Multitaper analysis of HRV power and its stress-related correlation to respiration frequency

LUND
UNIVERSITY

Zite He
Supervisor: Maria Sandsten
Mathematical Statistics
Lund University

June 13, 2018

Abstract

In this Master thesis, different multitaper methods are implemented to estimate the spectra of respiratory signals and HRV data, and further to estimate the correlation between the respiratory center frequency and the narrow-banded high frequency band of HRV power. The methods are applied first on ARMA-process data, then on the integrated pulse frequency modulation (IPFM) data simulations, where the evaluation is performed by calculating the bias and standard deviation of the narrow-banded HRV power and its correlation with respiratory frequency. The results show that the Thomson multipaper and the peak-matched multitaper has each own pros and cons on different type of signals and that the Thomson multitapers are the best giving the strongest negative correlation. The second aim is to check whether the correlation deviates between subjects with different level of stress. A total number of 47 individuals, divided into 3 groups had their respirations and heart rates recorded. A clear difference is found indicating that the more stressed the stronger the correlation.

Keywords: Multitaper (multiple windows); Heart rate variability (HRV); Spectrum analysis; Integrated pulse frequency modulation (IPFM); High frequency band (HF); Thomson multitaper (TH MW); Peak-matched multitaper (PM MW); Correlation.

Acknowledgements

I would like to express my gratitude to all the people who have contributed to this Master thesis. First of all I would thank Prof. Maria Sandsten at Mathematical Statistics for the useful comments, remarks and engagement through the whole process of this Master thesis. I would then like to say thanks to all my friends who have supported and encouraged me during the past few months, both by keeping me harmonious and helping me putting pieces together.

Contents

1 Introduction

As mental states and stress conditions are suggested to have diverse impacts on heart rate variability (HRV) [1], it is of great interest to investigate how HRV reflects differently the modulation to the rhythm of heart by the control of the sympathetic and parasympathetic nervous systems under different stress states [2]. In order to analyze the difference of HRV under different stress states, a good estimation of spectrum of HRV is necessary. Therefore the very first thing to do in this thesis is to evaluate the performances of several multitaper spectral estimation methods. As heart rate is regulated by body's automatic nervous system (ANS), which also controls how people breath, the influence of the respiration on HRV is well known as respiratory sinus arrhythmia (RSA) [3]. The power spectrum of HRV is divided into three main frequency ranges, i.e., high-, low- and very-low-frequency (HF, LF and VLF) band. A negative correlation has been found between the respiratory center frequency and the high-frequency band of the HRV power [4]. It is also interesting to see whether the correlation deviates between subjects with different levels of stress, which is kind of an extension of [4] since more participants and samples are available this time.

1.1 Background

Before analyzing the signals it is always better to know what the signal means biologically and how it is collected, together with the mathematical methods (introduced in the following section) which are going to be applied on the signal data. However this thesis is not going deep into explaining how human body nervous system functions. Only a brief introduction of some physiological terminologies will be presented in this section.

Heart rate variability (HRV)

Heart rate variability (HRV) measures the time changes between successive heart beats. The time between beats is measured in milliseconds (ms) and is called an "R-R interval" or "inter-beat interval (IBI)". Electrocardiography (ECG) is the most commonly used method of recording and visualizing the electrical activity of the heart over a period of time (see figure 1). HRV analysis is now widely used as probably the easiest and most useful non-invasive method to track health and fitness especially assessing overall cardiac health and the state of the ANS responsible for regulating cardiac activity [2].

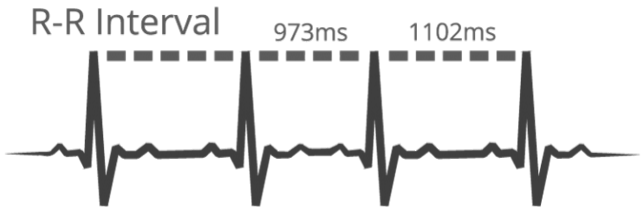

Figure 1: An example of ECG with time changes between successive heart beats.

As mentioned above, the frequency components of HRV could be broken down into three ranges (see figure 2). High-frequency (HF) band (usually $0.15 - 0.4Hz$), which has been shown to corre-

late almost exclusively to parasympathetic nervous system activity, is associated with a physiological phenomenon known as respiratory sinus arrhythmia (RSA).

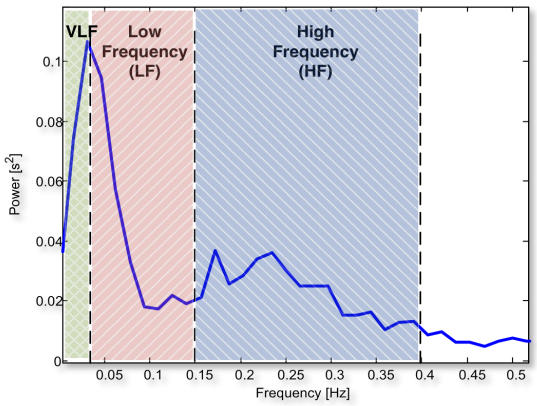

Figure 2: An example of the PSD to illustrate the HRV frequency range.

Respiratory sinus arrhythmia (RSA)

Respiratory sinus arrhythmia (RSA) refers to activity of the vagus nerve, an important component of the parasympathetic branch of the ANS. It is typically a benign, naturally occurring variation in heart rate that occurs during the breathing cycle, i.e., during inhalation (exhalation) heart rate increases (decreases) (see figure 3). This is mostly found particularly among young and healthy individuals and decreases with age [5].

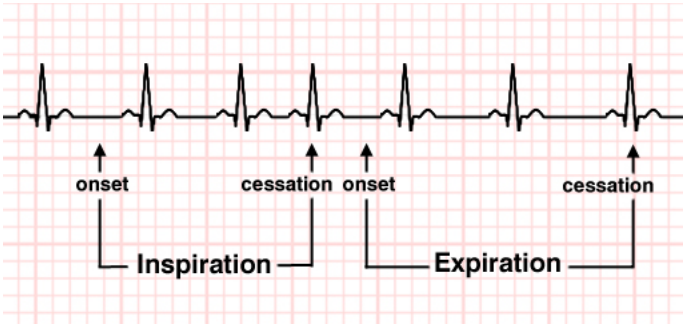

Figure 3: An example of RSA during a breathing cycle.

Respiratory and HRV signals

There are many ways to extract respiration signals, mostly from ECG signals. Three ECG-derived methods have been introduced and implemented in [6]. In this thesis the measurement of the respiration was done by using a strain gauge over the chest. The HRV measurement was done by using ECG as mentioned above. Figure 4 shows what a typical respiratory and HRV signal looks like in this thesis with sampling frequency $f_s = 4Hz$.

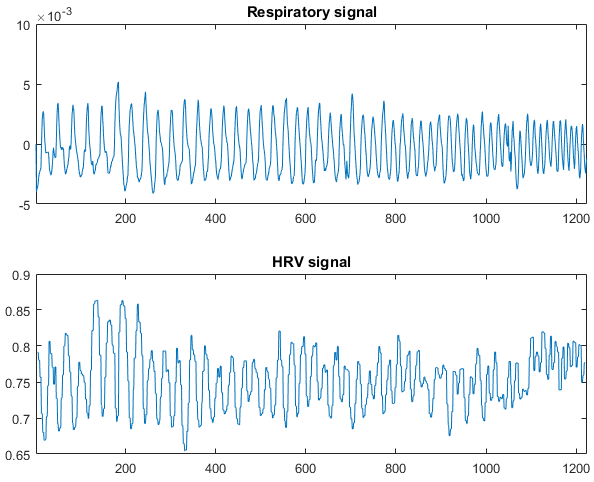

Figure 4: An example of respiratory and HRV signals. As one would notice that the amplitude of HRV decreases as the frequency of respiration increases.

1.2 Aim of the thesis

The aim of this thesis is to first evaluate the performance of different multitaper methods. In order to do so, the maximum of respiratory frequency is taken from multitaper spectral estimation of respiration, which is then used to calculate the narrow-banded HRV spectral power. The evaluation is made not by computing the averaged normalized mean squared error around the peak value [7] but by comparing the bias and standard deviation of the band power estimation. When true value of the band power is unknown, the correlation between the maximum of respiratory frequency and the band power of the HRV spectrum is estimated and used as another evaluation method. Secondly, as previously found that stress has negative impact on HRV power [8, 9], this thesis also aims to find what impact stress has on the correlation mentioned above, i.e., whether the correlation deviates between subjects with different level of stress.

2 Mathematical methods

2.1 Spectral analysis

Spectral analysis, also known as *frequency domain analysis*, was first used as a method to find out the periodicity of a time series data [10]. Such analysis has been found to be especially useful in many fields like communications engineering, management science, medical science, and biomedical science. In medical ultrasonography long ago for example, spectral analysis of ultrasonic reflections from biological tissues can be used to determine basic tissue parameters for use in differential diagnosis [11]. In this thesis, the spectral analysis is used mainly to extract information from HRV for further cardiac or neural diagnosis [12].

Power spectral density (PSD)

Power Spectral Density (PSD) is the frequency response of a random or periodic signal. It tells us where the average power is distributed as a function of frequency. The PSD of a random time signal $x(n)$ can be expressed in one of two ways that are equivalent to each other.

- The PSD is the average of the Fourier transform magnitude squared, over a large time interval

$$S(f) = \lim_{N\to\infty} E\{\frac{1}{N}|\sum_{n=0}^{N-1} x(n)e^{-i2\pi fn}|^2\}. \tag{1}$$

- The PSD is the Fourier transform of the auto-correlation function

$$S(f) = \sum_{\tau=-\infty}^{\infty} r(\tau)e^{-i2\pi f\tau}. \tag{2}$$

Spectrum estimation

There are mainly two different ways to estimate a spectrum, parametric and non-parametric spectrum estimation. The parametric spectrum estimation techniques are based on the use of models for the data, assuming the data is generated in a certain way. For instance, one could assume that the data is the output $y(n)$ of a linear time-invariant (LTI) system in response to a white noise input sequence $x(n)$ (see figure 5).

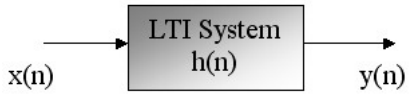

Figure 5: A simple LTI system with an impulse response $h(n)$.

The PSD of $y(n)$ is then obtained by doing the following computation

$$S_y(f) = |H(f)|^2 S_x(f) = |H(f)|^2\sigma_x^2, \tag{3}$$

where $H(f)$ is called the frequency function of the filter with impulse response $h(n)$ and σ_x^2 is the variance of the input white noise (for review see [13] chapter 6). There are several different

classes of systems that are typically used with this kind of approach to spectrum estimation such as auto-regressive (AR), moving average (MA) and auto-regressive moving average (ARMA). In this situation the problem is then to estimate the parameters in the model. A higher resolution might be achieved if a good model is assumed, however there is little one can say about most of the signals and the estimation might be useless if a wrong model is assumed.
In this thesis, only the non-parametric approaches are applied on the data, which make no assumption on the model but only rely on the direct use of the available data. Such methods might produce a less precise result, but could be widely applied for different cases.

Periodogram

In eq.(1), an infinite number of samples are needed to compute the PSD, however one would never have infinite number of samples in reality, which leads to an expression based on N samples collected

$$\hat{S}(f) = \frac{1}{N}|\sum_{n=0}^{N-1} x(n)e^{-i2\pi fn}|^2. \tag{4}$$

Eq.(4) is the definition of the periodogram, an estimate of the spectral density of a signal, which is the basic idea of non-parametric spectral estimation.
The periodogram is a very useful tool for describing a time series data set and is easy to compute. However as known, the periodogram is not a good estimate for its bias and especially its large variance. It can be rewritten as

$$\hat{S}(f) = \sum_{\tau=-N+1}^{N-1} \frac{1}{N} \sum_{n=0}^{N-1-|\tau|} x(n)x(n+|\tau|)e^{-i2\pi f\tau} = \sum_{\tau=-N+1}^{N-1} \hat{r}(\tau)e^{-i2\pi f\tau}. \tag{5}$$

The expected value of the periodogram is obtained by direct calculation as

$$E[\hat{S}(f)] = \sum_{\tau=-N+1}^{N-1} E[\hat{r}(\tau)]e^{-i2\pi f\tau} = \sum_{\tau=-N+1}^{N-1} (1 - \frac{|\tau|}{N})r(\tau)e^{-i2\pi f\tau}, \tag{6}$$

which tends to

$$\sum_{\tau=-\infty}^{\infty} r(\tau)e^{-i2\pi f\tau} = S(f) \quad \text{as} \quad N \to \infty, \tag{7}$$

showing that the periodogram is an asymptotically unbiased estimate of the spectral density.
The variance of the periodogram is asymptotically tending to

$$V[\hat{S}(f)] \approx \begin{cases} S^2(f), & \text{for } 0 < |f| < 1/2 \\ 2S^2(f), & \text{for } f = 0 \text{ and } \pm 1/2 \end{cases} \quad \text{as} \quad N \to \infty. \tag{8}$$

Therefore, the variance does not decrease as more samples are used in computation (for more detailed proof and derivations see [13] chapter 9). Based on the idea of periodogram, a lot of methods have been developed to reduce the variance. The idea of using multiple windows will be introduced later in this section.

Windows

For small values of N, eq.(6) can be rewritten as

$$E[\hat{S}(f)] = \sum_{\tau=-\infty}^{\infty} k_N(\tau) r(\tau) e^{-i2\pi f\tau}, \tag{9}$$

where $k_N(\tau) = \max(0, 1 - |\tau|/N)$ is called a *lag* or *triangular window*. Denote $K_N(f)$, also known as the *Fejer kernel*, as the Fourier transform of the window $k_N(\tau)$ by

$$K_N(f) = \frac{\sin^2(N\pi f)}{N \sin^2(\pi f)}, \tag{10}$$

which has a shape as shown in figure 6.

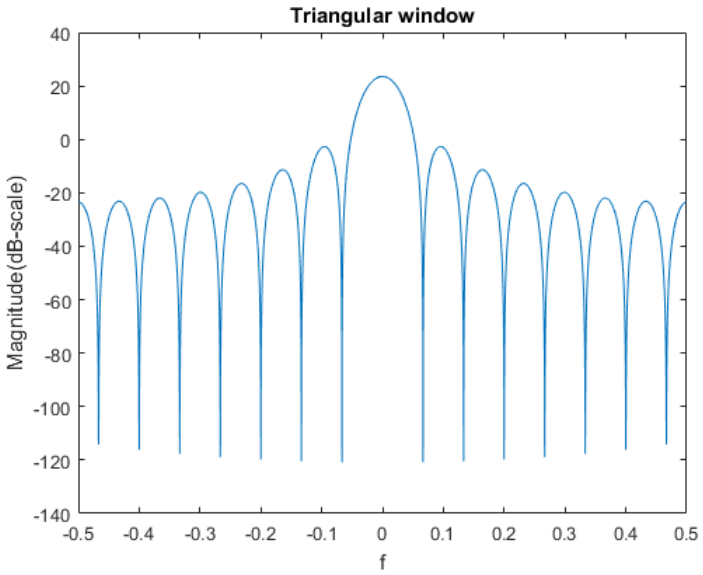

Figure 6: The Fourier transformed triangular window, also known as *Fejer kernel*.

The high sidelobes of the kernel cause severe power leakage in spectral estimation, leading to large bias, especially when the dynamical range of the spectrum is large. Windows other than triangular window are created and used to reduce the bias caused by leakage as well as the variance. However, estimations given by such windows will have a decreased resolution.
Hanning window is one of the mostly common used windows, it has lower sidelobes and a wider mainlobe compared to the triangular window. The Hanning window is defined as

$$k_N(\tau) = \frac{1}{2} - \frac{1}{2}\cos(\frac{2\pi\tau}{N-1}), \quad \tau = 0, \ldots, N-1, \tag{11}$$

which is often normalized as

$$w(\tau) = \frac{k_N(\tau)}{\sqrt{\frac{1}{n}\sum_{\tau=0}^{N-1} k_N^2(\tau)}}, \quad \tau = 0, \ldots, N-1. \tag{12}$$

Eq.(4) is then modified as

$$\hat{S}(f) = \frac{1}{N}|\sum_{n=0}^{N-1} x(n)w(n)e^{-i2\pi fn}|^2. \tag{13}$$

The Fourier transformed Hanning window $K_N(f)$ is shown in figure 7.

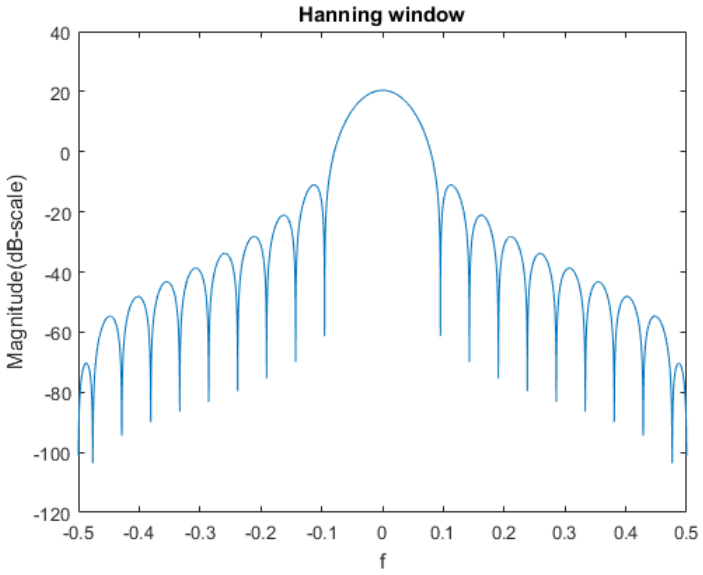

Figure 7: The Fourier transformed Hanning window.

Using *windowing* or *tapering* of the data is an essential part in non-parametric spectrum estimation. Different windows are chosen under different situations depending on the shape of the spectrum (large dynamics or not, peaks close or apart) or the need of researchers (resolution or smoothness).

2.2 Multitaper (*Multiple window*)

Other than using only one window, the multitaper spectral estimator uses several different data tapers which are orthogonal to each other, producing uncorrelated spectral estimates to obtain an averaged windowed periodogram with reduced variance.

2.2.1 Welch's method (*Welch*)

Welch's method is an approach to spectral density estimation, which is an improvement on the standard periodogram spectrum estimation in reducing noise in the estimated power spectra in exchange for reducing the frequency resolution.
The given signal data is first split up into possibly overlapping segments (usually 50% overlap). For each segment a windowed periodogram is formed (an example of overlapped Hanning windows is shown in figure 8). The Welch's estimation of the PSD is then obtained by averaging all the periodograms.

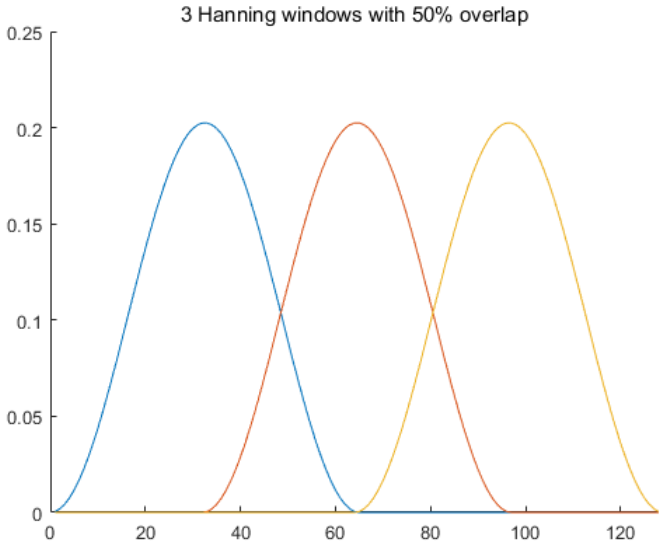

Figure 8: 3 Hanning windows with 50% overlap, window length L = 64 and data length N = 128.

2.2.2 The Thomson multitapers (*TH MW*)

A new method based on a "local" eigen-expansion to estimate the spectrum was first presented by Thomson in 1982 [14]. Given a real valued stationary discrete-time random process $x(n)$, the Thomson multitaper (TH MW) provides K orthogonal windows over N samples of the process to create K spectral estimates $\hat{S}_k(f)$ which are then used to compute an averaged $\hat{S}(f)$ with reduced variance

$$\hat{S}(f) = \frac{1}{K}\sum_{k=1}^{K}\hat{S}_k(f), \tag{14}$$

where

$$\hat{S}_k(f) = |\sum_{n=0}^{N-1} x(n)h_k(n)e^{-i2\pi fn}|^2. \tag{15}$$

Eq.(12) is the multitaper spectral estimate by using orthogonal and normalized tapers $\boldsymbol{h}_k = [h_k(0)\dots h_k(N-1)]^T$, where

$$\sum_{n=0}^{N-1} h_k(n)h_l(n) = \delta_{kl} \quad \text{for} \quad k \neq l, \tag{16}$$

and

$$\sum_{n=0}^{N-1} h_k^2(n) = 1 \quad \text{for} \quad 1 \leq k \leq K. \tag{17}$$

The Thomson multitapers are designed to give small correlation between subspectra for a rather smooth spectrum, especially for a white noise spectrum. The tapers $\boldsymbol{h}_k$ are given by the solution of the eigenvalue problem

$$\boldsymbol{R}_B\boldsymbol{q}_k = \lambda_k\boldsymbol{q}_k, \quad k = 1\dots N, \tag{18}$$

where $\boldsymbol{h}_k = \boldsymbol{q}_k$ is the eigenvector and λ_k is the corresponding eigenvalue, ordered in decreasing magnitude. The covariance matrix $\boldsymbol{R}_B$ has a Toeplitz structure with elements

$$r_B(\tau) = B\,\mathrm{sinc}(B\tau), \quad 0 \leq |\tau| \leq N-1, \tag{19}$$

where $\mathrm{sinc}(x) = \sin(\pi x)/(\pi x)$ and B is the predetermined band-width.
The solution is found as the Discrete Prolate Spheroidal Sequences (DPSS). An example of how the tapers look like is shown in figure 9.

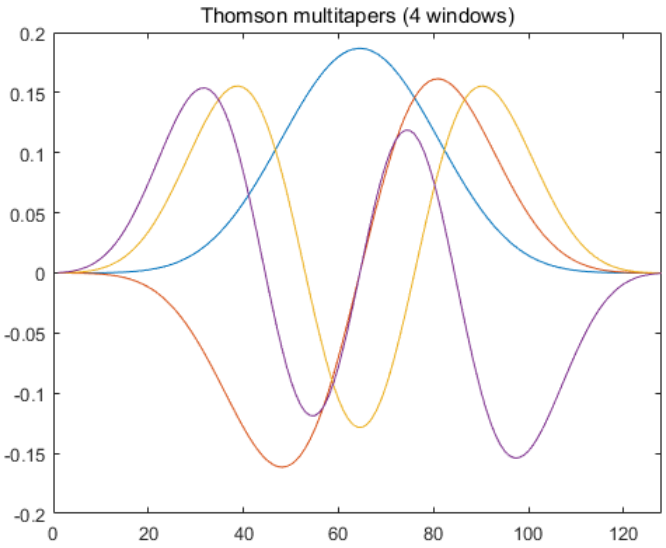

Figure 9: 4 Thomson windows (DPSS) with window length N = 128.

The tapers have low amplitude close to 0 at the edges to minimize the power leakage. Such tapers are chosen by looking at the eigenvalues. If the eigenvalues are close to 1 then their corresponding eigenvectors are good to use as tapers.

2.2.3 The peak-matched multitapers (*PM MW*)

For varying spectra, i.e., spectra with large dynamics, the peak-matched multitapers will be used to find the peaks and notches [15]. The tapers $\boldsymbol{h}_k$ are given by the solution of the generalized eigenvalue problem

$$\boldsymbol{R}_B\boldsymbol{q}_k = \lambda_k\boldsymbol{R}_Z\boldsymbol{q}_k, \quad k = 1\ldots N, \tag{20}$$

where $\boldsymbol{h}_k = \boldsymbol{q}_k$ corresponding to the K largest eigenvalues. The covariance matrix $\boldsymbol{R}_B$ has a Toeplitz structure with elements

$$r_B(\tau) = r_w(\tau) * B\,\mathrm{sinc}(B\tau), \quad 0 \leq |\tau| \leq N-1, \tag{21}$$

where $\mathrm{sinc}(x) = \sin(\pi x)/(\pi x)$ and $*$ denotes the convolution operator. Here $r_w(\tau)$ is the covariance function of a given zero-mean real-valued stationary random process $w(t)$ with peak located at $f = 0$. In this paper the chosen peaked spectrum is $S_w(f) = e^{(-2C|f|)/(10B\log_{10}(e))}$, $|f| \leq 1/2$ with $C = 20$ and $B = (K+2)/N$.

The suppression matrix $\boldsymbol{R}_Z$ also has a Toeplitz structure which corresponds to a penalty frequency function used to suppress the sidelobes of the windows. The function is as following

$$S_G(f) = \begin{cases} G, & B/2 < |f| < 1/2, \\ 1, & |f| \leq B/2, \end{cases} \tag{22}$$

where $G = 30$ is chosen in this paper. An exampled of PM MW is shown in figure 10.

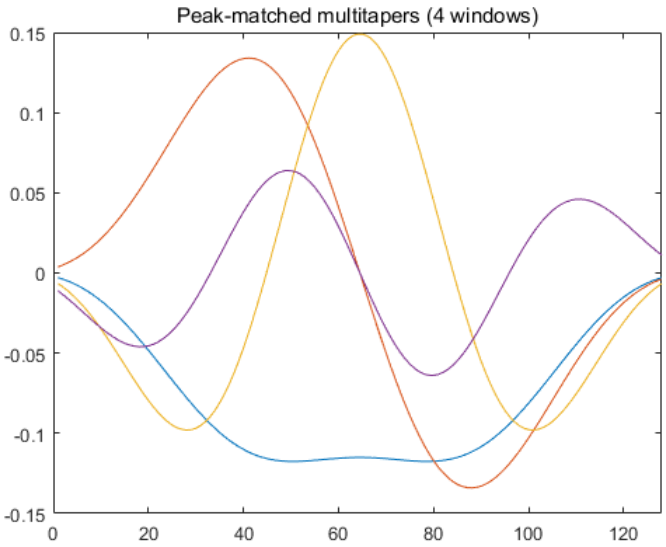

Figure 10: A number of 4 Peak-matched windows with window length N = 128.

3 Data simulation

Before the methods are applied to real data, it is necessary to evaluate the performance of each methods on the simulated data.

3.1 ARMA-process

Two kinds of ARMA-models are used to simulate a spectrum with either large dynamics (with a clear peak) or small dynamics (varying smoothly) in dB-scale. The reason why the ARMA-process is chosen is that it provides large dynamics in the PSD and the true PSD is also known through the parameters of the model which makes evaluation easy.
Two models together with plots of poles and zeros and PSD are given as following

- ARMA(2,2) with a clear peak in the PSD: $x_t - 1.2x_{t-1} + 0.85x_{t-2} = e_t + 0.5e_{t-1} + 0.75e_{t-2}$.

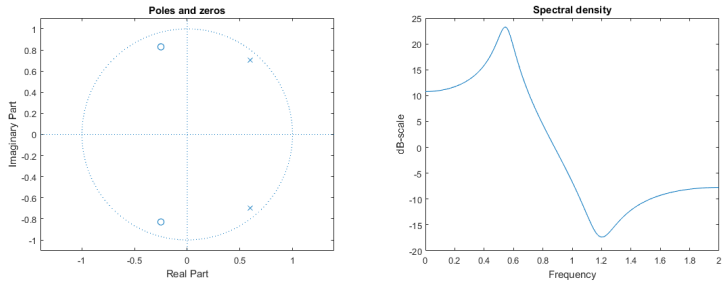

Figure 11: Left: poles and zeros. Right: the PSD in dB-scale.

- ARMA(2,2) which varies smoothly in the PSD: $x_t - 0.6x_{t-1} + 0.5x_{t-2} = e_t + 0.9e_{t-1} + 0.5e_{t-2}$.

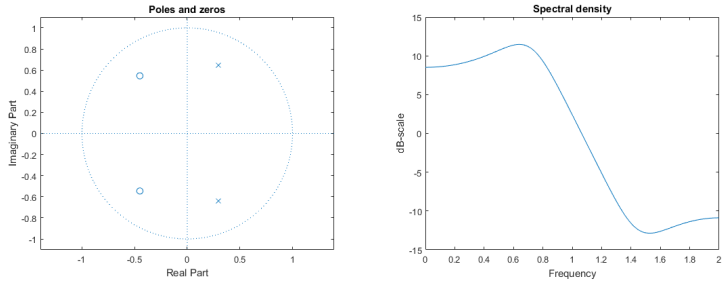

Figure 12: Left: poles and zeros. Right: the PSD in dB-scale.

The $\{e_t\}$ is a sequence of uncorrelated Gaussian noise with mean 0 and variance 1 in both cases. For each case, 400 samples are simulated and the simulation is iterated 100 times.

3.2 Integrated pulse frequency modulation (*IPFM*)

There are quite many ways to simulate an ECG-signal. The integrated pulse frequency modulation (IPFM) is one of such techniques used to generate a heart-beat event series. The IPFM model used in this paper is the one introduced in [4].
First the respiratory signal is simulated as a sinusoid signal with a time-varying frequency

$$x(n) = \sin(2\pi(\frac{f_r(n)}{1000})n + \phi) \tag{23}$$

where $f_r(n)$ is the increasing respiratory frequency and ϕ is a random phase which is equally distributed in the interval 0 to π.
The heart-beat signals are generated using an IPFM model based on respiratory frequency but is jittering in time. The jittering signal is simulated as a sinusoid signal as well without a random phase ϕ but with a time-varying amplitude $A(n)$ and disturbed by a LF noise $v(n)$,

$$y(n) = A(n)\sin(2\pi(\frac{f_r(n)}{1000})n) + v(n) \tag{24}$$

where the amplitude $A(n)$ decreases as the respiratory frequency $f_r(n)$ increases. The LF white noise $v(n)$ of standard deviation σ_v is filtered with a FIR-filter of order 200 which is cut off at frequency $0.12Hz$.
The heart-beat event series is then formed as

$$z(kP + y(kP)) = \begin{cases} 1, & k = 0, 1, 2, 3 \dots \\ 0, & \text{otherwise} \end{cases} \tag{25}$$

where the constant $P = 1000(60/heartrate)$. An illustration plot is shown in figure 13

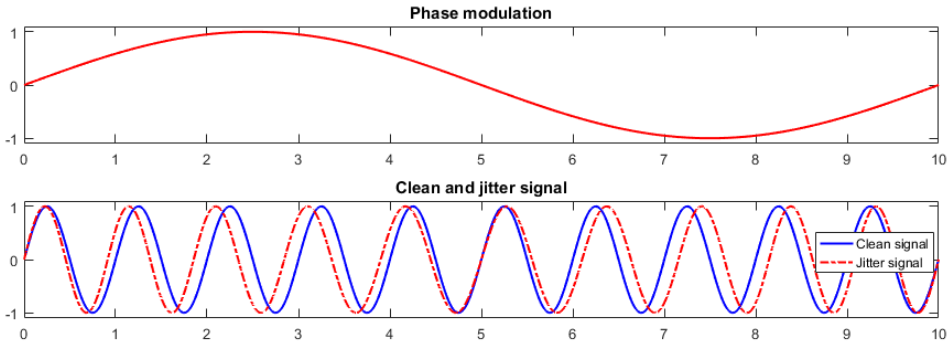

Figure 13: An illustration of how a jittering signal is simulated.

where the heart beat which was once counted as the maximum points of the blue curve without phase modulation, for example, now is counted as the maximum points of the red dash-dotted curve. The resulting signal is finally down sampled as in real experiments from $1000Hz$ to $4Hz$. 1200 samples are generated from each simulation with respiratory frequency increasing from $0.12Hz$ to $0.3Hz$ both linearly and quadratically. The standard deviation of the noise is the only parameter that changes in each simulation, increasing from $v(n) = 1$ to $v(n) = 10$ with step 0.01. The simulation is iterated 100 times for each value of $v(n)$.

4 Method evaluation

It is necessary to know how different methods work with different kinds of signals and noise. In this section all the methods will be applied on the simulated data generated using models introduced previously, from ARMA-process, where the true PSD is known, to the IPFM model where the spectrum is unknown, and finally tested on the recorded data in next section.
As mentioned in the introduction, the evaluation of the performance of different methods is based on the bias and standard deviation of the narrow-banded HRV spectral power within $f_r \pm 0.05Hz$, where $f_r = \max_f S(f)$ being the maximum of respiratory frequency and $S(f)$ being the PSD of respiration. When the true PSD is unknown, the standardized bias and standardized deviation is calculated and compared. Correlations between respiratory frequency and narrow-banded power of HRV are also calculated, used as a further evaluation approach.

4.1 ARMA-process evaluation

Two kinds of ARMA(2,2) processes are created from last section - one with a clear peak in spectral density and the other one rather smooth (both in dB-scale).

ARMA-process 1

The first ARMA(2,2)-process is built as following

$$x_t - 1.2x_{t-1} + 0.85x_{t-2} = e_t + 0.5e_{t-1} + 0.75e_{t-2}. \tag{26}$$

The spectral estimation of 100 simulations of this process is produced using the Welch, TH MW and PM MW with number of windows $K = 2$ and predetermined band-width $B = 0.01$. The mean value and standard deviation of the averaged band power is calculated and is then compared to the true averaged band power, plotted in box-plots, together with the plots of the averaged spectral estimation and bias in figure 14.

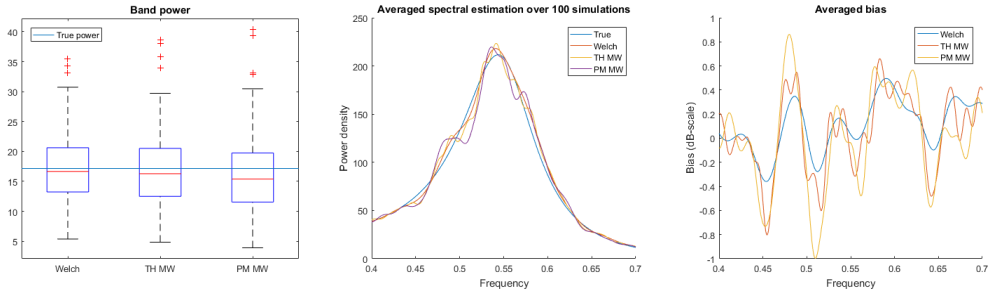

Figure 14: Box-plot of estimated band power, averaged PSD estimation and bias.

The Welch gives the smallest variation and averaged estimation most close to the true value as one can see from the box-pot and the bias. The PM MW seems to produce the largest variation shown in the last plot, while the TH MW produces a relatively small variance as expected. The reason why the PM MW does not perform well might due to that there only one peak in the spectrum and the dynamics of the spectrum is relatively small.

ARMA-process 2

The second ARMA(2,2)-process is built as following

$$x_t - 0.6x_{t-1} + 0.5x_{t-2} = e_t + 0.9e_{t-1} + 0.5e_{t-2}. \tag{27}$$

Same methods are performed on 100 simulations generated from this model and similar plots are shown in the figure 15.

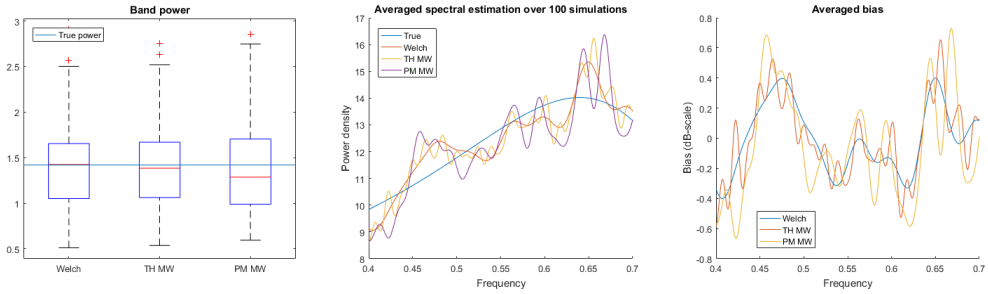

Figure 15: Box-plot of estimated band power, averaged PSD estimation and bias.

Almost the same conclusion is drawn here, however, as this is now more smooth than the first one, the performance of the TH MW is now very close to the Welch. The PM MW still performs 'worst' here for it does not show any advantage in a less dynamic case.

4.2 IPFM evaluation

Now the methods are applied on the simulated respiratory and HRV data. The data is generated with given respiratory frequency which increases over time (both linearly and quadratically). The spectrogram (with high resolution in time) is returned by Matlab (see figure 16 and 25), however, it is of more interest in this thesis to estimate the spectrum within a time period where signal is roughly treated as stationary. With the IPFM model, no true spectral density is available. Therefore in this section, each method will produce a so-called 'true' spectrum, i.e., an averaged estimation without LF white noise ($v(n) = 0$). The standardized bias and deviation of the band power will then be computed considering the scaling problem.

Linear case

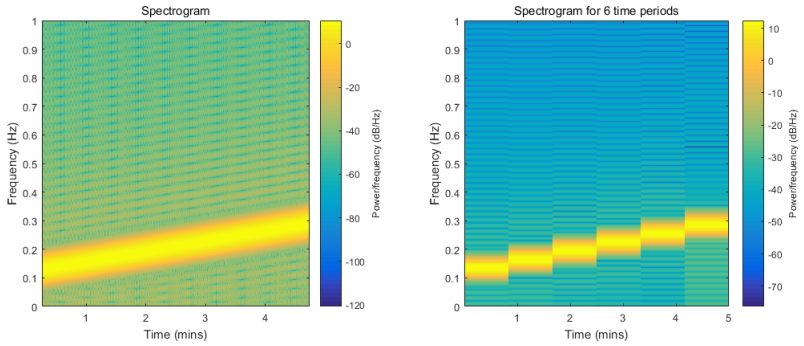

Figure 16: The spectrogram with different time-resolutions.

First the signal is generated with linearly increasing frequency. A number of 1200 samples with final sampling frequency 4 Hz of HRV and respiratory signals are simulated each time and segmented into 6 consecutive parts with 200 samples each. An averaged spectral estimation of each time period is produced. Figure 17 and 19 show an example of spectral estimations of both respiratory and HRV signal with the standard deviation of the LF noise $v(n) = 3$.

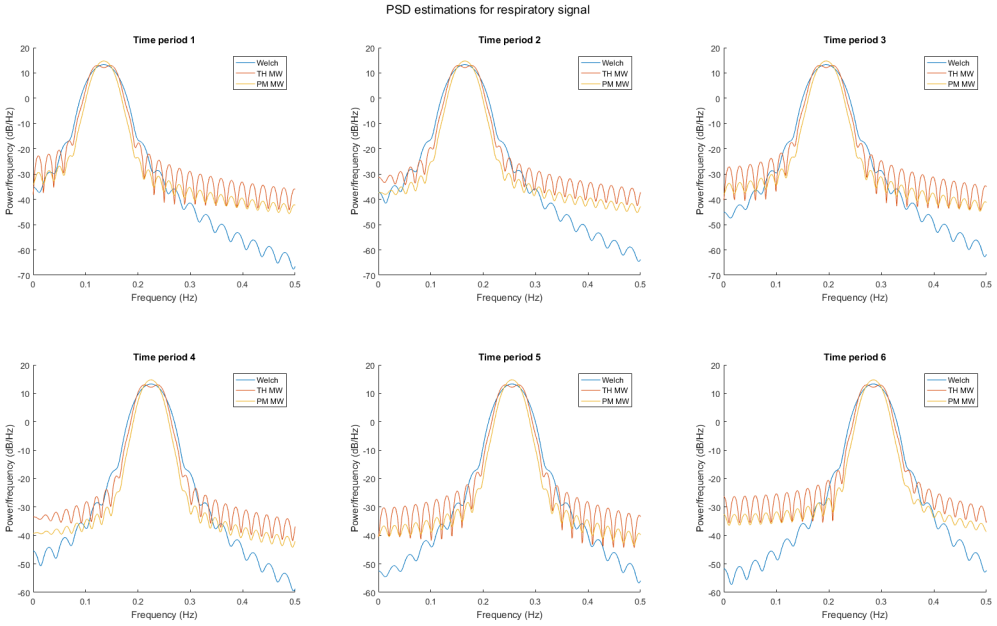

Figure 17: The PSD estimations of the respiratory signal. A linear growing trend is seen and the central frequency band is still a bit wide since the frequency changes quite much within each segment.

For the respiratory signal, the main aim is to locate the center frequency, i.e., to find out f_r =

$\max_f S(f)$, so the power leakage is not the problem that one needs to consider here. However, when it comes to HRV power estimation, one has to consider the choice of predetermined band-width B and the number of windows K to use.
Taking the TH MW for an example, figure 18 shows how estimation changes with different B and K. More power leakage is introduced when using more windows and it would become more difficult to locate the center frequency with more windows if applied on respiratory signal as one can see from the left plot. One obtains a wider range of HF power when increasing B as seen from the right plot.

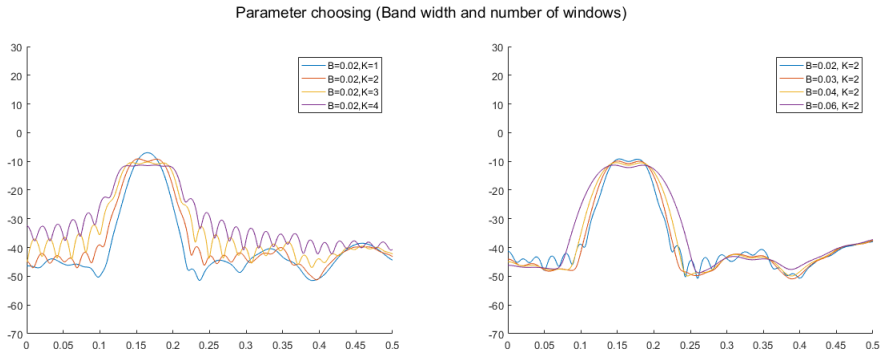

Figure 18: Different sets of K and B tested on one of the data segments.

After testing several sets of both parameters, K is set to 2 in all three methods and the predetermined band-width is set to $B = 0.02$ for the TH MW and PM MW, considering the width of the peak in the PSD.
As a conclusion of above all, a too wide band-width will result in a over smooth estimation around the peak, making it difficult to obtain an estimated band power close to what it should be (especially for the TH MW). If one only looks at the estimation around the peak, both the Welch and PM MW finds the peak frequency almost equally well. The TH MW would have a little trouble here to locate the maximum value considering the shape of its window spectrum with 2 or more windows. However the method does not go to trash bin since it has some other desirable features and still locates the HF part in HRV estimation quite well.

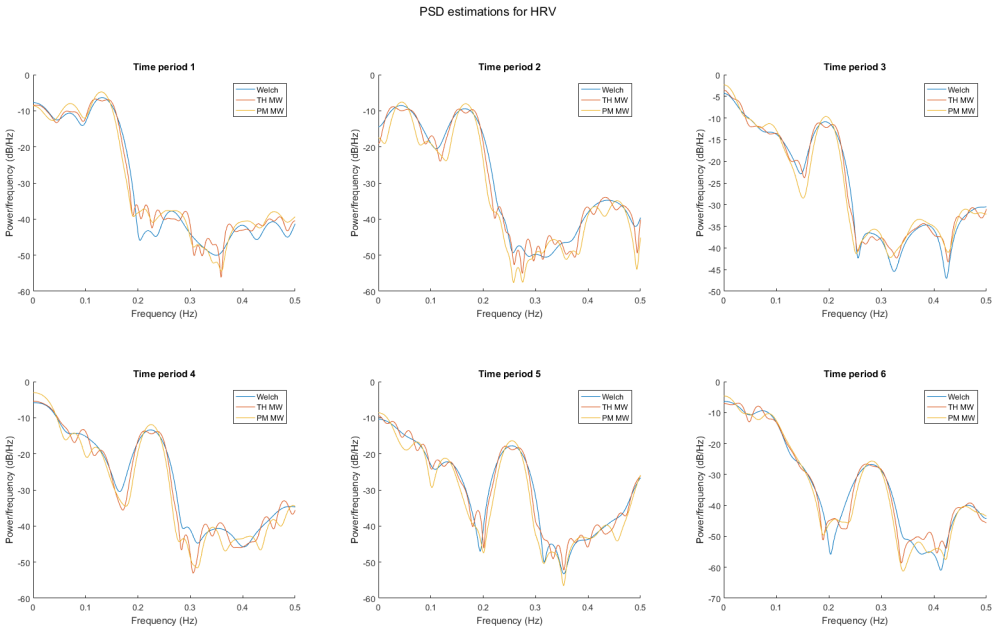

Figure 19: The PSD estimations of HRV. The standard deviation of the LF noise is chosen to be $v(n) = 3$, which results in quite high power in LF part. Band power estimations in the first time period are strongly affected.

The Welch gives the smallest variance over the whole spectral estimation, however it reduces the frequency resolution as stated in section 2, especially in time period 1 when the main respiratory frequency is close to the low-frequency range. The estimations get worse with increasing standard deviation of the noise form $v(n) = 0$ to $v(n) = 10$. Figure 20 is an example of how estimations get worse with larger deviation of the noise for time period 2.

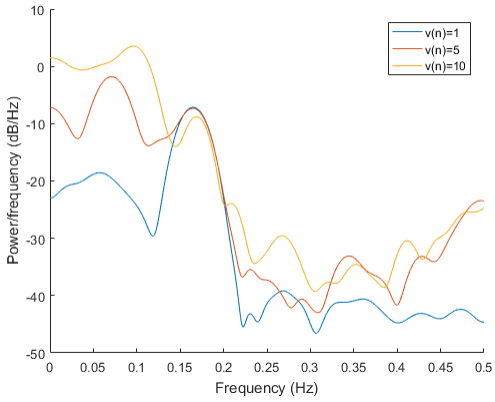

Figure 20: The LF part gets nasty when one increases the noise (one simulation using PM MW).

It is then to check how the estimations change with different noise, i.e., the robustness of each

method to noise (see figure 21).

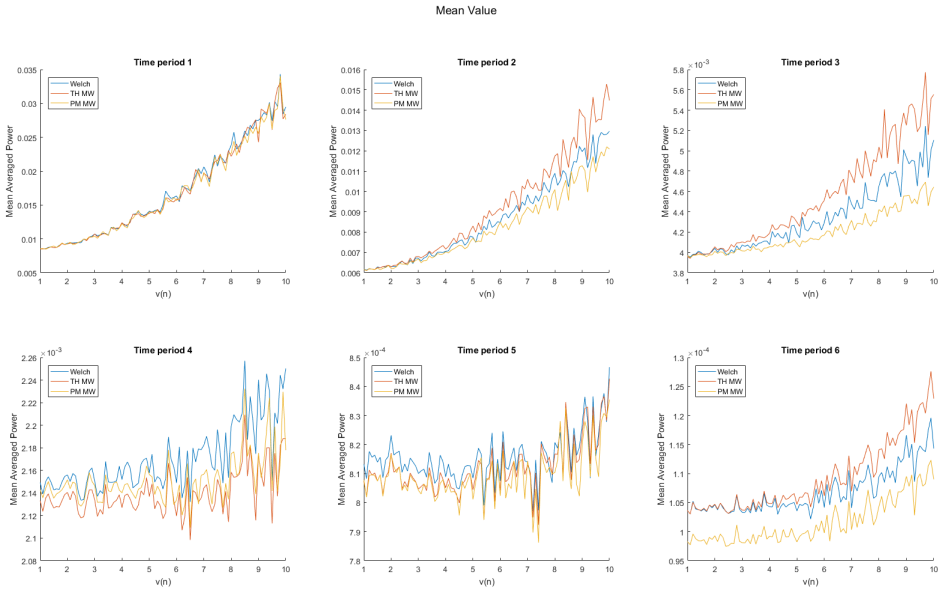

Figure 21: How the mean value of HRV band power estimation varies with increasing noise.

The mean value of the band power is computed and compared to each other, with increasing standard deviation of the noise. The noise has quite big effect on the estimation of the band power at first as one would expect since the PSD estimation for HRV looks a bit messy at low frequency part. One could see a clear increasing trend in mean value estimation. As the main respiratory frequency increases, the increasing trend with $v(n)$ gets weaker (note the different scale in figure 21).
Since there is no underlying truth for the band power, an averaged estimation without noise is taken as the 'true' spectral density for each method, and the bias of each method is calculated. However, as seen above, the estimated mean value of each method differs quite a lot as the respiratory frequency increases, so the bias might be compared under different scale. Thus it is better if one looks at the standardized bias (*StdBias*) instead, which is calculated as

$$StdBias_v = \frac{Bias}{Mean\ value} = \frac{\hat{P}_v - \hat{P}_0}{\hat{P}_0}, \tag{28}$$

where v is the deviation of the noise from 1 to 10 and $\hat{P}_0$ here represents the estimated band power with no noise in the signal. The *StdBias* for each time period is shown in figure 22.

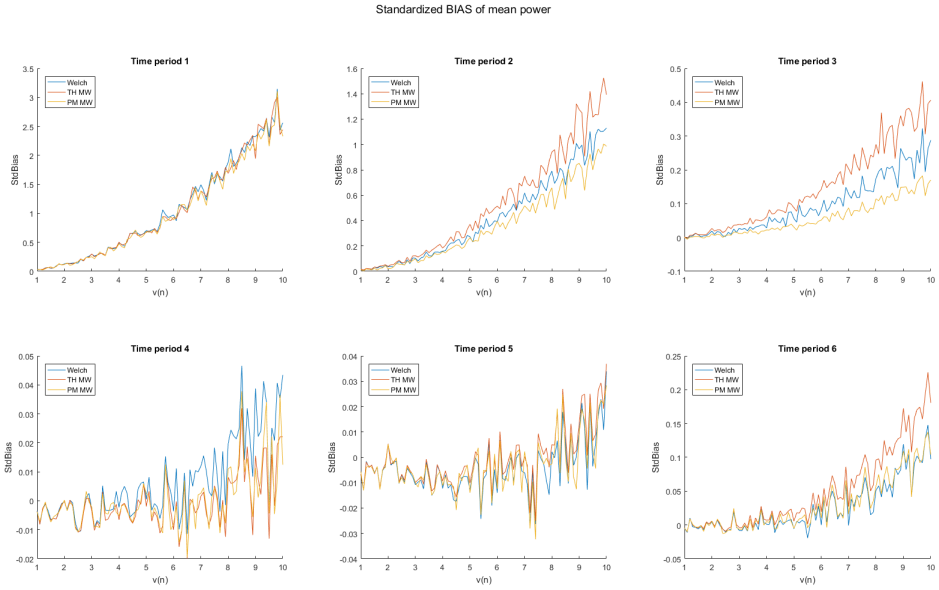

Figure 22: The standardized bias shows relatively how far the estimation is from the 'true' value.

Actually the *StdBias* does not really differ from the bias. Three methods reach similar results as the respiratory frequency increases (HRV HF power decreases). They follow the same curve at first since it is when the estimation is strongly influenced by the noise. Then as center frequency moves a bit away from LF part, the PM MW gives the smallest *StdBias*, while the TH MW gives the largest one as noise gets bigger which might due to that the TH MW introduces more power leakage.
The standard deviation of the band power estimation for each time period is also calculated against different level of LF noise to see how spread out the estimation is for each method. Again, to avoid scaling problems, standardized standard deviation (*SSD*) is calculated by doing

$$SSD_v = \frac{Standard\ deviation}{Mean\ value} = \frac{\mathrm{STD}(\hat{P}_v)}{\hat{P}_0}, \tag{29}$$

where v and $\hat{P}_0$ is the same as above. The *SSD* for each time period is shown in figure 23.

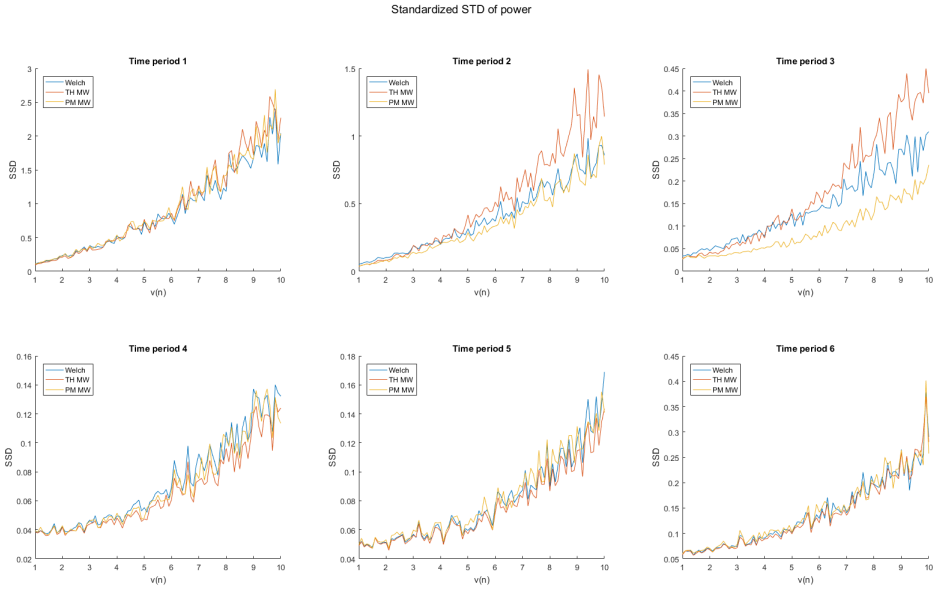

Figure 23: The standardized standard deviation.

None of the three methods is robust to noise as the deviation of the band power estimation increases with increasing deviation of the noise. The TH MW performs the worst in time period 2 and 3. It is hard to draw a conclusion here which method outperforms by only looking at the mean value, bias and the deviation, since no 'true' value is known here. However it is mentioned in previously cited paper that a stronger correlation between the respiratory center frequency and the HF band of the heart rate variability (HRV) power indicates a more robust estimate of the power [4]. The correlation is defined as

$$\hat{\rho} = \frac{C(\hat{f}_r, \hat{P}_h)}{\sqrt{C(\hat{f}_r, \hat{f}_r)C(\hat{P}_h, \hat{P}_h)}}, \tag{30}$$

where $\hat{f}_r$ is the estimated respiratory center frequency, $\hat{P}_h$ is the band power of HRV within $\hat{f}_r \pm 0.05Hz$ and $C(,)$ denotes the covariance which is calculated as

$$\begin{aligned} C(\hat{f}_r, \hat{f}_r) &= \sum_{k=1}^{K} (\max_f(S_k^r(f)) - \bar{\hat{f}}_r)^2, \\ C(\hat{P}_h, \hat{P}_h) &= \sum_{k=1}^{K} (\int_{\hat{f}_r^k - \delta_f}^{\hat{f}_r^k + \delta_f} S_k^h(f)df - \bar{\hat{P}}_h)^2, \\ C(\hat{f}_r, \hat{P}_h) &= \sum_{k=1}^{K} (\max_f(S_k^r(f)) - \bar{\hat{f}}_r)(\int_{\hat{f}_r^k - \delta_f}^{\hat{f}_r^k + \delta_f} S_k^h(f)df - \bar{\hat{P}}_h), \end{aligned} \tag{31}$$

where $\hat{f}_r^k$ is the center frequency of respiration given from $\max_f(S_k^r(f))$, $\delta_f = 0.05$ which gives a frequency interval of $\pm 0.05Hz$ around $\hat{f}_r^k$ for the power estimate of HRV and $\bar{\hat{f}}_r, \bar{\hat{P}}_h$ denote the mean value of respiratory frequency and HRV power respectively.

The correlation with increasing noise for three methods is plotted in figure 24. As one can tell from the figure, The Welch and PM MW have similar correlation estimations while the TH MW gives the strongest results for all values of $v(n)$.

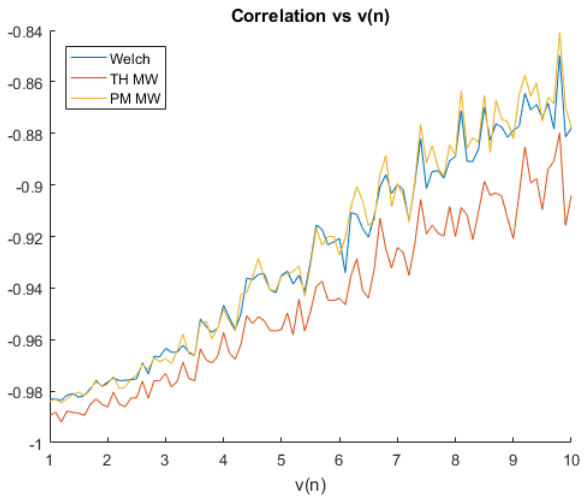

Figure 24: The performance of all the methods plotted in one figure showing that the TH MW gives the best result.

Quadratic case

Then 1200 samples of HRV and respiratory signals are simulated with quadratic increasing frequency starting at $0.12Hz$ and ending at $0.3Hz$. This nonlinear frequency change is more close to what is shown in real data (as one can see from the third spectrogram plot in figure 25).

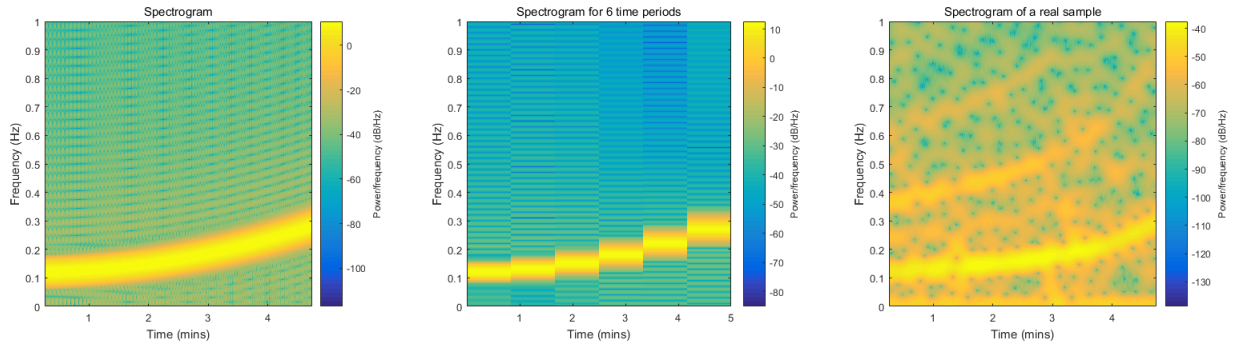

Figure 25: The spectrogram with different time-resolutions and the spectrogram on one of the real samples.

Figure 26 and 27 show an example of spectral estimations of both respiratory and HRV signal with the standard deviation of the LF noise $v(n) = 3$. Same methods are applied on these samples and the mean value, *StdBias* and *SSD* of the estimated band power of HRV are calculated and plotted in figure 28, 29 and 30 respectively.

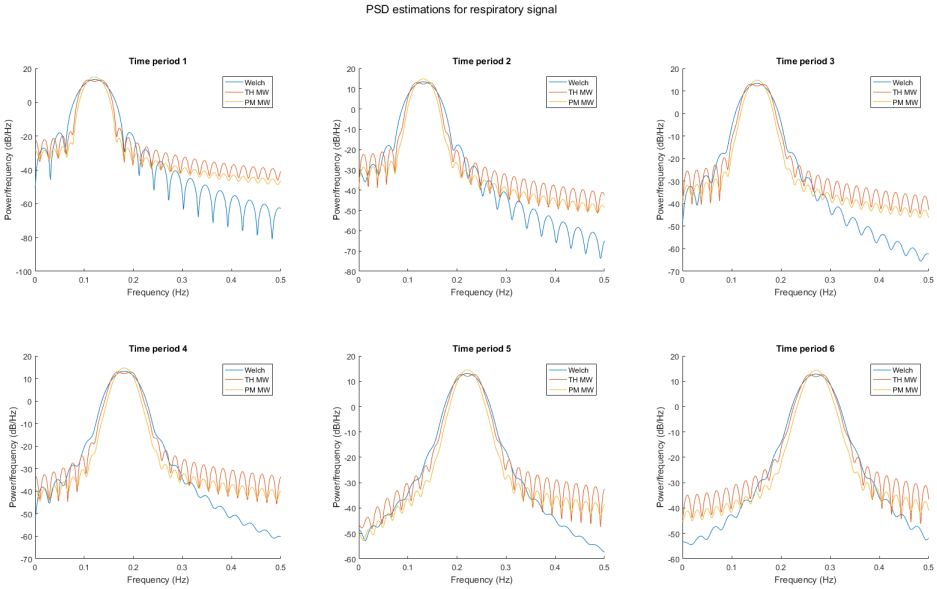

Figure 26: The PSD estimations of respiratory signal. A quadratic growing trend is seen and the central frequency band gets wider since frequency changes more rapidly as time goes by.

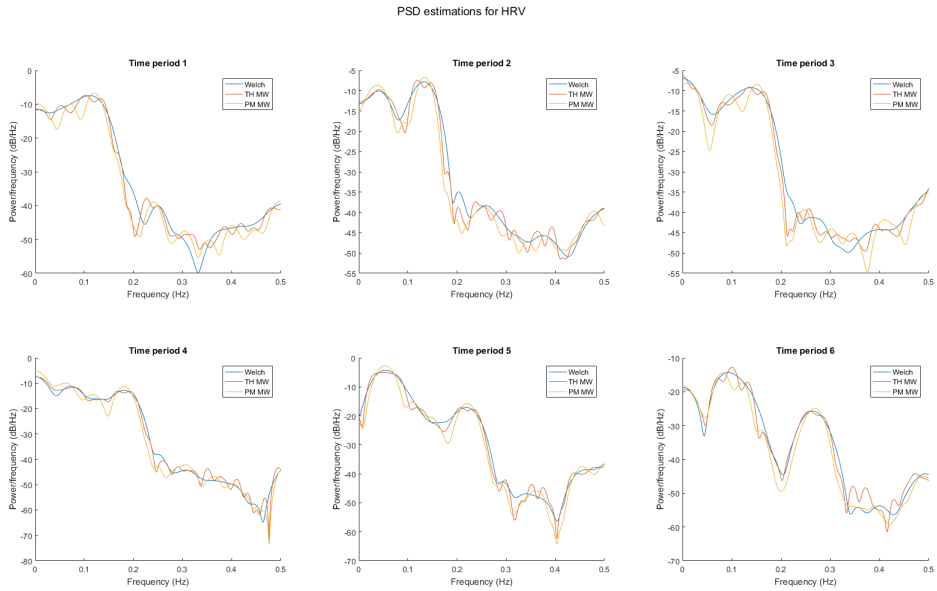

Figure 27: The PSD estimations of HRV. The standard deviation of the LF noise is chosen to be $v(n) = 3$, which results in quite high power in LF part.

If compared to what is obtained from the linear case, one could see that in figure 28 and 29 for

the first 3 time periods, the TH MW now gives the smallest mean value estimation which is due to the reason of slowly increasing frequency at first. After the respiratory frequency band moves away from the LF range (time period 6), the estimation becomes similar to linear case.

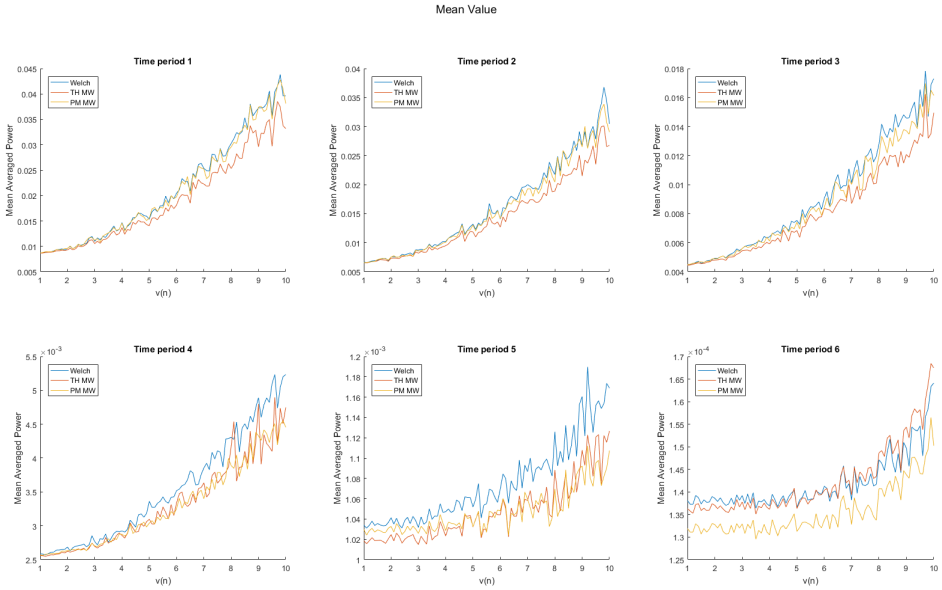

Figure 28: How the mean value of band power estimation varies over increasing noise.

The *SSD* in figure 30 looks quite the same as what is found in linear case, but this time the divergence appears in time period 4 while in linear case the divergence is seen clearly in time period 2 and 3. The respiratory frequency in these time periods being similar to each other is the reason here. Again, no method is robust to noise.

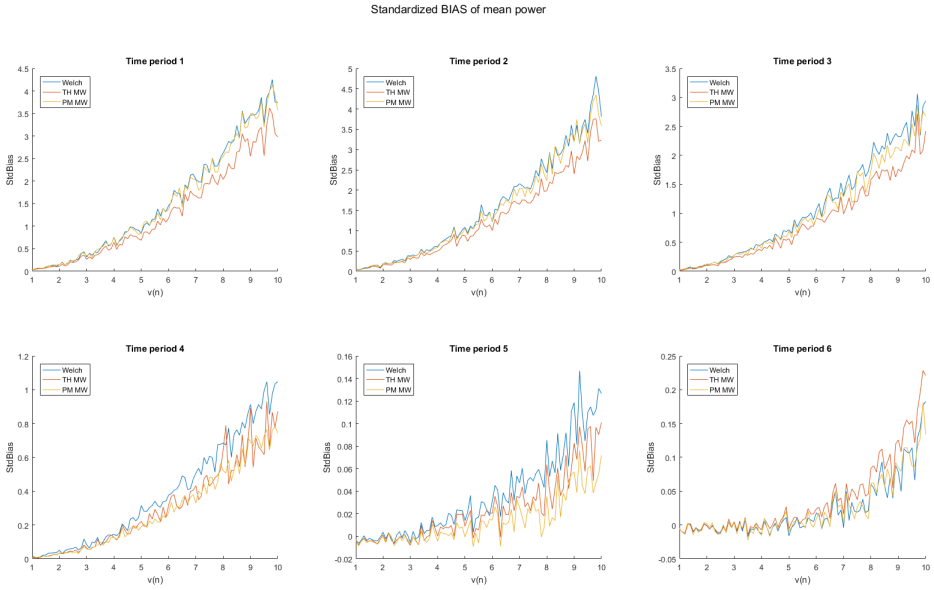

Figure 29: The standardized bias shows relatively how far the estimation is from the 'true' value.

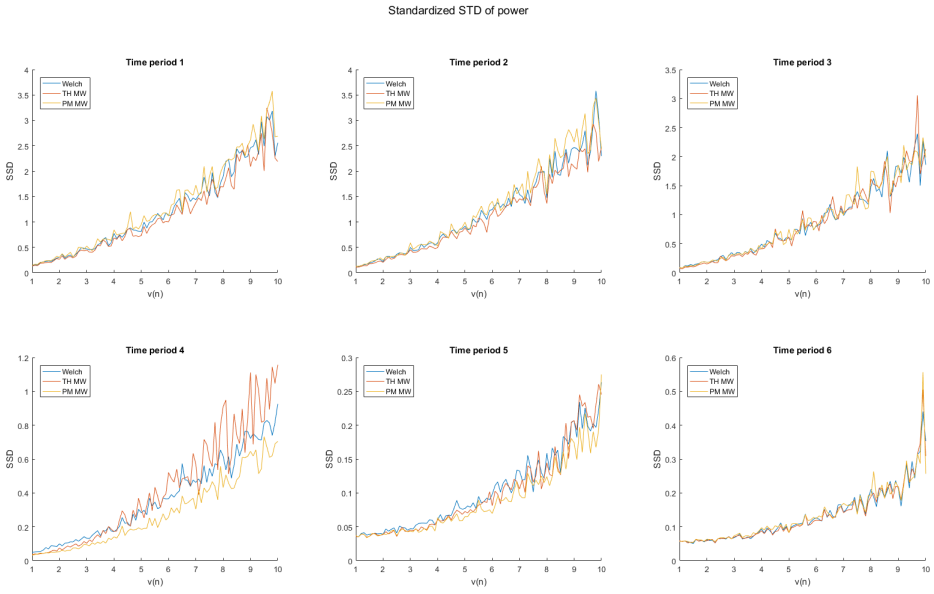

Figure 30: The standardized standard deviation.

The correlation between respiratory center frequency and HRV power is again used as an evaluation approach for the quadratic case. This time the correlation looks less tidy, but still one could tell that the TH MW gives slightly stronger correlation (figure 31).

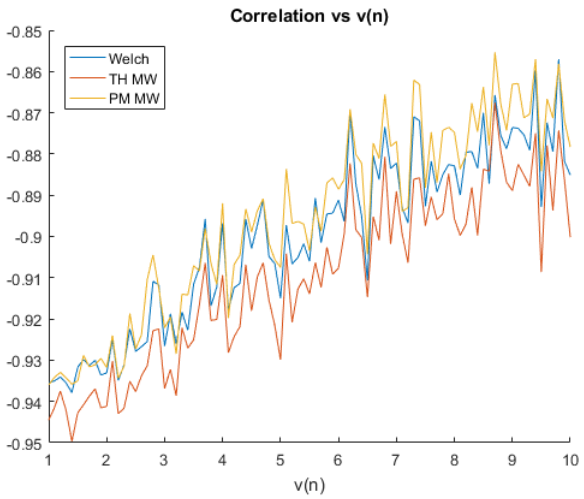

Figure 31: The performance of all the methods plotted in one figure showing that the TH MW still gives the best result.

From the above evaluations, one would conclude that at the beginning, when the center frequency is next to the LF range, all three methods produce results that are strongly affected by the noise, and such effects get larger with increasing level of noise. Then as respiratory frequency increases but still close to LF range, the TH MW seems to most sensitive to the noise for it produces the largest *StdBias* and *SSD* in time period 2 and 3 for linear case and time period 4 for quadratic case. The PM MW performs the best in these periods as it captures the shape better when there is a peak in the PSD estimation. As the frequency center moves away from LF range, the TH MW reaches the same level as the other two and there is no big difference spotted between them. However from the view of the correlation estimation, the TH MW seems to be the most robust method in HRV band power estimation producing the strongest negative correlations for all values of $v(n)$.

5 Real data

5.1 Data description

The data was collected from 21 women and 26 men, aging from 20 to 65 years old, with different levels of working related stress or burnout. Participants were told not to take any food, caffeine or tobacco during 2 hours of before the experiment, nor alcohol the day before. No participant has disease that affects cardiovascular system.
Participants were asked to breath following a metronome starting at $0.12Hz$ and slowly increasing to $0.3Hz$. Additional information on general health and stress level has been collected including age, gender, height, weight, BMI, STAI (State Anxiety and Trait Anxiety) and SMBQ (Shirom–Melamed Burnout Questionnaire) [17].

5.2 Classification

Before applying multitaper methods on the real samples, it is necessary to classify them into different clusters. Since part of the aim of this thesis is to study the difference in coherence between groups with different level of anxiety and stress, only the anxiety related indices are used to conduct an unsupervised classification (K-means clustering used in this thesis). According to information presented in [16, 17], where SMBQ score is used as a categorical variable to classify individuals who has previously experienced stress, it is reasonable to set 3 groups based on SMBQ, together with State Anxiety and Trait Anxiety. Age could also be taken as a stress-related variable but since there are too few samples in each age range given only 47 participants aging from 20 to 65, it is not included in the classification.
K-means clustering with 100 replicates is performed and figure 32 shows the pairwise plots and a 3D plot is shown in figure 33.

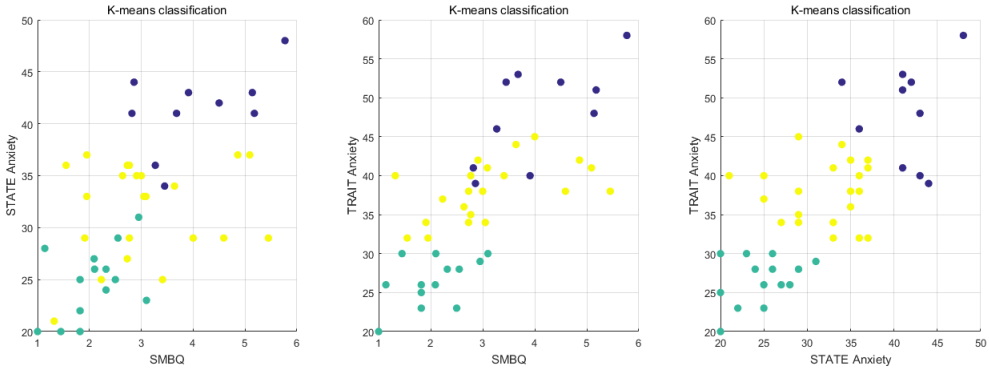

Figure 32: Pairwise scatter plots, where green, yellow and blue dots indicate low, median and high stress group relatively.

A positive linear relationship between each pair of variables is found, making the classification easy to understand - more stressed person, higher the anxiety level. Different from using only SMBQ to categorize, State and Trait Anxiety are helping set a more clear boundary between each groups.

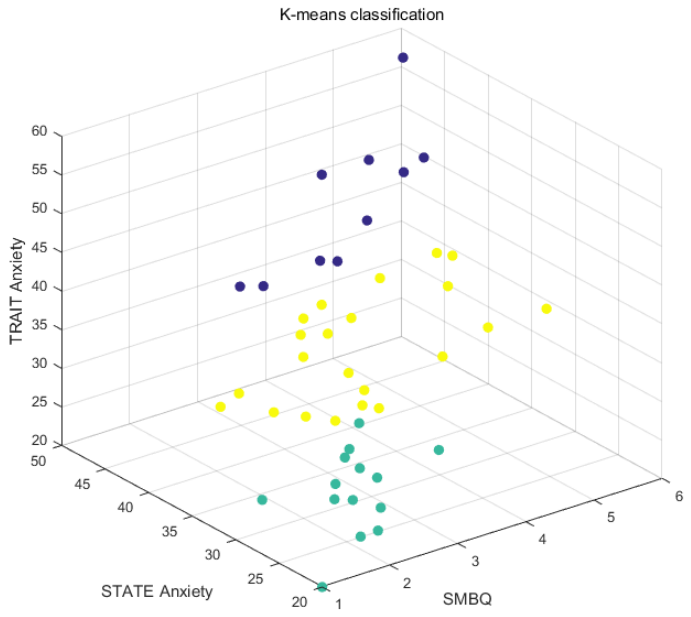

Figure 33: A 3D scatter plot where green, yellow and blue dots indicate low, median and high stressed group relatively.

5.3 Results

It is hard to make comparison between each individual and also meaningless to interpret the difference between them since difference does exist between individuals, let alone noise exists when recording data. However, persons in the same group have something in common, and the information could be extracted by calculating the values within each group.
Similar to what is done with simulated data, the comparison will be made mainly by presenting figures. No true value is available now, so bias could no longer be the criteria to evaluate the goodness of the methods. The standard deviation is available here, however, since there might be a scaling problem as mentioned when evaluating the simulation data and also the difference does exist between individuals as said, the deviation of the band power estimation is not considered as an approach for evaluation. Problem also exists when one wants to compute the *SSD* which could be used to match the values in simulation part to measure the noise level in real data. Again since there is no true value or even the so-called 'true' value introduced on real data, it is impossible to compute the *SSD*.
However checking the correlation between respiratory frequency and HRV power could reduce the deviation brought by individual diversity. It is then to check how strong the correlation is between frequency and power and whether it deviates between groups, which is stated in the introduction part as part of the aim. Different segments, either with or without sample overlap, and different number of windows K from 2 to 6 is evaluated. The correlations are calculated according to eq.(30) and eq.(31) for all combinations and are plotted in figure 34.

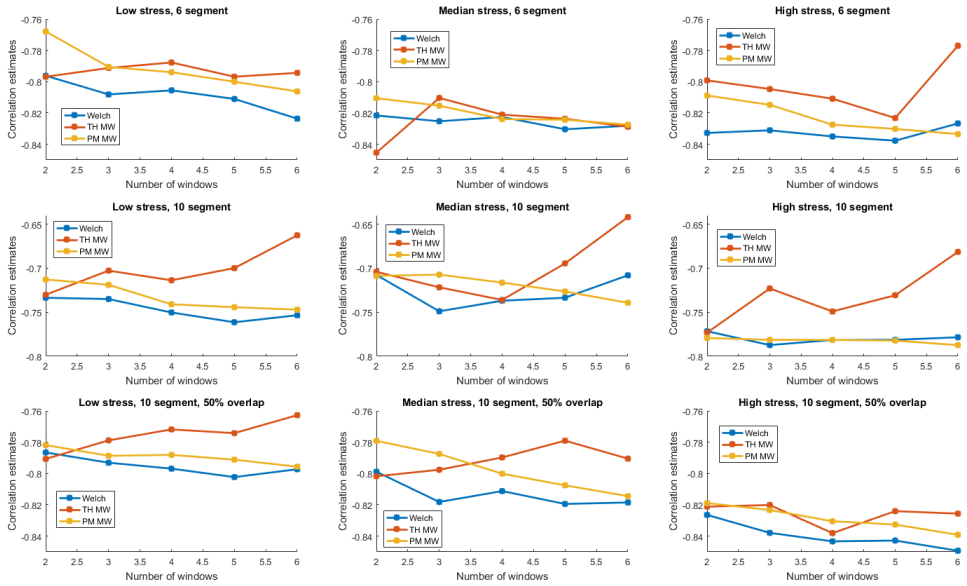

Figure 34: How each method performs when estimating the correlation between respiratory frequency and HRV power. In each subplot, different methods are compared on the same group of individuals. Row: different segmentations. Column: different groups.

The TH MW seems to be most sensitive to the K, as under most cases the correlation given by the TH MW decreases (weaker coherence) as K increases. This also explains why only 2 windows are used on simulation data especially for the TH MW. The values of the correlation differ quite much when the data is segmented into 10 non-overlap parts, with the window length relatively short. The reason for this could be that for a short data segment, signals have almost the same frequency within each window. The Welch and PM MW give quite the same result with the Welch outperforms a little bit. Both methods show a stronger correlation with more windows. When data is partitioned into 6 non-overlap parts or 10 segments with 50% overlap, the windows have similar length, resulting in similar plots though the values are no longer uncorrelated when there is overlap between segments. The Welch seems to be the most robust method since it indicates the most strongest correlation between respiratory frequency and HRV band power. Of course it is of more interest here to see how different the correlation might be between different groups, so the correlations are replotted in another way to see the influence of stress level in figure 35.

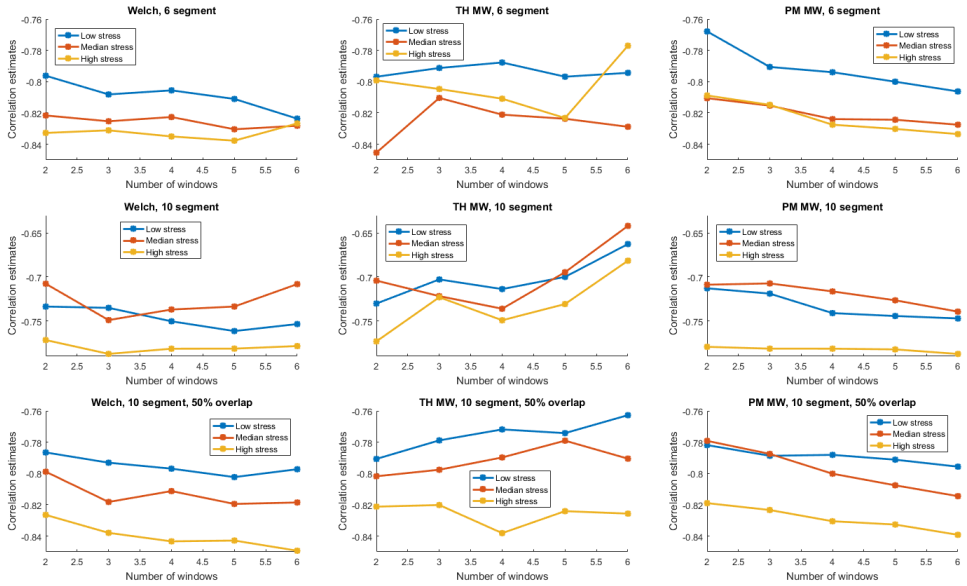

Figure 35: Another way to present the estimated correlation to see if there is difference between groups with different levels of stress using same method. Row: different segmentations. Column: different methods.

Stress is found to have a negative impact on HRV, i.e., high stress resulting in a reduced HRV power (in HF range) [8,9]. However, strongest negative correlation is seen in high stressed group in most plots in figure 35 (except the second plot when data is segmented into 6 parts using the TH MW). The Welch still seems to be the best one to differentiate different stress groups if data is segmented into 10 segments with 50% overlap. The TH MW could distinguish the difference between groups as well under same segmentation but less windows is preferred as it might produce incorrect estimation of HRV power using too many windows. The PM MW is quite robust to different segmentations, also indicating that the strongest correlation is found for high stressed group. It is hard to make a choice of the number of windows to use on real data since different correlation estimations are obtained with different methods and segmentations, however, less windows are preferred (for example using 2 or 3 windows). Also it is better to cut the signal data into 10 parts with 50% overlap since under this segmentation a more clear difference is found between groups.
The correlations between respiration frequency and HRV band power could possibly be further used as a way for classification of individuals with unknown stress levels if given only the respiration and HRV data. This could be a more reliable way to diagnose the stress or anxiety condition of a patient since physiological signals don't lie while people might hide true information when filling a questionnaire. However, more data is needed to support this idea since the clustering of participants is done also by using assessment data.

6 Conclusion

In the process of completion of this thesis, most time was spent on reading and studying multitaper methods and learning how to implement these methods onto data. The simulation of the IPFM signal and the evaluation on the IPFM signal consists most part of the thesis. The signal is modified to be as close to real respiratory and HRV data as possible. Three methods (Welch, TH MW and PM MW) are chosen to be implemented for estimating the spectrum of respiratory signal and HRV signal to obtain the center frequency of respiration f_r and narrow-banded power of HRV within $f_r \pm 0.05Hz$.
Two ARMA-processes were first created to obtain a quick look at the performance of three methods. Since it is a simple model with known PSD and without much structure in it, the Welch gives the best spectral estimation by its smoothness and low variance. The TH MW and PM MW would have captured more information from the signal, however here, both methods seem to overestimate, leading to higher bias and variance.
When it comes to evaluate methods on the IPFM signal simulations, the 'standardized' bias (*Std-Bias*) and deviation (*SSD*) is calculated, i.e., bias and deviation divided by mean power of each method when no noise is introduced in the signal to avoid scaling problems. It is hard to conclude which method gives the closest band power estimation for no 'true' value could be taken as criterion, but the PM MW seems to be the best method for spectrum estimation giving the least deviation. However the TH MW gives a slightly stronger correlation between respiratory frequency and HRV band power (especially in quadratic case which is more close to real data), indicating a more robust estimation. The robustness to noise of each method is also checked, which might not be what is needed. None of the three methods is robust to noise, and it is hard to draw a conclusion.
Finally the methods were applied to real data. The TH MW is most sensitive to the number of windows K, varying a lot and the more windows the weaker correlation one gets, while the Welch and PM MW show a growing trend of correlation (decrease in value). A number of 10 segments with 50% overlap seems to be the best way to cut signals, a clear difference of correlation is found between groups of different level of stress. It is interesting to find that under most cases, strongest negative correlation is found in high stressed group as one might think the other way round given that stress is found to have a negative impact on HRV power. The correlation could be further used as an approach of classification for individuals with unknown stress or anxiety state. Further discussion and research of what it means and its applications in physiology could be done, however, it is not the aim of this thesis.
At last, this thesis was based entirely on non-parametric methods. Though they are most useful and efficient approaches with these signals, there could be some parametric model families that perform well or even better. Moreover, the data is non-stationary that is cut into sections and then treated as stationary. It is not a good assumption, so alternative approaches could be studied to handle this kind of chirp signal better.

References

[1] G. Berntson and J. Cacioppo. (2007). "Heart Rate Variability: Stress and Psychiatric Conditions," in M. Malik and A. J. Camm (Eds.), *Dynamic electrocardiography* (pp. 57–64).

[2] U.R. Acharya, P.K. Joseph, N. Kannathal, C.M. Lim, and J.S. Suri (2006). "Heart rate variability: a review," *Medical and Biological Engineering and Computing*, 44, 1031-1051.

[3] J.D. Schipke, G. Arnold and M. Pelzer, "Effect of respiration rate on short-term heart rate variability", *Journal of Clinical and Basic Cardiology*, 1999; 2 (1), 92-95.

[4] M. Hansson-Sandsten and P. Jönsson, "Multiple Window Correlation Analysis of HRV Power and Respiratory Frequency," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 10, pp. 1770-1779, Oct. 2007.

[5] R.E. De Meersman, "Aging as a Modulator of Respiratory Sinus Arrhythmia," *Journal of Gerontology*, Volume 48, Issue 2, 1 March 1993, Pages B74–B78.

[6] E. Helfenbein, R. Firoozabadi, S. Chien, R. Carlson and S. Babaeizadeh, "Development of three methods for extracting respiration from the surface ECG: A review," *Journal of Electrocardiology*, Volume 47, Issue 6, 2014, Pages 819-825.

[7] M. Hansson-Sandsten and P. Jönsson, "Estimation of HRV spectrogram using multiple window methods focusing on the high frequency power," *Medical engineering & physics*. 28 (2006), 749-761.

[8] J. Taelman, S. Vandeput, A. Spaepen and S. Van Huffel. (2009). "Influence of Mental Stress on Heart Rate and Heart Rate Variability," *IFMBE Proceedings*, 22, 1366-1369.

[9] A. Lennartsson, I.H. Jonsdottir, and A. Sjörs (2016). "Low heart rate variability in patients with clinical burnout," *International journal of psychophysiology*, 110, 171-178.

[10] J. D. Cryer and K-S. Chan, *Time Series Analysis*, Ch.13, Introduction To Spectral Analysis, Springer, 2008.

[11] F. Lizzi, L. Katz, L. St. Louis and D.J. Coleman, "Applications of spectral analysis in medical ultrasonography," *Ultrasonics*, Volume 14, Issue 2, 1976, Pages 77-80.

[12] G.G. Berntson, T.J. Bigger, D.L. Eckberg, P. Grossman, P.G. Kaufmann , M. Malik, et al. "Heart rate variability: origins, methods, and interpretive caveats," *Psychophysiology*, 1997 Nov;34(6):623-648.

[13] G. Lindgren, G. Rootzén and M. Sandsten, *Stationary stochastic processes for scientists and engineers*, Chapman and Hall, 2013.

[14] D.J. Thomson, "Spectrum Estimation and Harmonic Analysis," *Proceedings of the IEEE*, Vol.70, No.9, September 1982.

[15] M. Hansson and G. Salomonsson, "A multiple window method for estimation of peaked spectra," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 778-781, Mar 1997.

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

[17] 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 *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies*, 2018, vol. 4, pp. 17–25.