Classifying Heart Rate Variability Data using Multitaper Spectrum Analysis

Heidi Mach

June, 2021

Contents

Populärvetenskaplig sammanfattning

Heart rate variability (HRV) mäts från tidsintervallet mellan två hjärtslag. Tidsintervallet benämns oftast som variation i hjärtslagen. Med hjälp av HRV går det att beskriva vissa av kroppens olika tillstånd. Det finns en naturlig variation i HRV som associeras med ett friskt eller neutralt kroppstillstånd. Ibland när det uppstår en abnormal variation, då kan det innebära att kroppen är i ett stresstillstånd. Det är skillnaden mellan dessa tillstånd som är av intresse. För att beräkna HRV, används oftast ett elektrokardiogram (EKG).

I rapporten kommer data från en studie i Kristianstad att användas. Sammanlagt var det 53 personer som deltog. Deltagarna fick först ha sin hand i kallt vatten, sedan gjorde en kontrollset där de istället fick ha handen i ljummet vatten. Detta gjordes för att kallt vatten var känt för att få igång stress stimulus, alltså, få deltagaren i stresstillstånd. I de två tillstånden producerades två olika HRV.

Baserat på det givna HRV vill vi veta i vilket av de två tillstånden personen är i. Vi kommer att utföra en binär klassificering, där HRV data antingen klassas som positiv eller negativ. I detta fall innebär positiv att från testet med kallt vatten klassas som kallt, medan negativ innebär att data från testet med ljummet vatten klassas som varm. Syftet med rapporten är att använda oss av frekvensanalys för att kunna utföra binär klassificering. Beroende på koncentration av energi vid viss frekvens domän kan vi då bestämma vad för HRV det är. I denna rapport kommer vi att använda oss av Thomson-metoden och Welch-methoden för vår frekvensanalys.

Det bästa och mest tillförlitliga klassificeringsresultatet i den här rapporten är 69%. Det uppmättes när vi använde Thomson-metoden och antagandet att det finns en skillnad i energidistribution i tidsdomänen.

Abstract

Heart rate variability (HRV) is the variation between two consecutive heartbeats. The irregular variability in this interval can indicate different health issues such as stress. The goal of this project is to correctly classify if a HRV signal comes from a resting state or a state which is affected by stress related stimuli. The analysis will be conducted using non-parametric multitaper spectrum analysis in the frequency bandwidth, 0.12-0.4 Hz. Two different multitaper methods will be tried; Welch method and Thomson method. For the binary classification of the HRV signal, it was assumed that there was a difference in energy distribution. In the pair-wise classification the assumption was instead that there was a difference in total energy. The highest and most trustworthy binary classification of the methods was 65% for Welch and 69% Thomson, with the assumption of there being difference in energy distribution. For the pair-wise classification, it was 74% and 77%, respectively.

1 Introduction

Now more than ever, technologies are dominating our daily life. It is becoming easier and easier to gather large information about yourself. It is now possible to keep track on our weight, blood pressure, number of steps, calories, heart rate, and blood sugar. Researchers have found heart rate variability (HRV) can also be used for assessment of your health. Heart rate variability is the time interval between two consecutive heartbeats [1]. Figure 1. shows data obtained from a electrocardiogram (ECG) signal, where the heartbeats are shown as the peaks, often referred to as R-peak, with the corresponding RR-interval between the peaks. The HRV is the variance between the heartbeats.

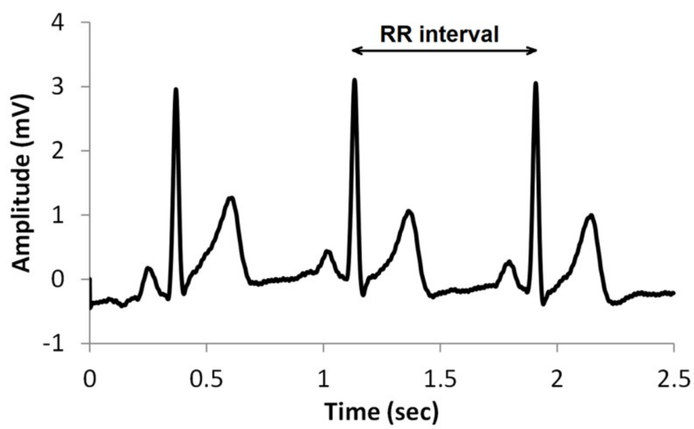

Figure 1: The plot shows a RR-interval between two R-peaks.

The HRV has also been shown to be a good qualitative marker of autonomic activity relevant to autonomic nervous system and cardiovascular mortality, including cardiac death. The HRV provides good measures of the health and has provided simple tool for researchers to use, hence, why it is of interests [2]. In the analysis of HRV, there are several different methods that can be used. The domain that the methods uses are either the time- or frequency domain. The frequency domain is usually of more interest than the time domain since it often provides more information of the HRV signal. The most common frequency range for adults used in analysis of the signal is the low frequency (LF: $0.04 - 0.12$ Hz) and the high frequency (HF: $0.12 - 0.4$ Hz). The frequency range that will be used for the analysis will be the HF domain. The HF domain has often been related to the parasympathetic activity, which describes the body's resting state. In the same frequency interval, it has been noticed in previous studies that exposing body to cold stimuli affects the HRV.
In this thesis, data from a "Cold Pressure Test" conducted in Kristianstad, Sweden, was used. The test lets its subject place a hand in ice-cold water for three minutes, and it also includes a control set, where the subjects places their hand in luke-warm water instead. The goal of this thesis is to be able to decide

whether the subject has its hand in cold or warm water based on the HRV data obtained from the test.

2 Spectrum Analysis

In the first part of this section some basic theory about the power spectrum will first be mentioned. Thereafter, the methods that will be used for the spectrum analysis such as the Welch method and Thomson method will be introduced. The different methods have proven to be good spectral estimates, however, there is a problem with trade-off between bias and leakage which will be mentioned.

2.1 Power spectrum

Given the zero-mean stationary process $x(t)$, for $t = \pm 0, \pm 1, ...$ We want conduct analysis of the process in the frequency domain. Assuming that the process has a covariance function, $r(\tau)$, and applying the Fourier transform on the function results in the spectral density forming. Thus, we can now conduct analysis in the frequency domain.

The discrete-time Fourier transform of the data sequence is:

$$X(f) = \sum_{-\infty}^{\infty} x(t) e^{-i2\pi f t}, \quad f \in [-1/2, 1/2]$$

Then, the inverse discrete-time Fourier transform is defined as:

$$x(t) = \frac{1}{2\pi} \int_{-1/2}^{1/2} X(f) e^{i2\pi f t} df$$

Given that $r(\tau)$ is the covariance function of $x(t)$, then

$$R(f) = \sum_{-\infty}^{\infty} e^{-i2\pi f \tau} r(\tau)$$

is the spectral density function where f is the frequency. The spectral density, $R(f)$, is assumed to be symmetric, integrable and positive. A method that is known to give a good spectrum estimate is windowing or tapering of the data. It is also sometimes called the modified periodogram and given by

$$\hat{R}_w(f) = \frac{1}{n} \left| \sum_{t=0}^{n-1} x(t) w(t) e^{-i2\pi f t} \right|^2$$

where $\hat{R}_w(f)$ is the estimated spectral density and $w(t)$ is the taper or window [3].

2.2 Non-parametric multiple window methods

The spectral density mentioned previously will now be used for the following multitaper methods: the Welch- and the Thomson methods. A typical taper that is used for the Welch method is the Hanning window with 50% overlap. For the Thomson method, Slepian and Hermite functions are usually used as windows for the method. In this section, all the methods and tapers will be introduced.

2.2.1 Welch Method

Given n data points, these are divided into K segments of equal length, $L = n/K$, with an overlap of p percent. The spectral density then becomes an average of K spectral estimates. Let $\hat{R}_{av}$ be the averaged spectral estimate, then it is given by,

$$\hat{R}_{av} = \frac{1}{K}\sum_{k=1}^{K}\hat{R}_{x,k}(f)$$

where $\hat{R}_{x,k}(f)$ is the k^{th} estimated spectral density, for $k = 1, ..., K$. The k^{th} spectral density is estimated using the modified periodogram with a given window that is the same length as the data segment. For the Welch method to be able to reduce the variance, the spectral estimates $R_{x,k}(f)$ has to be approximately uncorrelated. This is where the trade-off between the bias and variance occur. It has been shown that the Welch method with K tapers gives K times lower variance then only using one taper. Furthermore, the Hanning window with a 50% overlap has been known to give a good trade-off between variance and bias. This is because the segments with a 50% overlap are longer and leads to lower bias, since it now uses more data. The reason one wouldn't increase and use longer segments with 100% overlap to lower the bias even more is because the $R_{x,k}(f)$ will become more similar. When the spectral estimates are too similar, we will get a high and positive correlation. Averaging over correlated $R_{x,k}(f)$, will not decrease the variance efficiently [3]. Hence, the Hanning window with a 50% overlap will be used for the Welch method in this report.

The number of windows determined how smooth the spectral estimate eventually would be. The maximum of the optimal number of windows was computed as,

$$K_{max} = \frac{2n}{L} - 1$$

where K_{max} is the maximum number of windows, n the length of the data and L being the number of data points in two periods.

2.2.2 Hanning Window

The Hanning window is given by,

$$w(t) = \frac{1}{2} - \frac{1}{2}\cos\left(\frac{2\pi t}{n-1}\right).$$

The Hanning window reduces the bias compared to the periodogram without a window. In this thesis, a different number of Hanning windows was used for the Welch method. Figure 2 shows five Hanning windows with an overlap of 50% in the left plot and the window spectra for one Hanning window in the right plot.

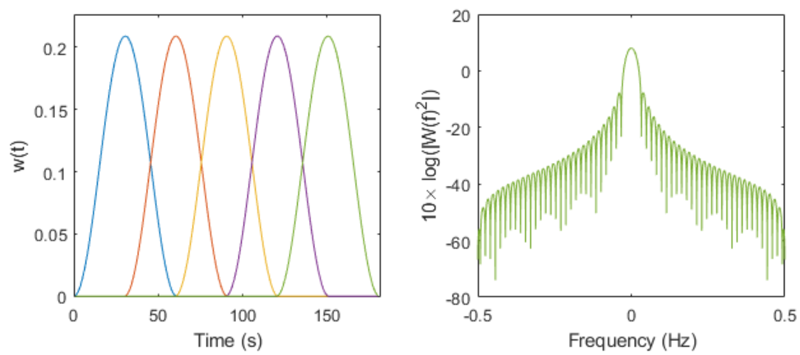

Figure 2: Representation of five Hanning windows with 50% overlap to the left and the corresponding window spectra to the right.

2.2.3 Thomson Method

Another multitaper method is the Thomson method. Its advantages compared to the Welch method is that instead of averaging over segments of the data, it averages over the whole data set.
Let $x(t)$ be the data sequence of length n, the multitaper spectral estimater is computed as,

$$S(f) = \frac{1}{K}\sum_{k=1}^{K} S_k(f) = \frac{1}{K}\sum_{k=1}^{K}\left|\sum_{t=0}^{n-1} x(t)w_k(t)e^{-i2\pi fn}\right|^2$$

where K is the number of windows, $S_k(f)$ is the modified periodogram with the k^{th} window function $w_k(t)$. The variance is reduced because the spectral densities $S_k(f)$ are uncorrelated. To achieve uncorrelated spectral estimates, the tapers should be pairwise orthogonal [4]. The tapers that are often mentioned and have proven to give good results are the DPSS sequences and Hermite functions, which will be described in the following sections.

2.2.4 Slepian sequences

To protect the data from leakage a set of K data tapers are needed such that each eigenspectra are nearly uncorrelated. This condition can be fulfilled if the windows are approximately orthogonal, such that $\sum_{t=1}^{N} w_i(t)w_j(t) = 0, \forall i \neq j$. The Slepian sequences, also called the discrete prolate spheriodal sequences (DPSS), are a set of orthogonal tapers. The sequences are defined as

$$w_k(t) \equiv \begin{cases} v_k(t-1; n, b), & \text{t} = 1, ..., n \\ 0, & \text{otherwise} \end{cases}$$

where $\{v_k(t; n, b)\}$ is denoted the k^{th} order DPSS, while $\{w_k(t)\}$ is the k^{th} order DPSS data taper. Here b is the half-time bandwidth. The order of the data taper is from $k = 1$ to K, if $K < 2nb\Delta t$ [5]. In this project the number of windows will be chosen as $K = 2nb - 2$ [4]. The DPSS, with the corresponding spectra window can be seen in Figure 3. It can be noticed that the main lobe in the spectra windows are narrower than the main lobe given by the spectra window of Hanning in Figure 2, although, the leakage increases with increasing k. The DPSS has been known to have a better trade-off between bias and variance compared to the Hanning window.

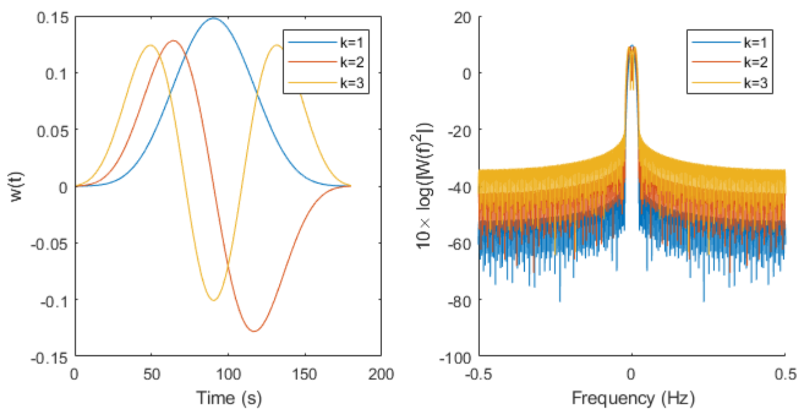

Figure 3: The left plot shows the DPSS sequences, for $k = 1, 2, 3$, with the corresponding windows spectra in the right plot.

2.2.5 Hermite Functions

The normalized Hermite functions of degree k, for $k \geq 0$ is defined by,

$$H_k(t) = \frac{1}{\sqrt{2^k k!}} e^{-t^2/2} h_k(t)$$

where $h_k(t)$ is the Hermite polynomial and $t \in \mathbb{R}$. The sequences $\{H_k\}_{k \geq 0}$ form an orthonormal system in $L^2(\mathbb{R})$, i.e.,

$$\int_{-\infty}^{+\infty} H_m(t) H_k(t) dt = \sqrt{\pi} \delta_{mk}$$

For $k \geq 1$ the recurrence relationships are given by,

$$H_{k+1}(t) = t\sqrt{\frac{2}{k+1}} H_k(t) - \sqrt{\frac{k}{k+1}} H_{k-1}(t)$$

where the initial values $H_0(t) = e^{-t^2/2}$ and $H_1(t) = \sqrt{2} t e^{-t^2/2}$ [6]. The $\{H_k(t)\}$ denotes the k^{th} order Hermite functions data taper. Therefore, we have that $w_k(t) = H_{k+1}(t)$. The functions can be seen in Figure 4. Observing the Hermite function, the sidelobes are noticeable lower than the DPSS. Although, the resolution is a bit worse.

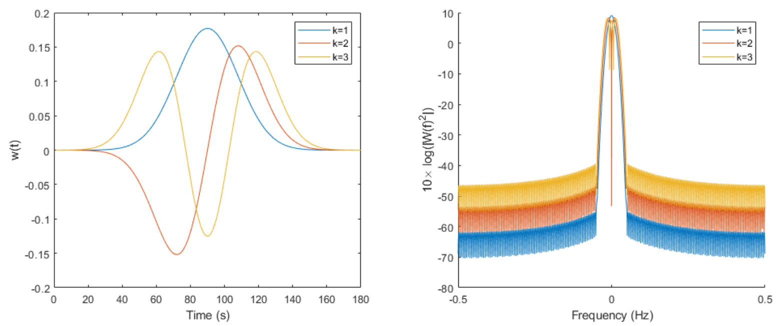

Figure 4: Hermite Functions over the domain $t \in [0, 720]$ for $k = 1, 2, 3$.

2.3 Classification

To be able to assess the quality of the classification, Matthews correlation coefficient was used. The ratio of the number of correct classifications over the total number of classifications will also be used for indication of the quality of classifications. In the "Evaluation" section, one of the classification methods will be based on implementing the line of best fit on two different data sets and then compare the coefficient. Thus, the interest in the line of best fit.

2.3.1 Matthews correlation coefficient

For the regular classification there are four possible states; true positive (TP), false positive (FP), true negative (TN) and false negative (FN). True positive and false negative is when data is correctly predicted as "positive" when it in fact is, and predicted false positive if the data belongs to "negative". The same goes for true negative and false negative, if data is correctly classified as "negative" and falsely classified as "negative" respectively. These values will then be displayed i a confusion matrix.
Matthews correlation coefficient was used in the classifying of the data. This is to be able to measure how "good" the algorithm is able to predict the data. The Matthews correlation coefficient uses the values in the confusion matrix and is defined as follows

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}},$$

for MC defined in [-1,1]. If the coefficient assumes 0, it means that there was a random classification. While 1 indicates perfect classification and -1 indicates worst classification.

2.3.2 Line of best fit

Let $y_1, ..., y_p$ be the response variable and $x_1, ..., x_p$ be the explanatory variables, then

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_p \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

is the line of best fit with the coefficient vector $\beta = (\beta_0, \beta_1)^T$.
Confidence intervals for the parameters β_0 and β_1 can also be computed. If $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimation of the parameters, and let the average $\bar{y}$ be estimated by

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

then the confidence intervals, I_{β_1} and I_{β_0} can be computed as,

$$I_{\beta_0} = (\hat{\beta}_0 - \frac{1}{\sqrt{m}}\lambda_{\alpha/2}, \hat{\beta}_0 + \frac{1}{\sqrt{m}}\lambda_{\alpha/2})$$

and

$$I_{\beta_1} = (\hat{\beta}_1 - \frac{1}{\sqrt{m}}\lambda_{\alpha/2}, \hat{\beta}_1 + \frac{1}{\sqrt{m}}\lambda_{\alpha/2})$$

where m denotes the number of subjects [7].

3 Evaluation

In Kristianstad, a study called 'Cold Pressure Test' was conducted. The test is about having a hand placed into ice-cold water for three minutes, corresponding to 180 s. Samples of the HRV was taken continuously during the time period, with a sampling frequency of 4 Hz. The number of samples amount to $n = 720$. In total, there were 90 participants between the ages 19 and 31 years old. However, only the data of 53 people were used due to people not finishing the test. The mean age of these people is 23.23 years, with variance 7.4. Except for the ice-cold water test (cold signal), there was also a control set where the participant instead had their hand in lukewarm water (warm signal). To construct the data set each participant took an ECG test, while at the same time the respiration data was measured. Note that, for the respiration data there was no guidelines given. However, the respiration data was not used in this thesis since it didn't provide better results. The HRV signal of the cold- and the warm signal were now used for classification. In this section different classification methods will be introduced. The data was sampled discretely over time, hence, the energies was computed as,

$$e_i = \sum_{j=1}^{n} \hat{R}_{av}^i(f_j)$$

where f_j is the frequency vector containing n entries and $\hat{R}_{av}^i$ is the spectral estimate of the ith set, which is estimated with either the Welch method or the Thomson method. For both method and for different number of sets, different number of windows was used.

3.1 Pre-processed data

In non-parametric spectrum estimation the data is assumed to be a zero-mean stationary stochastic process. To obtain a zero-mean process the mean was removed from the data and to compare the subjects to each other the data was also normalized. Let X be an $n \times m$ matrix of all the data with the mean, $\bar{X}$, then we have,

$$Z = \frac{X - \bar{X}}{\sqrt{(X - \bar{X})^T (X - \bar{X})}}$$

where Z is the normalized matrix. The normalized data will be used for all the methods mentioned throughout the thesis.

3.2 Classification measures

In this section, three different classification methods was tested on the data sets. The methods that will be tested are differences in energy distribution, the line of best fit and absolute comparison of energy. The two first methods will assume that there is a change in the energy distribution over time that uniquely distinguishes the cold- and warm signal. The last method is about pair-wise classification, where the assessment is based on comparing two signals from the same person, and determining which signal is from the cold-pressure test. In all the methods the energy of the sets of the signals was computed. Analyzing the signal, the biggest difference between the cold- and warm signal could be seen in the frequency band $0.12 - 0.4$ Hz. This frequency band will be used for all three methods.

3.2.1 Energy distribution

The classification in this method was based on there being a difference in the energy distribution in the signal. Observing the energy of the cold signal in the time domain, the energy seems to be increasing. The cold- and warm signal was then divided into two 90 s sets, and transformed into the frequency domain. To compute the energies an algorithm was implemented to find the optimal windows. For this method the optimal window was chosen such that the results lead to the highest correct classification of the cold signal, i.e., the window that resulted in most of the last 90 s having higher total energy than the first 90 s of the cold signal.
Let e_{c_1} and e_{c_2} denote the energy of the first 90 s and last 90 s of a cold signal, respectively. The optimal number of windows was chosen as,

$$k_{opt} = \max_k \sum_{i=1}^{n} \mathbb{1}_{\{e_{c_1} \leq e_{c_2}\}}$$

where k_{opt} is the optimal number of windows. The optimal windows were then used to compute the energies to classify the signal. Given a signal, if $e_1 \leq e_2$, then the signal was classified as "cold". Otherwise, the signal was classified as "warm". The computed sets of the cold- and warm energies can be seen in Figure 5. The plot to the left uses the data from the cold signal and is correctly classified since the energy is higher at the end compared to the beginning, and the plot to the right uses the warm data and is also correctly classified since, on the other hand, the energy is higher at the beginning compared to the end.

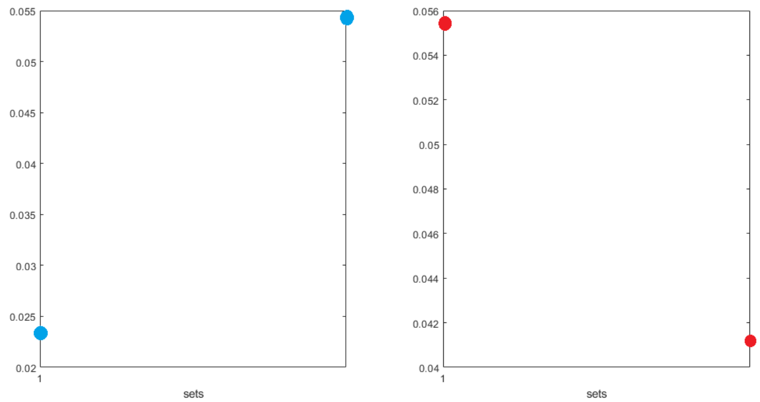

Figure 5: The plots shows the energies of subject 50 divided into two sets. The plot to the left uses data from a cold signal and plot to the right uses data from the warm signal. Both signals are correctly classified.

3.2.2 Line of best fit

In this method the classification is still based on there being a difference in the energy distribution of the signal. However, the signals is now divided $2, 4$ or 10 times. A linear model will be implemented for the sets. Different from before the optimal window is now based on both the cold- and warm signal. It is the one that leads to the most correctly classified cold- and warm signals. Let $e_{c_1},...,e_{c_p}$ and $e_{w_1},...,e_{w_p}$, for $p = 2, 4, 10$, be the computed energies of the sets of the cold- and warm signal, respectively. Implementing the line of best fit for both signals, let β_{c_1} and β_{w_1} be the slopes for each data set. The optimal number of windows will be computed as,

$$k_{opt} = \max_k \sum_{i=1}^{n} \mathbb{1}_{\{\beta_{c_1} \geq 0, \beta_{w_1} < 0\}}$$

where k_{opt} is the optimal number of windows. In the classification of the cold- and warm signal, a linear model was implemented. As before the optimal window will be used to compute the energies in the classification of the cold- and warm signal. For an unknown signal, if its slope is $\beta \geq 0$, the signal will be classified as "cold", and if $\beta < 0$ the signal will be classified as "warm".
Figure 6. shows plots using data from the cold signal to the left and data from the warm signal to the right. The data is divided into 4 sets. It can easily be seen that the slope in the plot to the left is positive, while the slope in the right plot is negative, indicating that both signals are correctly classified.

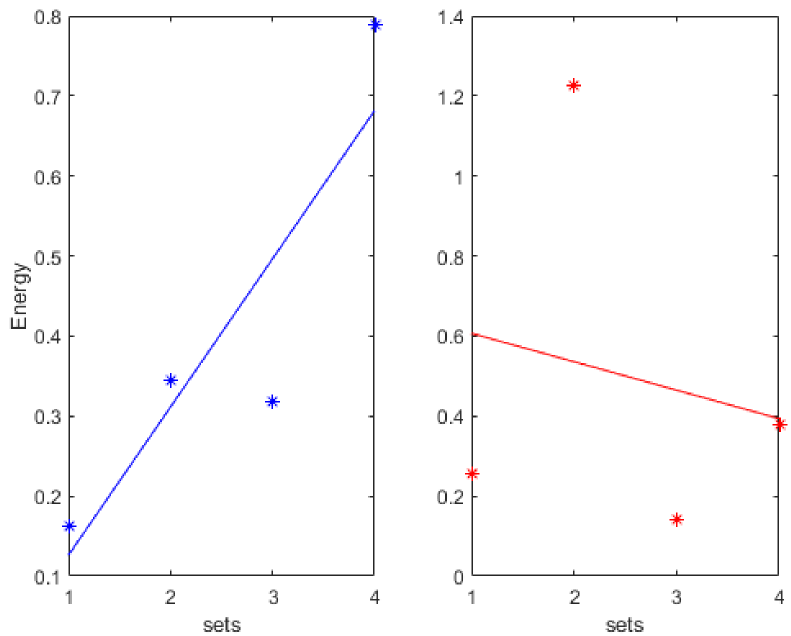

Figure 6: The plots shows data obtain from subject 50. The energies are computed with Welch using Hanning window with 50% overlap. The left shows values from the cold HRV signal and the right shows the values from the warm HRV signal. Observing the slopes of both the graphs, it can easily be seen that both signals are correcly classified.

3.2.3 Absolute Energy Comparison

For the absolute energy comparison method the classification will instead be about pair-wise classification. Given the cold- and warm signals, the classification is made based on comparison of energies of the signals. Based on the chosen frequency the total energy of both will be computed and then compared. Recall that the signals' energies are normalised, and this energy comparison is done in the frequency band $0.12 - 0.4$ Hz (which is also true for the two previous methods). Let e_c and e_w be the total energy of the cold- and warm signal, respectively. If $e_c < e_w$, then the signals are correctly classified as "cold" and "warm", else they will be falsely classified. For the absolute energy comparison, the whole data was used, as well as only the first 90 s of the sequence.

4 Results

In this section the results of the different classification methods will be shown. The tables show the number of sets that the signals are divided into. After the number of sets, the table shows the predicted value and the confusion matrix. These values are used for the MCC and the classification value. The number of sequences indicate the optimal number of windows that was used for generating the values for the classification. The tables for the Welch method using Hanning window shows the amount of overlap of the windows, while Thomson method instead shows the type of window that was used.

4.1 Energy distribution

The results for the energy distribution are shown below. The Welch method is shown in Table 1. and the values for the Thomson method is displayed in Table 2.

Table 1: Energy distribution of Welch using Hanning window.

Overlap	**# Sets**	**Pred.**	**P**	**N**	**MCC**	**Class.**	**# Seq**
50%	2	P	40	24	0.3086	0.6509	9
		N	13	29			

Comparing the different results, it can be noticed that Thomson method generated better results than Welch method. The Thomson method using Slepian sequences improved by about 3% and with the Hermite by 2% compared to Welch method.

Table 2: Energy distribution of Thomson with different windows.

Window	**# Sets**	**Pred**	**P**	**N**	**MCC**	**Class.**	**# Seq**
DPSS	2	P	41	21	0.3829	0.6887	11
		N	12	32			
Hermite	2	P	39	21	0.3426	0.6698	11
		N	14	32			

4.2 Line of best fit

A linear model was implemented for the signal over the HF. The results are shown in the following tables. For Welch method the optimal number of sets is 4, since that had the highest MCC as well as the highest number of classification.

For the Thomson method the optimal number of sets is 2 for both the Slepian and Hermite.

Table 3: Line of best fit with Welch using Hanning Window.

Overlap	**# Sets**	**Pred**	**P**	**N**	**MCC**	**Class.**	**# Seq**
50%	2	P N	37 16	17 36	0.3774	0.6887	2
	4	P N	44 9	21 32	0.4455	0.7170	4
	10	P N	39 15	26 27	0.2518	0.6226	2

In this method Welch performed slightly better, with the correct classification being approximately 72% compared to Thomson method where the best classification being approximately 70%.

Table 4: Line of best fit with Thomson method using different windows.

Window	**# Sets**	**Pred**	**P**	**N**	**MCC**	**Class.**	**# Seq**
DPSS	2	P N	40 13	19 34	0.3988	0.6981	7
	4	P N	39 14	20 33	0.3608	0.6792	2
	10	P N	39 14	29 24	0.1967	0.5943	1
Hermite	2	P N	39 14	19 34	0.3790	0.6887	8
	4	P N	40 13	21 32	0.3626	0.6792	3
	10	P N	27 26	26 37	0.2113	0.6038	2

After the classification the confidence interval was computed for the $\hat{\beta}_1$ of the cold and warm data. The confidence intervals of both multitaper methods are displayed in Table 5 and Table 6. Observing the mean of the coefficients, overall it had the expected sign, with the exception of the coefficient when dividing the cold data into 10 sets and using Thomson with Hermite window. All of the lower bounds had negative signs and the upper bound had positive sign, which is not really what we wanted. For the cold signal we expected the lower bound to also have nonnegative sign and for the warm signal we expected the upper bound to have negative sign.

Table 5: Confidence interval for Welch method

# Sets	HRV	$\hat{\beta}_1$	LB	UB
2	cold	0.1058	-0.1634	0.3751
	warm	-0.1722	-0.4415	0.0970
4	cold	0.0558	-0.2134	0.3251
	warm	-0.0258	-0.2951	0.2434
10	cold	0.0151	-0.2541	0.2843
	warm	-0.0031	-0.2723	0.2662

Table 6: Confidence interval for Thomson method using different windows

Window	# Sets	HRV	$\hat{\beta}_1$	LB	UB
DPSS	2	cold	0.0885	-0.1808	0.3577
		warm	-0.1263	-0.3955	0.1430
	4	cold	0.0852	-0.1840	0.3545
		warm	-0.0209	-0.2901	0.2484
	10	cold	0.0036	-0.2657	0.2728
		warm	-0.0065	-0.2758	0.2627
Hermite	2	cold	0.0817	-0.1876	0.3509
		warm	-0.1445	-0.4137	0.1247
	4	cold	0.0971	-0.1722	0.3663
		warm	-0.0169	-0.2862	0.2523
	10	cold	-0.0074	-0.2766	0.2619
		warm	-0.0395	-0.3088	0.2297

4.3 Absolute Energy Comparison

The method of absolute energy comparison deals with pair-wise classification. For the tables in this section, instead of dividing the signal into sets, we now instead looked at the part of the signal that was used. The optimal length for this method can be noticed is the whole HRV signal with 180 s for both methods.

Table 7: Absolute energy comparison with Welch using Hanning Window.

Overlap	**Length**	**Pred**	**P**	**N**	**MCC**	**Class.**	**# Seq**
50%	180s	P	39	14	0.4717	0.7358	5
		N	14	39			
	90s	P	36	17	0.3585	0.6792	3
		N	17	36			

For this method, Thomson with Slepian performed the best with 77% correct classification, as can be seen in Table 8. Overall, it can be noticed that Thomson method performed better when comparing the sets when comparing Table 8 with Table 7. The best result by Welch is 74% correct classification.

Table 8: Absolute energy comparison with Thomson method using different windows.

Window	**Length**	**Pred**	**P**	**N**	**MCC**	**Class.**	**# Seq**
DPSS	180s	P	41	12	0.5472	0.7736	15
		N	12	41			
	90s	P	36	17	0.3585	0.6792	3
		N	17	36			
Hermite	180s	P	40	13	0.5094	0.7547	11
		N	13	40			
	90s	P	37	16	0.3962	0.6981	9
		N	16	37			

5 Discussion

All the results that was used in this thesis is generated by the Welch method using Hanning window with 50% overlap and the Thomson method using Slepian window or Hermite functions. The number of windows for both data sets was chosen carefully. Having too many windows for the Welch method and for the Thomson method, while having a small data set, could result in in a far too smooth spectral estimate, implying that necessary information could be lost. Therefore, the number of windows was adjusted according to the length of each set. The limitation of this restriction could be seen when dividing the data set into 10 sets. All sets could only use one or two windows. Since there was only one or two alternative for the optimal number window, this indicated that one would not optimize the method. Comparing the classification using different number of sets for both methods, there was not much differences. When classifying based on difference in energy distribution between the cold- and warm signal the best result for the Welch method was 65% and 69% for the Thomson method using DPSS.

In the classification assuming there is differences in energy distribution over time using line of best fit, the best results for the Welch method was approximately 72% and for the Thomson method approximately 70%. The Welch method was slightly better, even though, one would expect the Thomson method to perform much better than the Welch method since each window uses the whole data set. A possible reason for this, is that the bell-shaped curve of the Hanning window in the Welch method was better at capturing the prominent features of the data when it is divided. For the line of best fit method, one could also use confidence intervals to asses the accurateness. It could be seen that the confidence intervals overlapped. This means that under the significant level 5%, for the parameter β, there is not a statistically significant difference between being "cold" or "warm". Hence, the results from the line of best fit might not be trustworthy.

For the pair-wise classification, the Thomson method performed better, with best result being 77% correct classification compared to the Welch method 74% classification. The possible reason that the Thomson method performed better in the pair-wise classification could be because the energy concentration of the method was more suited for the whole data sequence.

The spectrum analysis was only conducted over the the frequency bandwidth $0.12 - 0.4$ Hz. It has been known that finding an individual frequency band through the respiratory signals could improve the results significantly, however, the reason for not using the respiratory signal is because the results that was produced was much worse than not using it.
Additionally, it can be noted that the higher frequency band, 0.48 - 1.6 Hz, was also tested. However, it was decided to not be included in the thesis, since there was not enough information to support the choice of frequency interval. Although, the choice of frequency band is an interesting topic for further analysis. It could be noticed that the data was not divided into a test and validation set,

due to the data set being too small in combination with large variances between the subjects.

Furthermore, a different normalization technique can also be of interest for the data set. Different from how the data was normalized in the thesis, we also tried to normalize the data according to,

$$Z = \frac{X - \bar{X}}{\|X\|_2}$$

where $\|X\|_2$ is the 2-norm of matrix X. The 2-norm is computed as,

$$\|X\|_2 = \max_{i=1}^{n} \lambda_i^{1/2}$$

where λ_i denotes the eigenvalues of $X^T X$ [8]. The normalized data using the 2-norm improved the results for the absolute energy comparison considerably. Some of the results even lead to perfect classification. The reason for the exceptional improvement could be due to the normalization technique using all the data that was available. However, since the normalizing method uses information that is not available when conducting the classification of a single HRV the results was, therefore, not included. Even though the result is not included, it is still interesting since it indicates that a different normalizing techniques might improve the classifications.

6 Conclusion

In conclusion, it is possible to classify the cold- and warm HRV signal using different multitaper spectral methods on this specific HRV data set. There does seem to be an difference in energy distribution over time, as well as, difference in total energy over the high frequency between the two HRV signals. In general, the line of best fit method had better results than the energy distribution method. However, since the results was not trustworthy we thus would refrain from using it. The best method for this data set was therefore when using the Thomson method with the DPSS window. For the pair-wise classification, the Thomson method still performed a bit better than the Welch method and, therefore, also more preferable. All in all, the Thomson method was most optimal for this data set. However, whether the results will be the same on a similar data set is still unknown and the methods still need to be evaluated. Since the results between both the methods were so close to each other, it is therefore still best to try out both methods when dealing with another data set. Different method can be better suited for different data sets.

7 References

[1] G. Billman, “Heart rate variability – a historical perspective”, Frontiers in Physiology, vol. 2, p. 86, 2011.

[2] M. Malik et al. ”Heart rate variability”. In: *European Heart Journal* 17 (1996), pp. 354−381.

[3] G. Lindgren, H. Rootzén, and M. Sandsten. *Stationary stochastic processes for scientists and engineers.* CRC Press, 2013.

[4] I. Reinhold. *Spectral Analysis for Signal Detection and Classification.* Media-Tryck, 2021.

[5] D.B. Percival, A. T. Walden. *Spectral Analysis for Physical Applications.* Cambridge University Press, 1993.

[6] Encyclopedia of Mathematics. Hermite polynomials. (Electronic version available at www.encyclopediaofmath.org)

[7] D. Anevski. *A concise introduction to mathematical statistics.* Studentlitteratur AB, 2017.

[8] E.Süli, D.Meyers. *An Introduction to Numerical Analysis.* Cambridge University Press ISBN, 2003