a SVM can detect microsleeps in an unknown subject with an accuracy

97.3 ± 0.1 % [12]. ECG-based discrimination of KSS-rated drowsiness remains clearly behind such capabilities.

Author's Statement

Research funding: The authors state no funding involved. Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

References

[1] Lenis G, Reichensperger P, Sommer D, Heinze C, Golz M, Dössel O. Detection of microsleep events in a car driving simulation study using electrocardiographic features. Current Directions in Biomedical Engineering, 2016, 2(1):283–287.

[2] Manikandan MS, Soman KP. A novel method for detecting R-peaks in electrocardiogram signal. Biomedical signal processing and control, 2011:7(2), 118–128.

[3] Moody GB, Koch H, Steinhoff U. The physionet/computers in cardiology challenge 2006: Qt interval measurement. In: Computers in cardiology, 2006, pp. 313–316, IEEE.

[4] Chesnokov YC, Nerukh D, Glen RC. Individually adaptable automatic QT detector. In: Computers in cardiology, 2006, pp. 337–340, IEEE.

[5] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation, 2001, 101(23):e215–e220.

[6] Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical recipes: the art of scientific computing. 3rd ed, 2007.

[7] Lomb NR. Least-squares frequency analysis of unequally spaced data. Astrophysics and space science, 1976, 39(2):447–462.
[8] Scargle JD. Studies in astronomical time series analysis. II-Statistical aspects of spectral analysis of unevenly spaced data. Astrophysical journal, 1982, 263:835–853.
[9] Heinze C, Sommer D, Trutschel U, Golz M. Discriminating healthy condition from heart failure using relevance-weighted features of heart rate variability. Biomedical engineering, 2014, 59:S189–S192.
[10] Cortes C, Vapnik V. Support-vector networks. Machine learning, 1995, 20(3):273–297.
[11] Schäfer I. Neuroinformatische Analysen von Herzrythmusvarianzzeitreihen aus einer Nachtfahrtsimulationsstudie. Bachelorthesis, University of Applied Sciences Schmalkalden, Sep 2016.
[12] Golz M, Sommer D, Krajewski J. Prediction of immediately occurring microsleep events from brain electric signals. Current Directions in Biomedical Engineer., 2016, 2(1):149–153.