Classification of HRV-signals using Time-Frequency Analysis

by Henrik Kragh and Ante Forsgren

June 18, 2018

Populärvetenskaplig Sammanfattning

Heart Rate Variability, förkortat HRV, är en term inom kardiologi som beskriver den naturliga variationen i tidsintervallet mellan en människas hjärtslag. Högt HRV, det vill säga stor variation i tiden mellan hjärtslag, har tidigare visats vara kopplat till god hälsa. Det har även visats att låg variation är kopplat till långvarig psykisk påfrestning. Den starka korrelationen med en människas hälsa har gjort att HRV på senare år har blivit ett intressant studieområde där mycket forskning sker. Ett antal studier gjorts på hur långvarig stress och andra påfrestningar påverkar HRV, fokuserar denna avhandling istället på hur HRV förändras på kort sikt efter en intensiv påfrestning.

Mätningarna som använts är från 59 olika individer och framtagna på två olika sätt, i det första fallet har deltagaren fått sätta ner handen i rumstempererat vatten medans hjärtaktiviteten mättes. I det andra fallet har deltagaren istället fått sätta ner handen i isvatten för att framkalla fysisk smärta, och hjärtaktiviteten mättes på samma sätt. HRV beräknades sedan ur dessa hjärtaktivitetsmätningar.

Frågeställningen var om det fanns något sätt att separera på dessa två mätningar av HRV. Två studier gjordes, i den första studien hade man tillgång till båda mätningarna hos en individ, och målet var att klassifiera vilken mätning som var vilken. I det andra, lite mer komplicerade fallet, hade man bara tillgång till en mätning, och målet var att korrekt klassifiera vilken sorts mätning den var.

Båda studierna har gjorts med så kallad tidsfrekvensanalys, det vill säga att man uppmätt vilka frekvenser som uppkommer i HRV-mätningarna, samt hur dessa varierar över tid. Flera metoder har formulerats och kombinerats på olika sätt för att kunna separera på mätningarna, med goda resultat i båda studierna.

Slutsatsen man kan dra, eftersom man kan separera på mätningarna med sådan god säkerhet, är att det säkerligen finns en skillnad i en människas HRV väldigt fort efter den utsätts för påfrestning, vilket är kunskap som förhoppningsvis kan användas till fortsatta studier.

Abstract

Heart rate variability (HRV) is a term within cardiology describing the naturally occurring variation of the time interval between heartbeats, and high HRV activity has been linked to both cardiovascular and non-cardiovascular health. This bachelor thesis aims to find differences between HRV-signals where short term physical strain (pain) was induced and HRV-signals where no such strain was applied. This was done using time-frequency methods, primarily the spectrogram, with most of the analysis performed in the so called high frequency band (0.12 to 0.4 Hz). The estimated respiratory frequency was also used as input to limit the analysis to relevant frequencies. Multiple methods examining spectral power in these frequencies was then designed and applied. The most successful methods could classify a participants signal pair correctly 81.4% ($p<0.0001$) of the time, and a individual signal 72.9% ($p<0.0001$) of the time, indicating that the induced strain had a statistically significant impact on the HRV-measurements.

Contents

1 Introduction

1.1 Background

Heart rate variability (HRV) is a term within cardiology that describes changes in instantaneous heart rate between heartbeats. HRV is measured by the RR-interval, which is the interval between two heartbeats and is usually found by using an electrocardiogram (ECG), where the so called R-peak indicates a heartbeat. After the RR-interval is measured, the HRV-signal takes that value up until another heartbeat occurs and another RR-interval can be calculated, this leads to the block-like shape of the HRV-signal, as can be seen in figure 1 and figure 2.

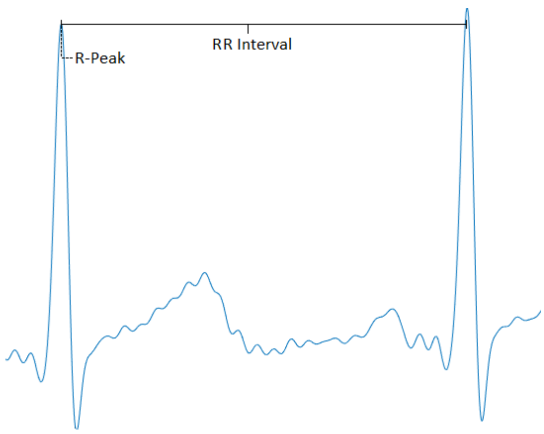

Figure 1: An ECG signal, with a R-Peak and the RR-Interval shown.

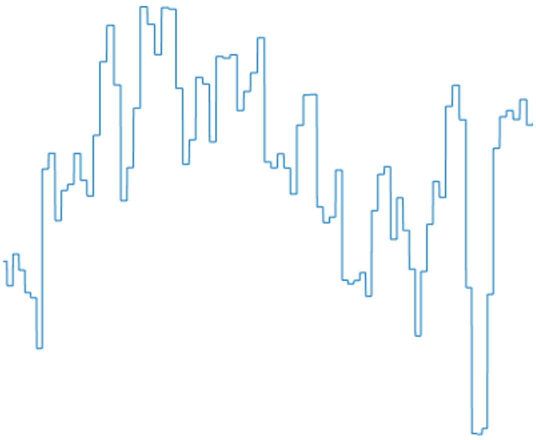

Figure 2: A HRV-signal.

The time-frequency spectrum of HRV is typically divided into the high frequency (HF) band from 0.12 to 0.4 Hz, the low frequency (LF) band from 0.04 to 0.12 Hz, and the very low frequency (VLF) band below 0.04 Hz [1][2]. These intervals are defined under relatively normal circumstances, and it is worth noting that the HF-band spectral peaks can shift up to 1.05 Hz in infants or during exercise. The analysis of the HRV spectra is usually confined to either the LF-band, the HF-band, or the ratio between the LF and HF spectral powers, known as the LF/HF-ratio [3].
HRV is, unlike arrhythmias, a naturally occurring phenomenon and high HRV is generally regarded as a sign of both cardiovascular and non-cardiovascular health [6]. On the other side of the spectrum several studies have shown that both short and long term psychological strain, such as stress or depression, is correlated with reduced HRV-activity [7][5].

HRV is also linked to respiration, being both positively correlated in frequency as well as negatively correlated in spectral power to the respiratory frequency. This occurs due to respiratory sinus arrhythmia (RSA), the process in which the heart-rate decreases during exhalation and increases during inhalation. RSA does not, despite the name, require regular breathing to occur [4].

The data used in this thesis are HRV and respiratory measurements from 59 male individuals between the ages 19-31. ECG as well as respiration were measured at 1000 Hz in a three minute interval. Two data sets were measured, one control set where the individual was asked to place their hand in tepid water during the entirety of the experiment, and one measurement where the individual was asked to place their hand in ice water during the entirety of the experiment to induce short term psychological strain. The HRV was then extracted from the ECG measured during the experiment and downsampled to 4 Hz. HRV-signals and respiratory signals from the first data set will be referred to as warm in this thesis, while HRV-signals and respiratory signals from the second data set will be referred to as cold.

1.2 Purpose

This thesis aims to develop an algorithm capable of differentiating between the warm and the cold data sets mentioned above. This will be done by using several methods of time-frequency analysis as well as regression models on the HF-HRV spectra, as well as on a frequency-interval surrounding the respiratory frequency.
Two classification problems were formulated:

- Two HRV signals $y_1(t), y_2(t)$ with corresponding respiratory signals $r_1(t), r_2(t)$ belonging to the same individual are given. One signal is warm, one is cold, and the objective is to determine which signal is which. This problem will be referred to as the Comparison Problem.

- One HRV signal $y(t)$ with corresponding respiratory signal $r(t)$ is given. The objective is to determine if the signal is from the warm or the cold data set. This problem will be referred to as the Categorization Problem.

The results on these two problems are what will be used to determine the effectiveness of all methods used.

2 Theory

2.1 The short time Fourier transform and the spectrogram

The short time Fourier transform (STFT) for a signal $x(t)$ is defined as

$$X(t,f) = \int_{-\infty}^{\infty} x(t_1)\omega(t_1 - t)e^{-i2\pi f t_1} dt_1$$

where $\omega(t)$ is a window function centered at t. Typically this window function is fixed, positive and even.

The spectrogram is a non-parametric spectral estimation method, defined as

$$S_x(t,f) = |X(t,f)|^2$$

meaning that for each time t, a spectral density is estimated using data filtered by the window centered at time t. These spectral densities are then combined to form a time-frequency plot. In this thesis the chosen window is the Hann (or Hanning) window, defined as

$$\omega(t) = \frac{1}{2}(1 - \cos(\frac{2\pi t}{N-1})), \quad -\frac{N}{2} < t < \frac{N}{2}$$

where N is the width of the window.

2.2 The Wigner distribution

The Wigner distribution is defined as

$$W_x(t,f) = \int_{-\infty}^{\infty} r_x(t,\tau)e^{-i2\pi f\tau} d\tau,$$

where $r_x(t,\tau)$ is the instantaneous auto-correlation function, defined as

$$r_x(t,\tau) = x(t + \frac{\tau}{2})x^*(t - \frac{\tau}{2}).$$

The smoothed Wigner distribution using a lag-independent kernel (referred to as the smoothed Wigner distribution) is an extension of the Wigner distribution.

2.3 The Hilbert transform

The analytic signal z(t),$-\infty < t < \infty$ is defined as

$$z(t) = x(t) + i\mathcal{H}(x(t))$$

where $\mathcal{H}(x(t))$ is the Hilbert transform of the real valued signal $x(t)$, defined as

$$\mathcal{H}(x(t)) = \mathcal{F}^{-1}((-i\ sgn(f)\mathcal{F}(x(t))).$$

A useful property of the analytic signal of a real signal is that its Fourier transform $\mathcal{F}(z(t))$, and thus its spectrum, are equal to 0 for all negative frequencies, without loss of information in the positive part. This can reduce aliasing in the time-frequency domain and is especially useful in the Wigner distribution in which aliasing results in additional cross-terms between each auto-term.

2.4 The Rényi entropy

The concentration measure being used is a time-frequency adapted version on the normalized Rényi entropy which, for a given time-frequency spectrum $W(t,f)$ is given as

$$R_\alpha = \frac{1}{1-\alpha}\log_2\Big(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Big(\frac{W(t,f)}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|W(t,f)|\,dtdf}\Big)^\alpha\Big)dtdf \quad , \qquad \alpha > 2, \in \mathbb{N}$$

resulting in a measure that is large for highly concentrated data, and lower for less concentrated data.

2.5 Linear regression - Ordinary least squares

The multivariate ordinary least squares (OLS) is used to estimate the coefficients β in the regression model

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & x_{21} & \cdots & x_{k,1} \\ 1 & x_{1,2} & x_{22} & \cdots & x_{k,1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,n} & x_{2,n} & \cdots & x_{k,n} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon$$

where $\mathbf{Y}$ is a dependent variable, β are the coefficients, $\mathbf{X}$ are the explanatory variables and ϵ is the error term. This is estimated by minimizing the sum of squared residuals i.e

$$\hat{\beta} = \arg\min_\beta (\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta),$$

which has the solution

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}.$$

The R^2 value of a regression is defined as

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2}$$

where $(\hat{y}_i)$ is the prediction of y_i, and $\bar{y}_i$ is the mean of y. The R^2 value is the proportion of variance in the dependent variable which can be predicted by the input variables. The adjusted R^2 value of a model is defined as

$$R^2_{Adj} = 1 - \frac{(1-R^2)(n-1)}{n-k-1}$$

where n is the sample size and k is the number of regressors. this has the same function as the normal R^2 value, but in addition penalizing you for adding regressors that do not fit the model.

2.6 Binary regression - The logistic model

Binary regression is regression in which the dependent variable $y \in \{0,1\}$, i.e. is binary. One of these are the logistic model, a regression model designed to estimate

$$P(y_i = 1|\mathbf{x}_i), \text{ where } y \in \{0,1\} \text{ and } \mathbf{x}_i \in \mathbb{R}^k.$$

We assume that $P(y_i = 1|\xi(\mathbf{x}_i)) = F(\xi(\mathbf{x}_i))$, where

$$F(\xi(\mathbf{x}_i)) = \frac{1}{1 + e^{-\xi(\mathbf{x}_i)}}, \quad \xi(\mathbf{x}_i) = \beta_0 + \beta_1 x_{i,1} + ... + \beta_k x_{i,k}$$

meaning that it estimates the probability that a given data point with given characteristics fall into a certain binary characteristic of the data, in this case that y=1. One can then make a prediction $\hat{y}_i$ on y_i by assigning that

$$P(y_i = 1|\mathbf{x}_i) > \frac{1}{2} \longrightarrow \hat{y}_i = 1$$

$$P(y_i = 1|\mathbf{x}_i) < \frac{1}{2} \longrightarrow \hat{y}_i = 0$$

.

2.7 Student's t-test

A t-test is any statistical hypothesis test where the statistic will follow a Student's t-distribution under the null hypothesis. The one-sample t-test for a given estimated mean $\hat{\theta}$ is used to test the hypothesis

$$H_0 : \hat{\theta} = \theta_0$$

$$H_1 : \hat{\theta} \neq \theta_0$$

and is performed using the statistic

$$t = \frac{\left|\hat{\theta} - \theta_0\right|}{\hat{\sigma}/\sqrt{n}}$$

Where $\hat{\sigma}$ is the estimated standard deviation of the data, and n is the sample size. This statistic will, under the null hypothesis as well as the assumtion that $\hat{\theta} \in N(\theta, V(\hat{\theta}))$, follow a Student's t-distribution with n-1 degrees of freedom, and can thus be compared to the quantiles of one. If the given t-value is greater than the quantile of the sought after level of confidence it is determined to be statistically significantly different from H_0, which then can be rejected.

The t-test can also be used to check the significance of a regression coefficient β_k on a multivariate OLS model. This is done via the hypothesis

$$H_0: \quad \beta_k = 0$$
$$H_1: \quad \beta_k \neq 0$$

and by using the statistic

$$T = \frac{|\beta_k|}{C_{kk}}$$

where

$$C = \hat{\sigma}^2 (X^T X)^{-1},$$

X is the matrix of all regressors and $\hat{\sigma}$ is the estimated standard deviation of the dependent variable.

3 Method

In an earlier project, an evaluation of spectral estimation methods for HRV-spectra was performed, using the negative correlation between respiration frequency and spectral power of HRV as an evaluation criteria. This was done both on a generated data set, but also on a real data set where the participants were instructed to breath with linearly increasing frequency. From this project, it was shown that the spectrogram was the better spectral estimator for HRV-signals. As such, most of this thesis also used the spectrogram as the spectral estimation method.

The experiment was conducted by Peter Jönsson at the university of Kristianstad. There were 59 participants in total, all being males ages 19-31 (mean 23.0, standard deviation 2.48). For the experiment, the participant was fitted with an ECG as well as a strain gauge across the chest to record respiration. For the control test, they were instructed to place their hand in tepid water and to sit in that state for 3 minutes while their ECG and respiration were being recorded. No instructions were given on respiration. The second test were constructed in the same manner, except their hand was instead placed in ice water, a so called cold pressor test, which is a test used both for pain tolerance tests as well as to measure changes in blood pressure and heart rate [8]. This is what acted as short term physical strain to affect the participants HRV. After the test was completed, the participant were asked to grade from a scale one to ten how painful they considered the experiment.

From the experiment, 180 seconds of ECG, HRV and respiratory data downsampled to 4 Hz was gained from the 59 participants. To remove, or at least notably reduce artifacts in the beginning and end of the spectrum, the signals were extended the following way:

$$y_e(t) = \left\{ \begin{array}{l} -y(-t) \text{ for } -90 \leq t < 0 \\ y(t) \text{ for } 0 \leq t \leq 180 \\ -y(360-t) \text{ for } 180 < t \leq 270 \end{array} \right\}$$

As a final step before the time-frequency spectra for these new signals were calculated, the Hilbert transform was applied. After the time-frequency spectrum for the transformed signals were calculated, they were then cut so only the original time-interval 0 to 180 seconds were analyzed further. The spectrograms of the HRV signals (one warm and one cold for every participant) were calculated with a window length of N=256 (64 seconds). Similarly, the corresponding respiratory time-frequency spectrum were calculated with a window length of N=64 (16 seconds). These window lengths were chosen the same as in the earlier spectral estimation method project.

3.1 Simple approaches

To denote where the analysis was performed, the concept of the transformed spectrogram ($S_y^{\bullet}(t,f)$) was introduced, where $\bullet$ denotes what method was used to create the transformed spectrogram. Multiple approaches to create the transformed spectrogram were formulated. The first, and simplest approach was to analyze the power found in the entire HF-band, this method was simply called the High Frequency Band (HFB) method. In this case, $S_y^{HFB}(t,f) = S_y(t,f)$ for $0.12 \leq f \leq 0.4$. However, an additional simple approach was also formulated.
Two mean spectrograms using all the HF-warm spectra and all the HF-cold spectra were created and are presented in figure 3.

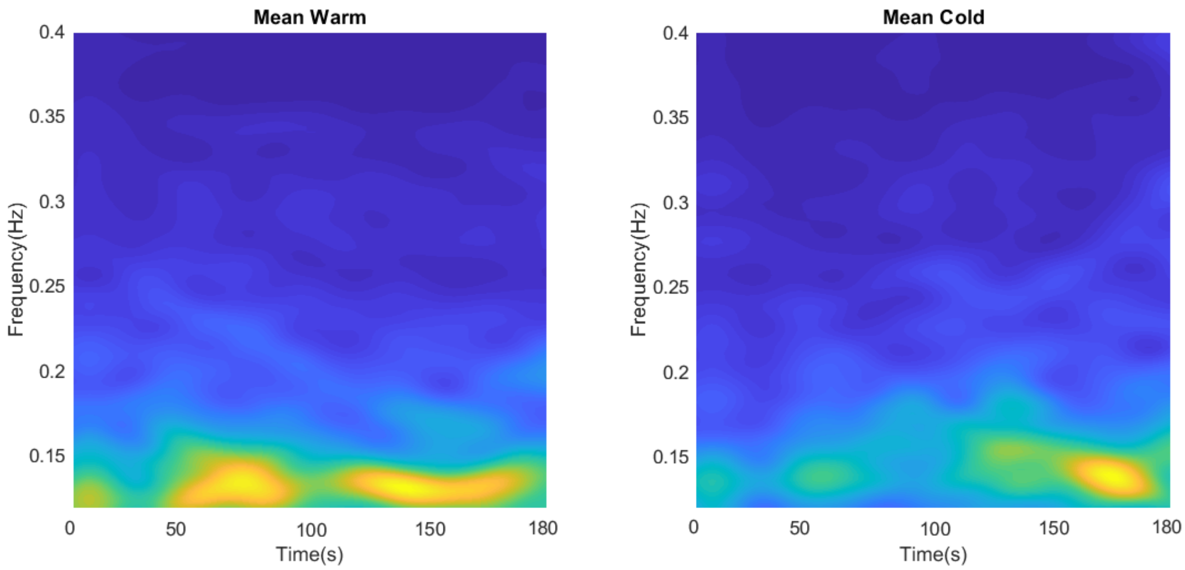

Figure 3: The mean spectrograms. As can be observed, most of the power is found between the frequencies 0.12 Hz and 0.2 Hz.

Observing where most of the power is located, analysis was also performed in a tightened band between 0.12 Hz and 0.2 Hz, i.e.,

$$S_y^{MB}(t,f) = S_y(t,f)$$

for $0.12 \leq f \leq 0.2$. This method was called the Mean Band (MB) method.

In addition to these methods, analysis using the respiration signal was also performed.

3.2 Respiration frequency methods

One way to reduce the sensitivity to noise in a HRV-spectrogram is to only analyze the spectral contents in a band around the respiratory frequency, which was also shown to improve the result even for the lower noise levels in the method evaluation project. As such, the respiration frequency was of great interest and had to be estimated.

Three different estimation methods for the respiratory frequencies were developed. To ensure the analysis of the HRV-spectrograms remained within (or close to) the HF-band, all respiration frequency finder methods were applied only on frequencies between 0.12Hz and 0.4Hz. The time-frequency spectrum of the respiratory signal was estimated using the spectrogram, the Wigner distribution and smoothed Wigner distribution using a lag-independent kernel.

3.2.1 The respiratory max method

The Respiratory Max Method (RMM) was the first method used, and the most simple one as well. The method takes the maximum frequency of the spectrogram for every point in time as the respiratory frequency, i.e

$$resp(t) = \arg\max_f S_r(t,f)$$

where $S_r(t,f)$ is the respiratory spectrogram. A visualization of the respiratory frequency found using RMM is presented in figure 4.

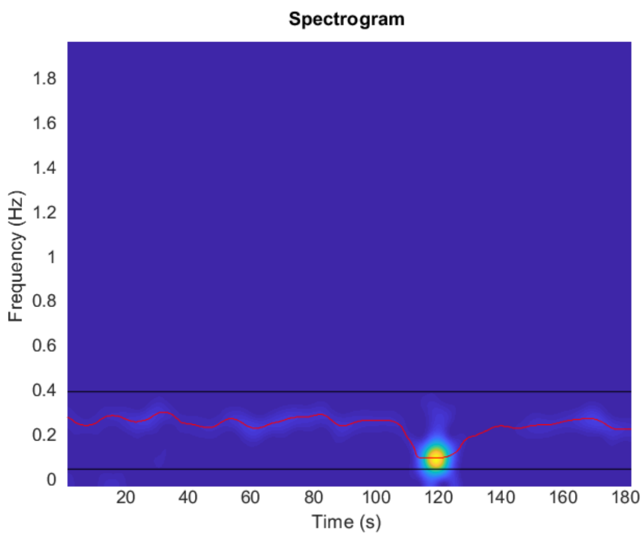

Figure 4: The respiratory spectrogram with the respiration frequency found using RMM.

3.2.2 The respiratory grid method

The Respiratory Grid Method (RGM) cuts the respiratory spectrogram into a number of rectangles, forming a grid with 50% overlap in frequency but no overlap in time. The total spectral power is then calculated for every rectangle, and the mean frequency for the rectangle with the highest spectral power for every point in time is then taken as the respiratory frequency. A visualization of this grid is presented in figure 5.

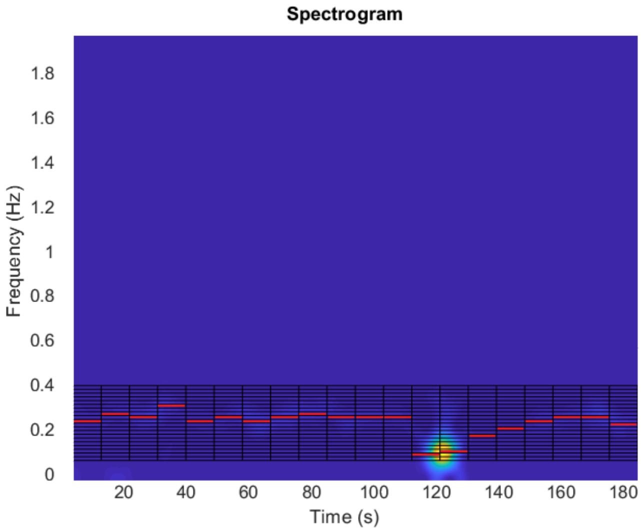

Figure 5: A visualization of the RGM. The rectangles containing the most power are in red. The mean frequency in these rectangles are then chosen as the respiratory frequency.

Mathematically:
Let m be the number of rectangles in time, n be the number of rectangles in frequency. Then define $f_0, f_1, ..., f_n$ and $t_0, t_1, ..., t_m$ such that $f_0 = 0.12$, $t_0 = 0$ with $f_{b+1} = [f_b + \frac{(0.4-0.12)}{n}]$ and $t_{k+1} = [t_k + \frac{180}{m}]$, then $resp(t) = f_{a(t)+1}$ where $a(t)$ is given by:

$$a(t) = \arg\max_b \int_{f_b}^{f_{b+2}} \int_{t_k}^{t_k+1} S_r(s,\nu) ds d\nu \ ,$$

where t_k is chosen such that $t_k \leq t < t_{k+1}$ and where $S_r(t,f)$ is the respiratory spectrogram. A visualization of the respiratory frequency found using RGM is presented in figure 6.

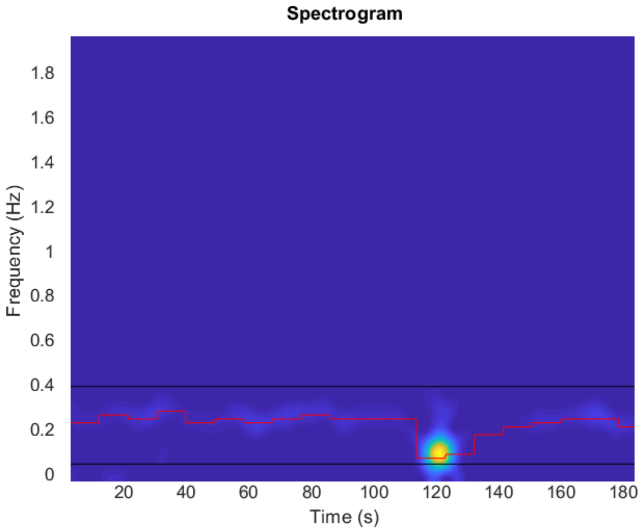

Figure 6: The respiratory spectrogram with the respiration frequency found using RGM.

3.2.3 The integrated respiratory method

The Integrated Respiratory Method (IRM) sums all the power in a rectangle around most points of the spectrogram to form the Integrated Spectrogram, RMM is then applied on this Integrated Spectrogram instead, i.e
let w, h be the width and the height of the rectangle, respectively.

$$IS_r(t,f) = \left\{ \begin{array}{ll} \int_{f-h}^{f+h} \int_{t-w}^{t+w} S_r(s,\nu) ds d\nu, & \text{for } 0.12 + h \leq f \leq 0.4 - h, w \leq t \leq 180 - w \\ S_r(t,f), & \text{for all other values of } f \text{ and } t \end{array} \right\}$$

where $S_r(t,f)$ is the respiratory spectrogram. A visualization of the respiratory frequency found using IRM is presented in figure 7.

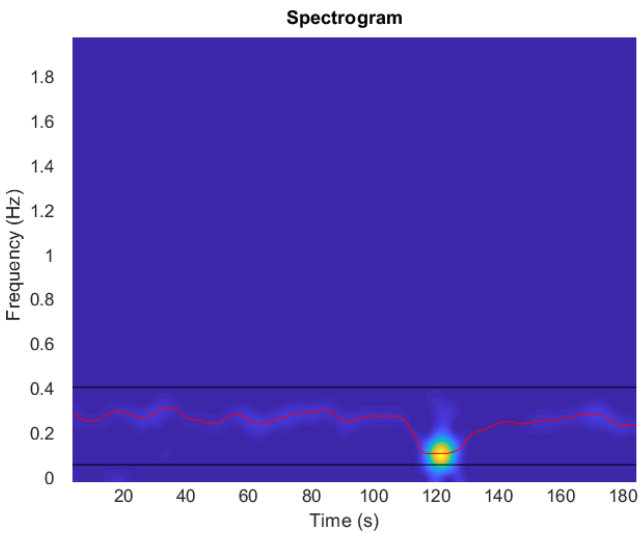

Figure 7: The respiratory spectrogram with the respiration frequency found using IRM.

After the respiration frequencies were found, the transformed spectrogram was formed by taking the parts of the spectrogram located 0.04 Hz around the respiration frequency, i.e $S_y^{RMM}(t,f) = S_y(t,f)$ for $0 \leq t \leq 180$ and $resp(t) - 0.04 \leq f \leq resp(t) + 0.04$ where $S_y(t,f)$ is the HRV-spectrogram. The following methods were applied on the transformed spectrogram.

3.3 Comparison tests for HRV

To find where the differences between the HRV-signals could be located, the transformed spectrograms were normalized such that their total sum of power was equal to one. After this was performed, the transformed spectrograms were split up in twelve time-intervals (for clarity reasons, these intervals will be referred to as epochs for the rest of the thesis with time-interval one being epoch 0.), and the total spectral power of each epoch was then calculated. Epoch 0 contained notably higher normalized power in both the warm and the cold signal, more than any other epoch, which was believed to be some sort of artifact that remained even after the data was treated. As such, epoch 0 was discarded, and not used in any analysis.

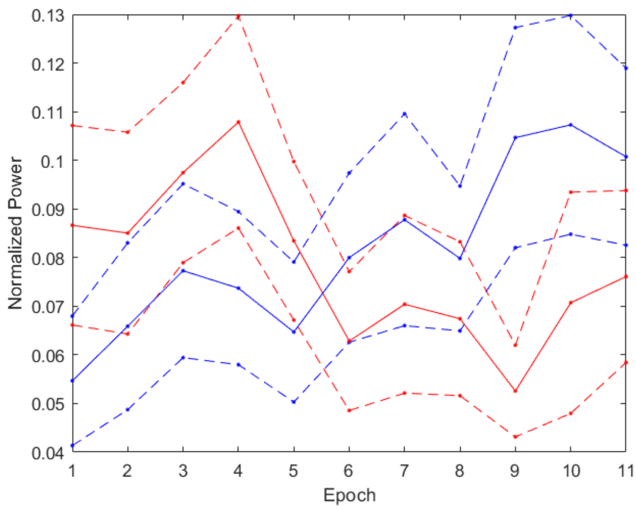

Figure 8: Normalized power, spread over eleven epochs. Red is the mean and confidence interval from 59 warm HRV-signals, blue is the mean of 59 cold HRV-signals.

From these eleven epochs, one could clearly see that the warm HRV-signal had most of its power in the first five epochs, whilst the cold HRV-signal had most of its power in the last six epochs, which is presented in figure 8. To see if this difference was statistically significant, a one sample t-test was performed on the difference in (normalized) power between the warm and cold spectra, on every epoch. Visualizations of this difference is presented in figures 9 to 13. Additionally, it was calculated whether the total power of the two signals over some interval was statistically significantly different. The intervals tested were, the first half (the first 90 seconds of measurements), the second half (the last 90 seconds of measurements), the entire 180 seconds of measurement, as well as the statistically significant epochs. After the total sum of power in the transformed spectrograms were calculated in these intervals, a one sample t-test for each interval was performed on the difference between the warm and the cold power. From this, a couple of comparison tests were formulated in an attempt to decide which, out of two signals (belonging to the same participant) was the warm one, and which one is the cold one. All of the tests uses the transformed spectrograms as their starting point, with all of the analysis methods used to create the transformed spectrograms.

3.3.1 The power sum test

After the transformed spectrograms were formed, the total spectral power was calculated over some interval for the two transformed spectrograms and then compared. Depending on the interval chosen, the choice of whether the signal with more power is decided to be the warm or the cold one will differ.

3.3.2 The interval shape test

The second comparison test formulated was the Interval Shape Test (IST). The transformed spectrograms were created as described before and then normalized so that the total power of each was equal to one. This normalized spectrogram was then split up into eleven epochs and the total power of each epoch was calculated. The power in the statistically relevant epochs were then compared, both individually, but also together. Figure 8 shows whether the signal with more power should be classified as the warm signal or the cold signal, depending on the epoch compared. If the epoch being compared is one of the first five, then the signal with the higher power in that epoch would be classified as the warm signal. If the epoch is instead one of the latter six epochs, the signal with lower power would be classified as the warm signal. In the case where all significant epochs were used, whatever signal that was classified as warm in the most epochs was also what the test concluded to be warm.

3.3.3 The total interval shape test

As an extension to IST, the Total Interval Shape Test (TIST) was also formulated. Using the same calculations as in IST, one instead uses all epochs, including the non statistically relevant ones. Figure 8 shows whether higher power in an epoch indicates the signal should be classified as warm or cold. If six or more epochs regarded a signal as warm, that signal was chosen as warm by the test.

3.3.4 The Rényi spread test

Another comparison test formulated was the Rényi Spread Test (RST). After the transformed spectrogram was formed, the Rényi entropy with $\alpha = 3$ was calculated and compared. By comparing the means it was decided that the signal with higher Rényi entropy should be classed as the warm signal. A t-test was also performed to see if the difference in Rényi entropy was statistically significant.

3.3.5 The combined comparison test

After the mentioned comparison tests were analyzed, an attempt to combine them was made. The first version used the results from PST and TIST. Whenever the tests agreed, that result stood. If the two tests disagreed however, RST was used as the decider. This version of CCT was called CCT1.
A second version of the Combined Comparison Test was also performed using scored versions of PST and TIST. After the results of PST was determined, a score was calculated and given in favour of what PST considered to be the warm signal. The same was performed for the TIST result. The scores were calculated as:
For PST: $1 - \frac{p_c}{p_w}$, where: p_c is the total power of the signal classified as cold, and p_w is the total power of the signal classified as warm. If there is a big difference in power between the signals, this score will be close to 1, and if the difference in power is small, the score will instead be close to 0.
For TIST: $\frac{(n_s - n_f)}{11}$, where: n_s is the number of epochs where the (according to TIST) warm signal was warm, and n_f is the number of epochs where the (according to TIST) warm signal was incorrectly regarded cold. If all epochs agreed that a signal was warm, this score would be equal to 1, if however six epochs considered a signal warm, while the remaining five epochs considered the signal cold, the score would instead be only $\frac{1}{11}$. If PST and TIST agreed on which signal was the warm one, that signal was chosen as the warm one. In the case of conflicting result however, the two scores were compared, with whatever test yielding the higher score being the test that was trusted. This version of CCT was called CCT2.

3.4 Categorization tests for HRV

The methods formulated to categorize a single HRV signal, were largely based on the comparison methods already formulated, but comparing against means instead of against another HRV signal. All of the tests uses the transformed spectrograms as their starting point, with all of the analysis methods used to create the transformed spectrograms.

3.4.1 The power sum test

In the Power Sum Test (PST) analogue, the total power over some interval was calculated. Next, the euclidean distance between the calculated power and the warm and cold means were calculated and compared. Whichever distance is lower, would be what the test chooses as its result.

3.4.2 The interval shape test

In the Interval Shape Test and Total Interval Shape Test analogues, the spectral power of the normalized transformed spectrogram was calculated for eleven epochs. The euclidean distances against the warm mean and cold mean was then calculated for every epoch. Whichever distance is lower, will be what the test chooses as its result for that epoch. When multiple epochs were used, whatever result is found in the most epochs would be what the test categorizes the signal as.

3.4.3 The combined categorization test

In an attempt to combine the newly formulated Categorization Tests, some scores was formulated, similar to the Combined Comparison Test. These scores were:
For PST: $\frac{\min(d_a, d_m)}{\max(d_a, d_m)}$ where d_a is the calculated power and d_m is the warm mean if the signal was chosen as warm and cold mean if the signal was chosen as cold.
For TIST: $\frac{|n_s - n_f|}{11}$ where n_s is the number of epochs the signal was deemed warm and n_f is the number of epochs the signal was deemed cold. If PST and TIST agreed on whether the signal was warm or cold, that decision stuck as the result. In the case of conflict, whatever test that yielded the higher score was trusted.

3.5 Regression models

3.5.1 logistic models

Lastly, a couple of logistic models were made, both for categorization and for comparison. A logistic model using only the total power of the transformed spectrogram as input was made, as was a model using only the statistically significant epochs of the normalized spectrogram. Lastly, a logistic model combining all of these inputs was made. For the comparison tests, whichever signal yielded the higher probability of being warm, was chosen as the warm signal. This regression was performed multiple times, first by using all 59 individuals to determine a model, then by using data from fewer participants. When less data was used, the individuals used in the regression were chosen at random, and the model was then evaluated both using only individuals that was not included in the regression, but also all 59 individuals. This was done using a minimum of ten participants for the regression and was repeated 1000 times for every number of individuals used in the regression.

3.5.2 OLS regression for pain

Two regular OLS regression models were also made to predict the self-reported pain level after the experiment was performed. Both models used total power over some interval of both transformed spectrograms as well as the individuals age and BMI. The first model used the normalized power in only the statistically significant epochs of both spectrograms while the second model used all epochs.

3.5.3 Repeated random sub-sampling validation

To somewhat validate if methods were not overly fitted to the data, a random sub-sampling was performed. A number of individuals was randomly chosen, and the models were built using only these individuals. This process was then repeated 1000 times for every amount of people. In the case of PST, the percentage of the time the difference in total spectral power over some interval was statistically significant was analyzed. For TIST, both the case where the mean normalized power of the warm spectra was higher than the mean normalized power of the cold spectra in exactly epoch 1-5 and the case where an error in one epoch was allowed was analyzed.

4 Results

In this result section, visualizations of the combined tests are shown. The results will be presented in the following manner:

- Visualizations of the paired difference in normalized power
- The comparison results, using all methods
  - Visualization of CCT2 Results
  - Results of all tests
  - Summary of most successful tests
- The categorization results, using all methods
  - Visualization of CCT Results
  - Results of all tests
  - Summary of most successful tests
- The regression model results
- The model results when using less data

Some tables containing results will also include a p-value. This p-value is calculated using the null hypothesis $H_0 : \hat{p} = 0.5$ by calculating $P(Y \geq x)$ where $Y \in$ Bin(59,0.5) and x is the result of the tested method. A one sample, one sided t-test was used to calculate the significance of the coefficients in the regression model.

4.1 Difference in Normalized Power

The mean difference in normalized power found in every epoch, for every method. The dashed lines are the 95% confidence interval. The epochs where the difference was found to be statistically significant are reported under each figure.

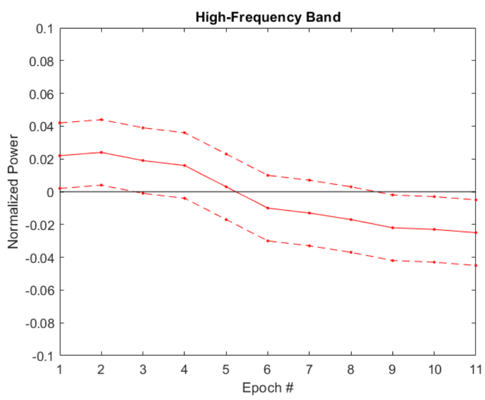

Figure 9: No difference was found to be statistically significant.

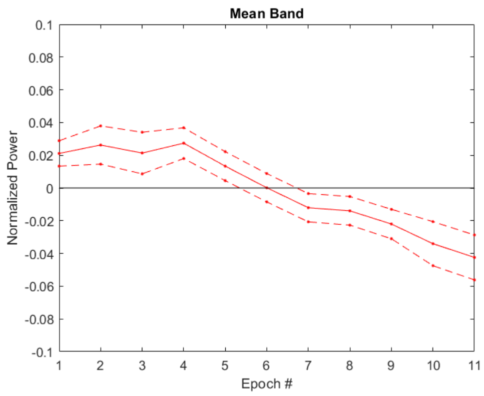

Figure 10: The differences found in epoch four, nine, ten and eleven was found to be statistically significant.

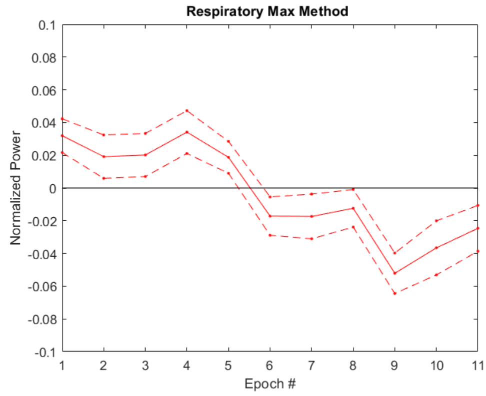

Figure 11: The differences found in epoch four, nine, and ten was found to be statistically significant.

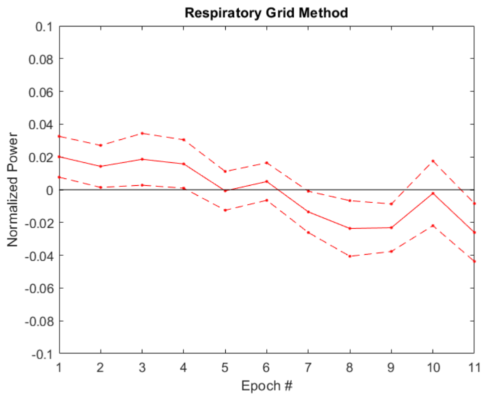

Figure 12: The difference found in epoch nine was found to be statistically significant.

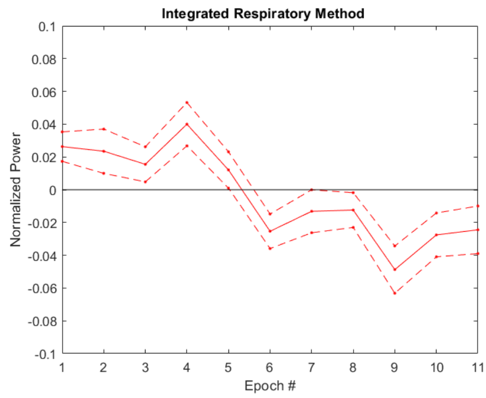

Figure 13: The differences found in epoch four, nine, and ten was found to be statistically significant.

4.2 Comparison Results

Multiple intervals was tried for PST, but only statistically significant results were found when using the first half of the spectra and using only epoch 4. The results using the first half of the spectra was better, and as such only these results are presented. The results for IST used different epochs depending on the method used, which epochs used can be seen in figure 14 to figure 18. Only the most succesful IST results are presented.

4.2.1 CCT2 Visualizations

A visualization of the CCT2 results using all methods. The calculated score of PST is in red, the calculated score of TIST is in blue, and the sum of these scores in purple. A score that is negative means that test has chosen the cold signal as warm, while a positive score means that the test correctly chose the warm signal as warm.

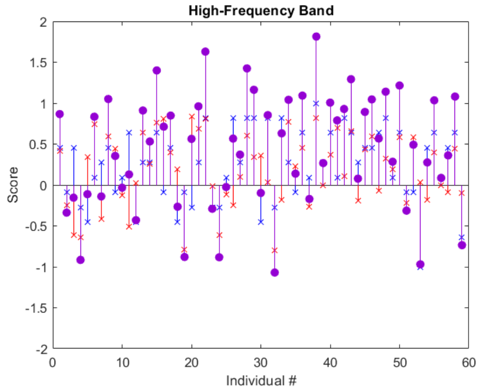

Figure 14

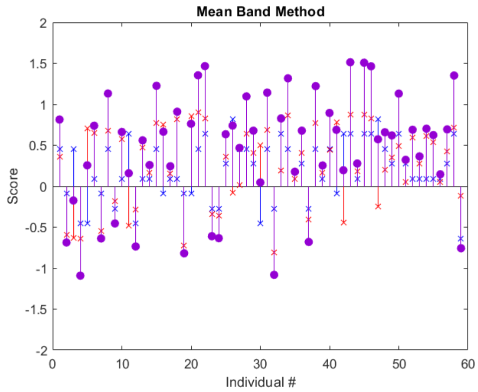

Figure 15

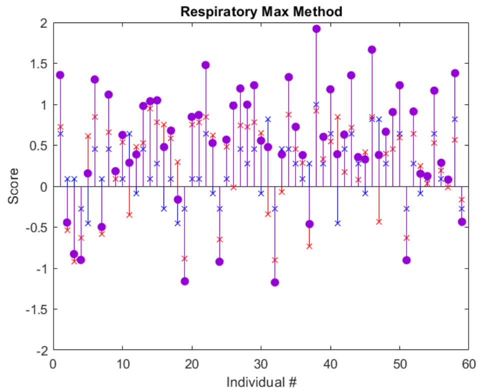

Figure 16

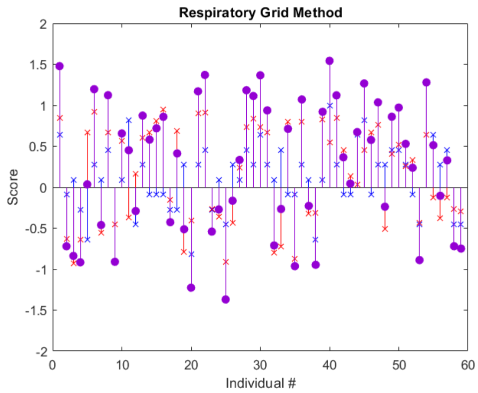

Figure 17

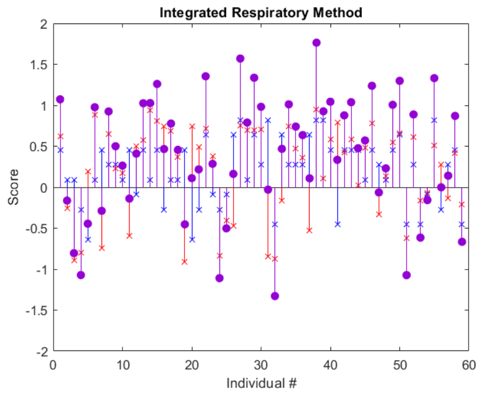

Figure 18

4.2.2 Results

Table 1: Percentages of correct comparisons.

	PST	IST	TIST	RST	CCT1	CCT2
HFB	62.7%	N/A	71.1%	62.0%	69.4%	64.4%
MB	71.1%	69.5%	71.1%	52.5%	71.1%	78.0%
RMM	76.3%	76.2%	74.6%	54.2%	76.3%	81.4%
RGM	57.6%	61.0%	61.0%	55.9%	59.3%	61.0%
RMM	67.8%	66.1%	71.1%	49.2%	79.2%	71.1%

Table 2: Number of correct comparisons.

	PST	IST	TIST	RST	CCT1	CCT2
HFB	37	N/A	42	37	41	38
MB	42	41	42	31	42	46
RMM	43	45	44	32	45	48
RGM	34	36	36	33	35	36
RMM	40	39	42	29	47	42

As can be seen from table 1 and table 2, RMM is the most successful method and CCT2 is the most successful test. However, all tests save for RST gave somewhat satisfactory results. Lastly, the MB method was very successful given that it did not use the respiratory signal as an input.

4.3 Categorization Results

As RMM was the clearly superior respiration method, categorization results are only presented using RMM. PST results are using the first half of the spectra and only the most successful IST results are reported.

4.3.1 CCT Visualizations

A visualization of the categorization CCT results using all methods. The calculated score of PST is in red, the calculated score of TIST is in blue, and the sum of these scores in purple. A score that is negative means that test has chosen the cold signal as warm, while a positive score means that the test correctly chose the warm signal as warm.

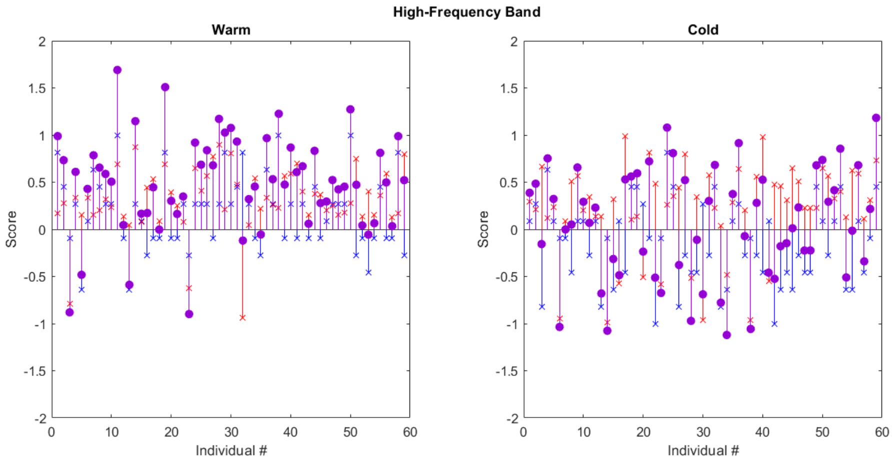

Figure 19

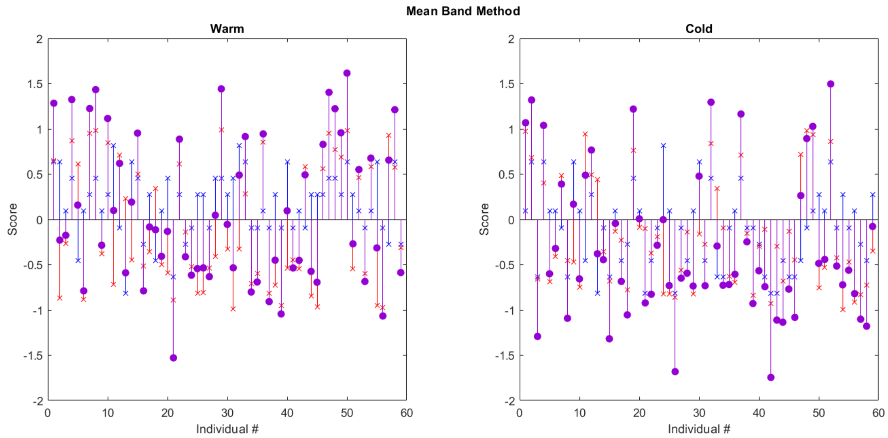

Figure 20

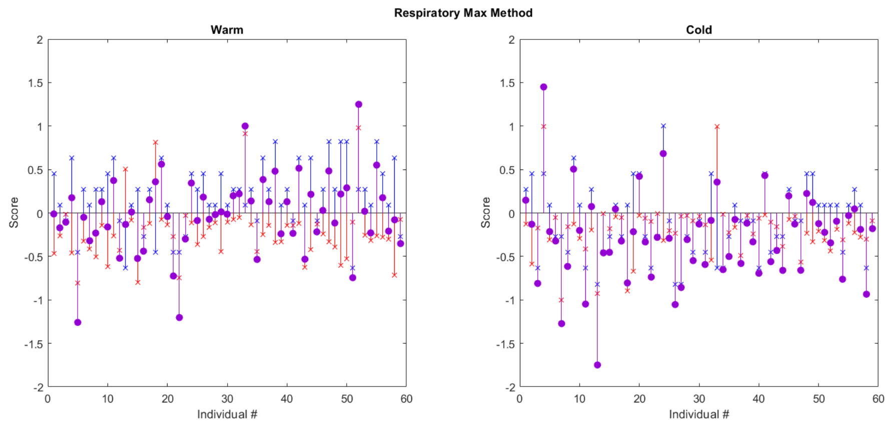

Figure 21

4.3.2 Results

Table 3: Percentages of correct classifications.

(a) Warm

	PST	IST	TIST	CCT
HFB	38.9%	N/A	67.7%	71.1%
MB	42.3%	72.8%	74.5%	88.1%
RMM	37.3%	78.0%	74.6%	71.1%

(b) Cold

	PST	IST	TIST	CCT
HFB	72.8%	N/A	59.3%	40.6%
MB	74.6%	57.6%	47.4%	39.0%
RMM	78.0%	50.8%	61.0%	40.6%

Table 4: Number of correct classifications.

(a) Warm

	PST	IST	TIST	CCT
HFB	23	N/A	40	42
MB	25	43	44	52
RMM	20	46	44	42

(b) Cold

	PST	IST	TIST	CCT
HFB	23	N/A	35	24
MB	44	34	28	23
RMM	42	30	36	24

As can be seen from table 3 and table 4, all tests are somewhat skewed to either categorizing a signal as warm (in the case of IST, TIST ans CCT) or cold (in the case of PST). The only method and test combination that even gets close to satisfactory results is RMM using TIST.

4.4 Regression models

Logistic models were also made with different inputs. Both when performing mean band analysis, as well as RMM.

4.4.1 Total power

Using the total power found in the first half of the signal as input for the logistic model yielded the results presented in table 5, table 6 and figure 22.

Table 5: Coefficients and p-values for the regression models using only total power.

Regressor	MB-Coefficient	MB p-value	RMM-Coefficient	RMM p-value
Intercept	-0.4613	0.1060	-0.2822	0.2782
Power	0.0094	0.040	0.008	0.1341

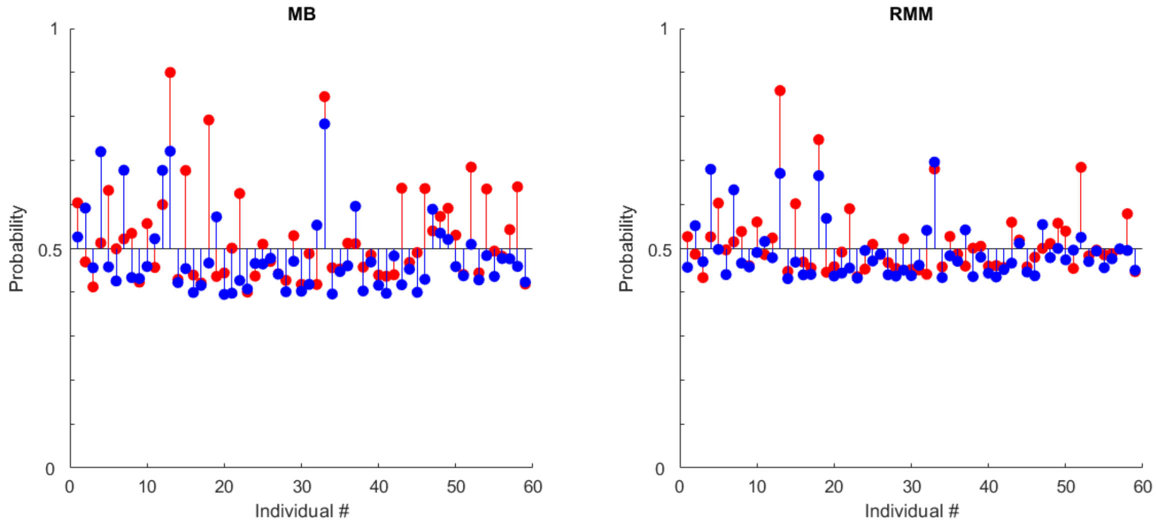

Figure 22: The estimated probability of a signal being warm using the power found in the first half of the signals spectrum as input of a logistic model for every participant. Stems in red are the probabilities for warm signals, while stems in blue are the probabilities for the cold signals. The left plot is the result using the mean band approach, and the right plot is the result using RMM.

Table 6: Results for the regression models using only total power.

Method	Signal	# of Correct	% of Correct	p-value
MB	Warm	27	45.7%	0.7825
MB	Cold	44	74.6%	0.0001
MB	Comparison	43	72.9%	0.0003
RMM	Warm	24	40.6%	0.9413
RMM	Cold	46	77.9%	<0.0001
RMM	Comparison	43	72.9%	0.0003

4.4.2 Normalized power

Using the normalized power found in epoch four, nine and ten as input for the logistic model yielded the results presented in table 7, table 8 and figure 23.

Table 7: Coefficients and p-values for the regression models using only normalized power.

Regressor	MB-Coefficient	MB p-value	RMM-Coefficient	RMM p-value
Intercept	0.0234	0.9649	0.7650	0.1365
Epoch 4	7.6664	0.0460	4.8591	0.1314
Epoch 9	-2.5772	0.6174	-14.8795	0.0017
Epoch 10	-5.1787	0.1774	-1.5948	0.5315

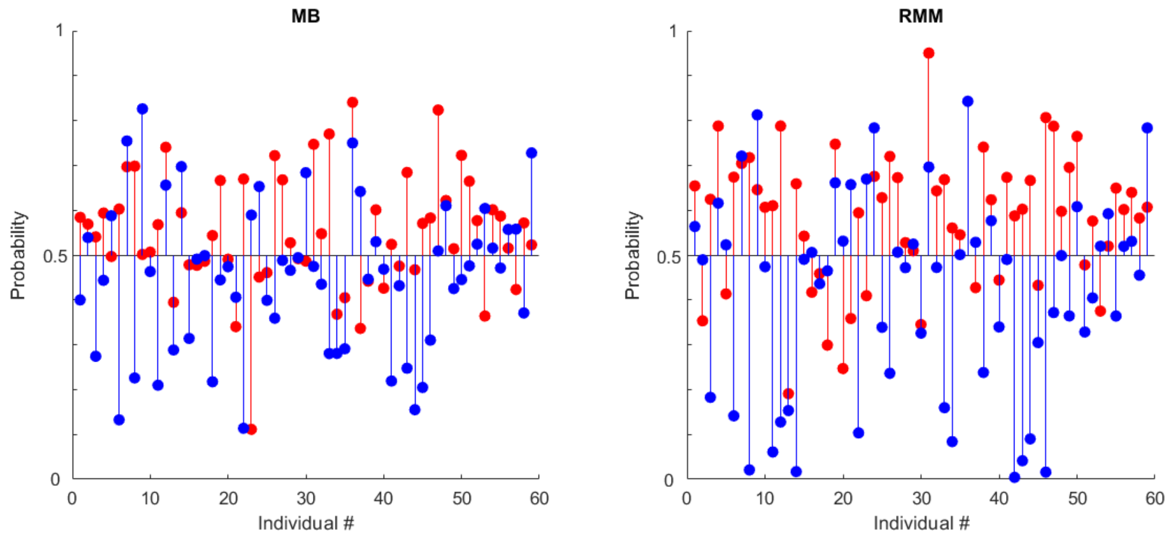

Figure 23: The probability of a signal being warm using the normalized power found in the fourth, ninth and tenth epoch of the spectrum as input of a logistic model for every participant. Stems in red are the probabilities for warm signals, while stems in blue are the probabilities for the cold signals. The left plot is the result using the mean band approach, and the right plot is the result using RMM.

Table 8: Results for the regression models using only total power.

Method	Signal	# of Correct	% of Correct	p-value
MB	Warm	38	64.4%	0.182
MB	Cold	39	66.1%	0.0092
MB	Comparison	41	69.4%	0.0019
RMM	Warm	44	74.6%	0.0001
RMM	Cold	35	59.3%	0.0963
RMM	Comparison	44	74.6%	0.0001

4.4.3 Total power and normalized power combined

Using the total power found in the first half of the signal and the normalized power found in epoch four, nine and ten as input for the logistic model yielded the results presented in table 9, table 10 and figure 24.

Table 9: Coefficients and p-values for the regression models using only total power and normalized power.

Regressor	MB-Coefficient	MB p-value	RMM-Coefficient	RMM p-value
Intercept	-0.1373	0.8027	0.6145	0.2445
Power	0.0077	0.0984	0.0103	0.0623
Epoch 4	5.9055	0.1400	3.9421	0.222
Epoch 9	-4.2502	0.4348	-17.3583	0.001
Epoch 10	-4.7178	0.2336	-1.4910	0.5689

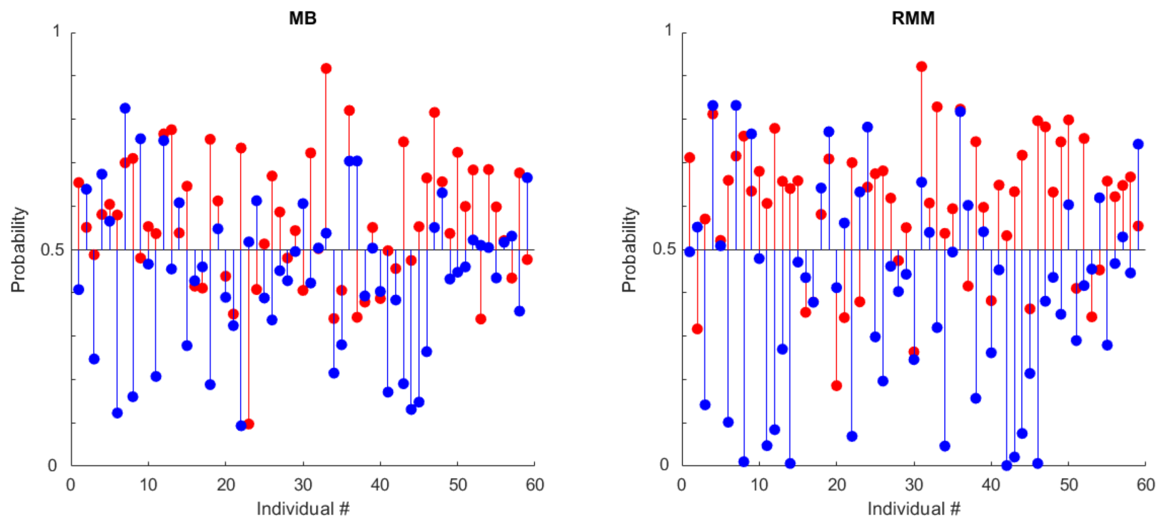

Figure 24: The probability of a signal being warm using the total power found in the first half of the spectrum and the normalized power found in the fourth, ninth and tenth epoch of the spectrum as input of a logistic model for every participant. Stems in red are the probabilities for warm signals, while stems in blue are the probabilities for the cold signals. The left plot is the result using the mean band approach, and the right plot is the result using RMM.

Table 10: Results for the regression models using only total power and normalized power.

Method	Signal	# of Correct	% of Correct	p-value
MB	Warm	38	64.4%	0.0182
MB	Cold	35	59.3%	0.0963
MB	Comparison	42	71.1%	0.0008
RMM	Warm	44	74.6%	0.0001
RMM	Cold	35	59.3%	0.0963
RMM	Comparison	44	74.6%	0.0001

Using the total power found in the first half of the signal and the normalized power found in all epochs as input for the logistic model yielded the results presented in table 11, table 12 and figure 25.

Table 11: Coefficients and p-values for the regression models using only total power and normalized power.

Regressor	MB-Coefficient	MB p-value	RMM-Coefficient	RMM p-value
Intercept	1.4839	0.7648	0.4612	0.7862
Power	0.0051	0.2885	0.0097	0.0936
Epoch 1	-1.3575	0.9358	5.0028	0.3581
Epoch 2	10.3070	0.1169	-0.1091	0.9763
Epoch 3	-15.6330	0.1829	1.3296	0.7572
Epoch 4	18.9232	0.0314	1.2609	0.7770
Epoch 5	-14.3364	0.2298	10.7420	0.0518
Epoch 6	9.6923	0.3076	-10.2171	0.0361
Epoch 7	-14.3860	0.1881	-2.2782	0.5174
Epoch 8	4.4304	0.6645	-1.4910	0.2938
Epoch 9	-8.5385	0.4241	5.3308	0.0022
Epoch 10	-1.6068	0.8391	-21.1065	0.0623
Epoch 11	-6.9580	0.2828	-0.1592	0.9644

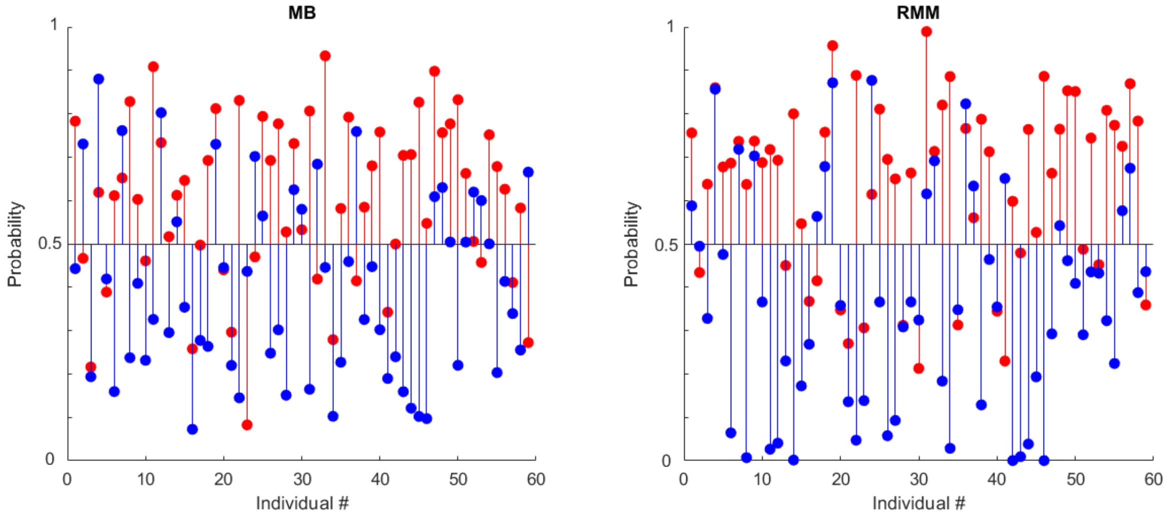

Figure 25: The probability of a signal being warm using the total power found in the first half of the spectrum and the normalized power found every epoch of the spectrum as input of a logistic model for every participant. Stems in red are the probabilities for warm signals, while stems in blue are the probabilities for the cold signals. The left plot is the result using the mean band approach, and the right plot is the result using RMM.

Table 12: Results for the regression models using only total power and normalized power.

Method	Signal	# of Correct	% of Correct	p-value
MB	Warm	41	69.5%	0.0019
MB	Cold	39	66.1%	0.0092
MB	Comparison	45	76.2%	<0.0001
RMM	Warm	43	72.9%	0.0003
RMM	Cold	43	72.9%	0.0003
RMM	Comparison	48	81.4%	<0.0001

4.4.4 Ordinary least squares for pain

The models yielded the result presented in table 13. The estimated coefficients are not presented as they are far too many. In addition to the regressors stated below, each model also used total power over some interval of both the warm and cold transformed spectrograms as well as the participants age and BMI.

Table 13

Regressors	R^2_{Adj}
Significant epochs	-0.0184
All epochs	0.2080

4.5 Models using less data

4.5.1 Logistic models

The following results are for the logistic Model using varying amounts of participants in the regression, using total power in the first half of the spectrum and the normalized power found in epoch four, nine and ten as inputs. This was performed both with mean band analysis and RMM, with results presented in figure 26 and figure 27 respectively.

MB method

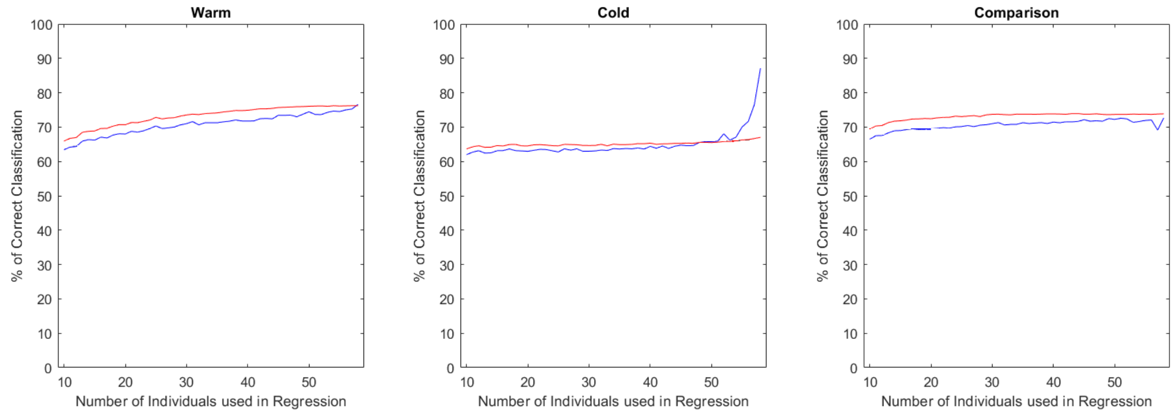

Figure 26: The percentage of signals receiving correct classification as a function of the amount of individuals used in the regression. The red line is all 59 signals, while the blue line is just the signals not used in the regression.

RMM

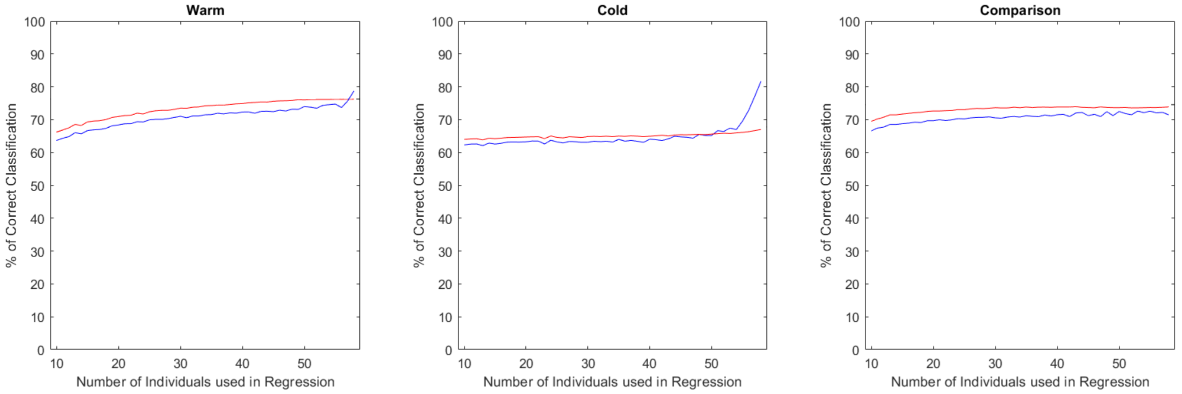

Figure 27: The percentage of signals receiving correct classification as a function of the amount of individuals used in the regression. The red line is all 59 signals, while the blue line is just the signals not used in the regression.

The Following results are for the logistic model using varying amounts of participants in the regression, using total power in the first half of the spectrum and the normalized power found in all epochs as inputs. This was done both with mean band analysis and RMM, with results presented in figure 28 and figure 29 respectively.

MB method

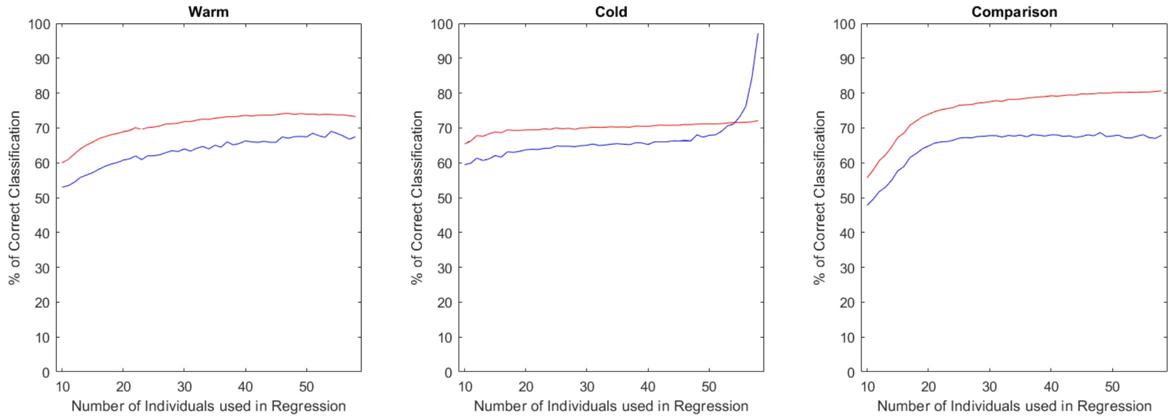

Figure 28: The percentage of signals receiving correct classification as a function of the amount of individuals used in the regression. The red line is all 59 signals, while the blue line is just the signals not used in the regression.

RMM

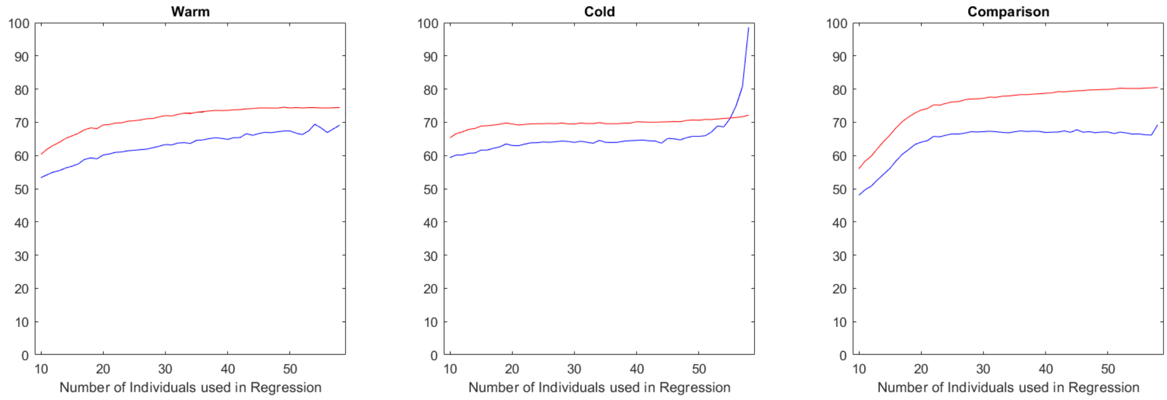

Figure 29: The percentage of signals receiving correct classification as a function of the amount of individuals used in the regression. The red line is all 59 signals, while the blue line is just the signals not used in the regression.

4.5.2 PST and TIST

To see if PST and TIST would have been formulated in the same manner if less data was used, a repeated random sub-sampling validation was done. In the case of PST, the first half of the spectra was chosen as the interval analyzed. This analysis was done using both mean band analysis and RMM, with results presented in figure 30 and figure 31 respectively.

MB method

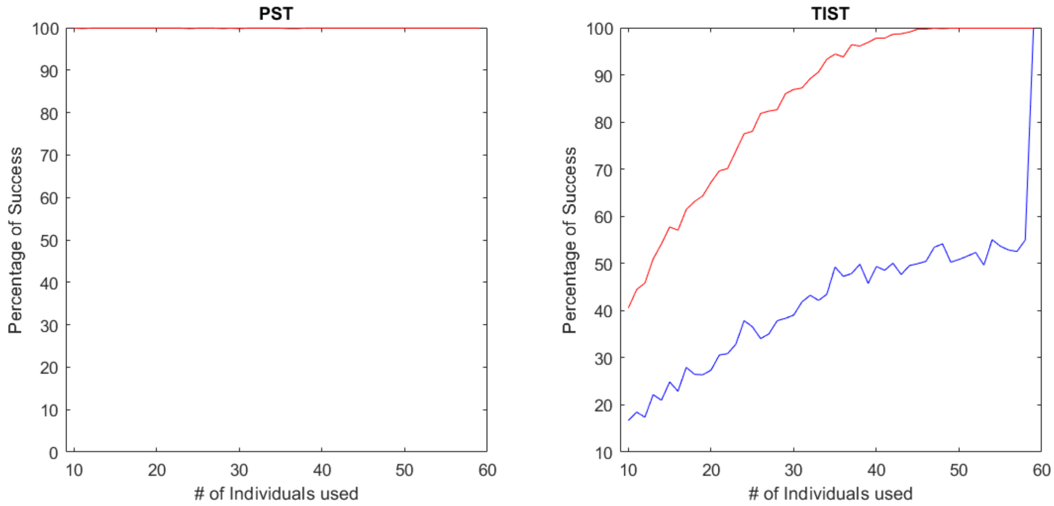

Figure 30: Left: The percentage of the time the difference in total power in the first half of the spectra was statistically significant as function of the number of individuals used. Right: The percentage of the time the normalized power of the warm spectra were greater in only epoch 1-5 in blue, and the same percentage in red, but when allow up to one epoch to not follow the trend.

RMM

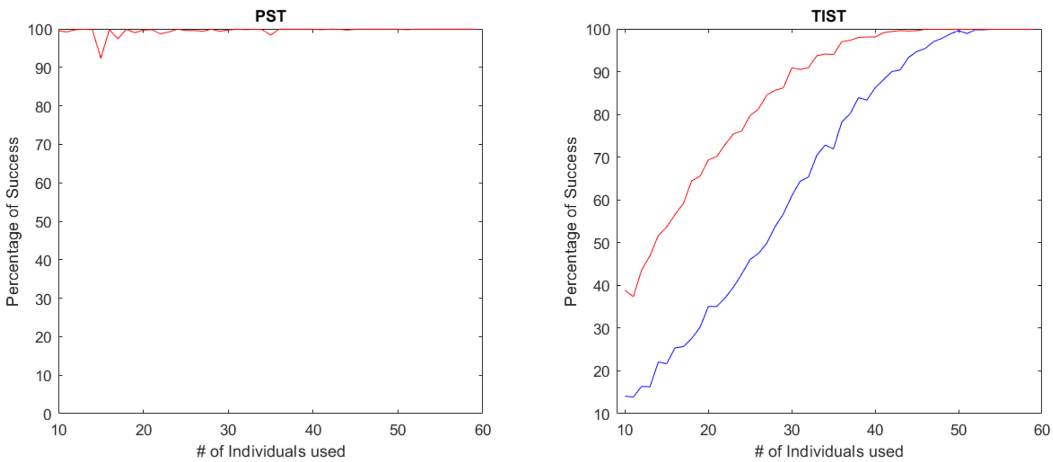

Figure 31: Left: The percentage of the time the difference in total power in the first half of the spectra were statistically significant as function of the number of individuals used.
Right: The percentage of the time the normalized power of the warm spectra were greater in only epoch 1-5 in blue, and the same percentage in red, but when allow up to one epoch to not follow the trend.

4.6 Summary

A short summary of the CCT and regression results for the most successful methods, with calculated p-values.

Table 14: Comparison Summary

Method	Test	# correct	% correct	p-value
MB	CCT2	46	77.9%	<0.0001
MB	Regression	45	76.2%	<0.0001
RMM	CCT2	48	81.4%	<0.0001
RMM	Regression	48	81.4%	<0.0001

Table 15: Categorization Summary

Method	Test	# correct	% correct	p-value
MB	CCT	75	63.5%	0.1156
MB	Regression	80	67.8%	<0.0001
RMM	CCT	80	67.8%	<0.0001
RMM	Regression	86	72.9%	<0.0001

5 Discussion

The spectrogram was concluded to be the optimal spectral estimator for the HRV-signals. The Wigner distribution produced result of similar quality as in the earlier method evaluation project, i.e, not satisfactory. The smoothed Wigner distribution produced significantly worse results on the data used in this thesis compared to the earlier project, likely due to the higher level of noise in this data. Thus, all HRV-spectra in this thesis is constructed using the spectrogram. It became clear that the choice of spectral estimator for the respiration signal was not a significant factor to help differ the data sets. For all methods, all three spectral estimators were applied, neither producing an increase in performance more than one correct classification and in most cases causing no difference. There was also no consistency in which estimator outperformed the others. Based on this result, the decision was made to use the spectrogram for all respiration methods, to keep consistent with the spectral estimator used for HRV.
Out of the attributes checked in the transformed spectrograms, total power and normalized power gave significant results when checked over some specific intervals, the spread of the time-frequency spectrum however, did not. The fact that differences existed in total power in the first half of the spectrum makes sense, since the initial "chock" of having your hand in ice water would dissipate over time, which means most of the difference

logically would be in the beginning. That the warm signal has higher spectral power could be explained by a pleasant feel of the tepid water lowering the stress of the individual while the feeling of ice water would increase stress. As stress-level is negatively correlated with power, this would mean that the warm signal should have higher power according to theory. After a while, the effect of the feeling of the water would diminish, meaning the total power in the latter half of the spectra become comparable.
That normalized power was different can be similarly motivated, as total power starts high and ends in the middle for the warm spectrum, while the cold spectrum starts low and climbs to end at the middle. Normalized, this means the result given is not surprising, the normalized warm spectrum has most of the power in the beginning, while the cold spectrum has most of the power in the end.
Surprisingly, the HF-method of analyzing the HRV-spectra gave significantly better results than what was expected. While the method consistently gave worse result than any other method, it still performed somewhat reliably in the comparison problem.
The MB-method of analyzing however, gave better results. This is advantageous, since it did not rely on the respiratory signal the way the respiratory methods did. However, since there is no guarantee that the HRV-spectra will lie in the band between 0.12 Hz and 0.2 Hz, this method might not necessarily work as well on a general HRV-signal.
Of the respiration methods used, RMM greatly outperformed the other methods, not only in results, but also in computation time. RMM also outperformed any method not using the respiratory signal. As such, if the respiratory signal is available, one should use RMM.
The comparison problem was the easier of the problems to solve, yielding statistically significant results for a majority of the tested methods. A likely reason for this is that it removes the huge individual variation in HRV that can make it difficult to analyze, as it is just about looking at differences on an individual level. This is confirmed by the fact that the PST model, the only model used which does not normalize in any way, only shows satisfactory results on the comparison problem.
The categorization problem required more advanced models to receive results of a similar level as the comparison problem. Only the TIST model and the regression models with several inputs resulted in satisfactory results. The CCT model could possibly also be able to reach similar results, had the scores been chosen to be less weighted towards the warm data set.
Unfortunately there does not seem to be a simple way to predict experienced pain using regression on HRV data. A likely reason for this is that experienced pain is very subjective. For example, the answers from the participants ranged from four to ten while all experiencing the same physical stimulation. Also it might simply be that even though psychological things affect HRV, it does not do it at a noticeable level compared to physical things that occur simultaneously.
A potential drawback to the regression models could be the multicollinearity issues that arise when using multiple epochs, especially in the case when every epoch is used. Since adding up the normalized power in all epochs would equal one, one could somewhat accurately predict what the normalized power in one epoch would be, using the other epochs. An unintended side effect of discarding epoch 0 however is that the grievous problems of perfect multicollinearity is avoided. Multicollinearity could possibly also

explain the very high p-values found for the coefficients.
While the regression worked well for this data set, it might not work for additional data, even if it was acquired in a similar or exactly the same way, as multicollinearity makes the coefficient estimates in the regression model very sensitive to the inputs given.
The fact that the regression models might not be completely suitable for new data can somewhat be seen when observing the regressions when not using all the data. Observing the percentage of correctly classified signals that was not used to build the regression model, it is seen that in the categorization case the percentage barely climbs above 60% when using normalized power in all epochs and 65% when only using normalized power in the significant epochs (in both cases also using total power of the first half of the spectra), these results sadly points to that the regression models are in actuality not that great in determining whether a signal is warm or cold.
Similar things can also be said when instead looking at the comparison tests, while the regression model using all participants and all data correctly classifies a signal pair over 80% of the time, signal pairs not used in the regression are not correctly classified even 70% of the time no matter how many participants were used in the regression. From these results, CCT2 might be the better comparison test as it gave similar results, but the test was formulated using only the knowledge of where the total and normalized power of the spectra is, instead of fitting already known results, as a regression model does.

5.1 Further studies

The regression model built to predict the self reported pain was very quickly performed, and is a part of this data set that certainly could be investigated further, with more sophisticated analysis and models. Linking HRV with perceived pain could also probably be useful in multiple fields, which further increases the interest in analysis of this relationship.
Furthermore, a possible improvement of the experiment could be made by letting the individuals being tested breathe with some predetermined and known frequency, so the problem of finding the respiratory frequency becomes almost trivial.
Lastly, since most of the models discussed in this thesis were built by using all data, it would certainly be of interest to see how well the models work on new data.

6 Conclusion

The purpose of this thesis was to develop a method that could categorize an HRV-signal as either warm or cold, both compared to its opposite, as well as independently. Several methods were successful in this, the most successful being a logistic model using the total power found in the first half of the signal and the normalized power of every epoch of the signal as input. This model predicted the correct category of a given signal with 72.8% accuracy ($p<0.0001$), and could correctly decide categorization of a given pair with 81.4% accuracy ($p<0.0001$).

References

[1] R. Anderson, P. Jönsson, and M. Sandsten. Effects of age, bmi, anxiety and stress on the parameters of a stochastic model for heart rate variability including respiratory information. In *BIOSTEC 2018*, volume 4, pages 17–25. SCITEPRESS, 2018.

[2] M. Hansson and P. Jönsson. Estimation of hrv spectrogram using multiple window methods focussing on the high frequency power. *Medical Engineering and Physics*, 28(8):749–761, 2006.

[3] A. Hernando, J. Lázaro, E. Gil, A. Arza, J.M. Garzón, R. López-Antón, C. de la Cámara, P. Laguna, J. Aguiló, and R. Bailón. Inclusion of respiratory frequency information in heart rate variability analysis for stress assessment. *IEEE journal of biomedical and health informatics*, 20(4):1016–1025, 2016.

[4] J. A. Hirsch and B. Bishop. Respiratory sinus arrhythmia in humans: how breathing pattern modulates heart rate. *American Journal of Physiology-Heart and Circulatory Physiology*, 241(4):H620–H629, 1981.

[5] P. Jönsson, K. Österberg, M. Wallergård, Å.M. Hansen, A.H. Garde, G. Johansson, and B. Karlson. Exhaustion-related changes in cardiovascular and cortisol reactivity to acute psychosocial stress. *Physiology & behavior*, 151:327–337, 2015.

[6] E. Kristal-Boneh, M. Raifel, P. Froom, and J. Ribak. Heart rate variability in health and disease. *Scandinavian journal of work, environment & health*, pages 85–95, 1995.

[7] A.K. Lennartsson, I. Jonsdottir, and A. Sjörs. Low heart rate variability in patients with clinical burnout. *International Journal of Psychophysiology*, 110:171–178, 2016.

[8] M.S. Menkes, K.A. Matthews, D.S. Krantz, U. Lundberg, L.A. Mead, B. Qaqish, K.Y. Liang, C.B. Thomas, and T.A. Pearson. Cardiovascular reactivity to the cold pressor test as a predictor of hypertension. *Hypertension*, 14(5):524–530, 1989.

[9] M. Sandsten. *Time-Frequency Analysis of Time-Varying Signals and Non-Stationary Processes.* 2016.