Jan-Dirk Janßen* and Thomas Schanze

Analysis and classification of ECG-waves and rhythms using circular statistics and vector strength

Abstract: The most common way to analyse heart rhythm is to calculate the RR-interval and the heart rate variability. For further evaluation, descriptive statistics are often used. Here we introduce a new and more natural heart rhythm analysis tool that is based on circular statistics and vector strength. Vector strength is a tool to measure the periodicity or lack of periodicity of a signal. We divide the signal into non-overlapping window segments and project the detected R-waves around the unit circle using the complex exponential function and the median RR-interval. In addition, we calculate the vector strength and apply circular statistics as wells as an angular histogram on the R-wave vectors. This approach enables an intuitive visualization and analysis of rhythmicity. Our results show that ECG-waves and rhythms can be easily visualized, analysed and classified by circular statistics and vector strength.

Keywords: circular statistics, vector strength, heart rate variability, ECG classification

1 Introduction

The heart rate variability (HRV) has become an important indicator for the relationship between the autonomic nervous system and the cardiac system. Many methods in time domain as well as in frequency domain and even non-linear methods have been introduced to describe the HRV [1]. Those methods often result in difficult to read diagrams and statistics.

Rhythmicity or periodicity is often better described on a circle than on a linear timeline. Consequently, we must replace linear statistics with circular statistics. We also try to make use of the von Mises vector strength (VS) [2] that measures the periodicity or lack of periodicity of a signal. Furthermore, we want to show how the angular representation of R-Waves can be used to classify different ECG segments.

2 Methods

2.1 Data preparation

To analyse the rhythmicity of the recorded ECG signal, we divided the signal into N non-overlapping, sliding windows segment, thus reducing the influence of slow heart frequency rate changes. We applied Sedghamiz's complete implementation of the Pan-Tompkins algorithm [3] in order to detect the R-waves. Furthermore, we stored the related points in time $r_{k\,n}$, where k denotes the index of the R-wave in the nth window. In addition, we calculated the RR-intervals for each window.

2.2 Vector strength

Assuming that the majority of the heartbeats are regular, we calculate the median RR-interval $\tilde{t}_{RR\,n}$ for each window, thus eliminating irregular outliers. The next step is to project the R-wave time points around the unit circle using the complex exponential function:

$$\vec{r}_{k,n} = \exp(j2\pi r_{k,n}/\tilde{t}_{RR,n}). \tag{1}$$

One rotation around the unit circle represents a regular heartbeat and therefore all regular heart beats point in the same direction (see Figure 1).

The von Mises VS for the nth window is defined as following [2]:

_____

*Corresponding author: Jan-Dirk Janßen: Technische Hochschule Mittelhessen, Dept. Life Science Engineering, Wiesenstr. 14, D-35390 Gießen, Germany, e-mail: jan-dirk.janssen@lse.thm.de
Thomas Schanze: Technische Hochschule Mittelhessen, Dept. Life Science Engineering, Wiesenstr. 14, D-35390 Gießen, Germany, e-mail: thomas.schanze@lse.thm.de

$$\vec{v}_n = \frac{1}{K_n} \sum_{k=1}^{K_n} \vec{r}_{k,n} \tag{2}$$

where K_n is the total number of $\vec{r}_{k\,n}$ found in window n. Consequently, a vector strength of one is for perfect rhythmicity and zero for total irregular heartbeats.

The first appearance of an R-wave changes from window to window. This is reflected in different phases of $\vec{v}_n$ for each window. Consequently, we normalize each window by subtracting the direction of the vector strength from the direction of $\vec{r}_{k\,n}$:
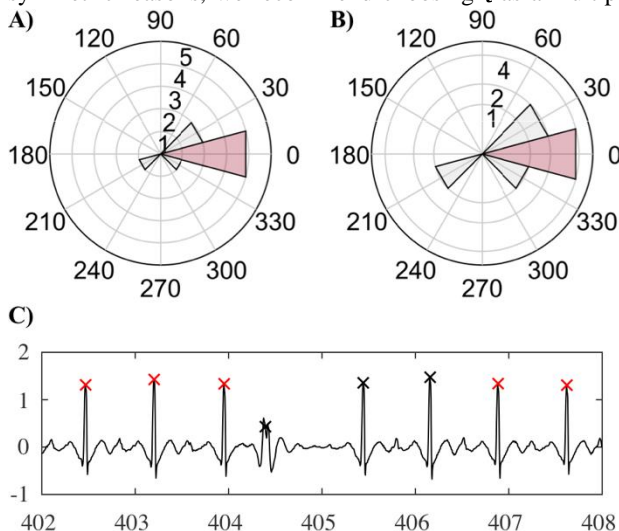
$$\arg\vec{\hat{r}}_{k,n} = \arg\vec{r}_{k,n} - \arg\vec{v}_n. \tag{3}$$

Hence, a normalized heartbeat vector should point in the 0° direction. Longer and shorter RR-Intervals lead to small positive and respective negative deviations from the mean direction (see Figure 1). Thus, the R-wave vectors from all windows can now be used to calculate the total VS $\vec{v}$:

$$\vec{v} = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{K_n} \sum_{k=1}^{K_n} \vec{\hat{r}}_{k,n}. \tag{4}$$

2.3 Circular visualisation and statistics

We obtain the circular distribution of the R-wave time points by splitting the unit circle into l circle sections of equal size and then counting the R-wave vectors in each section. For symmetric reasons, we recommend choosing l as a multiple

of four. However, the radius of each section in the angular histogram is proportional to the number of R-waves. Therefore, smaller sections appear more important than they actually are. Instead, we multiply the radius of every section with a correction factor:
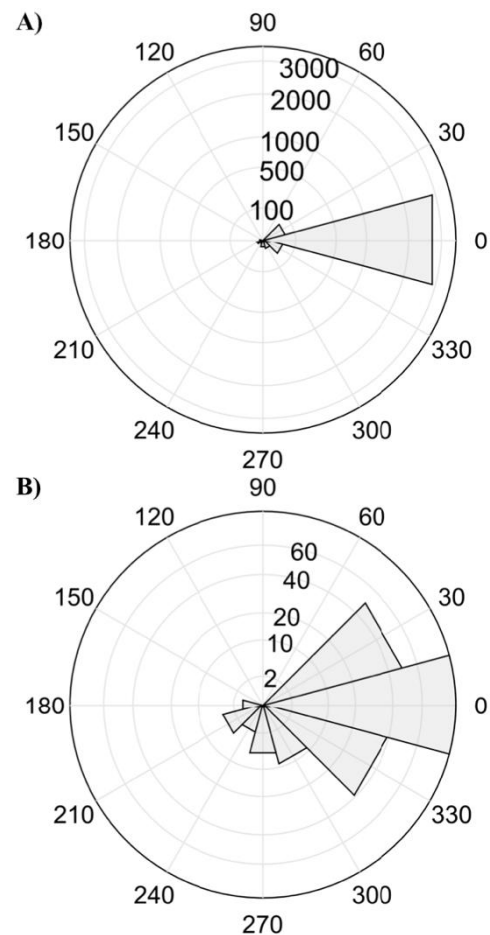
$$r_{corr} = \sqrt{2r/[\phi - \sin(\phi)]} \tag{5}$$

with $\phi = 2\pi/l$ and $l \in \mathbb{N}^+$ denoting the number of circle sections. As a result, the number of R-waves is proportional to the area of the section and therefore represents the accurate distribution of R-wave vectors (see Figure 1).

Because the data is not represented on a linear axis we cannot use linear statistics, instead we must use circular statistics. The Matlab Toolbox CircStat [4] was used to calculate the statistics.

The mean resultant length $\bar{R}$ and the mean direction $\bar{\theta}$ are defined similarly to the VS [5]:

$$\bar{\theta} = \angle \sum_{i=1}^{N} \exp j\alpha_i \tag{6}$$

Figure 1: A) Circular distribution of R-wave vectors; the radius of each section is proportional to the number of elements. The section pointing in the 0° direction and therefore containing regular R-waves, is marked red. **B)** Circular distribution of the same R-wave vectors; the area of each section is proportional to the number of elements in each section. **C)** A window containing the R-waves, corresponding to the circular distribution shown on the top. Note that the red markers indicate R-waves in the 0° section and the fourth and irregular R-wave translates into a vector pointing away from the 0° direction.

Figure 2: A) Angular histogram with a corrected r-axis sorting all R-wave vectors into 12 sections of 30°. **B)** Detailed view of the same angular histogram limiting the r-axis to 100 and therefore cropping the 0° section.

$$\overline{R} = \left| \frac{1}{N} \sum_{i=1}^{N} \exp j\alpha_i \right|. \tag{7}$$

Pewesy [6] has defined the skewness $\hat{s}$ and kurtosis $\hat{k}$ of a sample distribution as following:

$$\hat{s} = \frac{1}{N} \sum_{i=1}^{N} \sin 2(\theta_i - \overline{\theta}) \tag{8}$$

$$\hat{k} = \frac{1}{N} \sum_{i=1}^{N} \cos 2(\theta_i - \overline{\theta}). \tag{9}$$

3 Results

For this analysis, the 303rd recording from the MIT-BIH ST Change Database was used. The dataset has been recorded with a sample rate of 360 Hz, during an exercise stress test and exhibits transient ST depression [7-8]. We applied no additional filtering to the signal and divided it into 339 windows, 6 seconds in length. The Pan-Tompkins algorithm found 3007 R-waves in total.

3.1 Circular statistics and heart rate variability

Figure 2 displays the angular histogram of all R-waves vectors with a section width of 30°. Almost 3000 R-wave vectors point in the 0° direction and therefore, represent regular heartbeats (see Figure 2A). Both neighbouring sections contain about forty R-wave vectors each. This results in a very strong VS of 0.9856. The deviation of the mean direction from the 0° direction is almost non-existent. In contrast, the remaining sections contain only up to ten R-wave vectors each (see Figure 2B).

The kurtosis is very close to one and therefore indicating a very strong and peaked distribution of R-wave vectors around the mean direction. Additionally, the skewness of the sample distribution is very close to zero. Therefore, the number of positive deviations from the mean direction, is similar to the number of negative deviations.

## 3.2 Classification of ECG-waves

The angular histogram sorts the R-wave vectors based on the RR-interval for each window. Regular heartbeats point in the 0° direction and were thus assigned to the 0° bin. In contrast, irregular heartbeats that did not match the median RR-interval were not assigned the 0° section, but instead, to the sections between 195° and 315° (see Figure 2B).

Figure 3 displays an overlay of the QRS complexes corresponding to the R-wave vectors found in the 0°, 210°, and 240° section. Note the strongly different time courses, indicating different heart states.

4 Discussion and conclusion

We introduced a new method to analyse the HRV and rhythmicity of R-waves. Due to the phase normalisation of the VS in relation to the mean RR-interval in each window, we were able to lock the VS of each window to a relative heart frequency rate. Consequently, the combined VS is highly adaptive to low frequency heart rate changes.

When analysing the HRV in the frequency domain, low frequency changes of the heart rate can smear the amplitude spectrum. In contrast, the angular distribution of the R-wave vectors shows a very high concentration around the VS. A challenge for this method lies in finding good parameters describing the concentration of R-wave vectors around the VS and thus the HRV. Future research might consider finding a suitable angular probability model i.e. the Fisher or the von Mises distribution.
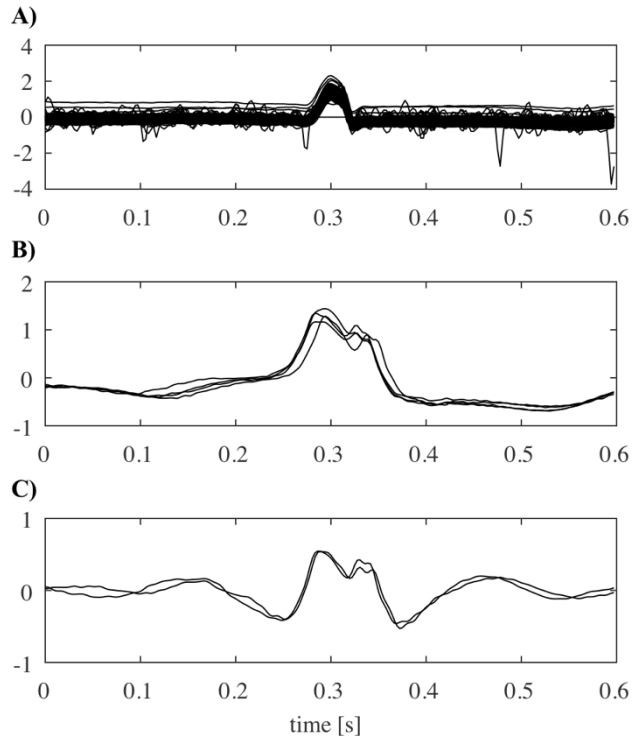
Figure 3: Ensemble of all QRS complexes assigned to a specific section. **A)** 0° Section: Ensemble of 2896 regular QRS-Segments, several segments polluted with noise or DC component. **B)** 210° Section: Ensemble of four irregular heartbeats with transient ST depression **C)** 240° Section: Ensemble of two irregular heartbeats with transient ST depression.

The analysis of arrhythmic ECG recordings shows that a large number of irregular heartbeats respond to a uniform distribution of R-Wave vectors around the unit circle, and therefore the VS is close to zero. Nevertheless, the collapsing of the VS does not conclude a failure of our method at all, instead it denotes the lack of rhythmicity of the ECG recording.

A weak point of this method is the influence of different window lengths on the VS. Extended simulations and analysis with different window lengths would be necessary to explore their relationship.

We tested this method using several data files from the MIT-BIH ST Change Database, as well as the MIT-BIH Arrhythmia Database, and ECG recordings from the authors.

In conclusion, this method allows for a detection, separation, and classification of ECG waves. We were able to separate regular heartbeats from irregular heartbeats with transient ST depression. Further investigations might test if the separation works for other ECG aberrations as well.

Author's Statement

References

[1] Task Force of the European Society of Cardiology and the North American Society for Pacing and Electrophysiology. Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use. Circulation 1996; 93:1043-65.

[2] van Hemmen JL. Vector strength after Goldberg, Brown, and von Mises: biological and mathematical perspectives. Biol Cybern 2013; 107:385-96.

[3] Sedghamiz H. Complete Pan Tompkins Implementation ECG QRS detector (v.17). MathWorks File Exchange 2014.

[4] Silva, I, Moody, G. An Open-source Toolbox for Analysing and Processing PhysioNet Databases in MATLAB and Octave. Journal of Open Research Software 2014; 2:e27.

[5] Fisher NI. Statistical Analysis of Circular Data. Cambridge: Cambridge University Press 1993.

[6] Pewesy A. The Large-Sample Joint Distribution of Key Circular Statistics. Metrika 2004; 56:25-32 Octave. Journal of Open Research Software 2014; 2:27.

[7] 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 2000; 101:215-220.

[8] Berens P. CircStat: A MATLAB Toolbox for Circular Statistics. Journal of Statistical Software 2009; 31:1-21.